Learning and Reasoning Strategies for User Association in Ultra-dense Small Cell Vehicular Networks

Shipra Kapoor

Ph.D.

University of York

Electronic Engineering

November 2018

Abstract

Recent vehicular ad hoc networks research has been focusing on providing intelligent transportation services by employing information and communication technologies on road transport. It has been understood that advanced demands such as reliable connectivity, high user throughput, and ultra-low latency required by these services cannot be met using traditional communication technologies.

Consequently, this thesis reports on the application of artificial intelligence to user association as a technology enabler in ultra-dense small cell vehicular networks. In particular, the work focuses on mitigating mobility-related concerns and networking issues at different mobility levels by employing diverse heuristic as well as reinforcement learning (RL) methods.

Firstly, driven by rapid fluctuations in the network topology and the radio environment, a conventional, three-step sequence user association policy is designed to highlight and explore the impact of vehicle speed and different performance indicators on network quality of service (QoS) and user experience. Secondly, inspired by control-theoretic models and dynamic programming, a real-time controlled feedback user association approach is proposed. The algorithm adapts to the changing vehicular environment by employing derived network performance information as a heuristic, resulting in improved network performance. Thirdly, a sequence of novel RL based user association algorithms are developed that employ variable learning rate, variable rewards function and adaptation of the control feedback framework to improve the initial and steady-state learning performance. Furthermore, to accelerate the learning process and enhance the adaptability and robustness of the developed RL algorithms, heuristically accelerated RL and case-based transfer learning methods are employed.

A comprehensive, two-tier, event-based, system level simulator which is an integration of a dynamic vehicular network, a highway, and an ultra-dense small cell network is developed. The model has enabled the analysis of user mobility effects on the network performance across different mobility levels as well as served as a firm foundation for the evaluation of the empirical properties of the investigated approaches.

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, thank be to God, who provided me with such an excellent opportunity in the world of science, and for giving me strength and perseverance to carry on with my work even at the most difficult times.

To be able to keep a strong and smooth sail, I would like to express my heartfelt and deepest gratitude to both my supervisors, Prof. David Grace and Mr. Tim Clarke for providing me with their persistent and valuable support, encouragement and guidance. Their countless suggestions on various research directions, boundless freedom to explore them and constructive feedback helped me pass through all the difficulties I encountered during my Ph.D. study.

I would also like to thanks my thesis advisor, Dr. Paul Mitchell for the motivating in-depth discussions, valuable comments and suggestions during the periodic research assessments that helped in shaping the research work. I would also like to thanks Prof. Muhammad Ali Imran (University of Glasgow) for agreeing to be my Ph.D. examiner.

Thanks to my postgraduate supervisor, Prof. Ghanshyam Singh, who introduced me to the field of research. Thanks to all my previous and current colleagues at the Department of Electronics for creating a friendly and helpful research environment. In particular, thanks to Jialu, Abimbola, Tautvydas, Kafi and Sumeiya for the wonderful memories in York. Many thanks to all my friends in India, especially Vinisha and Swati for their belief, support and love for me even when I was so far away.

With love and gratitude, I would like to sincerely thanks and dedicate this work to my parents, Mr. Santosh K. Kapoor and Mrs. Seema R. Kapoor who supported me in every decision of my life and who have always given me their unconditional love and support. I cannot forget to thanks my lovely beautiful sister, Shivani, and my husband, Pranjal, for their unconditional love, friendship and care that made my day-to-day life enjoyable and the journey of getting the Ph.D. less heavy.

Finally, I am very grateful for the financial support by the University of York that allowed me to pursue my studies.

Declaration

All work presented in this thesis is original to the best knowledge of the author. References to other researchers have been given as appropriate. This work has not previously been presented for an award at this or any other institution. The research presented in this thesis features in a number of the author's publications listed below.

Journal Articles

- S. Kapoor, D. Grace, and T. Clarke, "User Association in Ultra-Dense Small Cell Dynamic Vehicular Networks: A Reinforcement Learning Approach, accepted by Journal of Communications and Information Networks.
- S. Kapoor, D. Grace, and T. Clarke, "Heuristically Accelerated Q-Learning Approach for User Association in Ultra-Dense Dynamic Vehicular Networks", - prepared, planned to submit to IEEE Access.

Conference Paper

- S. Kapoor, D. Grace, and T. Clarke, "A base station selection scheme for handover in a mobility-aware ultra-dense small cell urban vehicular environment, in Personal, Indoor, and Mobile Radio Communications (PIMRC), 2017 IEEE 28th Annual International Symposium on, 2017, pp. 1-5: IEEE.

Chapter 1. Introduction

Contents

1.1 Background

Owing to the rapid growth of multimedia infotainment applications and increase in the number of high-end devices, a sharp increase in the global mobile data traffic has been witnessed in the past decade as outlined in the latest Cisco Visual Network Index (VNI) report [1]. The report presents that global mobile traffic will increase nearly threefold between 2016 and 2021, reaching 27.8 exabytes per month by 2021 as shown in Figure 1.1, wherein 82% of the traffic will be video, indicated in Figure 1.2. This requires increased wireless system capacity to provide high data rate whilst ensuring a guaranteed quality of service (QoS).

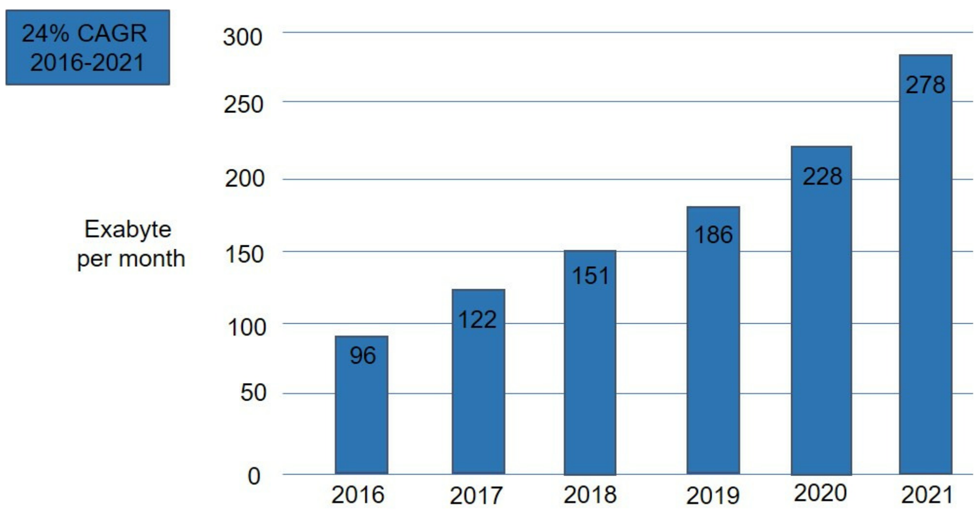


Figure 1.1: The mobile data traffic per month in exabytes by 2021 as reported by Cisco Visual Networking Index (VNI), redrawn from [1]

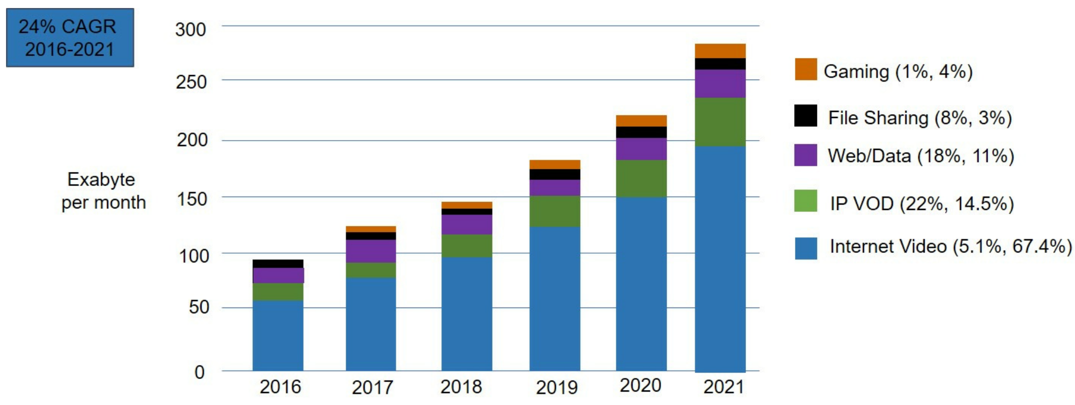


Figure 1.2: Global internet protocol (IP) traffic by application category, wherein IP video traffic will contributes 51% to the total traffic generated by 2021, redrawn from [1]

4G mobile networks are designed to provide mobile broadband, wherein currently the user experience relies on the infrastructure and varies significantly in response to the varying wireless channel conditions, particularly from the cell edge to the centre of the cell. The demand for highly reliable radio access network coverage is not only limited within the home or office but is also an essential requirement of future vehicular networks. It has therefore been collectively agreed by the researchers in the field of communications that a paradigm shift is needed to meet the dramatic increasing demands for mobile data services as incremental improvement in today's wireless broadband network that mainly consists of traditional cellular base stations cannot meet the foreseeable future data demands [2]-[4].

The fifth generation mobile network is envisioned to support a diverse range of new services and applications such as device-to-device communications, ultra-high definition video streaming, mission-critical services, and intelligent transportation services by providing high data rate and coverage, seamless mobility, and reduced end-to-end latency compared to the current 4G LTE networks. Some of the key technology enablers for the 5G communications as identified by the authors in [5] are massive Multiple-input Multiple-output (MIMO), millimetre Wave (mmW), and dense heterogeneous networks (HetNets). Although it still remains unclear whether or not the lightweight, low-powered small cell base stations (eNBs) will fully substitute the traditional macrocell networks, their progressive development has led towards the emergence of different types of BSs, resulting in Heterogeneous networks (HetNet) and Ultra-dense Small cell networks (UDN).

In the context of vehicular networks, it is foreseen that the 5G vehicle-to-anything (V2X) will enable safe and intelligent transportation services (ITS) by employing the traditional macrocell network for greater coverage and deployment of low power nodes to provide a better quality of service (QoS), quality of experience (QoE), ultra-high data rate and low end-to-end latency especially in those areas where traffic demand is high. Although further densification in a vehicular environment will assist in overcoming increasing traffic demands of moving users, however the challenge in an ultra-dense small cell environment is frequent user association to the most promising base station.

User association relates to adapting a radio-link for the transmission of data depending on the prevailing radio traffic environment. It is a critical element in communication networks as it substantially affects the network performance [6]. The user association decision in the existing LTE/LTE-A systems is taken by the radio admission control entity located in the radio control layer of the protocol stack [7]. The decision depends on the quality of service (QoS), priority level of the request and the availability of the resources. In existing systems, a UE is associated with a BS depending upon the measurement of the received signal strength (max-RSS) of neighbouring BSs, if the association is initiated by the user [8, 9]. Alternatively, a UE may also associate with a BS based upon signal quality, i.e. the maximum signal-to-interference-noise ratio (max-SINR) rather than signal strength [10]. However, these approaches do not consider the impact of vehicle mobility during the user association process. An association algorithm that does not take mobility into account in a dense small-cell scenario during scheduling may result in a higher handover frequency when compared to handovers in a conventional cellular network.

Handover is defined as a process in which the ongoing transmission is transferred from the current associated BS to a new target BS, depending on association policies [11][12]. The number of handovers per transmission is termed as handover frequency in the presented thesis. In dynamic vehicular environments, the number of handovers increases linearly with an increase in the vehicle speed, if the max-RSS user association approach is considered [13]. This may further increase the switching and signalling load resulting in undesirable end-to-end delay and possibly dropped transmissions [11] [14] . Therefore, more sophisticated and intelligent user association algorithms with the ability to consider the vehicular dynamics, thus, adapt to the environment are needed in order to improve the handover rate whilst delivering guaranteed network quality of service.

The key objective of the presented work is to apply known machine learning approaches to improve the handover rate in small cell dynamic vehicular environment. The user association algorithms that have been developed are with consideration to single base station association contrary to multi base station association. This is because multi cell association incur redundant signalling overheads in establishing and maintaining more than one association subsequently, in addition to intra-cellular as well as inter-cellular handover resulting in more complex evaluation of handover rate. Whereas, in single base station association no intra-cell handover is considered thus enabling an accurate assessment of inter cell handover rate with respect to vehicle speed. Moreover, in multi cell association geographically separated antennas are employed to receive signals from the active users to significantly improve the cell throughput whilst implementing scheduling decisions to control interference, as in the case of CoMP Uplink. However, in the presented scenario, omnidirectional antennas are employed that would lead to considerable interference if multiple cell association is considered. In addition to this, the fundamental objective was to develop machine learning inspired user association approaches that are able to adapt to the environment such that the handover rate is significantly reduced while delivering guaranteed network QoS across all investigated vehicle speed.

1.2 Hypothesis

The following hypothesis guides the research work presented in this thesis:

"Appropriate exploitation of heuristic information through reinforcement learning for user association in ultra-dense dynamic vehicular environments can significantly reduce the handover rate whilst delivering a guaranteed network quality of service."

Reinforcement learning (RL) is foreseen to adapt to the spatial-temporal irregularities of urban traffic flow in ultra-dense small cell scenarios enabling improvement in the network reliability, quality and latency by utilizing the heuristic information. The network adaptability and evaluation of improvement with the RL approaches include the amount of reduction in the number of handovers per transmission while delivering a guaranteed Quality of Service (QoS) at different vehicle speed. This thesis, therefore, focuses on developing intelligent user association schemes that effectively use radio and vehicular traffic information for decision making to adapt to the challenging dynamic radio environment. The developed algorithms are assessed using large-scale simulation on a highway scenario in an urban setting. The assessment could be extended to different network architectures or to additional aspects such as energy efficiency, user power consumption, and/or on-demand radio resource provisioning in future.

1.3 Research Contributions

The contributions of the thesis are as enumerated below:

- An extensive vehicular-communication traffic framework that allows system as well as user experiences to be modelled has been developed using MATLAB software environment (version R2015b) in Chapter 3. The model assist to investigate and empirically assess a range of user association approaches at dynamic urban traffic flow and constantly varying wireless channel conditions on overall network performance in a dense small cell highway scenario.
- The three-step performance metric dependent user association approach proposed in Chapter 4 selects an appropriate base station depending on desired performance metric, thus, demonstrate the impact of different performance metric on the overall network performance and user experience. The approach also contributes towards the significance of association range and the necessity to adapt to dynamic urban vehicular environment to deliver guaranteed network QoS whilst reducing the handover rate across different vehicle speed.
- A range of reinforcement learning based intelligent user association algorithms are proposed in Chapter 5 to adapt to the changing environmental conditions, thus, adjust the user association decisions accordingly. The developed adaptive,

model-free, online schemes exploit the state-of-the-art Q-learning framework to achieve an effective and reliable solution only through trial-and-error iterations in the considered multi-agent scenario. The schemes recursively reconstruct the current policy making it difficult for the learning agent to converge to local optimum, thus, encouraging to look for a better solution.

- The case-based reinforcement learning (CBRL) user association scheme proposed in Chapter 6 combines case-based transfer learning (CBTL) and heuristically accelerated reinforcement learning (HARL) on a conventional Q-learning framework to learn a reliable solution across different traffic conditions in dynamic environments. The scheme uses the network performance information to enable the learning agents to adapt to the dynamically changing environment, learn best association range, thus, stabilize their performance at different investigated vehicle speed.

1.4 Thesis Outline

The presented thesis is organised as follows:

- Chapter 2 provides a literature review to establish the background of presented work. First, the fundamental concepts of HetNets and vehicular networks are discussed, wherein small cell networks are envisioned as a promising solution to improve cell edge performance as well as provide ubiquitous connectivity and coverage for ultra-reliable low-latency services. Next, a comprehensive summary of the existing conventional user association strategies and a number of machine learning methods based on state-of-the-art reinforcement learning for user association is presented. Finally, in the last section of the chapter RL and associated techniques that are widely used in the wireless and the artificial intelligence domain are reviewed.
- Chapter 3 presents the scenario and the network architecture. Further, comprehensive explanation related to different modules and processes employed for the development of the presented simulator is provided. The experimental methodology used for empirical evaluation of the different user association algorithms

proposed in this thesis is discussed next.

- Chapter 4 begins by introducing user association mechanism. It further discusses the conventional user association technique that has been used as the baseline comparison scheme for the assessment of different proposed schemes later in this thesis. Following this, the performance metric based user association strategies that follow a three-step sequence for user association in vehicular networks is proposed. The results provide insights into understanding the influence of individual performance metric on the network performance and user experience.

- Chapter 5 proposes a class of adaptive user association algorithms in the vehicular network. First, a real-time control feedback user association approach that is a model-free computational algorithm, inspired by control-theoretic models and dynamic programming is proposed. Subsequently, a user association algorithm based on classical Q-learning technique is proposed. This scheme serves as a baseline approach for the investigated learning technique. In addition to this, the chapter introduces and employs the concept of variable learning rate and variable reward function for developing adaptive user association algorithms. Further, user association algorithms that integrate investigated RL techniques to develop variable reward, quality aware Q-learning (VR-QAQL) approach is proposed. This scheme significantly improves the network performance and user experience compared to the baseline approaches. Finally, the results obtained that evaluate the performance of the non-learning algorithms and other investigated Q-learning algorithms are demonstrated.

- Chapter 6 proposes a novel case-based reinforcement learning (CBRL) approach for user association; an extension to the VR-QAQL approach that was proposed in Chapter 5. The approach uses a combination of heuristically accelerated reinforcement learning (HARL) and case-based transfer learning (CBTL) technique on a VR-QAQL framework to achieve the tradeoffs between handover frequency and guaranteed network quality of service at all vehicle speeds in the considered bi-directional flow, multi-lane, urban traffic scenario. The chapter also provides an extensive review of the CBRL and HARL techniques and discusses their significance to accelerate and enhance the learning performance. Furthermore, the performance of the developed CBRL techniques is assessed using a number of simulations at different vehicle speed in the dynamically changing topology.

- Chapter 7 concludes the presented thesis, summarises the novel contributions and provides the recommendation for further extension of the presented work.

Chapter 2. Literature Review

Contents

2.1 Introduction

This chapter presents a review of the essential concepts, background information and extensive studies in the ultra-dense wireless networks, vehicular networks and artificial intelligence that relates to the scope of intelligent user association in ultra-dense small cell dynamic environments as per the hypothesis indicated in Chapter 1 of the presented thesis.

Due to the recent advancements in wireless technologies, vehicle infrastructure and automobile industry, vehicular ad-hoc networking is envisioned as a key enabling technology that has significant potential to ensure traffic safety, traffic management, infotainment, and other mission-critical applications. However, these applications demand reliable quality of communication that greatly depends on network architecture, advanced vehicular networking, and desired performance.

Consequently, to meet these essential advanced requirements of the vehicular networks, the 5G cellular networks are envisaged to be an essential technology and infrastructure facilitator. The goals of 5G are extensive, but, include vital advanced requirements of the vehicular networks. The primary technologies and approaches identified for 5G networks by [5] are heterogeneous networks (HetNets), massive MIMO, mm-waves, device-to-device (D2D), full-duplex, beamforming, energy aware, cloud radio access network (C-RAN), and virtualization of wireless resources. Among these technologies, heterogeneous networks (HetNets), massive MIMO, mm-waves have been identified as the "big three" technologies of 5G [2]. The indispensable advantages of the 5G cellular communications, therefore, appear to meet the advanced requirements of the vehicular networks. A fundamental task performed to ensure the connectivity of a user/user equipment (UE) to the network through a particular base station (eNB) before it starts data transmission, regardless of the technology adopted, is user association. In the case of dense small cell networks with dynamic vehicle mobility, the selection of an appropriate eNB for user association appears to be challenging due to frequent handover, that may lead to challenges such as network coordination, configuration and management [15].

One possible way to mitigate these challenges is the introduction of machine learning and artificial intelligence algorithms in the network. The introduction of intelligence to the network will assist the eNBs to self-adjust, synchronize and adapt to the changes in the environment such that an ultra-reliable low latency wireless communication for different vehicular applications is achieved [16]-[17]. The inherent advantages and prevalence of the dense environments and machine learning techniques motivated for the development of novel intelligent user association approaches for ultra-dense small cell dynamic vehicular environments that are proposed in the presented thesis.

The rest of this chapter is organised as follows: Section 2.2 reviews dynamic vehicular networks and the need of a standardized vehicular-communication traffic simulator. Next, an overview of the ultra-high capacity wireless networks is given in Section 2.3. Following this, user association mechanism that does not include machine learning techniques are discussed. Section 2.4 argues machine learning followed by an in-depth review of reinforcement learning and associated algorithms as applied in wireless communication domain. Finally, the chapter concludes in Section 2.5.

2.2 Dynamic Vehicular Networks

2.2.1. Vehicular Networks Overview

Vehicular networks is an emerging intelligent transportation system (ITS) technology that integrates wireless communications to vehicles to enable diverse applications related to road traffic [40]. The growing interests towards applicability of wireless communication to vehicles have resulted in the formation of several government organizations, standardization bodies, dedicated short-range communication (DSRC) in the US and the ITS-G5 in Europe, has been set up, both based on IEEE 802.11p technology. Moreover, different consortia such as vehicle infrastructure integration (VII) in the US, car-to-car communication consortia (C2C-CC) in Europe to formulate the guiding principles, requirements, architecture, and protocols for data transmission between vehicle to vehicle (V2V) and vehicle to infrastructures (V2I) [60]. The goal is to streamline the operation of vehicles as well as to support high mobile broadband access for infotainment, to improve traffic efficiency, road safety and management applications.

The two main vehicular communication modes are: (a) communication among nearby vehicles known as vehicle to vehicle communication, wherein, the data transmission takes place between the vehicles either via single hop or multi-hop, and (b) communication between vehicle and road-side infrastructure known as the vehicle to infrastructure (V2I) communication, wherein, the data transmission is one-hop. V2V communication is as important as the V2I communication. Few of the associated V2V and V2I applications considering latency as key parameter may be broadly split into latency tolerant and latency-intolerant applications. One of the examples of latency tolerant application is software update over the air, wherein, the end-to-end latency may be relaxed while a guaranteed network QoS is essential. On the contrary, real-time data applications are latency intolerant, therefore, in such use cases provision of highly reliable connectivity, ultra-low latency and guaranteed network quality of service is exceptionally important. The presented research focuses on the development of user association algorithms in the V2I scenario and disperses communication techniques and user association mechanism in V2V scenarios.

Considering the potential impact of 5G wireless communication on the automotive market, efforts to develop communication protocols, mobility models and network simulators related to vehicular communications has significantly grown. This is due to the importance and necessity to test, evaluate and analyze the performance of proposed algorithms prior to their implementation on the real test bed. The next section discusses the significance of mobility and network models to assess new algorithms in the field of vehicular communication. Furthermore, a mobility model framework is also argued that has been used as a guideline for the development of the comprehensive vehicular communication simulator in Chapter 3.

2.2.2. Vehicular Network Simulator

The assessment, evaluation and implementation of new strategies on a computer-based simulation are widely preferred in the research community over real test bed due to logistic difficulties, economic issues and technology limitations [61]. However, the key requirement for computer-based simulation is a standardized simulator. In the case of vehicular networks, the simulator is a combination of the mobility model that relates to the vehicle movement in the environment while the network model relates to the evaluation of network communication performance.

There are several excellent contributions in the literature towards the development of mobility models [77]-[82]. The authors in [61] presented an exceptional survey on available mobility models, their usability, merits, demerits and challenges for vehicular networking. Similarly, an extensive overview of major ITS programs and projects conducted in the USA, Europe and Japan is presented in [83]. This paper also discussed the different networking architecture and protocols implemented in these projects. Few of the mobile network operators and communication equipment vendors have developed their own proprietary simulators that are not open source or freely available [64]-[66]. Whereas, some of the simulators developed in academic-industrial collaboration are commercially available, but, lacks to provide the source code.

In [67], an LTE link-level simulator using MATLAB environment has been proposed. The main highlighted features of the simulator are adaptive modulation and coding (AMC), single and multiple user scenario, multiple input multiple output (MIMO) technique, and flexibility to implement different scheduling schemes. Moreover, the paper suggests that by employing parallel computing toolbox of MATLAB along with the simulator to significantly reduce the simulation time. Similarly, in [68], a MATLAB based system level simulator for LTE networks that employ the open loop spatial multiplexing and transmission diversity modes to evaluate the performance of shared channels of LTE SISO and MIMO networks in downlink flow is proposed. However, both the proposed simulator lack some of the important aspects relevant for vehicular communications such as a multi-cell environment, the uplink flow, dynamic vehicle mobility, and a complete LTE protocol stack.

Recently, a hybrid approach has been extensively used to assess developed strategies in vehicular networks. For example, Simulation of Urban Mobility (SUMO) is used to generate the vehicle mobility traces. The traces obtained are supplied into a standard network simulator, such as NS-2 or MATLAB to assess the proposed strategies. A study performed by the authors in [62] investigates the most popular VANET simulator between 2004 and 2007 at the ACM VANET workshop for vehicular networking. It shows that 70% of the papers out of the 51 papers that used simulators for accessing the proposed techniques, the most popular choice of network simulators were NS-2 [69], QualNet [70], and SWANS [71] while the mobility simulators used were SHIFT/SmartAHS [72], CORSIM [73], VanetMobiSim [74], VISSIM [75], and SUMO [76]. On the contrary, 16% of the researchers used a self-developed simulator. Despite the contrary, to the best of the authors' knowledge, at the present time, no standard, open source, *integrated vehicular-radio network* simulator which may serve as a common reference simulator for the comparison of results presented by different research groups is available [62, 63]. Therefore, one of the integral elements of the presented research work is the developed comprehensive, integrated vehicular-network simulator. It is exploited to assess the different user association strategies in this thesis. An in-depth discussion on the vehicular simulator is provided in Chapter 3. The next section discusses the mobility model framework that serves as a guideline for the integrated model.

2.2.3. Mobility Model Framework

Road traffic is a complex multi-agent system in which the agents i.e., the vehicles may or may not interact with each other. The way these agents move depends on different aspects of traffic flow operations, network topology as well as traffic flow modelling. In Figure 2.1 a mobility model framework is presented.

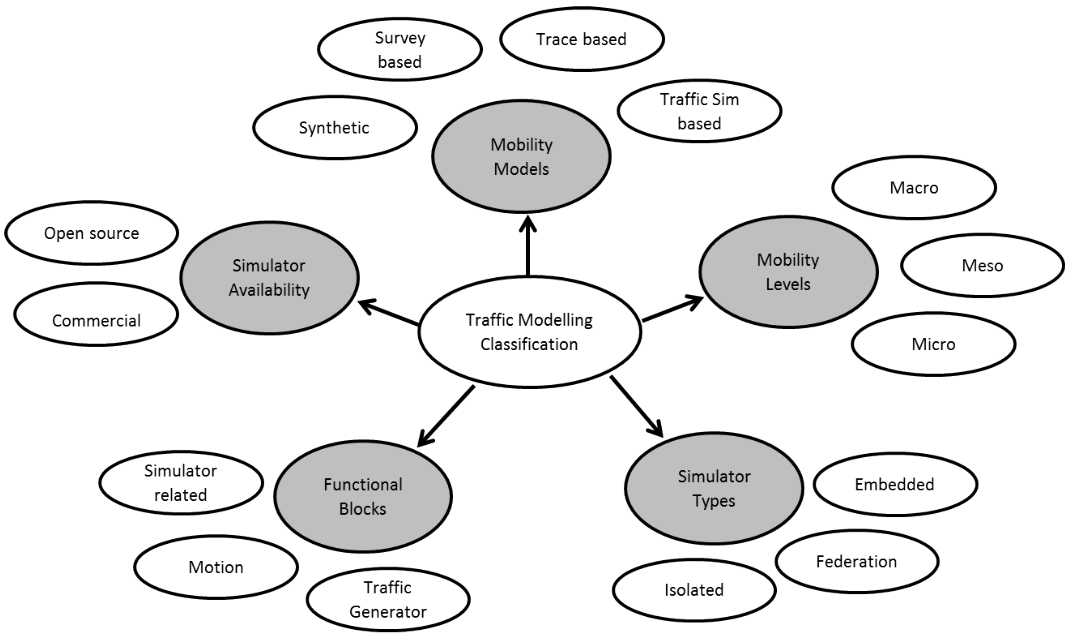


Figure 2.1: Mobility modelling taxonomy structure

One of the key factors while developing a mobility model is the modelling level. Considering the level of detail with which the vehicular flow is described, traffic modelling is characterized on the basis of time-spatial behaviour of individual vehicles under the influence of vehicles in proximity (microscopic models), the behaviour of individual vehicle without considerations of their time-space behaviour (mesoscopic models) or from a collective, vehicular flow viewpoint (macroscopic models) [90].

A microscopic model is a more detailed traffic flow modelling. The model captures individual vehicle location, trajectory, velocity, acceleration and architecture parameters such as the number of lanes, number of intersection and traffic light timings. The car following model by Wiedemann is an example of a microscopic model [90]. A mesoscopic model is a medium detail traffic model and defines the probabilistic behaviour of an individual agent. Whereas, a macroscopic level model is regarded as a low detail traffic model. It provides a viewpoint of collective vehicular flow and captures parameters such as the rate of traffic flow, mean speed, time headway and occupancy of traffic on the street. Although this is low detailed modelling, it is generally quick. The drawback of the macroscopic traffic model is its inability to record the behaviour of every single vehicle. An example of the macroscopic level is the cell transmission model that demonstrates the static density-flow relation [91] [92].

Mobility Models

As seen in the figure above, the mobility models for vehicular networking may be developed based on four main categories [61]. These categories are discussed next.

- **Synthetic models**: These models are based on mathematical models that have an ability to reflect a realistic mobility effect. According to the authors in [85], the synthetic models may further be classified into five subcategories such as stochastic models, traffic stream model, car following models, queue models, and behavioural models.
- **Survey-based models**: These models are developed by gathering the mobility information in real scenarios through surveys conducted by different government organizations. For example, the UDel Mobility Model [86] and Agenda-based Mobility model [87] are developed by the collaborative survey conducted by government and other research institutes in the US.
- **Trace-based models**: These models generate the mobility patterns based on real mobility traces. The mobility traces is gathered through various measure-

ment campaigns. Some of the developed trace-based models reported in [61] are DieselNet, the Cabspotting project and a collaborative model developed by the members of the Fleetnet project, the Network on Wheels project and the HWGui project.

- **Traffic Simulators-based Models**: These models are developed by refining synthetic models and verifying the outcome using real traces or behavioural surveys. Some of the popular traffic simulators-based models are PARAMICS [88], VISSIM [75], TRANSSIM [89] and SUMO [76].

The mobility traces generated by utilizing any of the above mobility model methods is fed into a network simulator using an interface for the further investigation of the proposed algorithm performance.

Simulator Interaction Types

The previous sections discusses the different methods to develop mobility model that are adapted to vehicular networks. However, to be used by the network community, the two models, mobility model and the wireless communication model, are required to interact with each other. The different interaction approaches are isolated approach, embedded approach and/or federated approach [61]. In the isolated approach, the two models are required to be generated prior to the simulation and translated by the simulator according to a predefined trace format for further processing and analyzing. Moreover, there is no interaction between the two models during the simulation. The embedded approach, integrates networking capabilities along with generation of vehicular mobility traces in a unique model. A bi-directional interaction between the vehicle mobility model and network simulator model persists. An example of embedded mobility model is MoVes [94]. Subsequently, the federated approach display potential of providing significantly advanced vehicular motion modeling and networking capabilities. Few of the examples of federated models are (a) TraNS that employ SUMO for mobility model and federating the ns-2 for networking capabilities [95] (b) VGrid project that integrates SWANS as network simulator and synthetic traffic model for vehicle mobility [96]. A comprehensive overview of the different interaction approaches as well as models under the three categories is presented in [61].

Functional Blocks

Another building block that substantially impacts the mobility model is the functional block that relates to the traffic flow. The traffic flow may be distinguished under three qualitatively distinct states. These are (a) free flow, for example, traffic flow on an unrestricted highway, (b) synchronized flow i.e., traffic flow in an urban scenario with traffic lights at intersections, and (c) restricted flow such as at traffic jams. The overall network performance is substantially affected by the traffic flow pattern as an instantaneous change in traffic flow significantly varies the communication traffic load on an associated eNB. However, the traffic flow pattern depends on different factors such as the network topology, time and day of the week, roadside infrastructure, traffic rules as well as technical restrictions imposed by the vehicle. Therefore, designing and developing a vehicular model requires careful planning and an ability to predict essential network traffic characteristics. Decisions based on poor predictions may adversely affect the network. The next section introduces some of the widely used mobility simulators.

Simulator Availability

Traffic simulators play an important role in vehicular communications research. They enable traffic flow simulation with the flexibility to make necessary changes in the road network, vehicle flow or vehicle density. CORSIM, VISSIM, TRANSSIM are few of the commercially available traffic simulators that require a purchase of license which in some cases may be quite expensive. On the contrary, some of the popular open source mobility models are SUMO, CARISMA and SHIFT. Some of the real-world test beds considered for developing the traffic simulator are presented in [93]- [99].

The next section overviews the second essential entity of the presented research work, the ultra-high capacity wireless networks that play an important role to provide connectivity and coverage to meet the advanced requirement of the vehicular networks.

2.3 Ultra-High Capacity Wireless Networks

The exponential rise in multimedia infotainment applications and high-end devices serves as a major driving force towards high data rate networks [18]. Some of the popular traffic-intensive applications that require high data rate include high definition video, wearable devices, virtual gaming, augmented reality, connected vehicles, etc. In order to meet the intensifying data requirements from these applications advanced technologies such as carrier aggregation (CA), multiple input multiple output (MIMO), and coordinated multipoint (CoMP) were introduced by the LTE-A standardization [19]. Despite these advanced technologies, the spectrum efficiency of the air interface in homogeneous networks has been reaching its capacity limit due to the increasing data traffic and user density [20] [21]. It was therefore agreed that a paradigm shift is essential for the emerging fifth generation mobile networks [7]. The goals of 5G are broad. Some of the notable key performance indicators (KPIs) envisioned by 5G-PPP for IMT-2020 are shown in Figure 2.2.

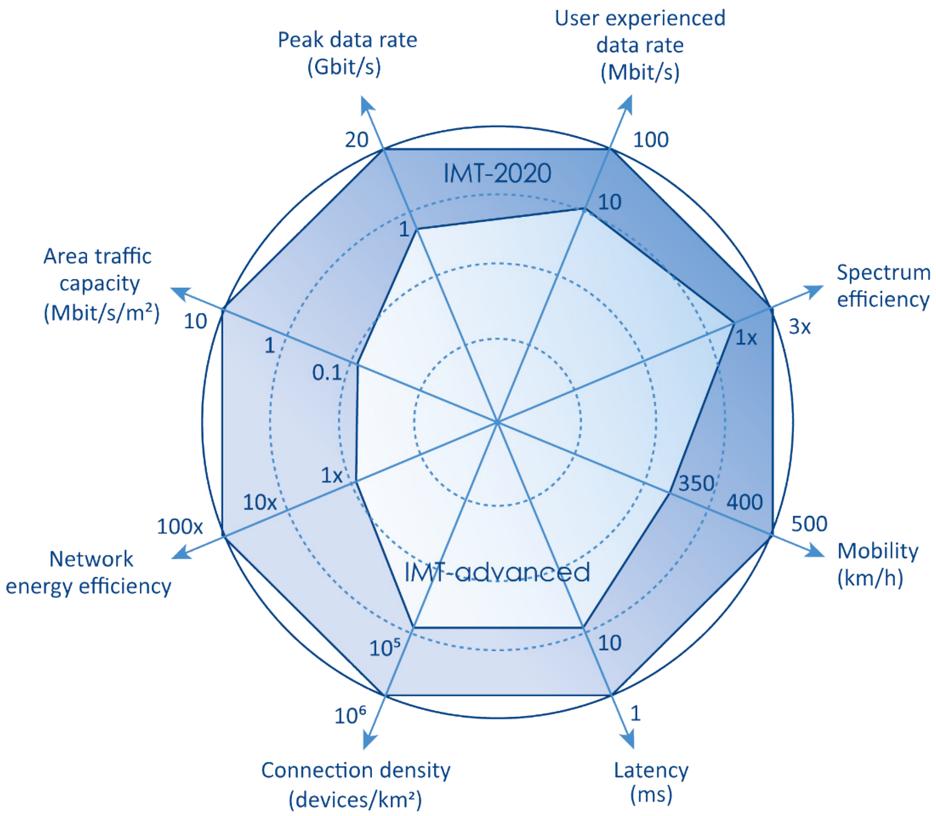


Figure 2.2: The key notable performance indicators envisioned by 5G-PPP for IMT-2020, directly reproduced from [23]

Among the highlighted KPIs in Figure 2.2, the KPIs that relates to the presented work are latency, mobility, user experience, data rate and the area spectral efficiency. The area spectral efficiency is defined as the information rate that can be transmitted over a given bandwidth per cell per unit area. It, therefore, depends on three key elements, which are data rate [in bps], available bandwidth [Hz/cell] and cell density [cell/area]. The challenges due to the data rate and available bandwidth, in the last decade, have been extensively addressed by implementing efficient modulation and coding schemes, employing multi-antenna techniques or adding new radio spectrum. However, the users at the cell-edge or at hotspot area still experience a low quality of service due to the distance between the eNB and bad channel conditions [24]. To mitigate this, one of the key enabling technologies envisioned by 3GPP in Release 12 and beyond is network densification, discussed next.

2.3.1. Ultra Dense Small Cell Networks

Network densification relates to the dense deployment of low power nodes also referred as small cells, in the existing macrocellular network. The objective is to reduce the distance between the base station and the active user to improve system throughput, spectrum efficiency and network performance especially in those areas where traffic demand is high [25]. An example of network densification is shown in Figure 2.3.

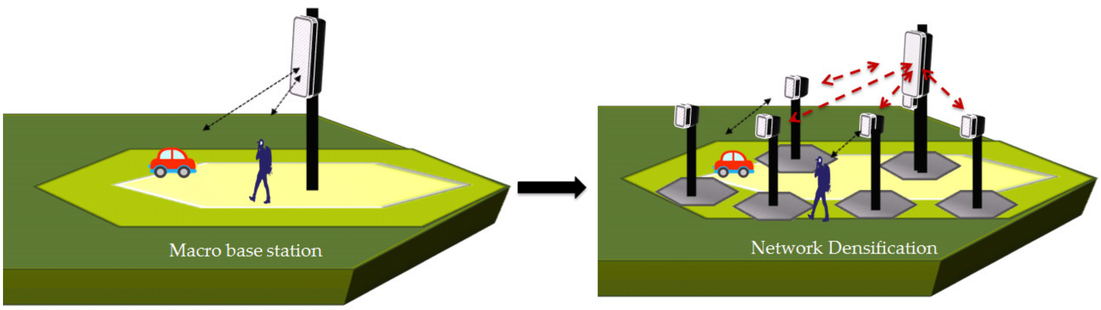


Figure 2.3: An overview of ultra-dense small cell network

A comprehensive survey on network densification is provided in [20], [26], [30], [31]. The authors in [26] defined ultra-dense networks as a network where the number of base stations per unit area is greater than the active user density. Subsequently, in [27], a quantitative analysis of the cell density per unit area for a network to be an ultra-dense network has been presented. In [28], an upper bound for the active user density in dense urban networks was evaluated. Nevertheless, the authors in [29], addressed the issues related to the fundamental limits of network densification as the network densification cannot expand indefinitely. However, further investigations are needed to further understand the fundamental limits explicitly.

Node Types

The small cells are categorized into two categories: (a) fully-functional base stations such as femtocell and picocell, and (b) extension access points such as relays or remote radio head. The fully functional base stations have the ability to perform all the macro base station functions with a limited transmission power. On the contrary, relays and remote radio heads are employed to effectively increase the signal coverage range. The different node types that contribute to the formation of heterogeneous networks are discussed next.

- Picocells are operator deployed small base stations with reduced transmission power and reduced cell size. They are able to perform all the functionalities of the macro base station. They are generally deployed at hotspots or at cell edge area to serve tens of users by offloading their traffic from the macro base station to improve the network capacity. The coverage range of a picocell is to a maximum of 100 meters. In the case of indoor deployment, the transmission power of a typical picocell is less than 100 mW, however, for outdoor deployment, it ranges from 250 mW to about 2 W [30].
- Femtocells belong to the fully-functioning base stations class and are usually deployed indoors by the users depending on individual requirements. The objective of a femtocell is similar to that of the picocell, i.e., to improve spectral efficiency. The coverage range of femtocell is between 10 m to 30 m and the transmission range is less than 100 mW. They are connected to Digital Subscriber Line (DSL), cable or fibre to provide backhaul [31]. Femtocell provides coverage to a small set of users. The users may access the network in three different modes; open, closed or hybrid [21]. The open femto access mode grants access to all the users in the coverage area. However, in the close access sce-

nario, only a defined group of users could access the network. In the hybrid scenario, few of the users are given higher priorities within the specified subscriber group.

- Relays are operator deployed, particularly at cell edge or areas with low or no coverage to improve the signal coverage range. Here, an in-band or an out-of-band air interface spectrum is employed to provide the backhaul for the transmission and reception of the user data from the macro base station. The coverage area, transmission power as well as the access scenario of a relay is similar to that of a picocell.

- Remote Radio Heads (RRHs) are radio frequency units that have similar functionality as relays. However, unlike relays, they are connected to the central base station using a fibre or microwave link [32]. RRHs could be deployed in a planned pattern to form centralized densification that may be an alternative to the distributed densification provided by picocells or femtocells.

Network Architecture

The small cell nodes may be deployed in the macrocellular networks following two different architectures; distributed and centralized. Figure 2.4 demonstrates the two different network architectures.

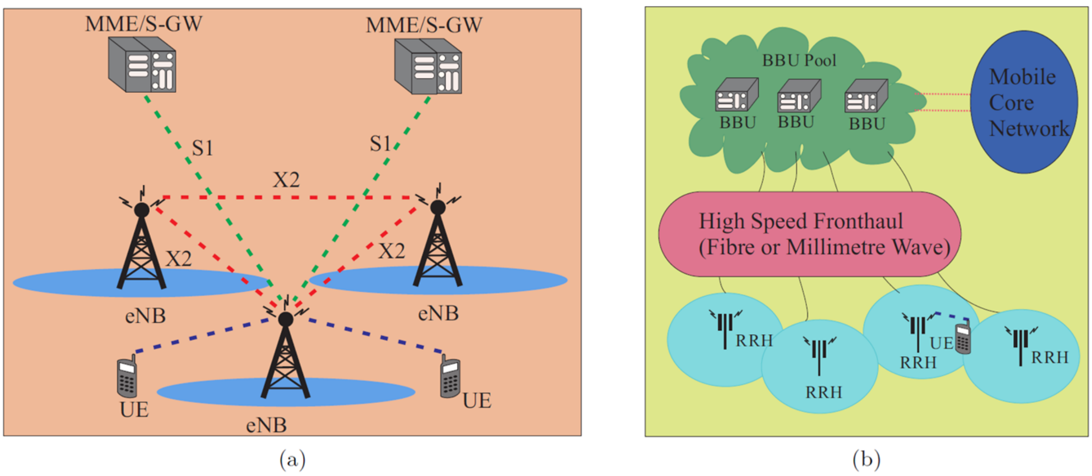


Figure 2.4: Distributed and centralized small cell architecture overview, directly reproduced from [23]

In the distributed network architecture, there is no central unit or controller. The users and the base stations interact with each other using the air interface and are allowed to make autonomous decisions related to network coordination, cooperation and management. Owing to the benefits due to distributed control such as reduced implementation complexity and the signalling overhead, the distributed architecture is suitable for large networks, especially networks involving autonomous nodes [26]. The radio frequency module, power amplifier and antenna module for each base stations in this configuration are located close to its baseband unit for the transmission and reception of signals within its coverage area. A notable example of the distributed architecture is the current Long Term Evolution Advanced (LTE-A) system.

Despite the advantages of distributed architecture, the dense deployment of small cell nodes requires a large infrastructure investment as well as operational expenditure to manage and maintain the network, resulting in high capital expenditure (CAPEX) and operational expenditure (OPEX). Moreover, as the users and the base stations are allowed to make autonomous decisions, effective radio resource management techniques would be required to address the critical issue of interference management. An alternative to the distributed network architecture is the centralized network architecture.

In the centralized network architecture, a central entity or controller governs the network responsibility of coordination, cooperation and management of the base stations under its coverage area. The information related to channel quality, interference level and channel demand is transmitted to the central unit before performing operations related to radio resource management or mobility management [26]. The centralized control based on the comprehensive traffic information has the ability to provide reliable network decisions, hence, resulting in quick convergence and better performance. The baseband unit in the centralized configuration is placed near the central controller. Moreover, a lower value for capital expenditure and operational expenditure in centralized network architecture are expected compared to its counterpart due to the reduced deployment of radio frequency module, power amplifier and antenna module for each base stations in the network. Cloud Radio Access Network (C-RAN) is an example of the centralized network architecture.

2.3.2. User Association Trends

User association is the fundamental mechanism in wireless networks that relate to the association of a user equipment to an appropriate base station for the transmission and reception of data. The user association mechanism with respect to traditional as well as heterogeneous networks has been extensively studied in the last decade [4] [33]-[37]. A comprehensive review of user association algorithms in 5G networks, specifically related to HetNets, massive multiple input multiple output networks (Massive MIMO), millimetre wave networks (mm-wave), and energy harvesting networks is performed by the authors in [7]. In [26], the authors presented an in-depth review of user association techniques in ultra-dense networks (UDN). This paper also discussed different driving factors for UDN along with a diverse range of modelling and interference management techniques. In [34] optimal user association for delay-power trade-off in HetNet with hybrid energy resources is provided. Likewise, energy-efficient user association for heterogeneous cloud cellular networks is provided in [35]. In [36], techniques for optimal user-cell association for massive MIMO networks are proposed. In [37], a comprehensive downlink SINR analysis for flexible user association in HetNet is performed. Whereas, [38] studies user association techniques for load balancing in HetNet.

In traditional cellular networks, user association is performed depending on the measurement of the received signal strength, max-RSS approach. An active mobile device search among the neighbouring base stations for the highest received signal power value using the cell selection criteria, thereafter clamps to the base station from which it receives the max-RSS value [4]. Alternately, user association may also be performed based on the signal quality, max-SINR, as in [33]. In this approach, the UE associates with a base station with the maximum SINR value. This approach is complex to employ in dynamic environments as the interference level varies constantly due to the frequent changes in network due to the changes in temporal-spatial vehicular distribution and active communication traffic load per base station, therefore, evaluation of accurate SINR level becomes uncertain. In addition to this, a problem that closely relates due to vehicle mobility in small cell dense environment is the handover. It is defined as a mechanism in which an ongoing transmission is transferred from one base station to another appropriate base station, as the vehicle moves along the path from the coverage area on one cell into the subsequent cell [11]. The number of handovers per transmission is directly proportional to the size and the number of the small cell deployed and the speed with which the vehicle moves. A vehicle moving at a moderate or high speed in an ultra-dense network will suffer a higher handover frequency compared to conventional macrocellular networks if the conventional max-RSS user association approach is employed [39]. The 3GPP technical report in [39] presents that the handover performance in case of a macrocellular network is significantly better than in HetNet.

A comprehensive survey on mobility and handoff management in vehicular networks is presented in [40]. This survey paper begins by presenting an overview of vehicular networks. Further, the two different mobility management protocols for V2I communications, namely Mobile Internet Protocol version 6 (IPv6) and network mobility (NeMo) are discussed. In [11], an extensive review of the trends in handover design is presented. Similarly, [41] and [42] discusses different handoff strategies and considerations in microcellular systems. A notable work addressing the control of handover initiations in microcellular networks is presented in [43]. A parametric model based on the vehicular traffic theory results for urban and suburban microcellular networks has been proposed in [44]. This model offers an insight into the teletraffic modelling for microcellular environments. Further, in [45], the authors analyzed the impact of user mobility and base station density on the handover rate in a multi-tier HetNet under stationary as well as dynamic conditions. Moreover, a speed-dependent bias factor user association approach to effectively improve the system performance has also been proposed. Nevertheless, the significance of user association in vehicular networks was not highlighted in these works. The development of user association approaches in UDN considering dynamic vehicle mobility is therefore extremely important as user association directly relates to handoff decisions. The classic max-RSS user association approach that is employed for user association in LTE-A is chosen as a baseline approach in the presented work and is discussed in detail in Chapter 4. This research proposes a series of user association approaches in vehicular networks in later chapters of this thesis. The performance metric based user association approaches are proposed in Chapter 4 followed by intelligent user association techniques in Chapter 5 and 6.

2.3.3. Emerging User Association Techniques

Cell Range Expansion

HetNet comprises of a diverse range of access nodes with different transmission power and coverage. Considering the conventional max-RSS user association approach most of the active users will associate with the macrocell base station as the signal associated with the macro base station is significantly stronger than the signal strength from eNB, yielding considerable load imbalance, hence resulting in inefficiency of small cell deployment [46]. To overcome this transmission power imbalance drawback, a load based user association approach is introduced in [47]. This approach is known as cell range expansion (CRE) or biasing - a preferred industry and 3GPP method.

CRE is a sub-optimal technique to increase the system capacity and improve the cell-edge throughput. In this technique, the association of the active users to eNB is achieved by adding an external bias value to the signal strength of an eNB. This bias value depends on a number of parameters such as initial eNB transmit power, the coverage range, the node type, the density of the active mobile users in the area as well as the desired network quality of service.

The authors in [48] investigated the downlink performance of HetNet by employing the CRE approach with a diverse range of bias values under three different combinations of users and eNBs. In this, a lightly loaded control channel transmission subframe technique was used to address the control channel interference issue due to inter-cell interference coordination (ICIC). The simulation results demonstrate that by adding a moderate bias SINR value to small cell base station the user throughput was improved. In [38], a low-complexity distributed algorithm that uses the load-aware user association approach to evaluates the impact of base station density and transmit power on the biasing factor is proposed to improve the throughput gain for cell-edge users. The authors in [49] employed distributed Q-learning algorithm to learn the optimal bias value for the cell range expansion. In this paper, each UE uses its past experience to learn an optimal bias value that would minimize the number of UE outages. The simulation results exhibit a marginal reduction in the number of UE outage, however, the network throughput improved by about 20%.

Coordinated Multipoint Processing

Coordinated Multipoint Processing (CoMP) is a 3GPP standardized user association approach that relates to coordinated transmission or reception of information between an active user and the base station to mitigate inter channel interference, improve throughput and cell edge performance. An extensive survey on coordinated multipoint is presented in [50] [51]. In the downlink, a number of BSs provide coordinated transmission to a single active user employing three different approaches that are (a) Joint Transmission, (b) Dynamic Point Selection and (c) Coordinated Scheduling/Beamforming.

The technique to transmit data by more than one base station using the same frequency and the subframe is termed as Joint Transmission. This approach improves the throughput as well as cell edge performance at the expense of high backhaul bandwidth and low latency requirement. Meanwhile, if the transmission from different base stations is scheduled at the different subframe, the technique is called Dynamic Point Selection. In this approach, the data is transmitted from the same eNB, however, depending on the availability of the resource and the current channel conditions, the serving eNB changes dynamically. In the Coordinated Scheduling/Beamforming (CB/CS) approach, only the the channel state information (CSI) is shared among the cooperating eNBs. The CB/CS technique reduces multi-user and multi-cell interference and requires lower bandwidth compared to joint transmission.

Coordinated Scheduling/Beamforming and Joint Reception are the CoMP approaches for Uplink [52]. Here, the active user is associated with multiple eNBs in the cluster leading to multi-cell association. The user scheduling and processing is performed jointly by the multiple coordinating eNB, however, only one eNB receives the user data. A comprehensive survey on CoMP clustering schemes is presented in [52]. The authors in [53] proposed a low-complexity joint reception CoMP algorithm combined with an effective antenna selection technique to improve cell-edge gain. Further, a comparative performance analysis of intra-site joint reception CoMP for LTE-A system with single-cell single user and multi-user MIMO for an uplink was performed. The disadvantage of the CoMP approach for uplink is increased complexity, high backhaul bandwidth and low latency requirements.

Dual Connectivity

Dual connectivity (DC) user association approach is introduced by the 3GPP in release 12 [54]-[56]. The objective is to increase the per-user throughput and improve the mobility performance in HetNet, by allowing the users to simultaneously associate with a macro BS and multiple eNBs. The macro BS and eNBs is connected together through the X2 interface and operate at different carrier frequencies. In this approach of user association, the control plane and the user plane are split. The control plane is responsible for the system information transmission and handles user connectivity. Whereas, the user plane manages the user data. Figure 2.5 presents dual connectivity architecture for HetNet.

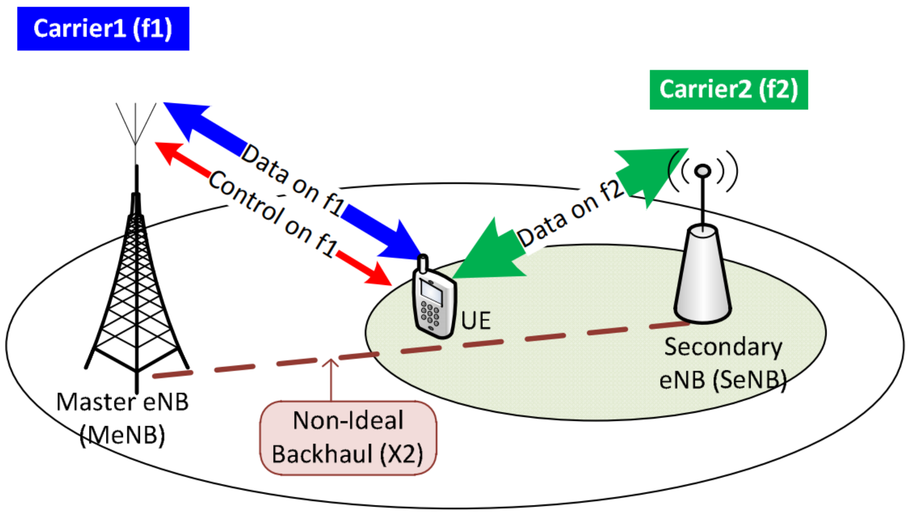


Figure 2.5: The dual connectivity architecture, directly reproduced from [57]

In [57], the authors present a detailed overview of dual connectivity in LTE. Further, potential benefits of dual connectivity in uplink and downlink such as an increase in per-user throughput, load balancing, better resource utilization, and mobility management were discussed. Moreover, in [58], a low-complexity user association model with dual connectivity and constrained backhaul was proposed to maximize the sum-rate of all the users. The benefits of DC are two folds: (a) the user is able to use the radio resources across the two base stations that lead to an improved spectrum efficiency (b) the mobility resilience is improved as a user does not has to initiate handover until it is in the coverage area of the associated macro base station.

2.3.4. Challenges and Limitations

Section 2.3 addresses network densification as one of the major breakthroughs for 5G cellular networks that would likely address the limitations of coverage, capacity, high data rate and low latency to the end users. However, new challenges due to heterogeneity and dynamic node mobility will emerge; some of them are now discussed.

Mobility Management - One of the dominant challenge associated with dynamic environments is mobility. The mobility management in traditional wireless networks depends on the base station location areas that are statistically configured. However, in the case of an ultra-dense network defining the cell boundary for different vehicle speed becomes difficult. Thus, it is crucial to modify the current static planning to dynamic cooperation so as to gain the neighbourhood knowledge for better user association. Chapter 4 proposes a three-step sequence user association approach wherein few nearest neighbours to the moving active user are identified. Following this, the proposed scheme, depending on diverse performance metrics, provide user-centric services such as adaptive user association, resource management, mobility and handover management.

Interference Management - Network densification significantly increases the interference level compared to a macro-cellular network. This is due to the presence of greater number of interference sources, such as eNB and active user equipment per unit area, as well as significantly strong interference from line of sight component. Interference management in ultra-dense dynamic environments is indispensable as it substantially affects the network performance and user experience. Therefore, the presented research work proposes advanced, machine learning inspired algorithms with distributed control for user association in dynamic environments that demonstrate ability to strike a trade-off between interference level and network performance at different traffic conditions.

Backhauling - It relates to the transmission of data from an eNB to the core network, either through a wired optical fibre link or by a wireless link. The provisioning of an ideal backhaul in a dense small cell environment in a rapidly changing radio environment appears to be quite challenging. The wireless backhauling techniques including the mm-Waves link, relay link and massive MIMO link appears to be a more popular alternative to wired backhauling in densely deployed networks [58] [59]. In the presented work, perfect backhauling is considered. However, there should be considerable investigation that should be done on backhaul configuration for vehicular networks as it predominantly impact the access network performance.

Network Configuration, Co-ordination and Management - In HetNets, enormous amount of data will be collected by each eNB to monitor the network performance and maintain network stability. This will intensify the complexity to configure and maintain the wireless networks. In the conventional 4G LTE network architecture, services such as service control and mobility management function are centralized in the core network. However, a centralized control in HetNet will lead to higher signalling load and a longer end-to-end delay. Therefore, it is essential to develop intelligent, distributed algorithms, as proposed in Chapter 5 and Chapter 6, to perform flexible networking thus support high throughput, network QoS and resource utility.

Energy Efficiency - Despite the deployment of low powered eNB, the aggregate power of an ultra-dense small cell network may be extremely large, resulting in a high operational expenditure. Energy efficiency relates to the number of transmitted bits per unit area. The variation in energy efficiency directly impacts the interference level, the link quality and eventually the overall network QoS. Therefore, the development of self-organizing energy efficient algorithms that considers user experience and network performance in dense networks would yield interesting results.

One possible way to mitigate these challenges is the *introduction of intelligence* in the network. This will enable the network not only to take autonomous decisions but also to learn and improve the network performance based on previous decisions. The next section provides an overview of machine learning, in particular, reinforcement learning. Following this, a range of reinforcement learning techniques that are widely used in both the wireless and artificial intelligence domain are discussed. Furthermore, these machine learning approaches are employed to develop intelligent user association approaches that are proposed later in this thesis.

2.4 Machine Learning

Machine learning (ML) is a field of artificial intelligence concerned with the development of algorithms that converge to an optimal solution and improves the system performance without any human intervention. The ML paradigm is broadly classified into three different categories; supervised, unsupervised and reinforcement learning [15] [100]-[102]. Supervised learning relates to learning from training sets, examples and/or instances provided by an external knowledgeable source. The objective of supervised learning is to train the model such that it is able to generalize or map its responses correctly in situations other than situations in the training set. Few popular supervised learning techniques are Baye's theory, k-nearest neighbour (k-NN), and neural network (NN). In the case of unsupervised learning, the learning agent has to uncover the correct behaviour, pattern or hidden structure in the collection of unlabelled data. K-Means, self-organizing maps (SON), feedback and fuzzy controllers are few examples of unsupervised learning. Lastly, reinforcement learning (RL) is a goal-directed learning technique wherein the decisions are learnt by recursively interacting with the environment whilst the reward mechanism is employed. The most widely used reinforcement learning techniques in artificial intelligence domain as well as in wireless networks are Q-learning and state-action-reward-state-action (SARSA).

2.4.1. Reinforcement Learning

Reinforcement learning (RL) is a branch of artificial intelligence, a class of machine learning, that employs a reward and punishment policy to enable an agent to learn a solution to a decision problem by interacting with its environment purely through trial-and-error such that the overall reward value is maximized [100]-[102]. Unlike other learning techniques, RL focuses on a goal-directed learning, therefore, depending on the consequences of the learnt action a reward is awarded to the learning agent in case of successful attempts else it is punished. The key merit of RL is its ability to learn a solution without any prior knowledge of the environment or the reward function. However, one of the challenges in RL is the trade-off between exploration and exploitation. A learning agent aims to maximize the reward by effectively employing an action that has proven promising in the past. But, to discover such an action, the learning agent has to try each available action. Therefore, the task of a learning agent is to *explore* all the available actions in order to learn and subsequently *exploit* the most efficient action in the future.

Figure 2.6 shows a basic diagram of the RL. The four key elements of RL identified in the literature are a) a policy, b) a reward function, c) a value function, and d) a model.

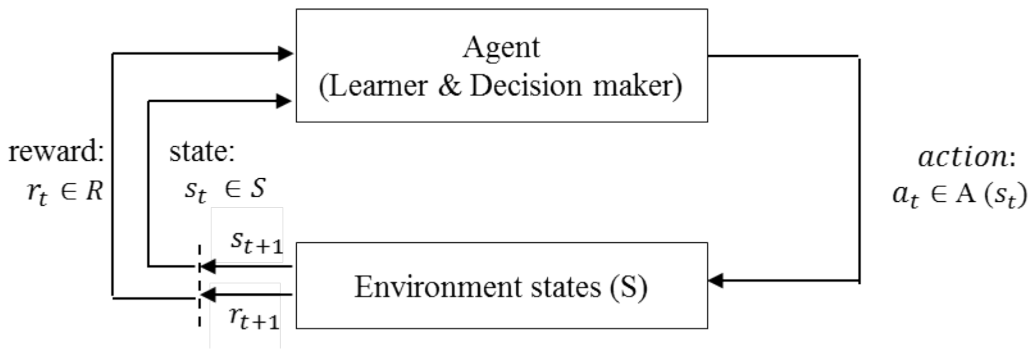


Figure 2.6: A fundamental block diagram of reinforcement learning [100]

- A *policy* relates to a learning agent's way to map the perceived *state* of the environments to *action* that should be taken in them at a given time. This is also termed as *policy function*. The RL agent interacts with its environment in discrete time steps. At each learning time step, the agent is in some state without having any information about the environment and selects an action purely through a trial-and-error approach to move to the next state. The outcome assists to formulate or to reconstruct the current policy.

- For each learnt action, a learning agent receives a *reward* depending on the consequence. A reward value received by an agent assists it to access the current choice of action and understand whether the action learnt was good or bad. The objective is to maximize the total reward it receives over the long run, therefore, if a low reward is received for the learnt action the learning agent reconstructs its current policy accordingly such that an appropriate action that leads to a higher reward is selected in that state in future.

- The *value function* indicates what may be an appropriate action in the long run, unlike, a reward function that assesses and informs the nature of a learnt action at every time step. The value of a state or a state-action pair represents the total amount of the reward an agent may expect to receive over the future if it starts from that state. The value function plays a substantial role during the learning process as it attempts to learn an action with maximum value rather than an action with the highest reward. The value function is also referred to as value table. However, in the case of Q-learning, it is termed as Q-table or Q-value. The value of a particular action a in state s is evaluated using the equation below

$$V^{\pi}(s,a) = E_{\pi}\Big\{\sum_{k=0}^{+\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a\Big\} \tag{2.1}$$

  where, r_{t+k+1} is the reward received, γ is the discount factor in the range [0,1] and E_{π} is the total expected reward under the policy π for the considered state-action pair.

- Lastly, *a model* is the representation of environment dynamics. It presents the correlation between the different RL elements that include state, action, reward and transition probabilities between state, action and future state. The class of RL that requires well developed mathematical model demonstrating an accurate and complete relationship between different elements is classified as model-based RL. This category of RL problems involves validation of results using the popular Bellman optimality equation. An example of model-based RL is the Dynamic Programming [103]. The inherent disadvantage of model-based learning is the growth in computational requirement which increases exponentially with an increase in the number of state variables; however, it is one of the most widely used techniques for solving conventional stochastic optimal control problems. Alternatively, the Monte Carlo RL and the Temporal Difference RL method do not require a model to evaluate the policy and are referred to as the model-free RL method.

2.4.2. Model-Based Reinforcement Learning

Model-based RL is an approach to solving the RL problem in order to compute optimal policy or a solution, given transition probability matrix (TPM) and transition reward matrix (TRM). The role of TPM is to specify the probability of being in a certain state, execute a particular action and progress to another state. While, TRM express the current or immediate reward received after a particular state-action transition is performed. The learning agent attempts to construct the model of the environment, by including a collection of RL elements such as state, action, reward and transition probability. The model of the environment usually presents a finite Markov decision process (MDP). Based on the transition taken by the learning agent from the current state to a future state, a probabilistic reward is received. A policy is then computed using dynamic programming (DP) algorithm to select the most appropriate action in the current state of the environment. In dynamic programming algorithms, an optimal policy is learnt after performing a series of policy iteration and policy improvement on the current policy. Policy evaluation relates to the assessment of the current policy, whereas, policy improvement reconstructs the current policy depending on the assessment. The primary objective of dynamic programming is the use of value function to learn good policies by solving the recursive Bellman optimality equation, given below:

$$Q^*(s,a) = \sum_{s'} P(s,a,s')[R(s,a,s') + \gamma \max_{a'} Q^*(s',a')] \tag{2.2}$$

where $Q^*(s,a)$ is the cumulative reward of taking an action a in state s. $P(s,a,s')$ is the probability of transition from current state s to a future state s' after selecting an action a. $R(s,a,s')$ is the expected immediate reward. γ is the discount factor that controls the importance of future rewards with respect to immediate rewards. $\max_{a'} Q^*(s',a')$ is the maximum Q-value which maximizes the reward for each state-action pair. It is derived by employing the greedy policy as follows:

$$\pi(s) = arg \max_{a} Q^* s, a \tag{2.3}$$

2.4.3. Model-Free Reinforcement Learning

An alternative to model-based learning is the model-free RL [100]. In this technique, a learning agent without any prior knowledge of the transitional probabilities or reward function learns a policy directly from the raw experience gained through recursively interacting with the environment. Monte Carlo (MC) methods and temporal difference (TD) learning are the two classes of methods under model-free RL. The Monte Carlo (MC) method updates the value functions episode-by-episode rather than step-by-step i.e., the value is updated only after an episode ends. The value function at a particular state is updated using the equation below:

$$V(s') \leftarrow V(s) + \alpha[G_t - V(s)] \tag{2.4}$$

where, $V(s')$ is the new value estimate, $V(s)$ is the old value estimate, α is a constant learning rate, G_t is the actual return received after the episode ends at time t. Moreover, an optimal policy is learnt by recursively performing policy evaluation and policy improvement to a fixed arbitrary policy π as:

$$\pi^0 \xrightarrow{\text{E}} Q^{\pi^0} \xrightarrow{\text{I}} \pi^1 \xrightarrow{\text{E}} Q^{\pi^1} \xrightarrow{\text{I}} \pi^2 \xrightarrow{\text{E}} \cdots \xrightarrow{\text{I}} \pi^* \xrightarrow{\text{E}} Q^* \tag{2.5}$$

where $\xrightarrow{\text{E}}$ denotes policy evaluation and $\xrightarrow{\text{I}}$ is policy improvement; making the policy greedy with respect to the current value function. The initial policy and Q-value are presented by π^0 and Q^{π^0} whereas π^* and Q^* relates to the optimal policy and Q-value.

Another classification for solving an RL problem is TD learning. It is a combination of DP and MC method. The TD methods (a) does not require any prior knowledge of the model of environments dynamics or reward function to learn the solution to a problem, similar to MC methods, and (b) the learning agent utilizes the value update rule every time step to learn a policy in order to select its next action to reach the goal state, as in DP. The value at a particular state is evaluated by using the equation below:

$$V(s') \leftarrow V(s) + \alpha[r_{t+1} + \gamma V(s') - V(s)] \tag{2.6}$$

where $V(s')$ is the new value estimate, $V(s)$ is the old value estimate, α is a small positive value known as the learning rate which influences the learning process. Since the learning is based on the difference of values, i.e., $[V(s') - V(s)]$, therefore this technique is also referred to as temporal-difference learning method. The TD learning methods are further classified as on-policy and off-policy TD control methods. The most widely used on-policy TD method is SARSA [104] while Q-learning [105] is the most popular off-policy TD method.

SARSA: On-policy TD Control

The objective of an on-policy TD control method is to learn an action-value function rather than a state-value function. In particular, the learning agent estimates the value following a policy for the chosen state and action at every transition in time. It then recursively updates the current policy towards an optimal policy by performing policy evaluation and policy improvement. As seen in Figure 2.7, the algorithm follows a particular sequence of RL elements (state, action, reward, state, action), to learn a policy or value function, giving rise to its name SARSA. The value function, under the current policy π in SARSA is updated at every time step using the following update equation:

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma Q'(s',a') - Q(s,a)] \tag{2.7}$$

where, $(Q(s,a)$ corresponds to the Q-value of the current state-action pair, α is the learning rate, γ is the discount factor, r is the reward received at each instance of time, and $Q'(s',a')$ is the Q-value of the previous state-action pair. The general framework of SARSA algorithm is shown in Figure 2.7

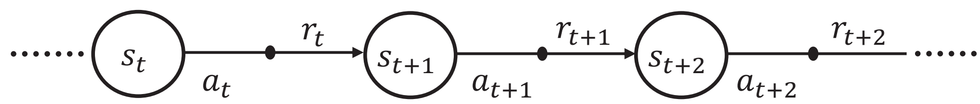


Figure 2.7: Work flow diagram presenting different stages in SARSA learning approach, redrawn from [140]

Q-learning: Off-policy TD Control

Q-learning (QL) proposed by Watkin is one of the most popular RL techniques in widespread use in wireless and artificial intelligence domain [105]. Moreover, it has been predominantly employed to develop the proposed intelligent user association techniques in the presented research work. The learning agent recursively interacts with the environment purely through trial-and-error iterations to gather the information related to the environment and learn the most appropriate solution. The key difference between Q-learning and SARSA is the action selection. In Q-learning, the learning agent in a particular state chooses an action based on greedy policy i.e., it selects the action with maximum Q-value. However, in the case of SARSA, the action selection depends on the current policy. The policy is improved gradually by performing policy evaluation and policy improvement repeatedly, thus, leading to the selection of an optimal action.

The learning agent in Q-learning uses a *learning policy*, such the ϵ-greedy policy, to learn an action with maximum value. Subsequently, the Q-value associated with each action is recursively updated using the following update equation:

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q'(s',a') - Q(s,a)] \tag{2.8}$$

where, a is the action taken in the current state s, s' is the next state, a' is the action that can be taken in the next state s, $Q(s,a)$ corresponds to the Q-value of the current state-action pair. The learning rate parameter, $\alpha \in [0,1]$ controls the convergence rate. The discount factor $\gamma \in [0,1]$ controls the importance of future rewards with respect to immediate rewards. r represents the reward value that is awarded for the learnt action, and $\max_{a'} Q'(s',a')$ is the maximum Q-value among the available actions in the next state s'.

In some of the learning problems the environment need not be represented by different states. The objective is to learn an appropriate action value by following a policy π. These learning problems are categorized as stateless learning or 1-dimensional learning problems [106] [107] . The Q-value associated with each action in stateless learning is updated using the following equation:

$$Q(a) \leftarrow (1 - \alpha)Q(a) + \alpha[r + \gamma \max_{a'} Q'(a')] \tag{2.9}$$

where, a is the action taken, $Q(a)$ corresponds to the Q-value of the current action. The learning rate parameter, $\alpha \in [0, 1]$ controls the convergence rate. r is the reward awarded for the learnt action a. The discount factor $\gamma \in [0, 1]$ controls the importance of future rewards with respect to immediate rewards. $\max_{a'} Q'(a')$ is the maximum Q-value among the available actions in the action space.

2.4.4. Multiagent Reinforcement Learning

Multiagent RL (MARL) is an extension to the traditional RL and relates to learning with multiple independent agents in the same environment. One of the prominent applications for multiagent systems is stochastic games [108] - [110]. Individual learning agents, within a similar task, co-operate and share experiences to learn faster and a better decision policy. The authors in [111] provide a taxonomy of MARL algorithms based on different types of stochastic games such as fully cooperative, fully competitive and mixed. Although due to the availability of more resource and the joint reward maximization, multiagent learning outperforms single agent learning, however, the need of coordination between the individual agents and growth of state-action pair for each new agent increases the computational complexity exponentially [112].

A popular way to specify a MARL goal and learn an optimal policy in the MARL environment is by the use of Nash equilibrium. Nash equilibrium is defined as a stable state of a system wherein multiple agents interact and update their individual policies such that the stability, convergence and reliability of a learnt solution in static as well as in stochastic game environments are maintained [113]. In [114], the authors investigated the reinforcement learning policy in the framework of stochastic games using the variable learning rate policy. The proposed approach demonstrated to overcome both of the shortcoming, rationality and convergence that were presented by conventional learning approaches in a multiagent environment. Moreover, the influence of learning rate variations on algorithms performance is also presented. Similarly, in [115], the authors proposed a model-free, distributed Q-learning algorithm for cooperative multi-agent decision processes. Experimental results show that the proposed technique was able to learn a rational policy based on the value function update in addition to coordination between the agents in the environment.

A recent noticeable contribution of application of reinforcement learning in a multiagent vehicular environment is presented by the authors in [116]. The authors proposed a distributed resource allocation approach for a vehicle to vehicle (V2V) communication by employing deep reinforcement learning. The preliminary results demonstrate the benefits of deep reinforcement learning to learn optimal spectrum and power for transmission whilst minimizing the end-to-end latency on V2V links and the interference to the vehicle to infrastructure (V2I) link. Another challenging application of multiagent reinforcement learning in vehicular networks is performed in [117]. The authors employed a conventional Q-learning approach in combination with a feed-forward neural network for scheduling traffic signals on a five intersection traffic network scenario. The results demonstrated minimization in the average queue delay across the intersections and reduced congestion in high traffic scenarios.

In traditional machine learning, the learning agent uses the same set of input features and data distribution to incrementally learn an appropriate policy. However, in certain scenarios where the learning domain changes, the learning agent rebuild its knowledge base from the beginning using the trial-and-error interaction process. This process of rebuilding the knowledge base is time-consuming, inefficient and expensive. Therefore, the state-of-the-art transfer learning techniques, discussed next, seeks to leverage previously learned knowledge from one domain into another domain thus improving the learner's performance.

2.4.5. Transfer Learning

Transfer learning is a machine learning approach that aims to employ the learnt knowledge from a source domain to a target domain in order to improve the learning performance of the target learner [118]. Transfer learning has proven to be a promising technique in cellular systems for topology management [121]-[123], caching [124], radio resource management [125], load and energy optimization [126]. An extensive

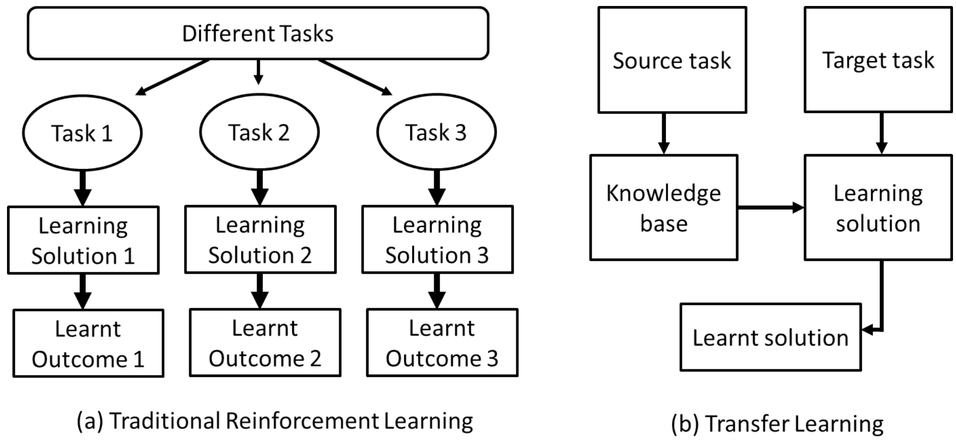


Figure 2.8: Figure demonstrating the traditional reinforcement learning and transfer learning procedure

survey of transfer learning techniques and their application to different real-world applications is presented in [118]-[120]. In [121], the authors investigated the use of transfer learning for topology management in a 5G aerial-terrestrial broadband access networks. The knowledge gained using conventional RL technique at link level for spectrum assignment was transferred using TL approach to build the knowledge base for user association. The system level simulation result demonstrated that the transfer learning-based user association approach was able to improve the energy consumption by (30 - 60) % compared to conventional max-RSS user association or QoS aware user association technique. Figure 2.8 demonstrates the two different learning processes; traditional reinforcement learning and case-based transfer learning.

The authors in [124] proposed a transfer learning algorithm to strategically cache contextual information at the network edge in 5G wireless networks. The contextual knowledge such as user's content viewing history, social browsing, etc. was gathered during device-to-device interaction and was regarded to as source domain knowledge. This gained knowledge was transferred at each small cell eNB to optimally cache strategic content. The results showed a significant gain in user QoE and backhaul offloading. Likewise, in [125], the authors propose transfer learning based radio resource management approach, wherein, by sharing the macro users scheduling information between small cell eNBs the excessive interference generated for macro users in future instances in a multi-user OFDMA networks was minimized. In [126], transfer learning is employed for cell selection to improve QoS, load balancing and energy efficiency in an opportunistic mobile broadband network by utilizing the knowledge learnt from resource allocation using conventional Q-learning approach. System-level simulations show that application of transfer learning to cluster UEs on selected eNB with better QoS reduces the energy consumption at medium traffic level. Transfer learning approach is exploited in Chapter 6 to improve the user association in vehicular communications. It has been demonstrated through a series of simulation experiments that the learning agent learns faster, converge to a better, consistent and a reliable solution with the application of transfer learning.

2.5 Conclusion

User association is a fundamental task in the wireless networks. It plays a substantial role to improve the spectral efficiency, end-to-end latency and overall network QoS in vehicular networks. The conventional max-RSS user association techniques associates the active user with the maximum signal strength base station. However, if employed in ultra-dense dynamic vehicular networks, it will lead to a range of challenges related to mobility, poor network performance and significantly high switching and signalling load. To reinforce and strengthen this argument, this chapter began with an overview of vehicular networks followed by an in-depth discussion of the role of mobility and network simulator that are essential for accessing the developed algorithms empirically. Moreover, current mobility and network simulators were introduced. A comprehensive description on mobility model framework that has been used as a guideline to develop the vehicular simulator in this work was presented thereafter. Next, a detailed overview of ultra-high capacity wireless networks with emphasis on dense small cell networks, conventional and emerging user association techniques was provided. Subsequently, the challenges related with dense deployment of small cells in vehicular environments were highlighted. This chapter also provided an overview of a range of single-agent and multi-agent reinforcement learning techniques that are found in literature specifically those employed in the wireless communication. Further, the chapter highlights the significance of intelligent user association techniques in regards with network performance and user experience in dynamic environments.

Chapter 3. System Modelling and Performance Evaluation Methodology

Contents

3.1 Introduction

This thesis proposes a number of intelligent user association approaches that aim to strike a balance between the number of handovers per transmission and system performance metrics whilst a guaranteed network quality of service (QoS) is delivered. In order to evaluate the empirical properties of the proposed approaches, a model that could accurately represent an ultra-dense small cell vehicular network is absolutely important. This chapter presents the modelling techniques used to develop the simulation model and key performance metrics utilized for assessment of these policies.

Considerable effort has been made to develop a sufficiently complex event-based system level simulator. It is an integration of a dynamic vehicular network, a highway, and an ultra-dense small cell network. It considers relevant aspects of LTE simulation, such as multi-cell environments with uplink flows, user mobility, handover procedures, cell planning, scheduling, interference calculation, and QoS management in a dynamic environment. Moreover, the developed simulator enables the analysis of the influence of user mobility on network performance across different mobility levels.

The simulator has been written in MATLAB, as it provides powerful matrix calculation functions as well as effective ways to produce graphical results. Moreover, programming and step-by-step debugging functionality in MATLAB provides more transparency and flexibility to model the environment at a system level. The other network simulators that could be used potentially include OPNET, NS2, NS3 for network communication and SUMO to obtain vehicular traces, as discussed in Chapter 2. However, these simulators are domain specific modelling simulators developed by domain experts, therefore, require careful integration to provide effective ways to evaluate proposed protocols and overall system performance.

The developed simulation model allows system and user experience to be modelled and serves as the firm foundation for the main research problem. At the beginning of a simulation, the event list is generated that includes the arrival time of vehicles (UEs) and data packets (files). During the simulation experiment, the event list is populated with more events such as vehicle mobility, vehicle location update, cell association, link performance evaluation, resource assignment, handover scheduling, etc. The events in the event list are sorted in an ascending order according to their timestamps, for an accurate and absolute execution.

The developed framework with multi-users, constantly varying radio environment and rapidly changing spatial-temporal vehicle distribution established an extremely complex scenario to be evaluated analytically. Therefore, extensive system modelling tasks are performed to assess and validate the proposed algorithms statistically. The next section presents the scenario and the network architecture.

3.2 Scenario and Network Architecture

The scenario is illustrated in Figure 3.1. Here, an ultra-dense small cell network is deployed over a linear urban stretch, a highway, to provide the UE on the move with ultra-reliable low latency communications as well as high capacity density. A user associates with the most appropriate base station for data transmission based on a variety of association strategies.

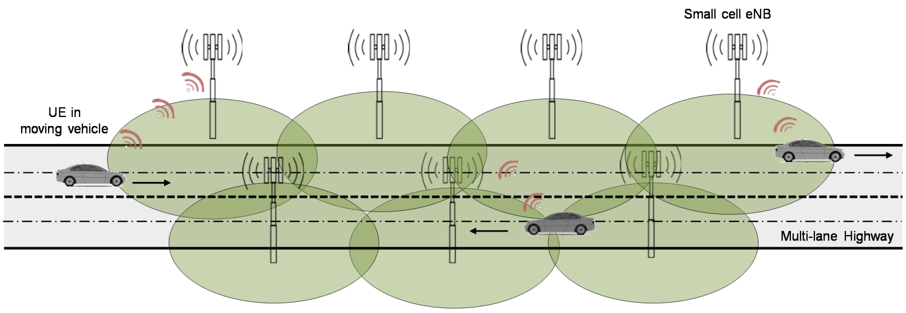


Figure 3.1: Ultra dense small cell vehicular scenario

Figure 3.2 illustrates the system architecture of the ultra-dense small cell vehicular network. It comprises two networks, a dynamic vehicular network with vehicles moving along the highway in a dedicated direction and an ultra-dense small cell network deployed over the service area in order to deliver improved coverage and capacity.

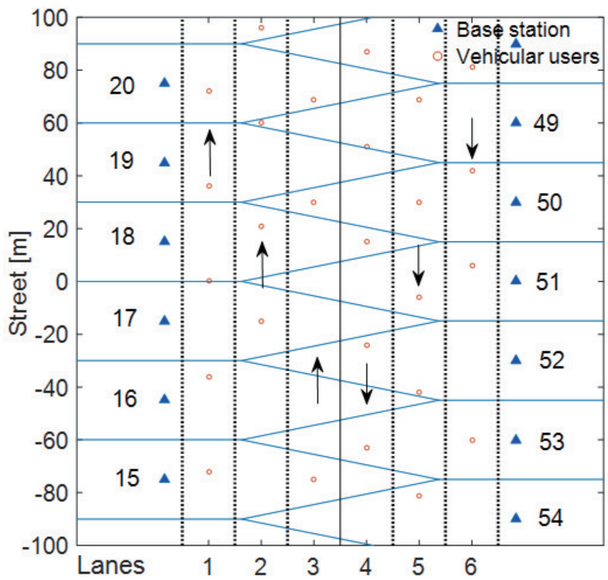


Figure 3.2: Ultra-dense small cell vehicular network architecture

3.3 System Models

This section presents the different modules employed in developing the simulation model that is exploited to empirically assess the adaptability and robustness of the proposed user association algorithms. As shown in Figure 3.3, the first is the architecture module. This includes the system scenario and the network architecture that reflects the distribution of the transmitters (UEs in the moving vehicles) and receivers (stationary eNBs) in an ultra-dense small cell network. The next is the traffic module that represents the traffic characteristics that relate to vehicle mobility as well as the radio communication traffic.

The radio propagation and mobility management module evaluates the attenuation of the radio signals and selects an appropriate link accordingly. The link module includes the radio resource allocation model as well as the channel capacity evaluation model. The radio resource allocation model is responsible for resource assignment to UEs for data transmission. The channel quality and channel capacity are evaluated in the channel capacity evaluation module for link level performance.

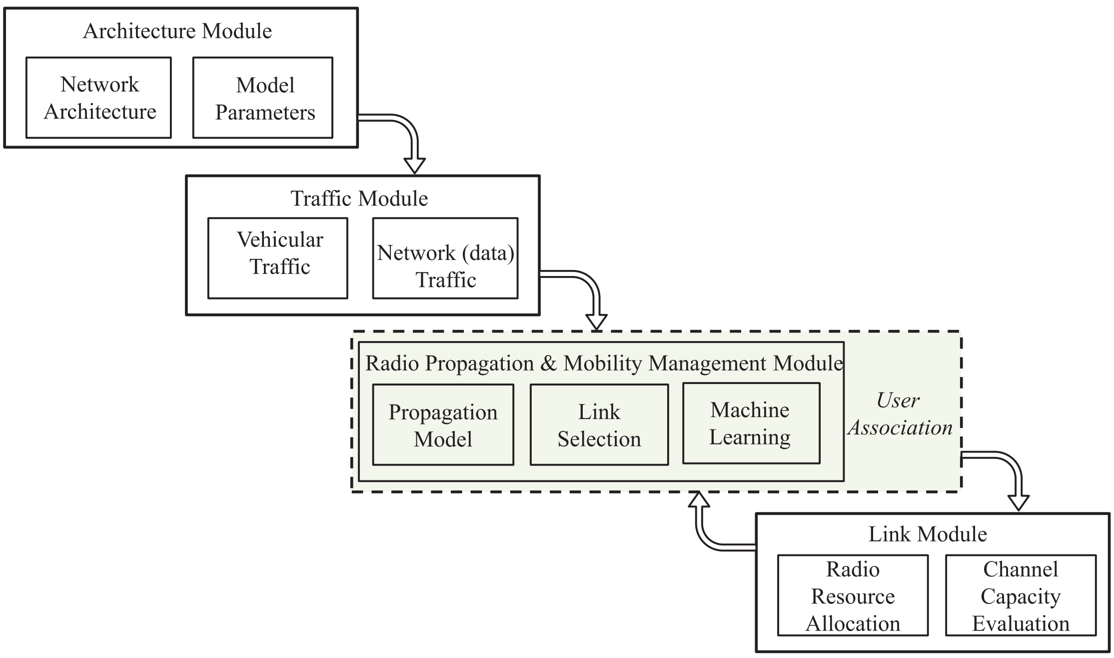


Figure 3.3: A reference model of the developed simulator

3.3.1. Architecture Module

Ultra-Dense Small Cell Model

The access network is constructed using a dense deployment of small cell BSs (eNBs) below the rooftop level of the surrounding buildings on roadside infrastructure such as on lamp posts, with omnidirectional antennas implemented. These lamp posts are assumed to be uniformly distributed on both sides of the street with an inter-site distance of approximately 30 m and height of 10 m.

The system architecture is derived from the FP7 BuNGee project architecture that aimed to provide a capacity density of 1 Gbit/s/km^2 in an urban area, as proposed in [128]. The eNBs are portable, low powered, light-weight devices that can be densely deployed as well as easily managed. The main goal of the ultra-dense deployment is to provide an extremely high data rate and ultra-low latency communication network to users on the move at all mobility levels. In the next section, a comprehensive description of the dynamic vehicular network model is presented.

Dynamic Vehicular Model

Road traffic is a complex multi-agent system in which the agents, vehicles, may or may not interact with each other. The way these agents move depends on different aspects of traffic flow operations, network topology as well as traffic flow modelling. A comprehensive review of vehicular mobility models and traffic flow is provided in Section 2.3.2. In the presented work, a multiple lane, highway model with a bi-directional vehicular flow is considered as the vehicular traffic model. The maximum number of vehicles in the system is assumed to be 60$\%$ of the maximum density at a particular mobility level. This is because the communication system saturates on assuming higher percentages. To understand the influence of vehicular mobility on network performance, all vehicles move at a constant speed whilst maintaining a safety distance from the vehicle in front [129]. This safety distance is assumed to be the maximum of the distance travelled by a vehicle in $2s$, or $1m$. The length of all vehicles is assumed to be 5m and the highway is 1020 m long. Vehicle speed ranges from 10

Table 3.1: Vehicular traffic model parameters

Parameters	Value
Grid layout	Linear
Simulation area	20x1020 m
Number of Base station	68
Number of lanes	6
Lane width	3 m
Pavement width	1 m
Cell size	30 m
Vehicle length	5 m
Vehicle width	2 m
Base station height	10 m
Mobile station height	1.5 m
Vehicle arrival rate λ_v	6 vehicle/s
Vehicle speed	[10-60] km/hr
Safety distance	$\max(2srule, 1m)$
Vehicular traffic model	Poisson distribution with negative exponential inter arrival time
Number of vehicles	60% of the max capacity at a particular mobility level

km/hr up to 60 km/hr. Furthermore, only one UE per vehicle is assumed. The different roadway physical characteristics as in [130] are listed in Table 3.1.

A vehicle arrival is modelled as a Poisson process with a constant mean arrival rate of λ_v (vehicle per second) and inter-arrival time following a negative exponential distribution. Upon vehicle generation, a lane is selected based on a random uniform distribution. Corresponding to each lane is a specific mobility direction. As seen in Figure 3.2, vehicles in lanes 1, 2, and 3 moves in a northerly direction while those in lanes 4, 5, and 6 travel south. Vehicle locations are constantly updated every t seconds using:

$$pos_i(t+1) = pos_i(t) + v_i(t) \times t \tag{3.1}$$

The index i is the i^{th} user, $pos_i(t+1)$ is the updated vehicle position. This depends on $pos_i(t)$, the current position of a vehicle and the distance travelled by it in t seconds. $v_i(t)$ is the speed of i^{th} vehicle at time t. The value of t is assumed to be equal to the time taken to traverse a distance equal to one-tenth of the cell size. The dynamic time update assists in maintaining consistency while updating the vehicle location leading to an appropriate evaluation of the number of handovers per transmission using different approaches. For example, if a static update time equal to 2s is assumed, then the next updated location for a vehicle moving at 60 km/hr would be two cells from the current cell which would result in a miscalculated number of handovers per transmission. A static update time assumption in the considered scenario will hinder the understanding of the influence of vehicle mobility on the system performance as well as affects the network performance evaluation.

To avoid border effects due to a limited simulated area, vehicles select a new lane following the uniform distribution to wrap around at the edges of the simulation area, i.e., a vehicle that leaves the service area from one edge of the model re-enters it from a randomly selected lane. The wrapping model ensures that the mobility pattern observed at the border cells is similar to that in the middle of the street. The flowchart in Figure 3.4 illustrates the vehicle generation, mobility pattern, and wrapping process in the considered scenario.

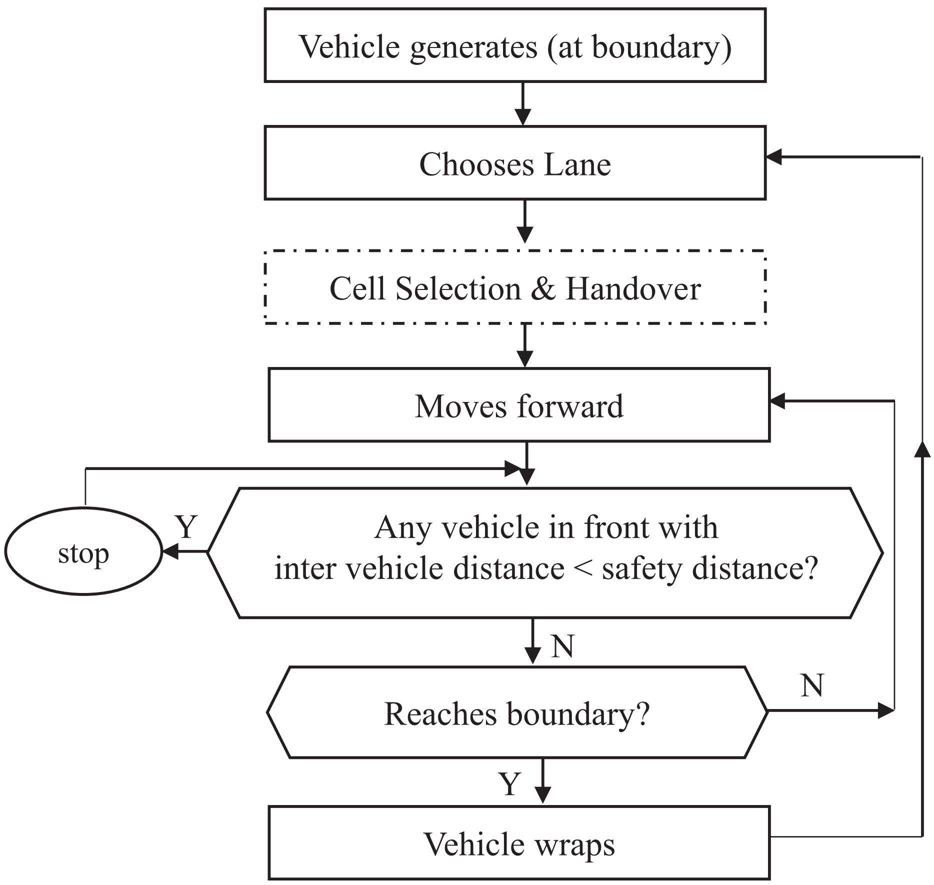


Figure 3.4: Flowchart presenting vehicle generation, mobility pattern and wrapping processes

3.3.2. Traffic Module

This module relates to the traffic characteristics of the network and models the behaviour of the traffic across the wireless network. Here, the mobility traffic is generated as explained in Section 3.3.1, whereas, the wireless traffic can be modelled as a call based service (circuit switching) or internet data at a session, burst or packet level. The file arrival is modelled as a Poisson process with a constant mean arrival rate of λ_f (files per second) and a fixed file size of 2MB. The inter-arrival time follows a negative exponential distribution [131]. The random traffic generated in bursts thus reflects the typical behaviour of mobile network traffic. The service time of each transmission depends on the channel quality. It is calculated by dividing the file size by the achieved link throughput.

3.3.3. Radio Propagation Module

The radio propagation model relates to the attenuation of radio signals. Radio signals, which propagate over a wireless channel, usually experience deterioration in signal strength due to three main effects: path loss, shadowing and multipath fading. The path loss is the attenuation of the transmitted power and is directly proportional to the distance between the transmitter and the receiver. Shadowing is the attenuation due to phenomena such as reflection and diffraction. Multipath fading is the fluctuation in received signal strength as a result of the transmitted signal arriving at the receiver through different paths with different attenuation and delay. Path loss and shadowing are large-scale effects and remain constant over time. Multipath fading and Doppler effect relates to small scale fading and changes quickly with time. To reduce complexity the small scale fading effect is not considered in the simulation.

Here, the communication traffic simulation uses a 20 MHz LTE channel in the frequency band of 2.6 GHz. The 20 MHz bandwidth is divided into 100 resource blocks, each having a bandwidth of 180 kHz [135]. According to LTE Type 0 resource allocation standardization, four consecutive resource blocks are grouped to form a resource group. The transmission bandwidth of each resource group is 720 kHz. A resource group is also frequently and interchangeably termed a channel in the current work. All the simulation reported focuses solely on uplink transmission i.e., from a user equipment to the base station.

The model used to calculate the path loss is the WINNER II B1 line-of-sight model proposed in [132] since the small cell radius is much smaller than micro-cells [133]. It is characterized by high vehicular speed of less than 20 ms^{-1} in an outdoor environment such as a local, metropolitan area with high user density. The frequency range is between 2-6 GHz, whereas, the height of the antenna at the eNB and at the MS is assumed to be below the rooftop of the surrounding buildings. The UE in the moving vehicle uses a fixed transmit power of 23 dBm. The UEs have line of sight (LOS) connection to the nearest base station with the possible exceptions of cases where the line of sight is blocked by traffic and the communication link becomes non-line of sight (NLOS). The breakpoint distance, also known as the critical distance is an important parameter as a dual slope propagation model is used for path loss calculation. The breakpoint distance is computed as:

$$d'_{BP} = 4h'_{BS}h'_{MS}\frac{f_c}{c} \tag{3.2}$$

where, d'_{BP} is the effective breakpoint distance, h'_{BS} = $h_{BS} - 1.0$ m and h'_{MS} = $h_{MS} - 1.0$ m are the effective height of base station and mobile station antenna. The actual antenna height, h_{BS} and h_{MS} are assumed to be 10 m and 1.5 m respectively above the ground. The effective environment height in urban environments is assumed to be equal to 1.0 m. f_c is the centre frequency in Hz, and c is the propagation velocity in free space. The following equation is used to determine the LOS path loss between the base station and the user equipment.

$$PL = \begin{cases} 22.7\log_{10}(d)+41.0+20\log_{10}(0.2f_c)+SFL, & 10m < d < d'_{BP} \\ 40.0\log_{10}(d) \ + \ 9.45 \ - \ 17.3\log_{10}h'_{BS} \ - & d'_{BP} < d < 5km \\ 17.3\log_{10}h'_{MS} + 2.7\log_{10}\frac{f_c}{5} + SFL, & \end{cases} \tag{3.3}$$

where PL is the path loss in dB, d is the distance between the BS and the UE in meters, f_c is the carrier frequency in GHz and SFL is the log-normal shadow fading loss with a standard deviation of the 3dB mean.

3.3.4. Link Module

The fundamental metric to determine the link quality is the signal-to-interference-plus-noise ratio (SINR). SINR is defined as the ratio of the received signal power to the sum of the received power from interfering transmitters together with noise power. The SINR on a particular resource group is calculated as follows

$$SINR = \frac{P_T^k G_T^k G_R PL_k^{-1}}{\sum_{i=1}^{N_I} P_T^i G_T^i G_R PL_i^{-1} + P_N} \tag{3.4}$$

where P_T^k is the transmit power from transmitter k, G_T^k and G_R^k are the respective transmitter and receiver gain equal to 0 dBi, as omnidirectional antenna is used, PL_k is the propagation loss between transmitter k and the receiver, calculated using Eq.3.3, N_I is the number of interfering transmitters on the same channel, and P_N is the receiver noise floor calculated using Eq.3.5 below.

$$P_N = 10log_{10}(kTB) + N \tag{3.5}$$

In this equation, P_N is the noise power in dBW, k is the Boltzmann constant; $k = 1.38 \times 10^{-23} m^2 kg s^{-2} K^{-1}$, T is the noise temperature in K; $T = 290$, B is the channel bandwidth in Hz; $B = 20 \times 10^6$ and N is the noise figure in dB; $N = 7$. The noise floor calculated using the equation is equals to -124 dBW, later converted into W, for SINR calculation.

A Truncated Shannon Bound model, proposed in [134], is used to determine the link throughput using Eq.3.6 below:

$$Throughput = \begin{cases} 0, & SINR < SINR_{min} \\ \alpha B log_2(1 + SINR), & SINR_{min} < SINR < SINR_{max} \\ \alpha B log_2(1 + SINR_{max}), & SINR \geq SINR_{max} \end{cases} \tag{3.6}$$

where $\alpha = 0.65$ is the attenuation factor, B is the link bandwidth, $SINR$ is the calculated signal-to-interference-plus-noise ratio at the receiver, $SINR_{min} = 1.8$ dB is the minimum SINR threshold to maintain a data transmission, and $SINR_{max} = 21$ dB is the SINR level that corresponds to the maximum link throughput.

3.3.5. Resource Allocation Module

A centralized radio resource assignment module is assumed for the allocation of the resources for data transmission depending on the channel quality information (CQI). The channel information changes rapidly with changes in the environment. Therefore, it is essential to evaluate the channel quality as the vehicle moves. Here, the central controller assigns the resource using the maximum SINR resource allocation approach. Adjacent channel interference is assumed to be negligible. The minimum SINR for accepting a transmission is 5 dB and the transmission is dropped if the SINR falls below 1.8 dB [131]. An ongoing transmission is assumed to occupy an assigned resource until it is dropped or handed over to a new BS. Moreover, a transmission is continuous until it is completed.

3.3.6. User Association Module

User association is a critical element in communication networks and substantially affects network performance. It relates generally to selecting a radio-link for the transmission of data depending on the prevailing radio traffic environment. The user association decision in the existing LTE/LTE-A systems is taken by the radio admission control entity located in the radio control layer of the protocol stack [7]. The decision depends on the quality of service (QoS), the priority level of the request and the availability of resources. Upon the arrival of each file transmission request, a UE associates with a BS using a defined scheduling strategy. For example, the UE may associate with a) the minimum distance BS to reduce end-to-end latency [136]; b) the maximum load BS for energy efficiency [38]; or c) the maximum distance BS to reduce the handover rate due to vehicle mobility [137]. It is assumed that a file is associated with only one BS i.e., a single-BS association is considered. However, the user association is based on different scheduling schemes such as maximum radio signal strength (max-RSS), minimum load or maximum distance BS depending on desired network performance.

3.4 Empirical Evaluation

To perform an empirical evaluation of the quality of service achieved employing the proposed user association schemes in ultra-dense small cell vehicular networks, different performance metrics that are used in this thesis are discussed now.

3.4.1. Performance Metrics

The key metrics used to evaluate the network performance in this thesis are the probability of blocking, the probability of dropping, the probability of retransmission, end-to-end latency, and the number of handovers per transmission. The probabilities of blocking and dropping are conventional parameters used for measuring the quality of service on a call based network, whereas, in a packet-based network, the probability of retransmission is used. The probability of blocking is defined as

$$P_B = \frac{N_B}{N_{tx}} \tag{3.7}$$

where P_B is the probability of blocking, N_B represents the total number of blocked transmissions, and N_{tx} is the total number of transmissions in the system. Similarly, the dropping probability is defined as

$$P_D = \frac{N_D}{N_{tx} - N_B} \tag{3.8}$$

where P_D is the probability of dropping, N_D represents the total number of dropped transmissions, and N_{tx} is the total number of transmissions in the system. The probability of successful transmissions could be evaluated using Eq.3.9 below.

$$P_s = (1 - P_B)(1 - P_D) \tag{3.9}$$

The probability of retransmission is the measure of a transmission being blocked or interrupted at least once. It is calculated using the following equation:

$$P_{retx} = \frac{N_{retx}}{N_{tx}} \tag{3.10}$$

where P_{retx} is the probability of retransmission, N_{retx} is the number of retransmissions and N_{tx} is the total number of transmissions in the system. The total number of transmissions is the summation of the number of retransmissions and the number of successful transmissions. The number of retransmission represents the transmissions which are interrupted at least once. In the case of an interrupted transmission, only the remaining part of the transmission that has not been transmitted is rescheduled.

The end-to-end delay comprises the transmission delay and the back off delay. The transmission delay depends on the link quality and is defined as the time taken to transmit a file across the wireless link. The back off delay follows a random exponential distribution and is defined as the time consumed by a file before a retransmission is scheduled. The average end-to-end delay of a file is calculated by the equation below:

$$Delay = \frac{1}{N_{tx}} \sum_{i=1}^{N_{tx}} \left(\frac{N_{bits}(i)}{C} + \sum_{i=1}^{N_{tx}} D_r(j)\right) \tag{3.11}$$

The number of handovers per transmission is defined as the total number of times a transmission is handed to a new eNB before it finishes.

3.4.2. Statistical Validation of Results

In order to ensure the validity of the results reported in this thesis, the following techniques are applied

- The data points for each performance measurements are plotted against the vehicle speed because it is the most effective way to demonstrate the behaviour of the system under different traffic conditions. These data points are obtained by averaging over 25 separate simulations with different random seeds.
- The results also include error bars showing the minimum and maximum value from the different simulation run for a particular data point. Error bars are used as a graphical representation of the variability of data.

3.5 Conclusion

This chapter provided a comprehensive overview related to the development of the integrated simulator model as well as described the methodology used for the empirical evaluation of a range of heuristic as well as learning based user association approaches, with respect to different performance metrics, that are proposed later in this thesis. A highway scenario, that involves dynamic vehicular traffic and an ultra-dense small cellular network, is used as the basis for the detailed system-level simulation model. The developed model served as a firm basis to understand the influence of mobility on network performance in an ultra-dense small cell scenario. The MATLAB software environment (version R2015b) has been used to develop the integrated model, while, the Monte Carlo approach was employed to generate statistically important results. Further, the key metrics that are used to evaluate the performance of the developed user association approaches are explained. The chapter concludes by presenting a discussion on statistically validating the results.

Chapter 4. Performance Metric based User Association in Ultra-dense Small Cell Dynamic Vehicular Environments

Contents

4.1 Introduction

As described in the previous chapters, user association relates to adapting a radio link for data transmission depending on the prevailing radio traffic environment. The user association decision in existing LTE/LTE-A systems is managed by the radio admission control entity located in the radio control layer of the protocol stack [7]. The decision depends on the quality of service (QoS), the priority level of the request and the availability of resources.

In the 3GPP Release 9 and later, when a UE is switched on for the first time it performs a cell selection procedure that involves searching and association with the strongest received signal strength eNB [4] [138]. This approach is the conventional user association approach referred to as the maximum radio signal strength (max-RSS) user association approach. In order to remain associated with an eNB, the channel quality criteria must be fulfilled. However, as UE moves along the street from the coverage area of one eNB into other, a cell reselection or handover results if the cell selection criterion is fulfilled. The cell selection criterion for LTE as specified in [139] is fulfilled when:

$$S_{rxlev} > 0 \tag{4.1}$$

Where:

$$S_{rxlev} = RSRP_{candidate} + Offset \tag{4.2}$$

where S_{rxlev} is the cell selection receiver level value in dB, $RSRP_{candidate}$ is the measured reference signal received power of the target eNB. Whereas, the offset is a combination of different hysteresis such as minimum required receiver level in the cell, offset to the signalled minimum required receiver level, maximum transmission power level and maximum radio frequency output power of UE.

In the case of ultra-dense small cell environments, due to cell densification, the distance between a UE and eNBs reduces significantly compared to a conventional micro or macrocell deployment scenario. A possible small cell scenario in a real network on a highway is shown in Figure 4.1.

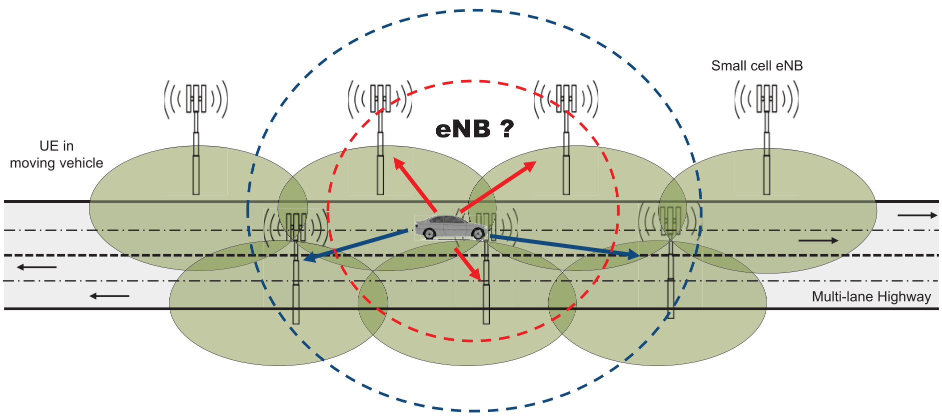


Figure 4.1: Ambiguity in eNB selection for user association in ultra-dense small cell vehicular environments

This distance between a UE and eNBs is termed as radio link length or association range, in the presented work. A decrease in radio link length results in acceptable radio signal strength from more than one eNB in the vicinity of the moving UE. This leads to an increase in the number of eNBs that may simultaneously satisfy the conventional user association criteria. Moreover, due to the dynamic vehicular traffic flow, frequent variations in the interference level may cause an unpredictable radio environment. Employing the conventional user association policy to associate moving UE to appropriate eNB such that a balance between different system performance metrics is established whilst providing a guaranteed network QoS in a continuously changing environment, is therefore extremely challenging.

This chapter, therefore, investigates several user association approaches depending on different performance metrics to understand the influence of individual metric on network performance in an ultra-dense small cell vehicular environment. In Section 4.2, heuristic user association schemes, max-RSS, that is used for baseline comparison, is presented. A performance metric based user association approach that employs a three-step sequence approach is proposed in Section 4.3. A range of results obtained by employing different user association techniques across different mobility level is shown in Section 4.4. Conclusion and future directions are provided in Section 4.5.

4.2 A Heuristic Scheme for Baseline Comparison

4.2.1. User Association based on Signal Strength

This section overviews maximum received signal strength user association approach. This scheme associates a UE to an eNB with the maximum received signal strength [4] [138]. In the following presented work, it has been assumed that all the UEs transmit at the same power level, and the cell size of each small cell is the same. The maximum received signal strength eNB to which the moving UE associates is the closest eNB among all the eNBs in its vicinity. Therefore, is also referred to as the minimum distance eNB. A detailed flowchart illustrating the user association and data transmission using the conventional user association approach is shown in Fig 4.2.

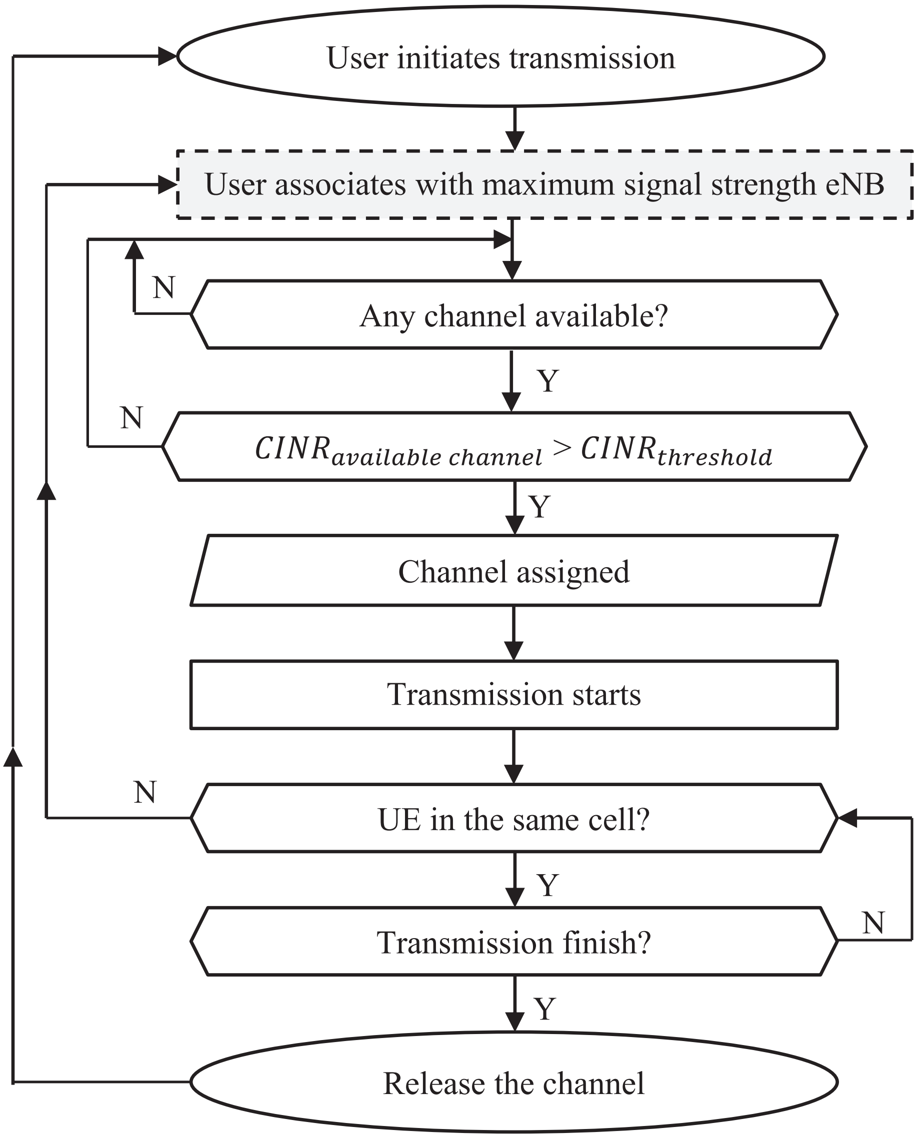


Figure 4.2: User association based on maximum received signal strength

The UE association status is defined by the association matrix as follows

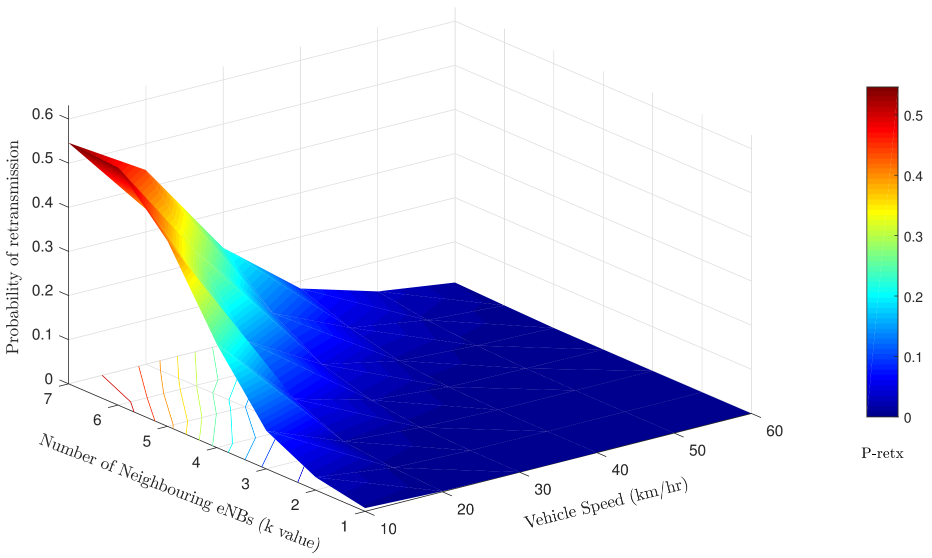
$$UE_i^j = \begin{cases} 1 & \text{if UE is associated with } eNB^j \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

where UE_i^j is the i^{th} UE associated to j^{th} eNB.

On associating with target eNB, the link quality of all the idle resources on this eNB is calculated using Eq. 3.4. A radio resource for data transmission is then assigned or denied, accordingly. If the channel quality of the assigned resource on an associated eNB drops below the minimum SINR threshold, then a new resource on the same eNB is identified and assigned to the ongoing transmission. If no resource with link quality above a threshold is available, then the transmission interrupts. So, a transmission is interrupted if a) there is no available resource on the associated BS; b) if the link quality of all the available resources is below the assumed threshold of 1.8 dB or c) if the new transmission is likely to drop an ongoing transmission in the system.

In the case of an interrupted transmission, a retransmission attempt takes place after a random exponential backoff time [8]. Furthermore, a maximum of 5 re-transmission attempts is performed before a file is categorized as an unsuccessful transmission and thereafter eliminated from the system. The received signal strength, the link quality, and the service time of the ongoing transmission is updated with an update in vehicle location.

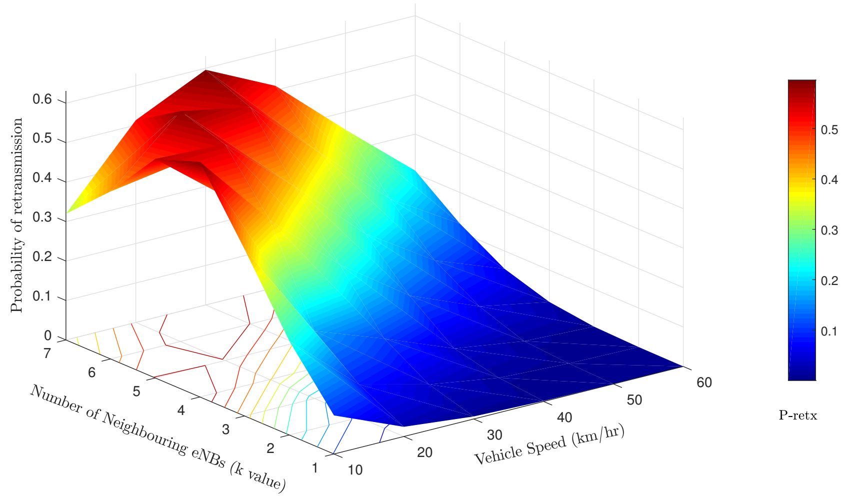
Figure 4.4 illustrates contour plots at relatively low (40%), medium (60%) and high (80%) vehicular traffic levels using a minimum distance user association scheme. Figure 4.3 shows that at low traffic loads, association with even the farthest eNB delivers guaranteed network QoS.



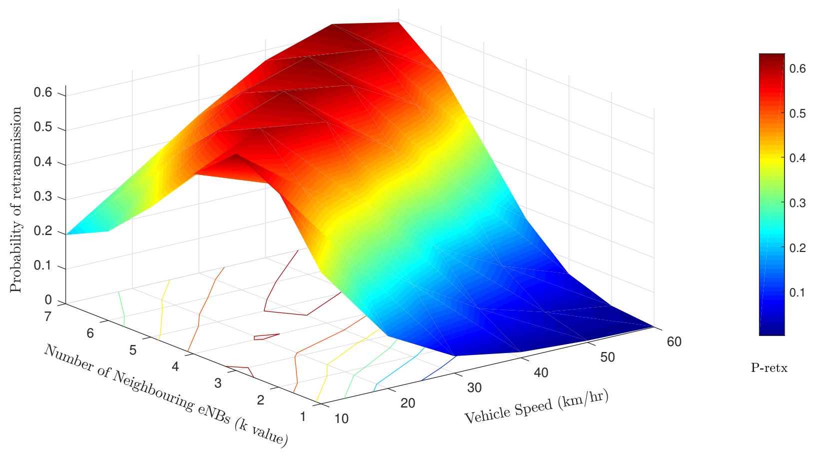
(a) Vehicular traffic: 40%, Communication traffic: 0.17 - 0.45 Gbps

Figure 4.3: Probability of retransmission using minimum distance user association approach at low offered traffic levels with a range of vehicle mobility level on a multi-lane bi-directional vehicle flow ultra-dense small cell highway scenario

However, at high traffic loads the communication system saturates even when the UE is associated with the closest eNB, thus a completely opposite pattern at high offered traffic is seen in Figure 4.4(b). Whereas, Figure 4.4(a) shows that at medium traffic load the network performance lies in a region between the two extremes. The three contour plots, therefore, demonstrate the influence of association range across different vehicle speed and offered traffic levels in dense dynamic environments. The results relating to minimum distance user association approach have been performed assuming medium offered traffic as explained in Chapter 3.



(a) Vehicular traffic: 60%, Communication traffic: 0.26 - 0.56 Gbps



(b) Vehicular traffic: 80%, Communication traffic: 0.35 - 0.60 Gbps

Figure 4.4: Probability of retransmission using minimum distance user association approach at medium and high offered traffic levels with a range of vehicle mobility level and number of neighbouring eNBs on a multi-lane bi-directional vehicle flow ultra-dense small cell highway scenario

4.3 Three-Step Sequence Scheme for User Association

This section investigates the selection of an appropriate base station for user association in a mobility-aware ultra-dense small cell vehicular environment. The proposed scheme is formulated as a search algorithm that aims to (a) identify and select the most appropriate eNB for the association and (b) to understand the influence of individual performance metric based user association on the network performance in ultra-dense, dynamic vehicular speed environments.

The eNB selection is based on different attributes such as (a) performance metric, (b) temporal-spatial distribution of moving UEs, (c) geographical location of eNBs, (d) user mobility level, (e) mobility direction, (f) number of neighbouring eNBs, represented by parameter k, and (g) the inclination and the azimuthal angle formed by the moving UE to prospective eNBs. The three steps followed to identify an appropriate eNB are enumerated below.

1. *Shortlist* k-nearest neighbouring eNBs in the vicinity of moving UE.
2. *Select* eNBs in the direction of vehicle mobility from the shortlisted list.
3. *Choose* most appropriate eNB from selected list based on performance metric.

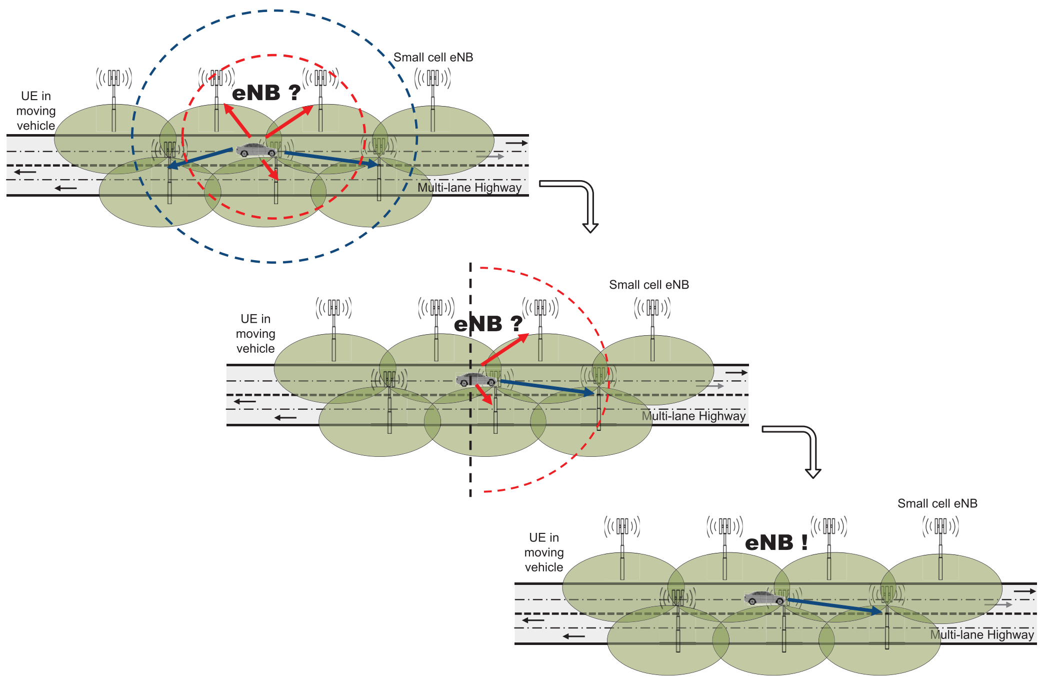


Figure 4.5: A three-step sequence rule for performance metric based user association

Step I: ***Shortlist k-nearest neighbouring eNBs in the vicinity of moving UE.***

The first step involves constructing a list of eNBs in the vicinity of the moving UE by employing the k-Nearest Neighbour (k-NN) search algorithm. The k-NN approach belongs to the non-parametric instance-base learning methods category as it constructs a hypothesis directly from the input parameters [140]. Here, the input parameter, k, defines the neighbourhood range i.e., the number of eNBs that will populate the k-NN list.

Conventionally, the k value is manually configured depending on received signal strength [141]. However, it can also be learnt online [142] [143]. Chapter 5 and Chapter 6 investigates intelligent schemes that possess the ability to learn an appropriate k value to automatically self-configure the k-NN list as well as continuously self-optimize it by utilizing vehicle mobility level and network performance.

In this chapter, the value of k is *fixed* across all mobility levels. However, it is not the same for every performance metric i.e, the k value assumed for spectral efficiency is 1, i.e., only the closest eNB to the moving UE will form the k-NN list, whereas in the case of network load balancing and handover optimization, $k = 4$. These values are assumed on the basis of results obtained in Section 4.2. As the vehicle moves along the path, the eNBs in the constructed k-NN list updates accordingly. Algorithm 1 describes the first step of the proposed user association scheme.

Algorithm 1 Three-Step Sequence User Association Approach - **Step I**

1: **Input:**
    The UE and eNBs location.
    The neighbourhood range, k value.
2: **Output:**
    Shortlisted k-NN list.
3: **Step I: Construct the nearest neighbour eNB list**
    Employ the k-Nearest Neighbour rule.
    Shortlist the closest k eNBs around the UE.
    Construct the shortlisted k-NN list.

The utilization of the k-NN rule, therefore, leads to the selection of prospective eNBs from among the eNBs in the neighbourhood of the moving vehicle. Moreover, shortlisting of eNBs appears to reduce the processing time, enhance the processing speed as well as enhance the device battery life during the user association process.

Step II: ***Select candidate eNBs in the direction of vehicle mobility.***

In the second step, shortlisted eNBs are categorized based on the direction of vehicle mobility. The eNBs in the direction of vehicle mobility is subsequently selected, omitting the options that are in the opposite direction of mobility. This is performed by calculating the inclination and the azimuth angle formed by the moving UE to all the shortlisted eNBs in the k-NN list. The motivation is to maximize UE dwell time per cell whilst delivering best-effort network performance.

In order to calculate the inclination and the azimuth angle, the vehicle is chosen as the datum point. A datum line is drawn from this datum point. The datum line is a perpendicular drawn from the vehicle mobility direction towards the left, as shown in Figure 4.6. The inclination and azimuthal angle to all the shortlisted eNBs from the vehicle is thereafter calculated using appropriate trigonometric formulae.

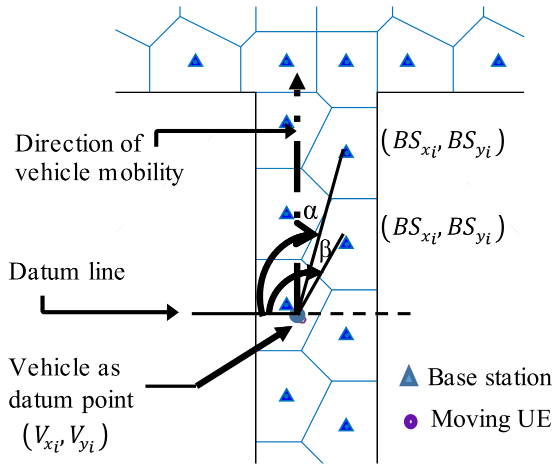


Figure 4.6: Classifying eNBs based on vehicle mobility direction, shortlisted eNBs location and azimuth angle

The eNBs that make an azimuth angle within the range $[0, \pi]$ with the UE are identified as the eNBs in the direction of mobility. The remaining eNBs are removed from the k-NN list that was constructed in Step I.

Subsequently, in addition to azimuth and inclination angle, active network load on the selected eNBs is computed. The active communication traffic load per eNB assists to *choose* an appropriate eNB for user association based on network load, in the last step of the proposed scheme. Algorithm 2 describes the second step of the three-step sequence user association scheme.

Algorithm 2 Three-Step Sequence User Association Approach - **Step II**

1: **Input:**
  Shortlisted eNBs list.
2: **Output:**
  Selected eNBs list.
3: **Step II: Select candidate eNBs in the direction of vehicle mobility**
  Compute the azimuthal and elevation angle.
  Identify eNBs in the direction of vehicle mobility.
  Identify eNBs opposite to the direction of vehicle mobility.
  Compute the active traffic load on each shortlisted eNBs.
  Construct list of eNBs in the direction of mobility.

Step III: ***Choose an appropriate eNB for user association or handover from selected eNBs, depending on the desired performance metric.***

In the final step, depending on the desired performance metrics, the most appropriate eNB from selected k-NN list is chosen. The different performance metrics investigated include (a) spectral efficiency, (b) network load, and (c) handover frequency. The third step of the proposed scheme is detailed below in Algorithm 3.

Algorithm 3 Three-Step Sequence User Association Approach - **Step III**

1: **Input:**
  Selected eNBs list.
  The desired performance metric.
2: **Output:**
  The appropriate eNB.
3: **Step III: Select most appropriate eNB depending on distinctive attributes**
  Spectral Efficiency - Maximum SINR eNB.
  Network Load - Minimum & Maximum load eNB.
  Handover Rate - Maximum Distance eNB.
  Forward Handover
    Same Side eNB: Azimuth angle: [0, π/2]
    Opposite Side eNB: Azimuth angle: [π/2, π]
  Backward Handover
    Same Side eNB: Azimuth angle: [3π/2, 2π]
    Opposite Side eNB: Azimuth angle: [π, 3π/2]

In the next section, user association based on the three distinct performance metrics using the three-step sequence scheme is discussed.

4.3.1. User Association based on Spectrum Efficiency

Spectrum efficiency is defined as the maximum data that can be transmitted over a given bandwidth in wireless networks. It can be quantified by parameters such as (a) end-to-end delay (b) user throughput (c) radio signal quality, and (d) probability of blocking/dropping. The maximum capacity i.e., the data rate at which the information could be reliably transmitted over a given bandwidth can be calculated using the Shannon capacity equation written below:

$$C = \begin{cases} Blog_2(1+SNR) & \text{if interference is not considered} \\ Blog_2(1+SINR) & \text{if interference is modelled} \end{cases} \tag{4.4}$$

where C is the channel capacity in bits per second, B is the channel bandwidth in Hz, SNR is the signal-to-noise ratio and $SINR$ is the signal-to-interference-noise ratio.

A content-recommended proactive cell association framework focusing specifically on multimedia data services is proposed by the authors in [144]. The collaborative filtering approach exploits the user context information and predicts the QoE level a UE-eNB association could deliver. The simulation results demonstrate 19% improvement in spectrum saving compared to the conventional maximum SINR based user association approach. Another contribution to the admission control and mobile association optimization in dynamic vehicular environments is reported by the authors in [145]. The overall system performance optimization was performed by utilizing a Semi Markov Decision Process (SMDP) framework. Whereas, the authors in [146] formulated the uplink user association problem as a college admission game, wherein the eNBs seek to recruit the UEs. It is a distributed algorithm that combines the concepts from matching theory and coalitional games. Depending on the packet success rate, the end-to-end delay and coverage area of each eNB, UEs and eNBs rank each other in order to optimize UEs utility whilst maintaining QoS. The simulation results demonstrate that the average utility per user improves by 23% relative to the conventional packet success rate algorithm. In the work presented, the three-step sequence user association approach employs the minimum distance eNB to achieve maximum spectral efficiency. The results obtained are shown in Section 4.4.

4.3.2. User Association based on Network Load

Achieving a high spectral efficiency is one of the key performance indicators (KPI) of the 5G and beyond wireless networks. User perceived data rate is the achieved instantaneous rate multiplied by a number of assigned resources. However, the number of resources in case of the proportional fair, as well as round-robin scheduling schemes, depends on the number of active users on the eNB, i.e. the active load on the associated eNB, which in case of dynamic environments varies both spatially and temporally. Therefore, the perceived data rate in dynamic environments not only depends on the channel quality but also on the active load on the associated eNB. Following this, user association based on network load and its influence on QoS are now discussed.

Load balancing is defined as a distribution of network traffic load among the eNBs such that the demand for radio resources is matched to the supply i.e., network capacity [147]. The authors in [148] utilized the massive MIMO properties and operational characteristics to develop practical load balancing methods to maximize the network-wide utility metric by optimization of the UE-eNB pair activity factor. Moreover, in [149] the authors utilized the stochastic geometry concept and the truncated channel inversion power control method to model user association for uplink cellular networks. The developed model is load aware and is based on per user power control scheme to balance the network load among the eNBs as well as to improve the cell edge user performance. The simulation results demonstrate that the proposed load aware model outperformed in terms of overall SINR performance.

In [150], the authors proposed a context-aware user association algorithm. The association problem was formulated in the framework of the matching theory. The different parameters such as transmission service time, probability of handover failure and different QoS requirements of users were taken into account to differentiate between the users while designing the user association algorithm. The simulation results demonstrate a better network traffic load balance among eNBs whilst delivering guaranteed QoS. A QoS aware user association approach with dual connectivity in heterogeneous wireless networks for load balancing is proposed in [151]. The authors formulated the user association problem as multi-objective optimization model and used the guaranteed bit rate, probability of blocking and throughput of the cell edge users as the performance metrics for performance evaluation of the proposed algorithm. Simulation results demonstrate that the dual connectivity user association algorithm delivered a better QoS compared to the max-SINR conventional approach.

The biased received power based user association also known as cell range expansion is one of the commonly used approaches for load balancing, also discussed under section 2.3.3. Under the biasing scheme a bias, i.e. an artificial value is added in the RSSI to divert the UEs to associate with the lower tier eNBs using the equation below

$$k^* = arg \max_{j=1,2...n} B_i P_{rx,i} \tag{4.5}$$

where B_i is the i^{th} tier bias and $P_{rx,i}$ is the received power from tier i. Thus, a UE device is diverted toward a small cell eNB for association rather than a macrocell base station. In order to achieve a high perceived data rate, load balancing is as important as maintaining a high channel quality.

Here, the two scenarios considered are a) association of UE with a maximum loaded eNB b) association of UE with minimum loaded eNB. The network load based user association approach carefully identifies an appropriate eNB for UE association such that the available resources on the deployed eNBs are successfully utilized whilst providing a guaranteed network QoS and user QoE.

The above-mentioned user association schemes do not consider the impact of user mobility during association. A user association algorithm that does not consider mobility into account in dense dynamic environments may result in higher handover rate compared to conventional cellular networks. An increased handover rate may further increase the switching and signalling load that may result in undesirable end-to-end delay and possibly dropped transmissions [152, 153]. UE mobility is thus a critical parameter which needs consideration while developing user association algorithms for dynamic environments, as it substantially impacts the overall network and user performance.

4.3.3. User Association based on Handover Rate

In vehicular environments, as the UE moves along the street from the coverage area of one eNB into other, user re-association also known as handover occurs, as discussed in Section 4.1. At moderate or high vehicle speed, the dwell time per cell may be so small that even before all the signalling and switching overhead signals were processed a UE might have passed through the coverage area of a candidate cell into the subsequent cell. Thus, multiple handovers within a very short time ensue. The dwell time per cell is also known as sojourn time and is defined as the time a UE remains in a cell before experiencing a handover. It depends on the UE velocity and the BS density. A vehicle moving at high speed may exhibit a small sojourn time per cell compared to a vehicle moving at low mobility level.

Moreover, there is a higher probability that a non-ideal eNB for handover in a small cell scenario might be selected. A non-ideal eNB is defined as (a) an eNB with fluctuating radio signal level that may cause the ping-pong effect, or b) results in a failed handover either due to unavailability of resources or the desired channel quality. Therefore, unacknowledged vehicle mobility will not only result in handover failure but also in unnecessary handovers. An increased number of handovers, therefore, may result in an increased number of interrupted transmissions, leading towards a poor network performance.

A moving direction user association strategy is proposed in [154] that utilizes a one-dimensional micro-cellular environment framework. Simulation results demonstrate that the moving direction strategy provides a lower probability of forced termination compared to nearest neighbour (NN), nearest neighbour+1 (NN+1) and fixed channel allocation schemes. In [153], vertical handover in an LTE-A multi-tier network utilizing a mobility-aware network selection method is proposed. Simulation results demonstrate an improvement by 20% in packet delivery ratio and about 14% in end-to-end latency.

In [155] a sojourn time-based cell selection scheme is proposed. The proposed scheme utilizes the user velocity and the BS density to determine the cumulative distribution function of the sojourn time in heterogeneous networks. Further, the sojourn time is obtained by analyzing the radio link length and user trajectory in the coverage of the associated eNB. Simulation results were compared with max-RSS cell selection scheme in which the ping-pong rate reduced by 50%, the number of handovers by 60% and the handover failures by 30%.

In [156] sojourn time-dependent mobility-aware caching strategy in heterogeneous networks is proposed. An upper bound for the mean download time of a transmission was derived by using different parameters such as user mobility, sojourn time and BS density. Similarly, in [152] the authors utilized the network topology, sojourn time and cooperative communication among BS to skip the unnecessary handovers to maintain the user QoS.

In [157] the authors use tools from stochastic geometry to analyze the mean handover rate and sojourn time and derive an expression that provides design guidelines for small cell deployment in 5G networks. The derived expression is a function of BS density, UE mobility and transmission probability. The paper concludes by claiming that the interruption ratio through the presented analysis is equal to the square root of the macro and small cell BSs densities.

The authors in [45] obtained the UE handoff rate and derived the probability of coverage in an irregular multi-tier heterogeneous network to analyze the impact of UE mobility on the overall system performance. The user association was again based on a biasing scheme in which maximizing the coverage is considered as the utility factor. Later the authors claim that the proposed speed-dependent bias factor approach has the ability to adjust the tier association parameter to improve the network coverage and system performance.

In order to optimize the number of handovers per transmission, user association based on maximum dwell time is employed. In this work, a UE associates, in the direction of its mobility, with the maximum distance eNB among all the eNBs in its vicinity. The eNBs in the direction of vehicle mobility are identified in Step II of the proposed three-step sequence algorithm. After associating with an eNB, the maximum SINR resource allocation policy is used to assign an appropriate resource to UE for data transmission. The signal strength, channel quality and service time of the ongoing transmission is continuously updated as the vehicle location updates.

The flowchart in Figure 4.7 illustrates user association approach based on handover rate. This approach is also referred to as maximum distance association approach in this thesis.

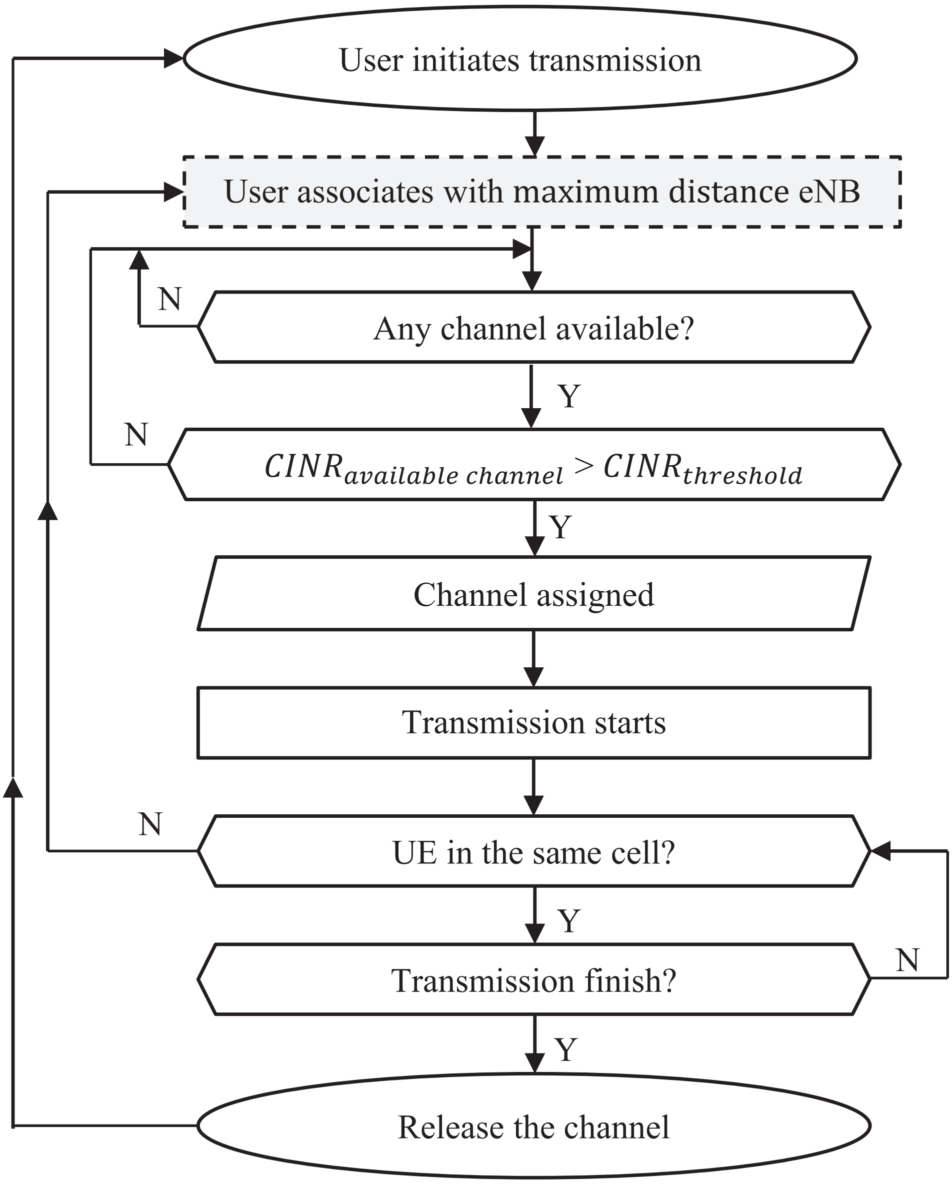


Figure 4.7: User association based on maximum distance approach

Figure 4.8 illustrates the number of handovers per transmission at different mobility levels with a range of the number of nearest eNBs, k value. It is seen that the number of handovers per transmission is directly proportional to vehicle speed and inversely proportional to the association range; k value.

The maximum distance user association policy used for handover optimization is also

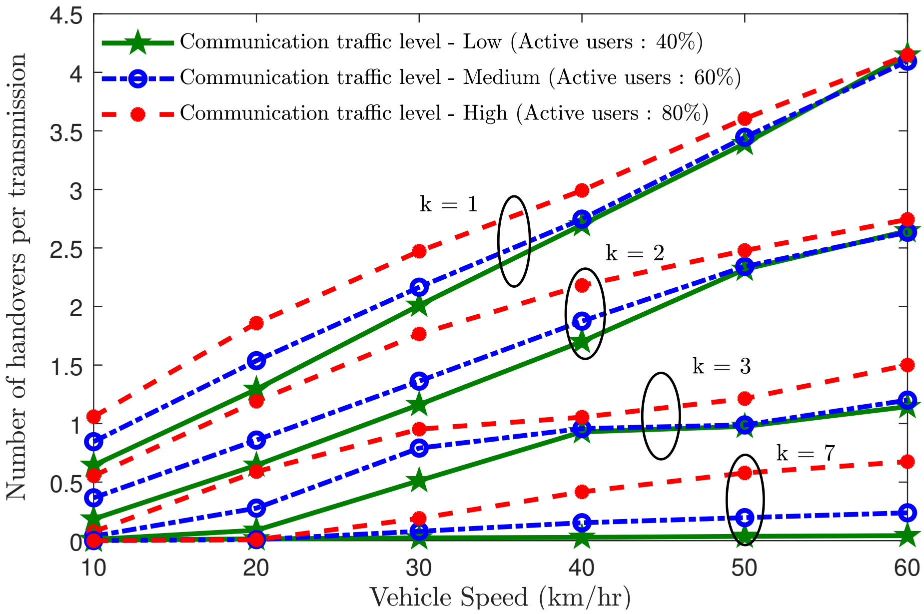


Figure 4.8: Number of handovers per transmission with a range of k values at different traffic level using maximum distance user association approach

tested based on the location of the eNB deployment at the roadside infrastructure. The two test cases using the maximum distance location-based user association approach are (a) Forward Handover - Near side (FH-NS) (b) Forward Handover - Far side (FH-FS). In FH-NS approach, the UE associates with the maximum distance eNB on the near side of the street while in FH-FS association with maximum distance eNB on the far side of the street, in the direction of vehicle mobility is performed.

4.4 Results

This section discusses the results obtained during simulation experiments using the investigated user association approaches. The results demonstrate the impact of each performance metric on the network QoS, at different vehicle speeds. The network performance is analyzed using the network throughput, the probability of retransmission, end-to-end latency, and the number of handovers per transmission. All investigated approaches were simulated at medium traffic level and using same file size. The number of neighbouring eNBs, k, is *fixed*, as explained in Section 4.3, to ensure a better understanding of the influence of individual performance metric under the same environmental setup.

In order to deliver high spectral efficiency, the minimum distance user association scheme; k=1 is employed. Whereas, to balance the spatially distributed network load and to maintain a minimum number of handovers per transmission, the user association scheme based on network load and the maximum distance is used respectively. The maximum number of neighbouring eNB, k, in case of spectrum efficiency is 1, whereas, for network load balancing and handover optimization is assumed as 4.

Figure 4.9 shows the overall network throughput achieved by employing different association approaches at different vehicle speed.

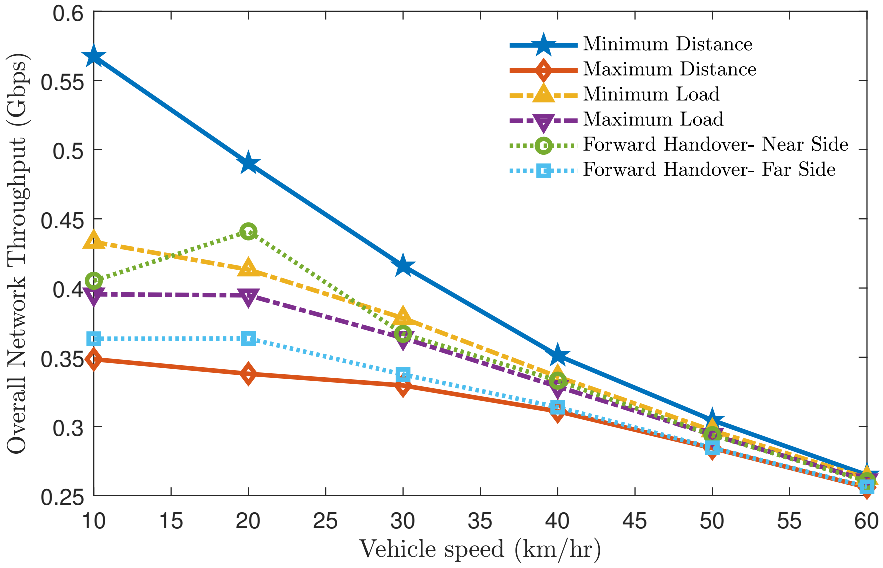


Figure 4.9: Network throughput response of the performance metric based user association technique at different vehicle speed and corresponding traffic load

It can be observed that the maximum distance user association scheme starts from a lower throughput because it can hardly avoid the strong cross-interference from the surrounding eNBs and delivers minimum network throughput compared to all the other investigated approaches. On the other hand, the minimum distance user association scheme shows to effectively control the interference due to a tight cell packing even at high-density conditions. The other investigated association schemes achieved network throughput between the upper and lower bounds due to the flexibility to identify and choose appropriate eNB based on active network load and/or the location from among the selected eNBs in Step III of the proposed algorithm.

Figure 4.10 and Figure 4.11 show the impact of different user association approaches on the probability of retransmission, and the probability of blocking respectively. In Figure 4.10, irrespective of the user association approach employed, the probability of retransmission decreases as the vehicle speed increases. This is due to the assumption of number of active users in the system. As explained in section 3.3.1, the maximum number of vehicles in the system is assumed to be 60% of the maximum density at a particular mobility level as the communication system saturates on assuming higher percentages, as seen in Figure 4.3 and Figure 4.4. Therefore, as the vehicle speed increases, the number of active users in the network reduces, thus, reducing the offered communication traffic load and probability of retransmission.

Alternately, drawing a comparison between different user association approaches, it is evident that at low vehicle speed, a poor network performance results across all mentioned approaches. This is due to the greater number of active users per unit area that leads to an increased network load per eNB compared to the network load per base station at high mobility levels. However, among the investigated approaches, the minimum distance user association approach outperforms all other techniques by providing the minimum probability of transmission.

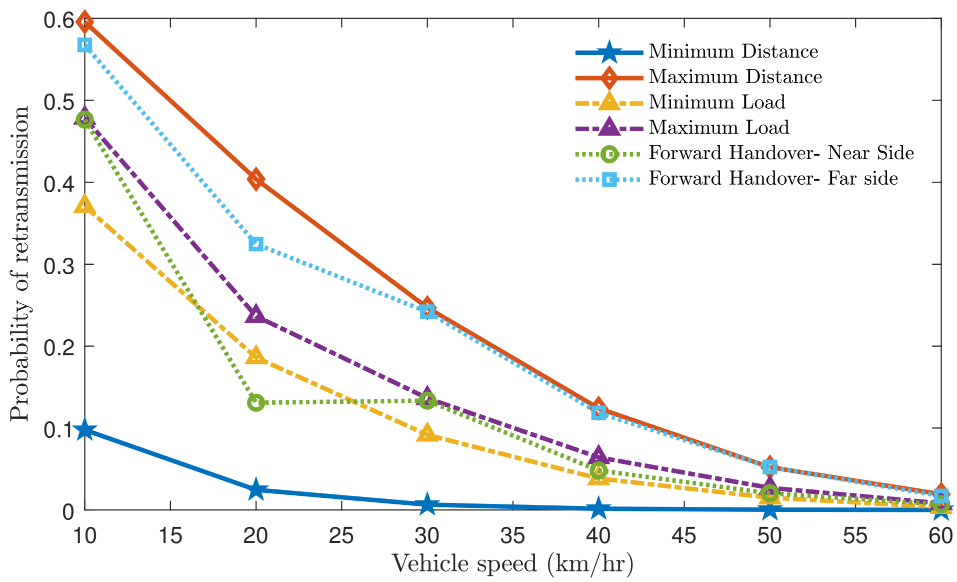


Figure 4.10: Average probability of retransmission response using the performance metric based user association at different vehicle speed and corresponding traffic load

In addition to this, as seen in Figure 4.11, employing the minimum distance and/or minimum load approach, at low vehicle speeds i.e., high user density per cell, yields better network performance compared to the other investigated approaches. This is because with minimum distance approach, a tight cell packing is observed that restrains the interference level and the probability of interrupted transmissions, thus, a better signal strength is delivered.In the case of minimum load, an improved channel quality results in better network performance. On the contrary, in the case of maximum distance user association scheme, the distance between the base station and the active user increases leading to a degraded signal quality hence a higher percentage of transmissions are interrupted compared to minimum distance approach, resulting in a poor overall network performance.

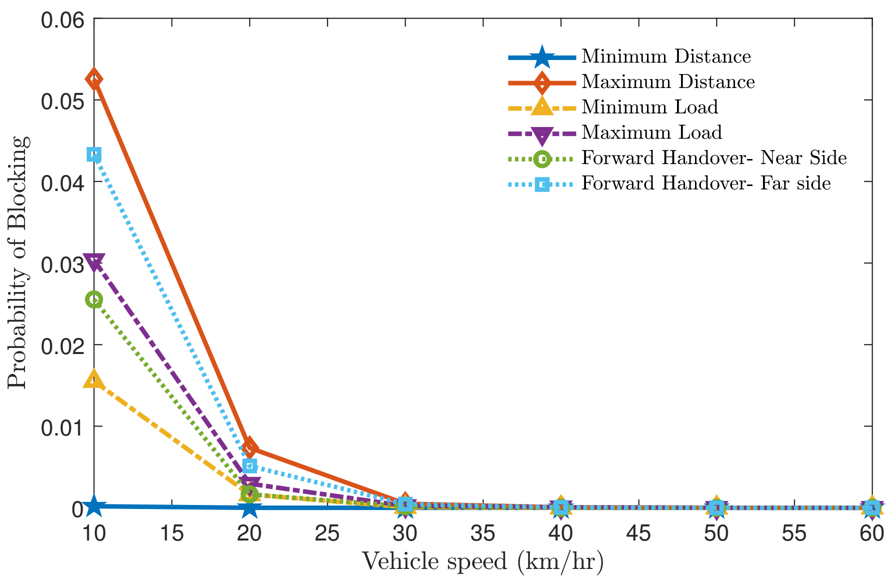


Figure 4.11: Probability of blocking response using performance metric user association at different vehicle speed and corresponding traffic load

The total network delay is presented in Figure 4.12, is the accumulated average delay of all the transmissions. At lower vehicle speed the communication network saturates leading towards higher network delay compared to that at higher vehicle speed. Association with the nearest eNB by using the minimum distance user association approach yet again outperformed the other investigated approaches, due to strong received signal strength and lower number of interrupted transmissions.

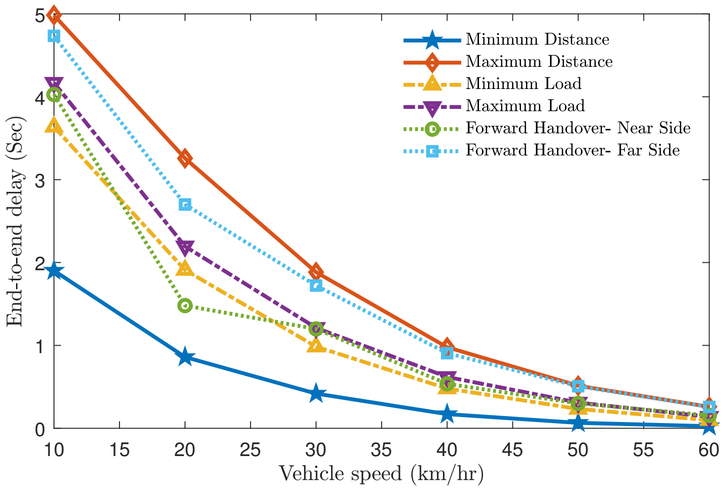


Figure 4.12: End-to-end delay response using performance metric user association at different vehicle speed and corresponding traffic load

Figure 4.13 presents the influence of user association approaches on the number of handovers per transmission at different vehicle speeds. The maximum distance user association approach outperforms all the other investigated approaches as it chooses the farthest eNB in the direction of its mobility. However, delivers a poor network QoS. This is because user association with the farthest eNB cause a greater overlap in the coverage range of each transmitting user, therefore, each user suffers from considerably strong interference. On the contrary, the minimum distance user association approach delivers the best network QoS, but, performs poorly to reduce the number of handover per transmission.

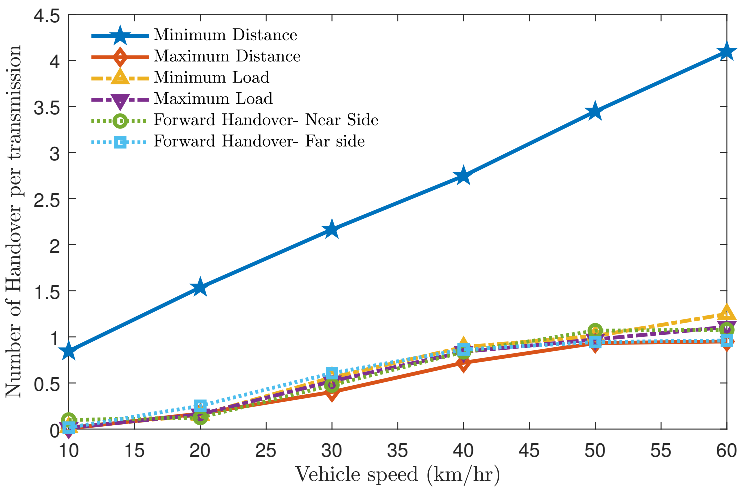


Figure 4.13: The number of handovers per transmission response using performance metric user association at different vehicle speed and corresponding traffic load

4.5 Conclusion

This chapter investigated the influence of performance metric dependent user association approaches on the network performance at different vehicle speeds in mobility-aware ultra-dense small cell environments. The chapter began with describing the user association and the handover mechanism. Following this, a conventional LTE user association approach, max-RSS has been discussed. This approach has been chosen to serve as a baseline approach for comparison with the other investigated techniques. Further, a novel user association approach, the three-step sequence is developed and tested. This approach assists to discover the nearest neighbours, eNBs, around the moving active user, thereafter, identify the most appropriate eNB for the association depending on specific performance metric to be optimized. Subsequently, the results presented assists to understand the influence of individual performance metric on the network performance and user experience at various vehicle speed. It has been observed that by employing max-RSS user association, i.e., the minimum distance user association approach in an ultra-dense environment, a better QoS results at the expense of significantly high handover frequency per transmission. However, the other technique such as the load based user association policy delivered better QoS with a marginal compromise on the handover frequency. Meanwhile, the maximum distance approach outperforms other investigated schemes in terms of handover optimization as the active user associates with the farthest eNB in the cluster, but delivered unsatisfactory network QoS. The investigated work thus provides insights into the trends related to user association based on different performance metrics in ultra-dense small cell vehicular environments at different mobility levels.

Chapter 5. Reinforcement Learning Based User Association in Multiagent Environments

Contents

5.1 Introduction

In Chapter 4, the influence of three step sequence user association approach on network performance, with fixed k value assumption, across different vehicle speed was investigated. Figure 4.10 and Figure 4.11 demonstrated that minimum distance user association approach delivered an effective QoS across all mobility levels, but, at the expense of high handover rate. Whereas, Figure 4.13 exhibit that the maximum distance user association approach outperformed all the other investigated schemes in case of handover optimization. However, none of the approach was able to strike a balance between the number of handovers per transmission and network QoS. Moreover, the investigated scheme was unable to adapt to the environment or learn appropriate k

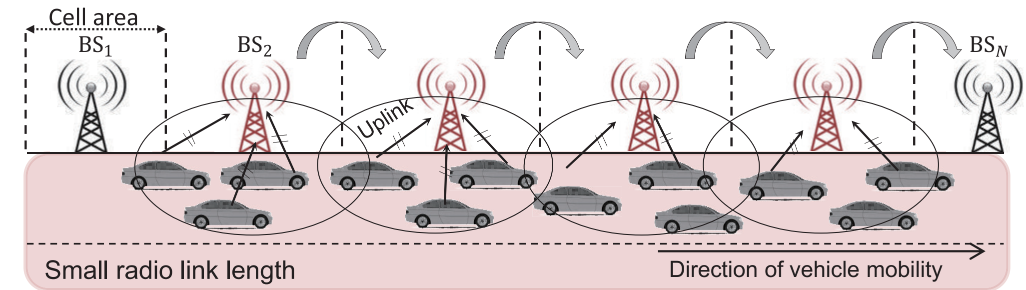
value, during the simulation, to self-configure and self-optimize the k-NN list so as to maintain an equilibrium between different performance metrics.

The purpose of this chapter is, therefore, to extend the three-step sequence user association approach by integrating known machine learning approaches, discussed in Chapter 2, to considerably reduce handover rate whilst network QoS is guaranteed across the investigated vehicular and communication traffic load in the considered ultra-dense dynamic vehicular environment. Furthermore, the developed intelligent user association algorithms should be able to demonstrate the ability to learn appropriate k value to self-configure and self-optimize the k-NN list by utilizing variation in the radio environment.

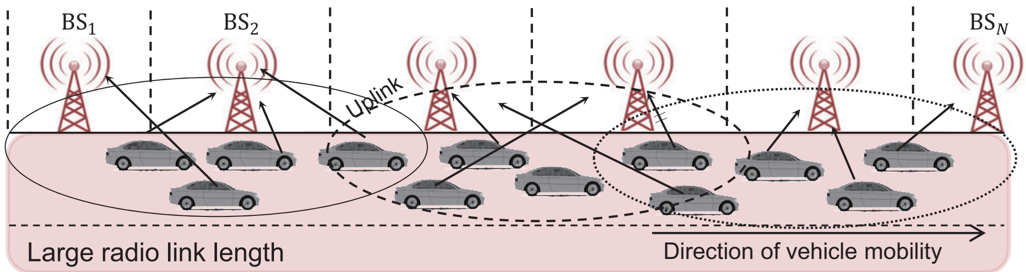
The rest of this chapter is organized as follows: Section 5.2 address the motivation to develop the intelligent user association approaches. Section 5.3 revisits the baseline approach. Further, a real-time control feedback user association approach is proposed. In Section 5.4, a series of intelligent user association approaches based on Q-learning are proposed. The result comparisons are performed in Section 5.5. Finally, conclusions and future directions are provided in Section 5.6.

5.2 Motivation

When considering vehicle mobility and potential for severe cross-interference in an ultra-dense topology, the *quality of decision* for frequent association of UE to the most promising eNB, as it moves along the street, is a significant metric. The quality of decision is governed by an essential parameter, the number of nearest eNBs, k, that populate the $k - NN$ list for the identification of an appropriate eNB. Conventionally, k value is manually configured depending on received signal strength [141]. However, in continuously changing environments it becomes absolutely challenging to configure an optimum value of k, due to the rapid fluctuations in radio signal strength and dynamic vehicle speed [45][159]. Considering a scenario in which a smaller k value is configured. Here, only a few eNBs in the neighbourhood of the moving UE will form the k-NN list. The association with closest eNBs will result in an improved QoS. However, a linear increase in the handover frequency, as well as high switching overhead and signalling load due to vehicle mobility and short coverage area of small cells, is monitored, as seen in Figure 5.1(a). On the contrary, if k is configured to be too large then most of the eNBs in the neighbourhood of the moving UE will be included. However, as shown in Figure 5.1(b), the association with farthest eNB would significantly reduce the handover frequency, but, at the expense of poor network QoS [6].



(a) Frequent handover in case of small scanning diameter.



(b) Severe cross-interference with large scanning diameter

Figure 5.1: Figure demonstrating the relationship between association range, handover rate and network QoS in ultra-dense small cell highway scenario

An emerging state-of-the-art technique for intelligent user association for uncertain environments is Reinforcement Learning (RL). There is evidence in the literature for the application of RL in continuous environments. Particularly, Watkin's Q-learning, as proposed in [105], is one of the most widely used RL algorithms in both wireless communication and other artificial intelligence domains [160]. However, there is no evidence in the literature for using RL to learn the best k value for user association in ultra-dense small cell dynamic vehicular environments, such that an equilibrium between handover frequency and guaranteed network QoS across different mobility levels may be achieved.

The researchers in [161] presented a speed-dependent bias factor approach to address the need for optimum cell association and hand-off management in a dynamic environment. It has been claimed that, by using a multi-level threshold handoff scheme that depends on the mobile velocity, a better performance could be achieved. The results demonstrate the performance metrics obtained using different quantizer thresholds. In [162], velocity adaptive algorithms for a micro-cellular scenario are investigated. Subsequently, the researchers in [163] derived the probability of coverage and co-operation between eNBs using cooperative transmission scheme. Moreover, vehicle handoff rate and a vehicular overhead ratio are employed to evaluate the vehicular mobility performance in a co-operative small cell environment. The paper concludes by claiming that an optimal overhead ratio can be achieved by adjusting the cooperative threshold. A common drawback of all of these schemes is that they lack the capability to self-configure and self-optimize the current policy online. Thus, it is imperative to develop intelligent algorithms for user association in ultra-dense small cell dynamic vehicular environments that have ability to continuously self-configure and self-optimize the nearest neighbour list by utilizing variation in the radio environment such that the numbers of handovers per transmission are significantly reduced whilst a guaranteed network QoS is delivered.

5.3 Heuristic Approaches for User Association

This section briefly revisits the baseline approach; maximum distance user association with a fixed k value, in the LTE uplink, using the framework of an urban vehicular environment. Following this, a real-time controlled feedback user association approach that is motivated by control-theoretic models and dynamic programming is discussed.

5.3.1. Maximum Distance Approach

A flowchart in Figure 5.2 illustrates the user association, re-association and resource assignment mechanism using the fixed k value. In the case of minimum distance approach, the value of $k = 1$. Meanwhile, for approaches such as minimum load, maximum distance, etc. the value of k depends on the cluster size. In this approach,

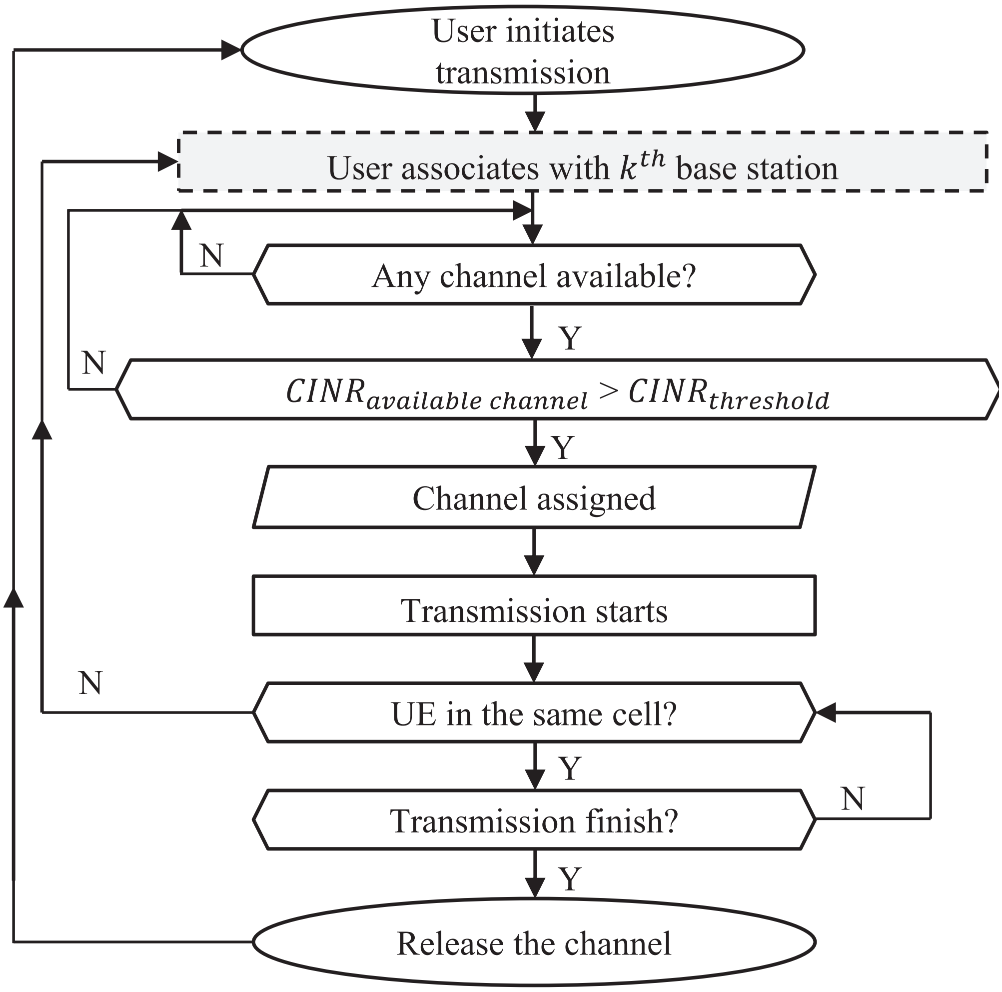


Figure 5.2: Flowchart of conventional user association approach.

upon arrival of a file, a k-NN list is configured. The number of eNBs that form the k-NN list is fixed and/or pre-defined, therefore, this scheme is also referred to as the fixed range user association approach. A three-step sequence algorithm utilizing **the maximum distance approach** in Step III of the algorithm, that has been discussed in detail in Chapter 4, is used to choose the most appropriate eNB for UE association as this performance metric **reduces the number of handovers per transmission considerably** compared to the minimum distance user association approach. UEs associate with the farthest eNB among the selected eNBs in the direction of vehicle mobility. This eNB may either assign a resource to start a new transmission or block it depending on the availability of ideal channel and channel quality. Equation 3.4 is used to evaluate channel quality of all the available channels on associated eNB.

As the UE moves along the street a) the k-NN list is continuously updated with new eNBs b) it is recursively checked if the current eNB to which the UE is associated is still an option in the k-NN list. If the current eNB is still one of the options in the updated k-NN list, the UE remains associated with the same eNB. Consequently, frequent handovers do not occur. Nevertheless, the channel parameters such as the SNR, the SINR and the data rate are updated. If the SINR of the assigned resource on the associated eNB drops below the threshold, then a new resource on the same eNB is identified and assigned to the ongoing transmission. If no resource is available, then the transmission interrupts. So, a transmission is interrupted if a) there is no available resource on the associated eNB; b) if the link quality of all available resources is below the assumed threshold of 1.8 dB or c) if the new transmission is likely to drop an ongoing transmission in the system. An interrupted file will re-transmit following a random exponential back-off time [132]. A maximum of five re-transmission attempts is performed before a file is categorized as an unsuccessful transmission and thereafter eliminated from the system. However, if the previous eNB is not an option in the k-NN list then a handover occurs. The ongoing transmission is handed over to a newly identified eNB in the direction of vehicle mobility. If the vehicle reaches the system boundary with no eNB in the direction of vehicle mobility, then the minimum distance user association rule holds true. The drawback of the fixed k user association approach is that it lacks autonomy. It is unable to restructure its current policies in a changing environment such that an equilibrium between different performance metrics and guaranteed network QoS may be established. To address this problem, a real-time control feedback user association approach is discussed next.

5.3.2. Real-time Control Feedback Approach

The distance-dependent path loss and the high attenuation due to the temporal-spatial vehicle behaviour significantly affects k value selection in dynamic interference-limited vehicular environments. Therefore, using a fixed k value to provide guaranteed QoS with minimum handover frequency per transmission at different mobility levels in such environments is extremely challenging. The real-time control feedback policy is a *computational* algorithm that does not need any prior network information for k value evaluation. It is a single-step input-output mechanism inspired by control-theoretic models and dynamic programming [177]. In [168], it is suggested that dynamic programming is more amenable to incremental planning as it considers only one action at a time rather than the entire action space. Our approach, shown in Figure 5.3, represents such a one-step input-output scheme. The algorithm derives the system probability of re-transmission, P_{retx}, as the input and produces the association range, k, as the output.

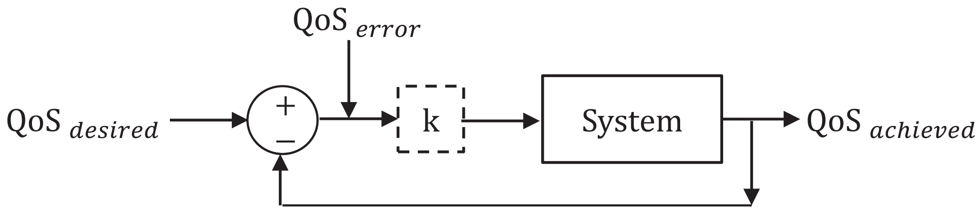


Figure 5.3: Block diagram of the real-time control feedback approach

The system P_{retx} is chosen as the performance metric to control the k-value variation. To compute an appropriate k-value, the P_{retx} was measured continuously after every successful transmission, P_{inst}, and periodically over a batch of successful transmissions, P_{mean}. The periodic P_{retx} calculated over a batch of successful transmissions successfully eliminated the impact of short-term variations and computed a more effective k value.

Upon arrival of the first transmission in the system, a k-value is randomly selected from the assumed action value set. The action value set is an array of an assumed number of eNBs, wherein, the assumption is based on the results obtained in Chapter 4. Depending on the value selected, the neighbourhood of the moving UE is scanned to construct the k-NN list. The UE associates with the maximum-distance eNB, in the direction of vehicle mobility, as in the conventional maximum distance association approach. A transmission retains the evaluated k-value until it finishes, irrespective of the multiplicity handovers it may experience as the vehicle moves along. Algorithm 4 represents, in detail, the developed real-time control feedback approach to compute an appropriate k-value for user association.

Algorithm 4 Real-time control feedback approach for user association

Input: The probability of re-transmission threshold value, $P_{threshold}$
Output: The computed SD range, k-value
while $now_{time} < end_{time}$ **do**
  for the first transmission **do**
    Randomly select an action value k, from the assumed action space
  end for
  for all subsequent file transmission arrivals **do**
    Exploit the computed k-value
    Associate UE to the k^{th} eNB
  end for
  Calculate the overall network probability of re-transmission, P_{retx}
  if $P_{retx} > P_{threshold}$ **then**
    Decrement the k-value by 1 towards the minimum assumed association range
  else
    Increment the k-value by 1 towards the maximum assumed association range
  end if
end while

Although it is a computational algorithm rather than a learning one, it still accomplishes better system performance compared to the fixed SD range policy, across different traffic conditions, based on metrics such as minimum handover frequency and a guaranteed QoS. An emerging state-of-the-art technique that even more effectively addresses the problem is reinforcement learning.

5.4 Reinforcement Learning Approaches for User Association

A common disadvantage of heuristic user association algorithms, such as the fixed k-value and the real-time control feedback approach is their inability to learn the best k value as they lack **autonomy**. These schemes, therefore, are unable to construct hypotheses directly from a changing environment. An emerging state-of-the-art technique for intelligent user association for such uncertain environments is Reinforcement Learning (RL) [102].

RL is a branch of artificial intelligence: a class of machine learning that employs a reward and punishment policy to enable an agent to reach a good solution during its interaction with the environment through trial-and-error [100]-[102]. In our presented multiagent environment, the task of RL is to learn an action, the k value, solely through trial-and-error and with no prior knowledge of the network itself or of network performance. All the agents in the environment are independent learners with no coordination between each other, but with an aim to learn, jointly, a common best k-value.

The four key elements of RL are a) the environment b) an action c) a state d) a reward function. In this thesis, the following terms have the following specific meanings:

- Environment - The dynamic vehicular network.
- Action - The number of neighbouring base station, k value.
- State - Here we use a simplified, stateless Q-learning model.
- Reward - The positive or negative reinforcement signal.

One popular RL technique in widespread use is Q-learning (QL). Here, a centralized array known as the Q-table is maintained. The values in the Q-table are called the Q-values and are initialized to zero, allowing the agent to start to learn with an equal choice among all available actions. The agent uses the *learning policy* to learn an action, whereas, the *update rule* is employed to update the Q-value associated with each action. The Q-table, therefore, presents an analysis of the choice of behaviour of all the individual agents, whereas, the Q-value represents the expected cumulative reward the agent receives by learning an action.

The learning policy: A learning agent adopts a policy or a value that guides it to learn the best solution. A ϵ-greedy exploration approach is used to learn the best action among the available actions in the action space. The approach states that an exploratory random action is picked with probability ϵ otherwise a good policy action (greedy) is selected with probability 1-ϵ. The greedy action is selected using Eq.5.1.

$$k = \arg\max(Q(k)) \tag{5.1}$$

The learnt action, k, has the highest Q-value in the Q-table [169]. A persistent exploration learning policy is assumed throughout the simulation experiment which helps the learning agent to continuously update its policy in a changing environment.

The update rule: A learning agent recursively updates the Q-value of each learnt action k, using the stateless Q-learning update equation:

$$Q(k) = (1 - \alpha)Q'(k) + \alpha r \tag{5.2}$$

k is the learnt action value from the action space. $Q(k)$ corresponds to the updated Q-value of the learnt action value k. $Q'(k)$ is the previous Q-value of the action k. The learning rate parameter, α, controls the convergence rate, and r is the reward awarded to an action determined by the reward function. The discount factor γ is assumed as 0, in this algorithm, therefore, the component is not included in Equation 5.2.

5.4.1. Conventional Q-Learning Approach

In this section, learning an appropriate association range k for user association using the conventional QL approach is discussed. Being model-free, QL does not require any prior information of the model or the reward function. The analysis of the action selection k explicitly represents the influence of reward and action selection history on future choices. The authors in [170] proposed a single-state QL scheme for the dynamic resource management in a 5G high capacity density network. The scheme utilizes spectrum access information through spectrum sensing to learn an optimal assignment decision. Their results demonstrate that, without any frequency planning, the algorithm converged quickly towards an optimal solution, delivering improved quality of service and system capacity. In [171], a case-based RL scheme is proposed for dynamic spectrum assignment in a cellular network. The scheme is a combination of RL and case-based reasoning. It achieves improved temporal performance for each base station as well as delivered an improved network QoS. The application of a classical Q-learning approach to learning an effective k-value across different mobility levels in an ultra-dense small cell dynamic vehicular environment is discussed next.

Upon arrival of a file, the learning agent performs ϵ-greedy action selection strategy to learn an action value from the available action space. The learnt action value is utilized to scan the neighbourhood of the moving UE to form the k-NN list. The UE then associates with the maximum distance eNB in the direction of vehicle mobility.

The flowchart in Figure 5.4 outlines the learning process using classical Q-learning.

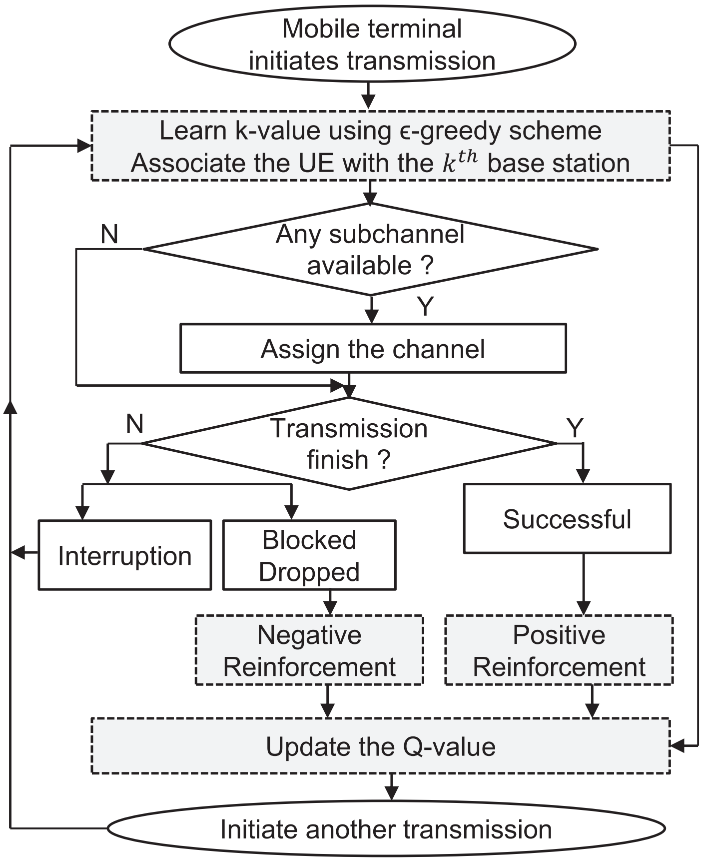


Figure 5.4: A classical Q-learning approach to learn scanning diameter range for user association in a small-cell vehicular environment

The Q-value of each learnt action is updated recursively using Eq. 5.2. The reward function for the Q-value update is

$$r = \begin{cases} +1 & \text{if a transmission is successful} \\ -1 & \text{if a transmission is not successful} \end{cases}$$

The two key parameters in the stateless Q-learning update equation are the learning rate α and the reward function r. A comprehensive study on the relationship between the different RL model parameters has been reported in [172]. The influence of reinforcement history on the choice behaviour was also discussed. In the next sections, these two parameters are investigated discretely, as Win or Learn Fast Q-Learning (WoLF-QL) and Variable Reward Q-Learning (VR-QL) approach, to understand their impact individually on the learning process in the considered scenario.

5.4.2. Win or Learn Fast Q-Learning Approach

The Win-or-Learn Fast Q-learning (WoLF-QL) is an agent-aware method to encourage learning convergence [173] [114]. The principle of "win or learn fast" states that, with a variation of the *learning parameter*, α, an agent can learn quickly in case of a learning trial failure and slowly when winning. The authors in [174] investigated the WoLF policy in a multi-agent learning problem environment. The two different properties examined were the rationality and convergence. The paper discussed how the RL failed to simultaneously meet both the criteria. Therefore, proposing the WoLF policy hill-climbing approach that proved to be rational and demonstrated convergence. The WoLF-QL is a proven technique to achieve convergence in many stochastic games.

Here, upon arrival of a file, an agent learns a k-value using a low learning rate of α_{win} if transmission is successful. To learn an effective k-value for user association across a range of vehicle speeds, in the considered scenario, using the WoLF principle, the learning rate is split into two parts α_{win} and α_{lose} ensuing the WoLF principle that is $\alpha_{win} < \alpha_{lose}$.

$$\alpha = \begin{cases} 0.01 & \text{if transmission is successful} \\ 0.05 & \text{if transmission is not successful} \end{cases}$$

The authors in [175] presented a detailed explanation of the significance of choosing the correct learning rate and its impact on convergence by presenting a performance comparison and sensitivity analysis considering different values of learning rate. Their simulation results demonstrate that using the WoLF variable learning rate algorithm, it is possible to achieve a robust and consistent QoS using distributed QL when compared with a conventional opportunistic spectrum sensing approach. The assumed learning rate value for the presented research are assumed based on their work.

An action value is rewarded with an award r and learning rate equal to α_{win}, if a transmission is successful. However, a learnt choice will be penalized if that results in an interrupted transmission using α_{lose} regardless of the system QoS that is calculated after every successful transmission. The WoLF-QL approach in the presented scenario, therefore, investigates the influence of the learning rate on the handover frequency and the system performance whilst learning an appropriate k-value. The scheme learns an appropriate association range value that delivers successful transmissions rather than learning the global optima. This is because the WoLF-QL approach is unable to perceive continuously changing network performance information, hence converges to a local optimum.

5.4.3. Variable Reward Q-Learning Approach

The impact of the *reward* function to learn the highest possible association range for user association is investigated using variable reward Q-Learning (VR-QL) approach. The approach is based on the High-risk, High-reward scheme. The reward value r in the VR-QL approach is directly proportional to the learnt action value k i.e., if an agent learns $(k = 3)$ as an appropriate range for user association and the data transmission with this learnt action value resulted in a success, then this learnt action is awarded with a reward $(r = +3)$. However, it is punished with a reward, $(r = -3)$, if it results in an interruption.

This scheme is also model free so *a priori* knowledge of the model or the reward function is not essential. The authors in [176] proposed a variable reward approach to solving a transfer learning problem to achieve faster learning in a task under hierarchical settings. An assumption of the different reward functions in a set of reward features was made. Their simulation results demonstrate that the proposed scheme compactly stores the optimal value functions for several Semi-Markov Decision Process (SMDP) that were subsequently used to optimally initialize the new SMDP value function.

Here, upon arrival of a file, a ϵ-greedy action selection approach is used during the learning process. The Q-value of each learnt action is recursively updated using Eq. 5.2. If the learnt action delivers a successful transmission then the learnt action is awarded a reward equal to the learnt action value. However, if the learnt action k results in an interruption, then this action is penalized with a value equal to the learnt action value; $(r = -k)$. The reward function for the Q-value update is as follows

$$r = \begin{cases} +\text{k} & \text{if transmission is successful} \\ -\text{k} & \text{if transmission is not successful} \end{cases}$$

A common drawback of the conventional QL, WoLF-QL and the VR-QL algorithm is that they learn the k-value only through trial-and-error. To learn an appropriate solution in a model free, continuously changing environment takes a large number of trials, a characteristic undesirable in real-time applications. Also, the investigated schemes lack the ability to *assure* that the learnt action is the global optimal or not. To mitigate this problem, the variable reward quality-aware Q-learning (VR-QAQL) algorithm for user association in an ultra-dense small cell dynamic vehicular environment has been proposed in the next section.

5.4.4. Variable Reward, Quality Aware Q-Learning Approach

The Variable Reward, Quality Aware Q-Learning (VR-QAQL) algorithm is a model free, intelligent approach that follows *learn-execute-evaluate-formulate-improve* principle to learn the best k value for user association. The proposed algorithm recursively performs policy evaluation and policy improvement processes throughout the simulation experiment. The policy evaluation is performed using the VR-QL approach while the real-time control feedback scheme is utilized for policy improvement.

In [177], the authors proposed a heuristically accelerated Q-learning approach to improve initial network performance, achieve quick convergence and enhance steady-state performance during dynamic spectrum access in the LTE downlink. The proposed scheme is an integration of distributed RL and standardized inter-cell interference coordination signalling. The simulation results show that the heuristic information guides the learning agent to deliver better QoS, support higher network throughput densities and speed up convergence.

The key difference in VR-QAQL approach from the other investigated approaches is in the action evaluation process. A learnt action is awarded a reward if both the criterion is satisfied; a) the transmission is successful b) the overall network probability of retransmission is less than the assumed threshold. The heuristic information, i.e. the network probability of re-transmission, is computed after each transmission during the simulation. This guides the evaluation function to evaluate every learnt action with respect to the transmission status and network QoS. The flowchart in Figure 5.5

demonstrate learning the best action value using the VR-QAQL approach.

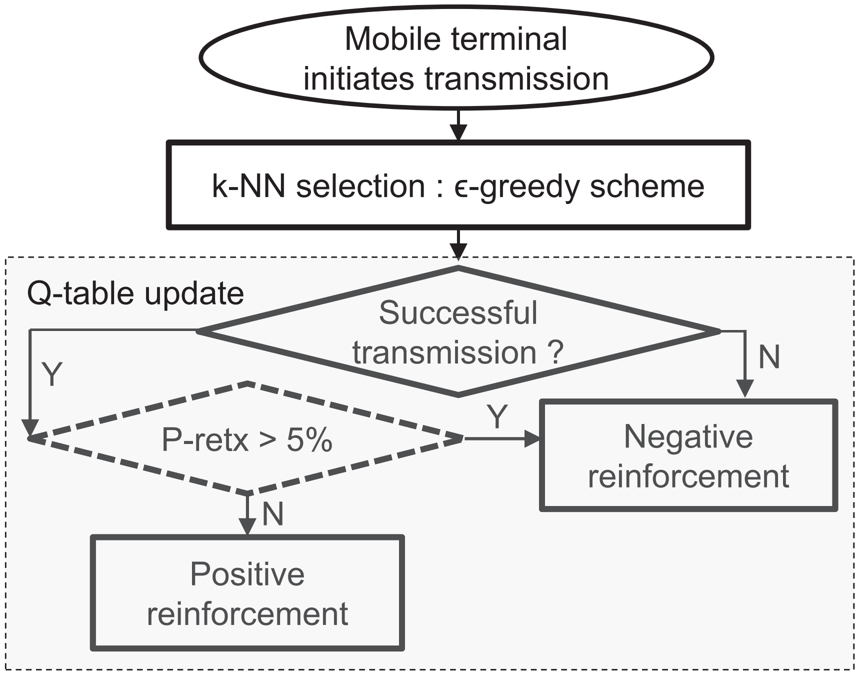


Figure 5.5: Variable reward quality aware Q-Learning approach to learn an appropriate k value for user association in ultra-dense dynamic vehicular environments

The VR-QAQL approach, therefore, evaluates a learnt action as follows

- Punish the action learnt if it results in an interruption or it increases the overall system P_{retx} above a threshold.
- Do this, but award it if and only if a) the transmission is successful without any interruption b) the overall system P_{retx} is less than the assumed threshold of 5%.

In the proposed VR-QAQL user association approach, evaluating a learnt action based on its transmission status as well as the overall network QoS pushes the learning agent to improve its policy recursively and not to select an action value which might have been proven to be promising in the early simulation phase. The Q-value of each learnt action is updated using the stateless Q-learning at Eq.5.2. The strong correlation between the policy evaluation and policy improvement process results towards learning the best action value i.e. the user association range.

5.5 Simulation Results

In this section, performance comparison between the non-learning and Q-learning based user association schemes are presented. The performance metrics used to evaluate the investigated algorithms are the probability of re-transmission, the end-to-end delay, the learnt k-value, and the number of handovers per transmission. Results are based on an average of 25 different random seed simulations to present a statistically representative temporal response. Error bars show the maximum and the minimum probability in those simulations. The UE in the moving vehicle employs the maximum distance approach to associate with an appropriate eNB for data transmission, as in Section 5.3.1 it was seen that the scheme demonstrated to reduce the number of handovers per transmission among other investigated association metrics. The two test scenarios that are used to represent the baseline results for user association, using the fixed k approach, in a vehicular environment are

1. UE association assuming fixed association range, k= 1.
2. UE association assuming fixed association range, k= 4.

In the first test case, when the k-value assumed is small (k=1), only one eNB is selected. The aim is to achieve high spectral efficiency. This eNB is the closest eNB to the moving user, therefore, may also be termed as the minimum distance eNB. Association with the minimum distance eNB results in a significant reduction in the path loss and cross-interference level, leading to better QoS at the expense of a linear increase in the number of handovers per transmission, as seen in Figure 5.6 and Figure 5.7 respectively. A linear increase in the handover frequency is due to the frequent handovers experienced by the UE due to a short coverage area of the small cells and increasing vehicle speed. The minimum distance approach, therefore, outperforms all the other explored user association approaches in terms of QoS. However, in terms of the number of handovers per transmission, it performs poorly.

On the contrary, a larger k value is assumed, ($k = 4$), in view to decrease the number of handovers per transmission. Here, the 4 nearest eNBs to the moving UE form the k-NN list. On initiating a transmission, the UE associates with the farthest eNB in the

k-NN set in the direction of its mobility. It remains associated with this eNB until it is no longer one of the options in the k-NN list. An increased k value dramatically reduces the handover frequency at the expense of increased cross-interference level that results in a degraded network performance. The drawback of using a larger k-value is that it fails to provide a guaranteed QoS across a majority of the mobility level. This is because an increase in the association range increases the interference level that results in an existing transmission being interrupted; consequently degrading the network QoS.

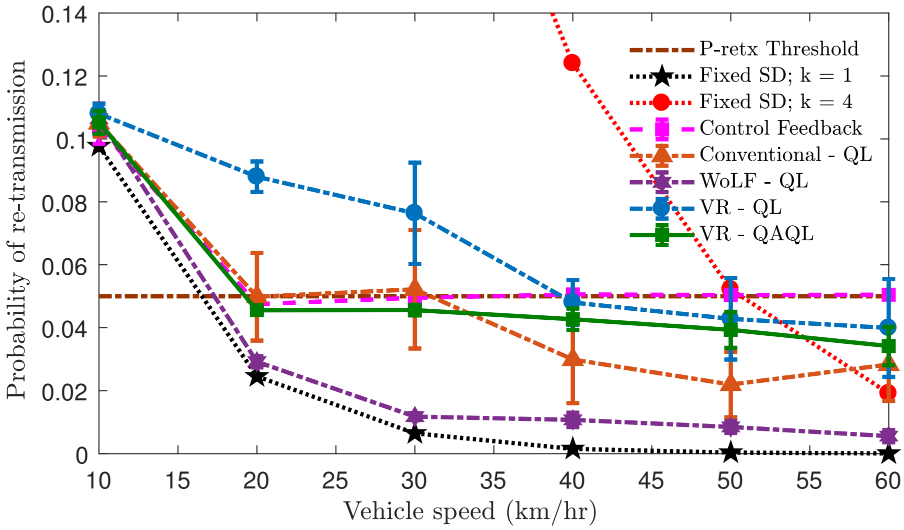


Figure 5.6: The probability of retransmission performance vs vehicle speed using the different learning and baseline schemes

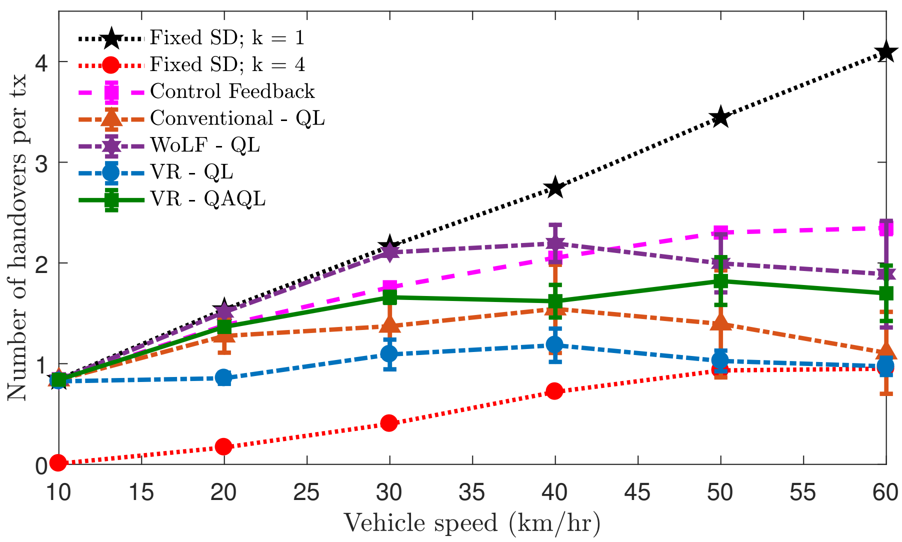


Figure 5.7: The number of handovers per transmission at different vehicle speed using learning and baseline schemes

Meanwhile, in the case of real-time feedback approach, the k-value is not assumed, but, computed during the simulation by using the current network QoS as heuristic. The temporal variation in the k-value utilizing the real-time control feedback approach is shown in Figure 5.8, while the CDF plot of learnt association range is presented in Figure 5.9. At high mobility levels, the vehicle spatial distribution is sparse - the vehicle density per cell is low as the safety distance between vehicles increases. This reduces the cross-interference level resulting in better network QoS that guides the learning agent to compute a higher k-value. However, learning a higher k value, causes the network QoS to degrade due to strong interference. Therefore, a continuous P_{retx} evaluation is performed that guides the algorithm to compute a smaller k-value such that a guaranteed QoS is maintained. Likewise, at low mobility level a vice versa procedure takes place.

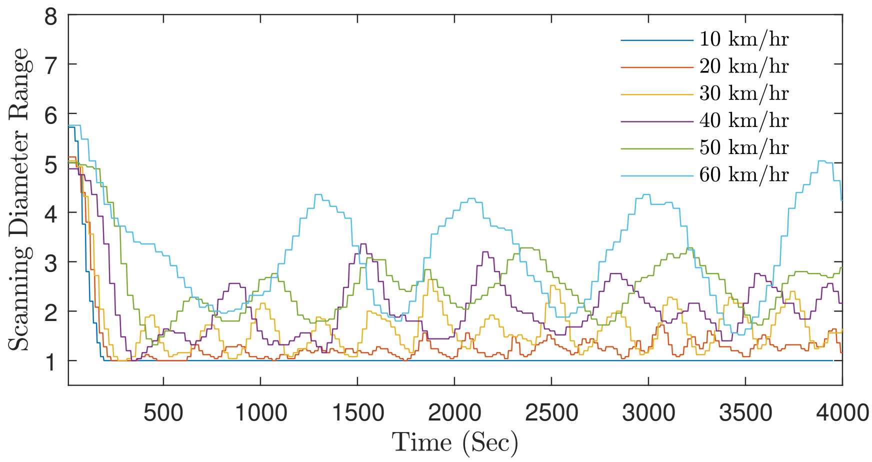


Figure 5.8: Temporal performance plot demonstrating k value variation.

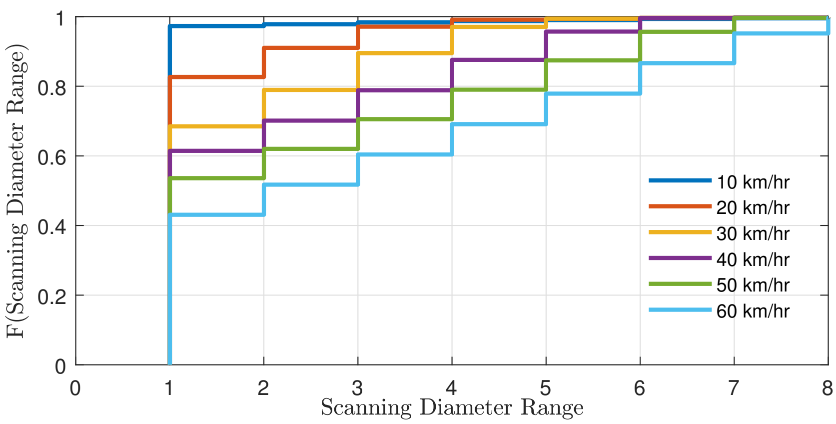


Figure 5.9: CDF plot of learnt k value using the real-time control feedback user association approach.

Figure 5.10 presents the k value learnt across different vehicle speed under distinct user association approaches. Here, the real-time control feedback approach depending on the current network performance was able to successfully computed an operational k-value such that a better network performance as well as reduced handover frequency may be achieved. The advantage of the feedback approach over the fixed range method is that it evaluates and reconstruct its association strategy recursively in accordance with the changes in the environment. In addition to this, as seen in Figure 5.6, the feedback algorithm across all the mobility levels demonstrated to deliver a guaranteed QoS. Subsequently, the result in Figure 5.7 demonstrate that at a vehicle speed of 60 km/hr, the number of handovers per transmission reduced by 40-42% when compared to the minimum distance user association; k=1.

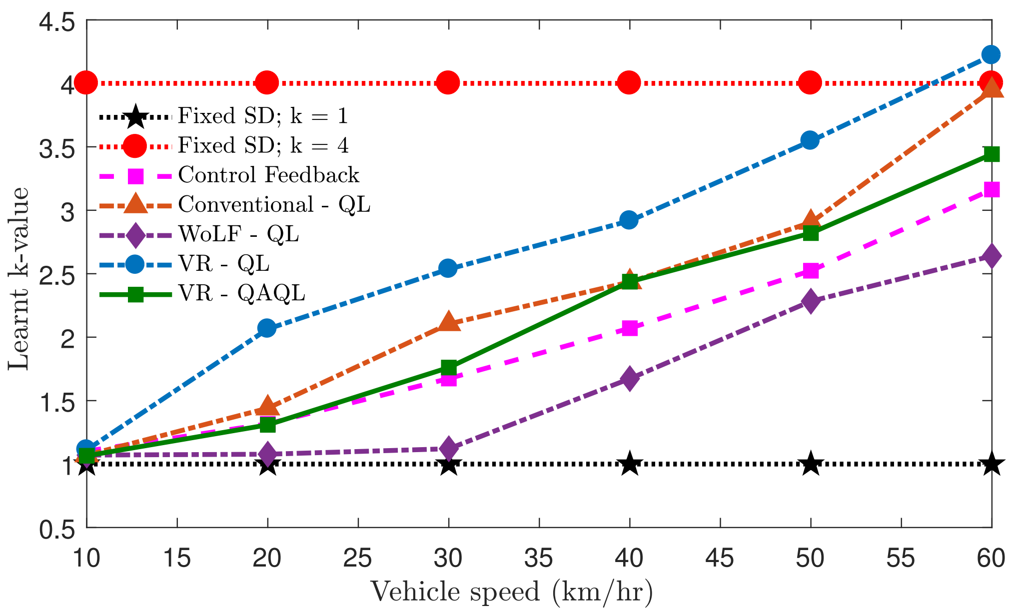


Figure 5.10: Comparisons of the investigated Q-learning approaches and the heuristic algorithm on k value learnt at different vehicle speed in an ultra-dense dynamic vehicle environment

Meanwhile, the conventional-QL, the WoLF-QL and the VR-QL approach learn an action value purely through trial-and-error interactions. As seen the conventional Q-learning scheme especially at higher vehicle speed significantly reduces the number of handovers per transmission, however, it fails to deliver a guaranteed QoS at low mobility levels. This is because the scheme relies only on the status of the transmission to learn the appropriate solution over a considerably high number of trials. It learns an action value that delivers successful transmissions but is unable to assure the reliability of the learnt action value. The conventional-QL, therefore, fails to simultaneously meet both criteria; i.e., to achieve minimum handover frequency and deliver a guaranteed network QoS. The WoLF approach, on the other hand, utilizes variation in the learning parameter (α) to learn the best k-value. The scheme successfully delivers better network QoS, almost identical to the k=1 association approach, but fails to reduce the handover frequency at the high mobility levels. This is because the WoLF algorithm aims to guarantee a quick convergence rather than learning the highest possible association range value.

The comparison of the VR-QAQL approach with these schemes is most appropriate. It is evident from the obtained results that by utilizing additional heuristic information to learn an effective k-value compared to the pure trial-and-error learning schemes, the VR-QAQL approach significantly outperforms all the other algorithms by learning a stable solution that reduces the number of handovers per transmission significantly whilst delivering a guaranteed network QoS. Figure 5.11 presents the CDF plot of learnt association range.

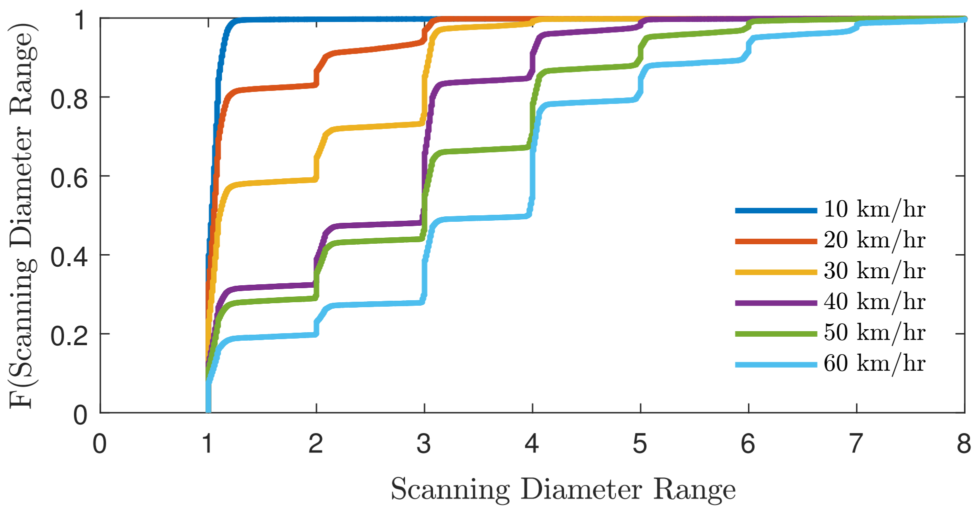


Figure 5.11: Cumulative Distribution Function of learnt k value using VR-QAQL approach.

The temporal variation in the k-value utilizing the VR-QAQL user association approach in 5.12 demonstrates that the algorithm adapts to the dynamic environment quickly by recursively reconstructing its policy, thus, is able to converge to a good solution in less than 500 sec from the start of the simulation across different investigated vehicle speed. This is due to the derived additional network information, during the simulation experiment, guides the exploration process to evaluate every learnt action rationally; meanwhile the variable reward function forces the learning agent towards the global best. On the contrary, as seen in Figure 5.8, the learning agent based on control feedback mechanism oscillates between the upper and lower bound, thus, is never was able to learn a good solution as the approach *computes* an appropriate solution depending on network dynamics.

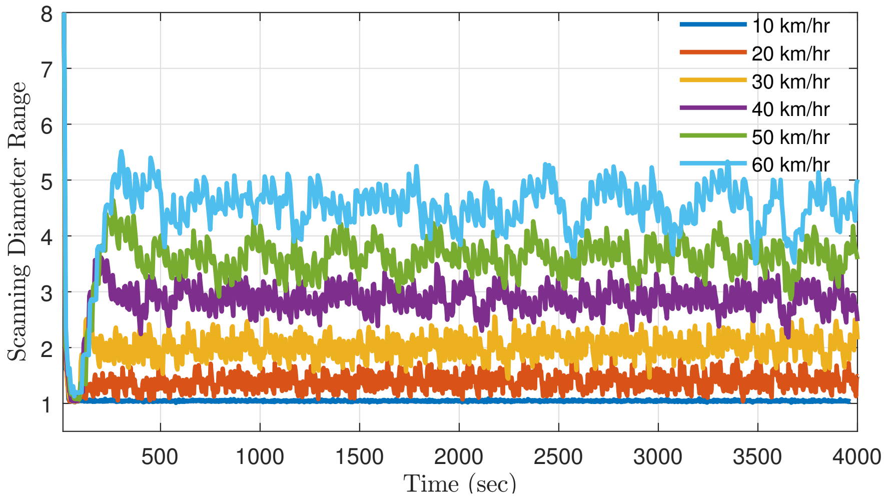


Figure 5.12: Temporal variation in learnt k value using VR-QAQL approach.

5.6 Conclusion

In this chapter the significance of application of reinforcement learning approaches for user association in vehicular networks was presented. Following this, the stateless Q-learning, the Win-or-Learn-Fast Q-learning, Variable Reward Q-learning and the variable reward quality aware Q-learning approach were introduced. The simulation results demonstrated that integrating machine learning approaches with heuristic user association algorithm resulted in significant reduction in the number of handovers per transmission. Among the different learning techniques, the conventional Q-learning, the WoLF-QL and the VR-QL approach learnt k-value purely through trial-and-error, irrespective of the network QoS, therefore, were unable to learn an optimal solution. On the contrary, the VR-QAQL algorithm outperforms all the other investigated user association approaches as it combines the variable reward approach with real-time control feedback policy in a conventional Q-learning framework. The VR-QAQL user association approach, thus, learnt an effective association range thereby reducing the number of handovers per transmission significantly whilst delivering a guaranteed network QoS across all mobility levels. Statistically, the VR-QAQL scheme reduced the number of handovers per transmission by about 70-74% at 60 km/hr with marginal compromise on the network QoS when compared with max-RSS approach.

Chapter 6. Case-Based Reinforcement Learning for User Association

Contents

6.1 Introduction

The purpose of Chapter 5 was to investigate improvements to the conventional user association approaches by employing intelligent Q-learning algorithms. The model-free, offline learning-based user association learnt appropriate values corresponding to different vehicle speed in the rapidly changing environment, but, lacks to provide an assurance that the learnt action value is a reliable global solution. This was due to a number of factors such as continuous variation in traffic, vehicular distribution, active network load per eNB, fluctuating channel quality, rapidly changing network topology that significantly limited the learnt solution reliability in dynamic environments.

The self-organization and self-configuration of network architecture, as discussed in Chapter 2, therefore is substantially important for an effective user association in dynamic networks. Few promising solutions to this issue, proposed in the artificial intelligence domain, is case-based transfer learning (CBTL), and heuristically accelerated reinforcement learning (HARL). In the definition given by Broad Agency Announcement (BAA) 05-29 of Defense Advanced Research Projects Agency (DARPA) in 2005, transfer learning aims to extract the knowledge from the source task and transfers the gained knowledge to a target task with an aim to converge to a better, consistent and a reliable solution [118]. Whereas, heuristics as defined in [180], is information that is either provided externally or derived during the simulation experiment in order to improve the learning agent performance.

The authors in [121] investigated the use of transfer learning for user association in a 5G aerial-terrestrial broadband access network. The scheme developed formulates a user association solution from perceived spectrum assignment knowledge in order to improve network QoS and reduce latency. Moreover, the authors in [181] proposed a graph-based method motivated by transfer learning for the identification of previously encountered games in order to automate domain mapping for value transfer function and speed up the RL on the different variation of previously played games. In [182], a Q-learning based network access selection scheme is proposed that uses a heuristic approach motivated by the concepts of simulated annealing to provide maximum convergence to mobile users, thereby, reach Nash equilibrium in dynamic environments. A comprehensive survey on the theory of implementation in Nash equilibrium is presented in [183]. In [184], Q-learning based cell selection method is proposed that utilizes past knowledge and behaviour of cells to predict their future behaviour, hence reducing handover frequency.

However, in the context of user association in ultra-dense small cell dynamic vehicular environments, there appears to be no evidence in the literature that implements case-based reinforcement learning technique (CBRL); a combined CBTL and HARL approach on a conventional Q-Learning framework, to self-configure and self-optimize the nearest neighbour list for identification of an appropriate eNB, such that a balance between different performance metrics whilst delivering guaranteed network QoS at different mobility levels may be achieved. This chapter, therefore, aims to exploit the CBRL approach to learn a reliable solution for user association in ultra-dense dynamic networks.

The rest of this chapter is organized as follows: In Section 6.2, the heuristically accelerated reinforcement learning is discussed. Section 6.3 presents an overview of case-based transfer learning and the motivation to implement it in the presented work. Furthermore, novel user association approaches that employ CBTL and HARL to learn the best k value for user association in dynamic environments are discussed in Section 6.4. The results obtained using the proposed schemes are compared with the conventional, max-RSS approach and other investigated RL schemes using some QoS metrics, such as the probability of retransmission, the handover frequency and the end-to-end delay in Section 6.5. Finally, in Section 6.6 conclusions are provided.

6.2 Heuristically Accelerated Q-Learning

The key elements of the heuristically accelerated Q-Learning (HAQL) approach are derived heuristics, that influence the choice of action, and Q-learning technique. Heuristics, as defined in [118], is an information that is either provided externally or derived during the simulation experiment in order to improve the learning agent performance. The heuristic information derived or extracted during the simulation experiment is termed as *Heuristic from Exploration* by the authors in [181] [185]. In the presented work, a policy is formulated and improved based on derived heuristic to accelerate learning of a global optimum. This information is derived recursively after a batch of successful transmissions using Eq.6.1.

$$H_t(s,a) = \frac{N_{retx}}{N_{tx}} \tag{6.1}$$

where, $H_t(s,a)$, is the heuristic information that relates to the probability of re-transmission, N_{retx} is the number of retransmissions and N_{tx} is the total number of transmissions in the system. The number of retransmissions represents the transmissions which are interrupted at least once. Whereas, the total number of transmissions is the summation of the number of successful transmissions and number of retransmissions.

The authors in [186], have used the heuristic policy with learning algorithms in Ant colony optimization to find solutions on symmetric travel salesman problem. The results obtained using this technique demonstrated to be at par to the results that would have resulted from using specialized heuristic approaches based on neural networks or local searches. The authors in [181] demonstrated that using a simple heuristic significantly improve the performance of reinforcement learning algorithm. Similarly, in [187], the authors use a case-based approach to transfer knowledge between different domains to accelerate the RL process. This algorithm utilizes different heuristic schemes to speed up the RL performance and build the knowledge base. The information from the constructed knowledge base is then transferred to a target domain resulting in a better network performance compared to traditional Q-learning.

Figure 6.1 shows the block diagram representation of HAQL. The top loop represents the conventional Q-Learning scheme that assists an autonomous agent to learn an action policy based on explicitly exploring its environment purely through trial-and-error interactions. However, the network performance information derived during the simulation that is used as a heuristic for the policy evaluation and improvement assist to evaluate the learnt action, therefore, reconstruct the current policy accordingly.

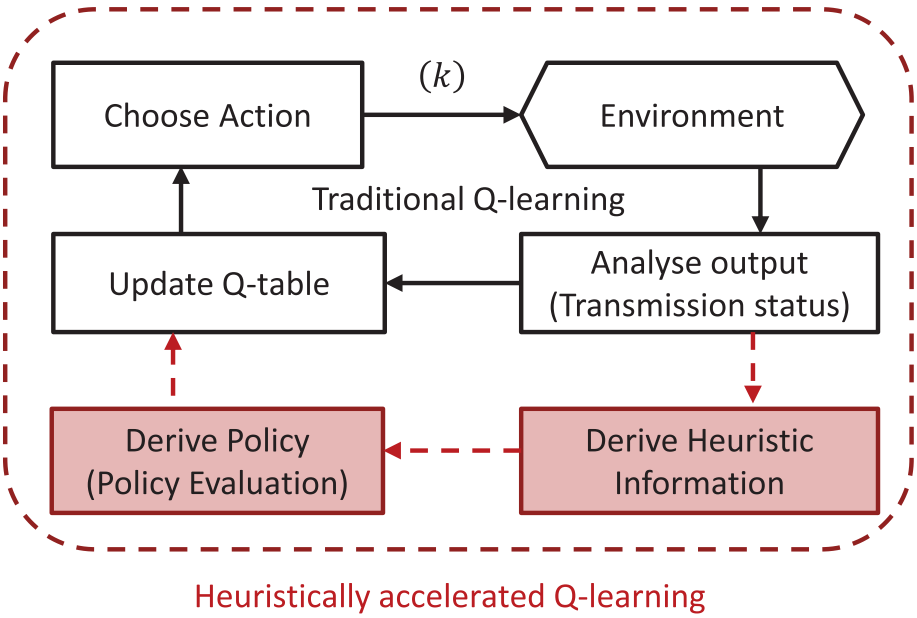


Figure 6.1: Block diagram of heuristically accelerated reinforcement learning

As emphasized in Chapter 5, a good policy $\pi : S \rightarrow A$ is learnt by observing action(s) in a particular state, analyzing the outcome and updating Q-table for further iterations. The improved policy, $\pi_{impr}(s, a)$, therefore depends on the heuristic derived. Moreover, a positive and negative reinforcement, $Q_{+}(s, a)$ and $Q_{-}(s, a)$ is used to update the policy using the following equation:

$$\pi_{impr}(s, a) = \begin{cases} Q_{+}(s, a), & H_t(s, a) < P_{retx}(threshold) \\ Q_{-}(s, a), & H_t(s, a) > P_{retx}(threshold) \end{cases} \tag{6.2}$$

As the heuristic information only influences the choice of action in a particular state, therefore, the policy is different from the conventional Q-learning scheme that specifically performs exploration by employing $\epsilon - greedy$ scheme for action selection. However, it does not modify the conventional Q-learning characteristic, i.e, the algorithm still exploits the free choice of training actions. The heuristic policy aims to speed up the learning process and thus learn a good action, k, that would not degrade the network QoS. Algorithm 5 details the HAQL approach.

Algorithm 5 Heuristically accelerated Q-learning algorithm

```
Input: Q-table, Q(s,a)
Output: The learnt action, k-value
while now_time < end_time do
    for the first transmission do
        Select an action value k, from the assumed action space using the action-
        choice rule, Eq.5.1
    end for
    for all subsequent file transmission arrivals do
        Select an action, a, using the action-choice rule, Eq.5.1
        Analyze the outcome, receive the reinforcement, r(s,a)
        Update the Q-table
    end for
    for number of successful transmissions > assumed batch size do
        Update the value of H_t(s,a) using Eq.6.1
        Update the Q-table, Q(s,a) using Eq. 5.2
    end for
end while
```

The next section discusses the case-based transfer learning technique that learns a solution based on previously gained knowledge.

6.3 Case-Based Transfer Learning

Figure 6.2 demonstrates the two different learning processes; traditional reinforcement learning and case-based transfer learning. A traditional reinforcement learning aims at learning the best solution using a *delayed reward* process. However, the goal of case-based transfer learning is to speed up the initial learning phase by *transferring* some of the gained knowledge from the source action, previous situations or cases, to a target action or case [188].

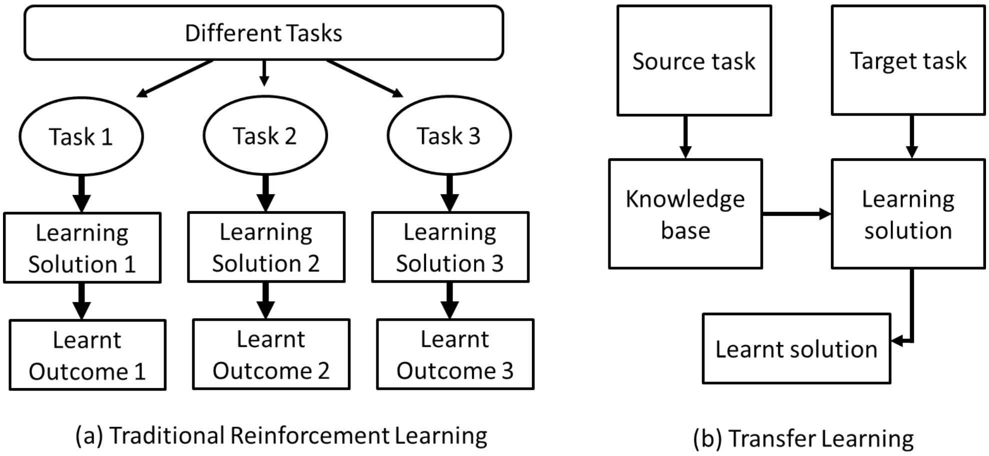


Figure 6.2: Figure demonstrating the difference between the traditional RL and the TL

In the case of the conventional RL scheme, as also discussed in previous chapters, the learning agents usually have no or very limited information about the policy or the environment when they are initially activated. The learning agent learns the solution on an almost random basis via trial-and-error iterations, as there is no concrete information available. The goal is to learn policy or value that maximize the positive reward in order to achieve global maxima, therefore by the principle of conventional RL, a good action have continuously increasing Q-values whereas the action value that leads away from the global optimum experience a decreasing Q-value. However, given a source and target domain/task, a case-based transfer learning approach as proposed in [189] [118] aims to accelerate the learning process by transferring some of the knowledge achieved from source task to target task. The source and the target task in the presented work are defined as the current and the prospective action value.

The three main research issues while employing TL are: (1) When to transfer; (2) What to transfer, and (3) How to transfer [118]. The first issue is addressed by the situation identification, i.e., when is transferring skills required. What to transfer relates to which part of the knowledge should be transferred across the domains/actions. Whereas, the last issue of how to transfer relates to the development of appropriate learning models that would be required to transfer the knowledge acquired. In the presented work, the three issues are addressed as follows:

1. *When to transfer*: The heuristically accelerated reinforcement learning policy is used to identify whether it is necessary to use transfer learning. If required, then the gained information is transferred using case-based transfer learning after periodic network performance assessment.

2. *What to transfer*: A conventional Q-learning approach is used to continuously update the Q-table after each transmission to build the knowledge base that guides towards learning an appropriate action value. The updated Q-table and the learnt action value is transferred to an appropriate target action value accordingly.

3. *How to transfer*: This relates to the development of models or algorithms to transfer the gained knowledge. This is performed by splitting the development process into two steps; value training method and value mapping method. The value training method is the first step wherein a model is trained such that it adapts to the environmental changes. Whereas, the value mapping method relates to transfer of knowledge using case-based reasoning to achieve the global maxima in the considered multi-agent, multi-action scenario.

The objective of case-based transfer learning is thus to accelerate the learning process by appending the Q-table with the transferred Q-value such that the Q-value on the good action is maximized whereas that on the bad actions is minimized. The learning agent, therefore, experiences a lower impact from the dynamic environment and is expected to learn a more adaptive, consistent and reliable solution. Moreover, a key benefit of employing transfer learning is that it requires less memory for storing the knowledge base as only one Q-table throughout the learning process is adequate.

6.3.1. Value Training Method

A value training method is designed to train a learning agent to adapt to the dynamic environments by constructing an initial knowledge base or source action. Here, the source action is the current *local* action value that is learnt using the conventional Q-learning and the target action is the *global* optima. From the perspective to learn solution with varying spatial-temporal distribution of the vehicles and thus active offered traffic load, a conventional Q-learning approach is used that continuously interacts with the environment, purely through trial-and-error policy during the value training loop to update the knowledge base; Q-table. A policy $\pi(s)$ that maximizes the expected reward, represented by Eq.6.3.

$$\pi(s) = \max_a \hat{Q}_\theta(s) \tag{6.3}$$

Where $\pi(s)$ is the policy in state s for a learnt action a, s represents the vehicle speed, $\hat{Q}_\theta$ is the parametrised Q-function that could be a linear function in the context of Q-learning or a non-linear parameter as in case of neural networks. In the presented work, the parameter θ is a linear parameter that represents the number of nearest eNBs. Eq. 6.3 therefore transforms into

$$\pi(s) = \max_a \hat{Q}_k(s) \tag{6.4}$$

The framework of the transfer learning value training method is illustrated in Figure 6.3.

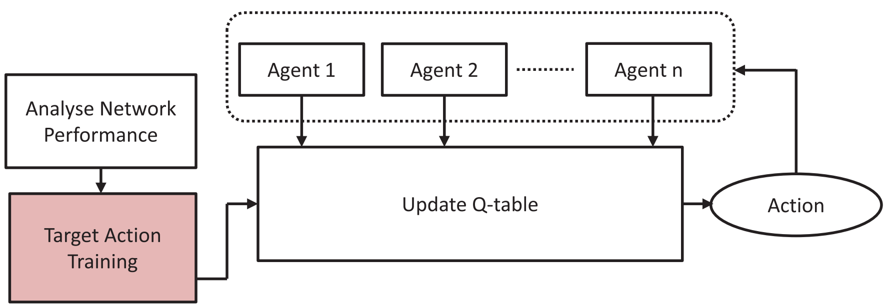


Figure 6.3: Block diagram of transfer learning value training method [190]

A continuous policy evaluation is performed using HARL after every transmission to monitor the network QoS. The heuristic obtained guides the learning agent to evaluate the current policy more carefully. Moreover, a periodic policy evaluation is performed after a batch of successful transmissions. The outcome of which determines whether or not a policy improvement is required. If further improvement to the current policy is needed then the value mapping loop is used to intelligently learn an appropriate solution by transferring the knowledge gained from the source action to the target action such that $\hat{Q}_\theta$ is close to Q^*, the optimal action value.

6.3.2. Value Mapping Method

The value mapping method is designed to map or transfer the knowledge gained during the value training method to an appropriate target action. The method not only accelerates the initial process of identification of the optimal solution but also reduces the impact of dynamic variations in the network performance due to temporal-spatial vehicle distribution and frequent radio environment changes. The identification of local maxima, k', in the Q-table is performed using Equation 6.5.

$$q(a(k')) = \max_a \hat{Q}_k(s) \tag{6.5}$$

The obtained q-value is then mapped or transferred from source action, k' to the target action, k, using the equation below:

$$q(a(k)) = q(a(k')) \tag{6.6}$$

where k is the prospective global maxima, $q(a(k))$ is the mapped q-value, k' is the action with maximum q-value, $q(a(k'))$, in the Q-table. The output is an updated Q-table that learns the best solution for user association, in the considered scenario, under the constantly changing environment. This updated table is used for learning during the conventional Q-learning loop before the next transfer learning loop initiates. The framework of value mapping method in the context of transfer learning is illustrated in Fig 6.4.

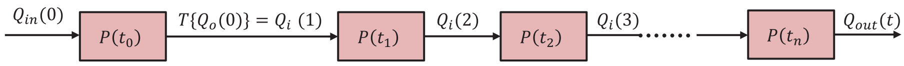


Figure 6.4: Block diagram of transfer learning value mapping method [190]

where, $Q_i(0)$ is the initial Q-table, t_n represents the number of iterations a policy undergo before a learning agent learns an optimal value, $P(t)$ represents the policy evaluation and improvement process, $T(Q_o(t_n))$ is the updated Q-table from the first iteration that serves as source task for the next iteration, $Q_{out}(t)$ is the final Q-table. The action with the maximum q value in this table represents the optimal solution.

6.4 Case-Based Reinforcement Learning Approaches

In this section case-based reinforcement learning approaches for user association in ultra-dense small cell networks are discussed. The algorithms combine the case-based transfer learning approach and heuristically accelerated reinforcement learning approach to learn the best solution. Figure 6.5 demonstrates the framework of the case-based reinforcement learning algorithms that follow a three-stage process.

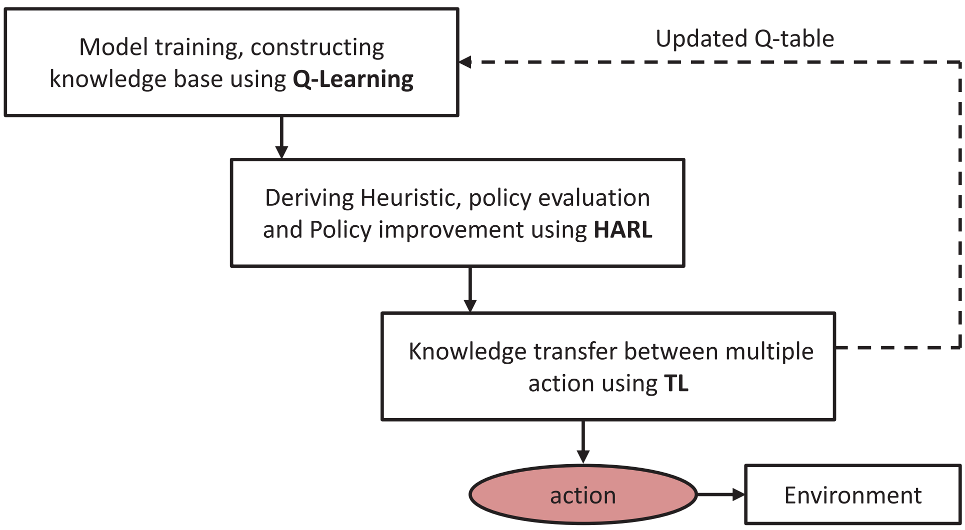


Figure 6.5: Block diagram of case-based reinforcement learning

In the first stage, the model is trained using Q-learning with an aim of maximizing the reward an agent receives during its interaction with the environment, thus, storing the values in a case base. In the next step, HARL is employed to derive the heuristics that performs policy evaluation and improves the current policy leading towards a global maximum. In the final step, the knowledge acquired in previous situations is used as heuristics and transferred between appropriate actions using the value mapping scheme of transfer learning following which the knowledge base, Q-table, is updated for future iterations.

Here, the two algorithms developed using the case-based approach are heuristically accelerated, variable reward, quality aware Q-Learning (HA-VR-QAQL) and heuristically accelerated, win or learn fast, quality aware Q-Learning (HA-WoLF-QAQL). The key difference between the developed strategies is that in the case of HA-VR-QAQL the reward function varies with the learnt action value while learning parameter, α is kept constant. Whereas, in the case of HA-WoLF-QAQL, the learning parameter is split into α_{win} and α_{loss} while the reward function is constant.

6.4.1. Heuristically Accelerated Variable-Reward Q-Learning

The heuristically accelerated, variable reward, quality aware Q-learning approach (HA-VR-QAQL) employs the case-base transfer learning and heuristically accelerated reinforcement learning to learn an appropriate solution for user association in dynamic environments. The model or reward function is not known to the learning agent; hence, the learning algorithm is classified as model-free. Moreover, the system relates to a multi-agent environment due to the co-existence of many independent learning agents in the same environment that simultaneously contributes to learning a common goal; an action value.

The concept of combining case-based reasoning to speed up RL by transferring some of the knowledge gained in the previous case to a new case was proposed for the first time by Drummond, [185]. The author emphasizes that extraction of heuristics at some abstract level in a related task may reduce the extensive re-learning effort and speed up the learning process. The heuristics obtained were used to formulate functions/policies as well as construct case-based knowledge base to produce a close approximation to the solution of a new task. The author concludes by stating that the combined function approximation algorithm produced a better solution compared to the basic reinforcement learning.

The authors in [188], developed a case-based HARL, (CB-HARL) technique, an extension of HARL algorithm, to solve a problem in target domains which include the 3D mountain car and the RoboCup 3D soccer simulator. The developed approach first identifies similarity in cases, a source and target case, based on a certain threshold prior to deriving heuristics using case-based reasoning and thereafter transferring the knowledge to learn a good solution. The heuristics were derived only when the similarity was above an assumed similarity threshold. The results demonstrated that the developed scheme outperforms other investigated learning algorithms such as SARSA(λ), TD(λ), and the conventional Q-learning that was used as a baseline approach.

In [191], the authors proposed an algorithm that uses case-based learning and reinforcement learning to enable distributed learning of behaviour policies as heuristics in a cooperative multiagent domain to learn a common goal without any communication between the learning agent. The approach was developed and tested by a series of experiments on a reactive job-shop scheduling domain. The outcome demonstrated that employing case-based reasoning with reinforcement learning the developed approach was able to produce a better solution.

Similarly, the concept of case-based reinforcement learning has been applied to the wireless communication domain in [171]. The paper presents the significance of the case-based reinforcement learning algorithm for spectrum assignment in a cognitive cellular network with dynamic topologies. The problem is classified as a distributed dynamic spectrum assignment with non-communicating eNB. This transforms the environment into a model-free, independent learning multi-agent environment. The developed learning based spectrum assignment algorithm was assessed empirically by measuring performance metrics such as the network blocking and dropping probabilities. The results demonstrated an improved temporal network performance when compared to the fully distributed spectrum assignment scheme.

Figure 6.6 presents a detailed HA-VR-QAQL user association algorithm flowchart that presents the integration of variable reward quality aware Q-learning approach and case based transfer learning policy to learn the best association range in a highway scenario.

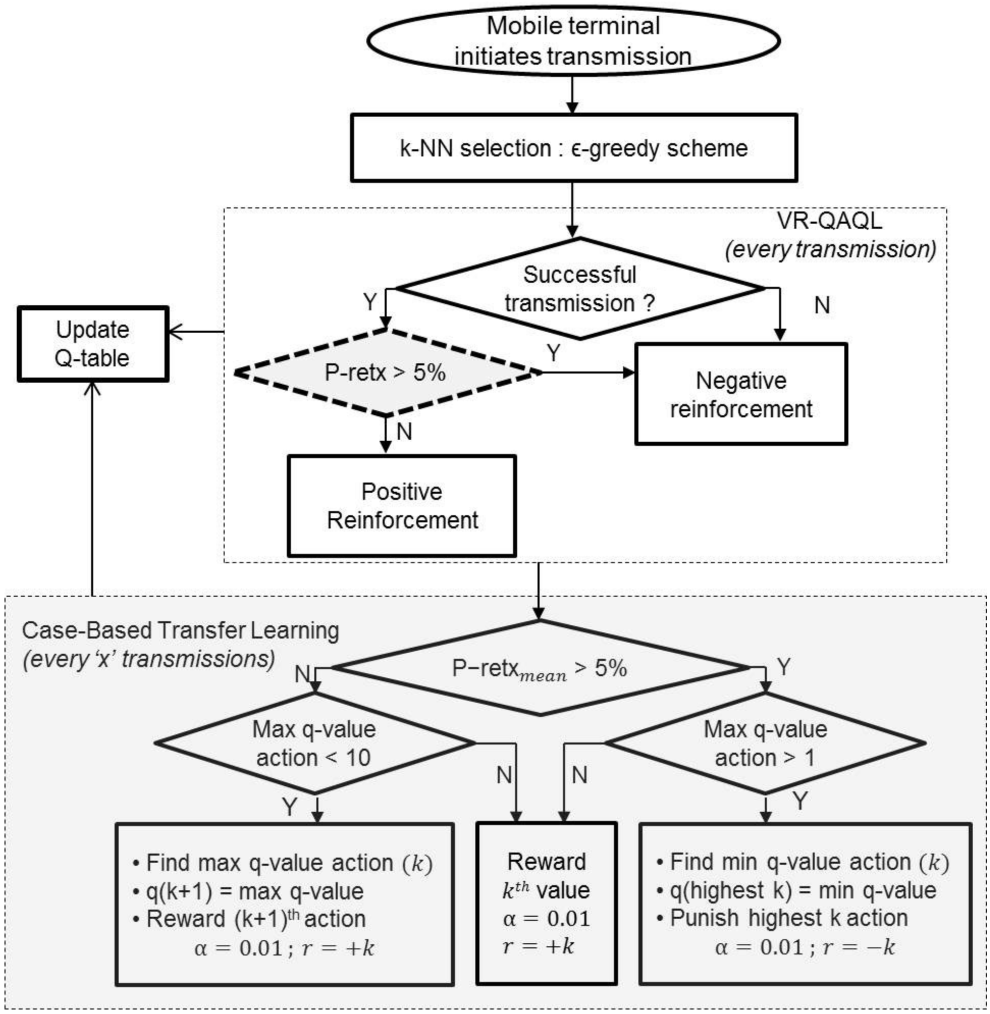


Figure 6.6: A comprehensive flowchart presenting the HA-VR-QAQL approach to learn the appropriate action value for user association in dynamic environments

In the first step of the developed algorithm, the variable reward Q-learning scheme is employed that uses reward shaping also known as variable reward function, as discussed in Section 5.4.3. The scheme rewards or punishes a chosen action with a reward equal to the action value depending on the success or failure of a transmission. An action that degrades the overall network performance is punished harshly even if it has been a promising solution during the initial learning phase.

The second step of implementing case-based transfer learning initiates after a batch of successful transmissions. In this step, the HARL scheme performs policy evaluation by deriving the heuristic information. The outcome is analyzed to identify and decide whether the third step should be exploited to transfer or map the gained knowledge between the actions using the value mapping method. The periodic evaluation of heuristics aims to improve the learning policy thus guiding the learning agent to learn a good solution by reconstructing its current policy. The proposed HA-VR-QAQL algorithm, therefore, aims to preserve the RL advantage of learning a good action policy that maximizes the expected reward through trial-and-error interaction with its environment as well as improves the stability and reliability of the learnt solution.

The third step relates to transferring a case base of heuristics using the value mapping function of case-based transfer learning to influence the action choice, i.e., allowing the retrieval and reuse of heuristic information from case-base between multiple actions. Furthermore, in this step, the learning agent believes that the current learnt action is a local maximum and that the immediate successor/predecessor action, depending on the network QoS, is the global optimum. Therefore, as shown in Figure 6.7, the value mapping function uses the q-value as heuristic and transfers the gained knowledge to a target action during the policy evaluation and improvement process.

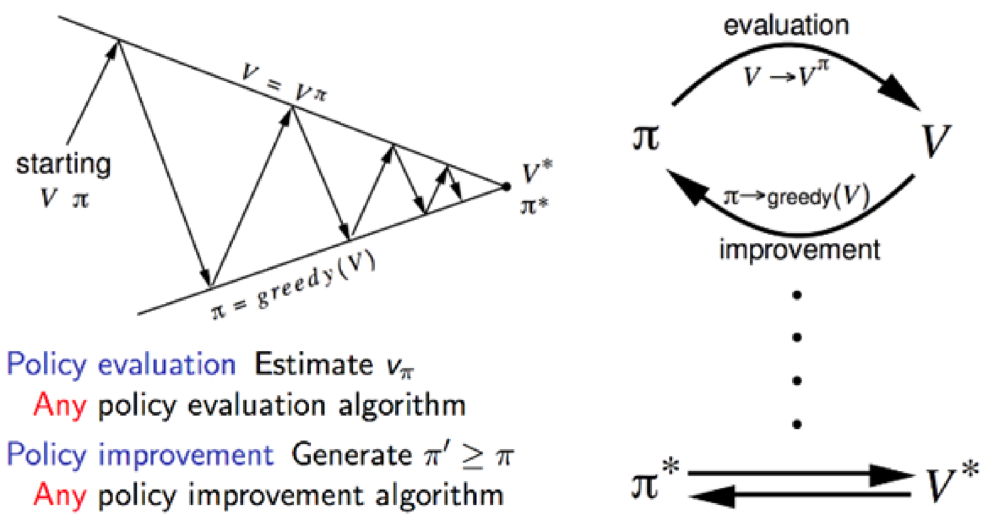


Figure 6.7: The policy evaluation and policy improvement procedure exploited in HA-VR-QAQL approach, directly reproduced from [100]

6.4.2. Heuristically Accelerated WoLF Q-Learning

Learning an optimal policy in dynamic multiagent vehicular environments is difficult due to (a) the presence of other learning agents who simultaneously contribute towards updating the knowledge base, and (b) the dynamic variations in the radio environment due to vehicle mobility. This section proposes a heuristically accelerated, Win or Learn Fast, quality aware Q-learning approach, HA-WoLF-QAQL that combines the WoLF concept, the case-based transfer learning and the heuristically accelerated reinforcement learning to learn a reliable solution across different traffic conditions in a dynamic environment. A detailed flowchart of the HA-WoLF-QAQL algorithm demonstrating the learning dynamics achieved using the case-based reinforcement learning is shown in Figure 6.8.

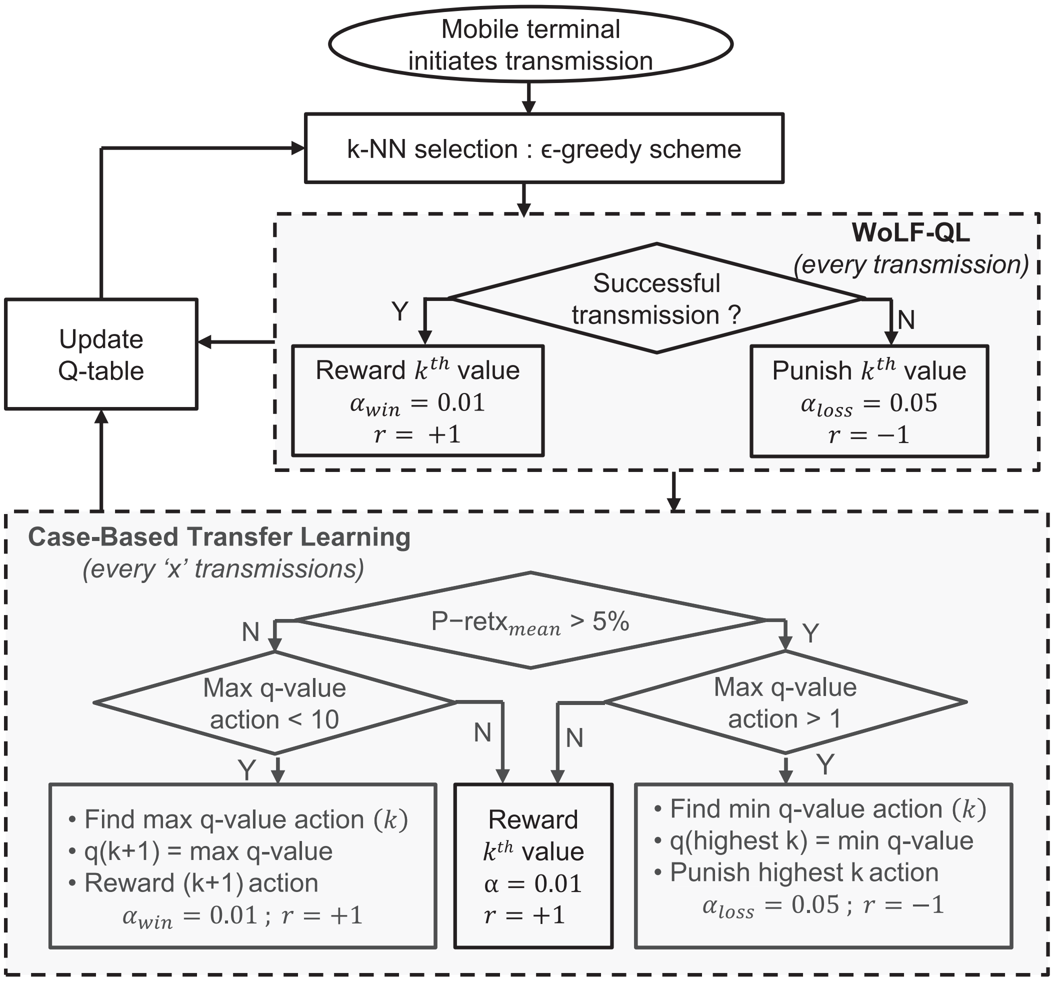


Figure 6.8: A comprehensive flowchart presenting the HA-WoLF-QAQL approach to learn the appropriate action value for user association in a dynamic environment

The developed scheme relates to a multi-agent environment similar to the HA-VR-QAQL approach, in the previous section, and is also model free, i.e, the learning agent is unaware about the environment or the reward function. The WoLF technique has been explained in detail in Section 5.4.2. In this section, the concept of WoLF is investigated to assess if the learning agent could adapt to the dynamics of vehicular as well as a wireless environment in a multi-agent environment to learn a reliable solution for user association. The Q-table update rule in conventional Q-Learning is as expressed in Chapter 5, re-written below:

$$Q(k) = (1-\alpha)Q'(k) + \alpha r \tag{6.7}$$

where k is the learnt action value, $Q(k)$ corresponds to the updated Q-value of the learnt action value k. $Q'(k)$ is the previous Q-value of the action k. The learning rate parameter, α, controls the convergence rate, and r is the reward awarded to an action determined by the reward function. The discount factor γ is assumed as 0, in this algorithm, therefore, the component is not included in the Eq.6.7.

Applying the WoLF principle to Eq.6.7, i.e., splitting the learning rate into α_{win} and α_{loss}, the equation transforms into:

$$Q(k) = \begin{cases} (1-\alpha_{win})Q'(k) + \alpha_{win} r, & r = 1 \\ (1-\alpha_{loss})Q'(k) + \alpha_{loss} r, & r = -1 \end{cases} \tag{6.8}$$

Equating the reward function $(r = \pm 1)$ in the above equation and rearranging the terms yields:

$$Q(k) = \begin{cases} Q'(k) + \alpha_{win}(1 - Q'(k)), & r = 1 \\ Q'(k) - \alpha_{loss}(1 + Q'(k)), & r = -1 \end{cases} \tag{6.9}$$

The magnitude of change in the Q-value $\triangle Q$ = $Q_t(k) - Q_{t-1}(k)$ is given by

$$|\triangle Q| = \begin{cases} -\alpha_{win} Q'(k) + \alpha_{win}, & r = 1 \\ \alpha_{loss} Q'(k) + \alpha_{loss}, & r = -1 \end{cases} \tag{6.10}$$

The above equation demonstrates a linear relationship between the Q-value and the magnitude of its change. It can, therefore, be concluded that if $\alpha_{win} < \alpha_{loss}$, then the output Q-value is negative, i.e., the learning agent learns quickly if it fails and vice versa. This property of the WoLF technique assists the learning agent not to quickly converge to a solution which would have proved promising in the initial learning stages. The results obtained are in coherence with the in-depth validation provided by the authors in [175]. The knowledge base developed is utilized by the learning agent similarly as in HA-VR-QAQL algorithm.

One of the advantages of employing the WoLF approach to learn a good solution is that the technique enables a thorough exploration during the initial phase of the learning. The knowledge base constructed is utilized by HARL, as in the previous learning algorithm, to initiate the case-based transfer learning that leads the learning agent towards learning a global solution. In conventional RL, an action is learnt irrespective of the transmission status at equal learning rate. This varies the Q-value in the Q-table considerably. However, in the case of WoLF if a learnt action results in several successful transmissions then its Q-value will keep on increasing. This action will be continuously used for future iterations due to the greedy action selection policy. Meanwhile, if this learnt action starts to result in interrupted transmissions, the learning algorithm will take fewer failed trials compared to a constant learning rate policy, thus, reconstructing its policies faster.

6.5 Results

6.5.1. Comparison of HA-VR-QAQL with other RL Approaches

This section presents the results obtained during a large-scale simulation experiment. Based on the investigated user association schemes, performance metric results such as the learnt k value in Figure 6.9, the number of handovers per transmission in Figure 6.10, the probability of re-transmission in Figure 6.11 and the end-to-end delay in Figure 6.12 are demonstrated. To present a more statistically valid temporal response, graphs present the average of 25 different random seed simulations.

In Figure 6.9, the action learnt by employing different association algorithm is demonstrated. The action value plotted at different traffic level represents an average of the different learnt action value during the simulation experiment. The baseline comparison approaches are the maximum-RSS, $k = 1$, and the maximum distance, $k = 4$, user association approach. The assumed k value remains constant throughout the simulation experiment for all the vehicle speed. Under the max-RSS user association approach, the closest eNB to the user is selected. Associating with the minimum distance eNB reduces the path attenuation, therefore, a minimum probability of retransmission and end-to-end delay are achieved across all mobility levels. However, a linear increase in the number of handovers per transmission as the vehicle speed increases is monitored as shown in Figure 6.10. This is due to unnecessary handovers that occur at every cell boundary. On the contrary, with $k = 4$, the UE associates with the furthest eNB in the k-NN list, the signal propagation attenuation increases due to an increase in the propagation range, resulting in poor QoS and end-to-end delay but with a significant decrease in the handover frequency.

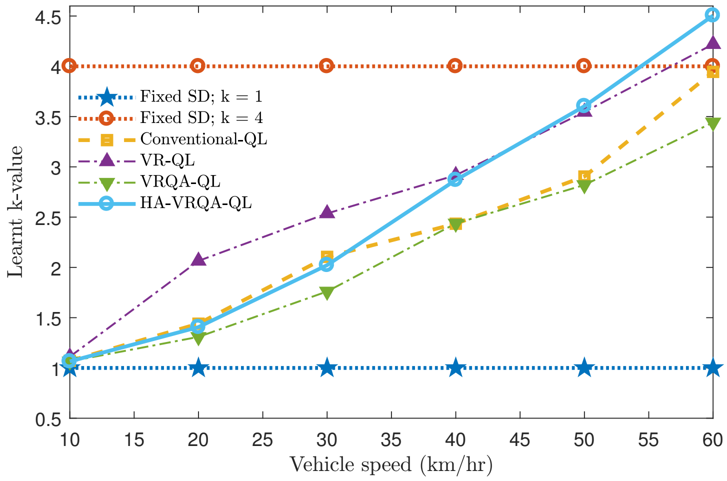


Figure 6.9: Performance comparison based on learnt appropriate action, k value, utilizing different schemes at different vehicle mobility

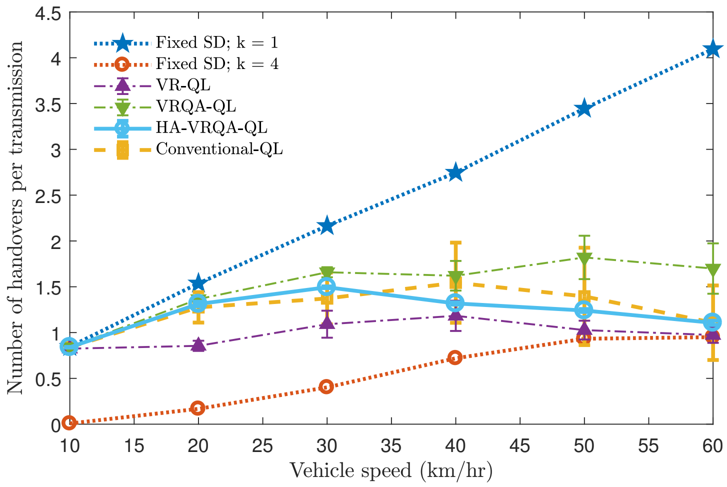


Figure 6.10: The handover frequency performance with different schemes

On the contrary, the learning schemes are adaptive, therefore, have the potential to evaluate, improve and restructure their policies as soon as any change in the environment is encountered. The classical Q-learning approach successfully learns an appropriate action value, only through trial-and-error and performs better than the non-learning user association schemes, especially at high mobility, but fails to deliver guaranteed QoS at low speed, seen in Figure 6.11 and Figure 6.12. It is because the scheme learns action based only upon the transmission status rather than the overall system QoS.

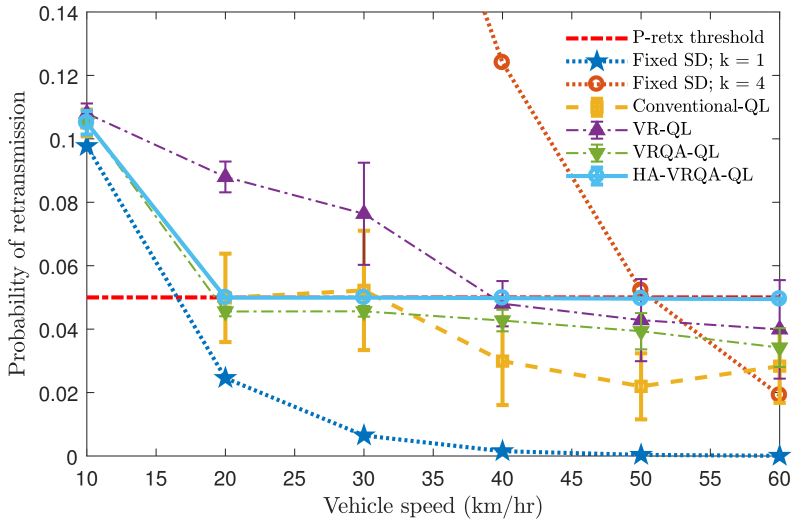


Figure 6.11: The probability of retransmission performance using different user association schemes across different vehicle speed

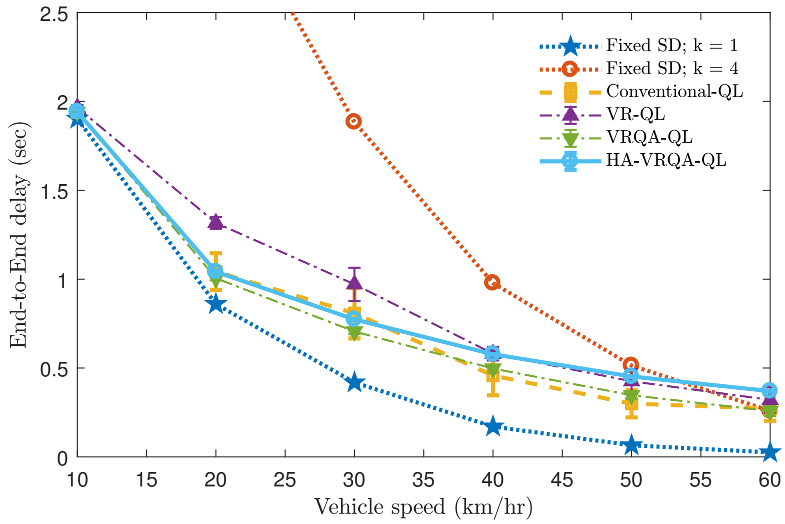


Figure 6.12: The end-to-end delay profile achieved using different schemes across different mobility levels

Similarly, the VR-QAQL approach exhibits better performance with respect to overall network QoS compared to the classical Q-learning approach. This is due to the incorporation of additional heuristic information which guides the exploration process to evaluate a better-learnt action with a marginal compromise in handover frequency, as seen in Figure 6.10. Despite learning an effective action value whilst providing a guaranteed QoS, as seen in Figure 6.11, there is still an opportunity for enhancement. The HA-VR-QAQL approach utilizes the variable reward function, heuristic information and case-based transfer learning to learn best action value, achieve a minimum number of handover per transmission and provide guaranteed QoS by (a) utilizing the transfer learning mechanism on a variable reward Q-learning framework during the periodic action-value evaluation that forces the learning agent towards the best solution, and (b) deriving heuristic information that guide the exploration function to carefully evaluate the action value. The HA-VR-QAQL algorithm, therefore, outperforms all the other association schemes because of the strong correlation between the action selection process and the policy iteration process.

6.5.2. Comparison between Case-Based RL Approaches

The temporal response of the network in terms of the probability of retransmission at varying offered traffic levels, using the case-based reinforcement learning approaches for user association is compared in this section. The graph in Figure 6.13 shows the average of 25 simulations with different random seeds in order to produce more statistically valid temporal results. Firstly, the graph shows that both the developed approaches deliver a guaranteed network QoS. Secondly, the developed schemes achieved a significant improvement compared with the baseline technique; max-RSS approach due to the guided exploration process of the case-based reinforcement learning that demonstrated to be highly effective. The network quality of service achieved by the HA-VR-QAQL algorithm in comparison with the max-RSS approach shows an improvement by a factor of 81-82% at a low mobility level of 10 km/hr.

In Figure 6.13 it is also seen that HA-WoLF-QAQL has a significantly lower probability of retransmission compared to the HA-VR-QAQL thus imply that it could adapt to the frequent changes in the environment considerably faster. This characteristic is more desirable with time-varying traffic distributions wherein the motivation is only to provide the best QoS with ultra-reliable low latency. However, the motivation of the presented work is not only to provide a guaranteed QoS or ultra-low latency but to learn the maximum association range for user association such that an equilibrium between guaranteed network QoS and number of handovers per transmission at all mobility levels are achieved. The results shown in Figure 6.14 presents the end-to-end delay profile achieved using the two case-based RL schemes and correlates with the result obtained in Figure 6.13. The graph demonstrates that the HA-WoLF-QAQL approach outperforms the HA-VR-QAQL largely because of a variable learning rate.

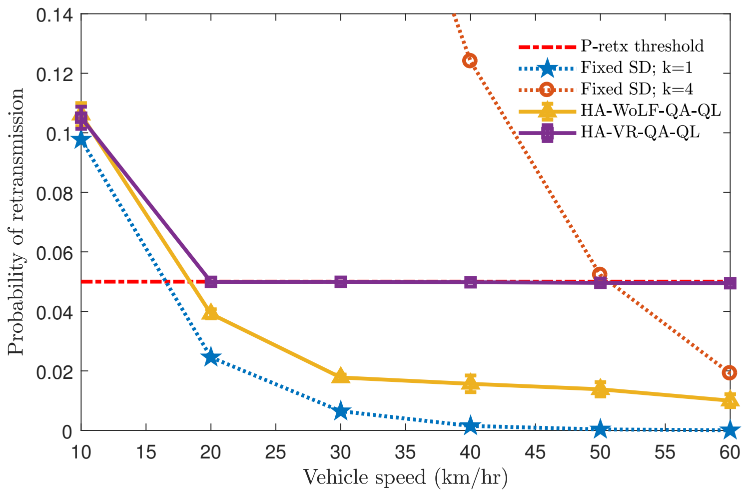


Figure 6.13: The probability of retransmission performance using different user association schemes across different vehicle speed

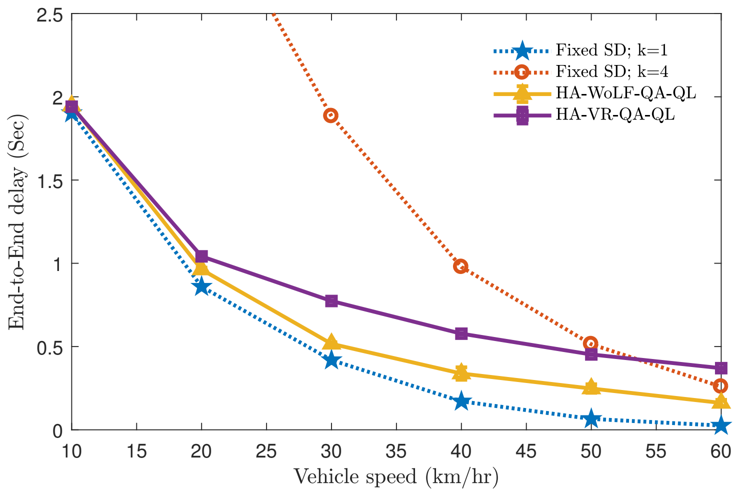


Figure 6.14: The end-to-end delay profile achieved using different schemes across different mobility levels

However, from the graphs demonstrated in Fig 6.15 and Figure 6.16, it could be concluded that the HA-VR-QAQL approach performed significantly better than HA-WoLF-QAQL in the dynamic environment due to the use of variable reward function.

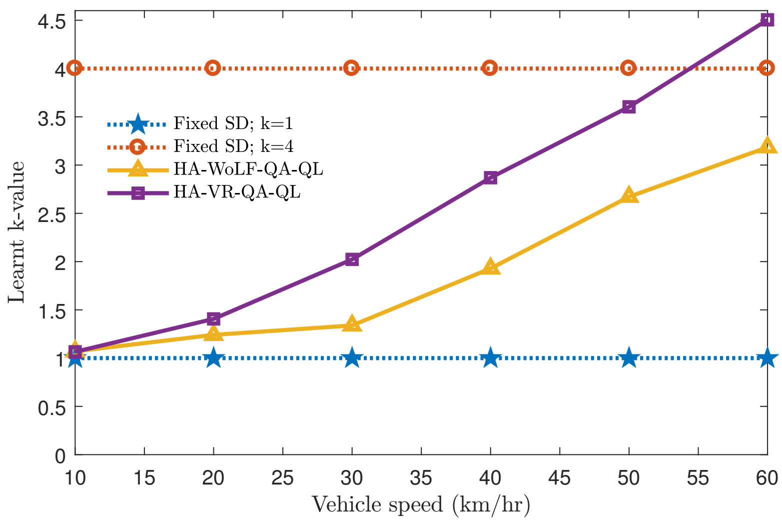


Figure 6.15: The performance of a learning agent to learn an appropriate action utilizing different investigated schemes at different vehicle mobility

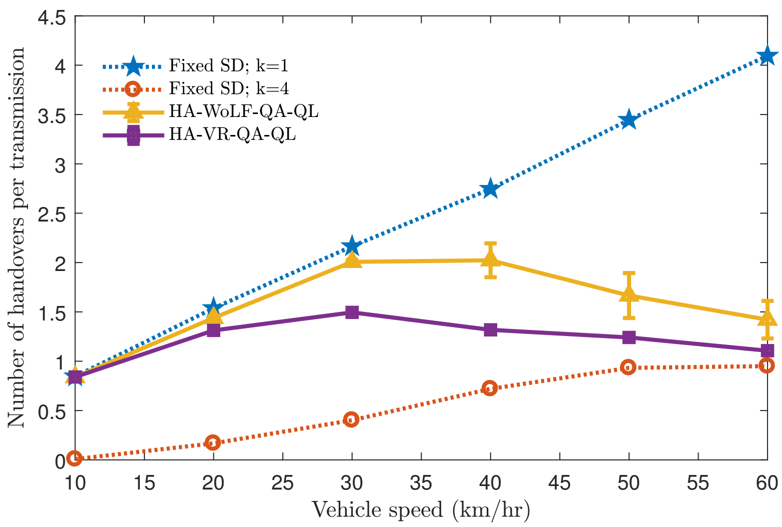


Figure 6.16: The handover frequency performance with different schemes

The obtained results collectively confirm that the HA-VR-QAQL technique outperformed all the other proposed user association techniques as it learns an optimal action whilst achieving both of the initial set goals across different mobility levels.

6.6 Conclusion

This chapter investigates the case-based reinforcement learning strategy to learn the optimal action at different vehicle speed in a highly dense small-cell dynamic vehicular environment. The proposed CBRL schemes have been tested across different mobility levels in a bi-directional multiple-lane highway scenario. The large-scale simulation results show that the HA-VR-QAQL scheme outperforms the baseline max-RSS user association approach as well as other reinforcement learning approaches by demonstrating the ability to strike a balance between handover frequency and system performance whilst providing guaranteed network QoS. The HA-VR-QAQL approach reduced the number of handovers per transmission by about 78-80% with marginal compromise in the network QoS when compared to max-RSS user association approach at a vehicle speed of 60 km/hr. The results, therefore, provide valuable insights into intelligent user association in dense dynamic environments as well as serve as a guideline for the development of more advanced strategies related to the vehicle communication optimization and topology management in future.

Chapter 7. Conclusions and Future Work

Contents

7.1 Conclusions

This thesis has investigated the application of heuristic as well as intelligent learning algorithms for user association in ultra-dense small cell vehicular network, in order to mitigate the mobility-related concerns whilst achieve a guaranteed network performance. It has been realized that the velocity and the density of vehicle nodes plays a crucial role in network performance and user experience in densely deployed low powered small cell networks. The two extreme cases identified in the presented thesis were (a) high mobility scenario with characteristics such as short network lifetime, few vehicle nodes, sparse vehicle distribution and low traffic congestion (b) low or no mobility scenario with traffic congestion, high vehicle density, and longer network lifetime.

Firstly, a conventional, performance metric based user association algorithm that follows a three-step sequence was proposed to understand the affect of individual metric on network performance and user experience. Subsequently, the results motivated to develop a real-time control feedback technique, an effective computing algorithm for user association in dynamic environments. The results provided an insight into the influence of vehicle speed on the network performance. Following this few reinforcement learning based user association algorithms were developed to learn appropriate solution purely through trial-and-error iterations regardless of the network topology and/or traffic dynamics. The empirical assessment of the proposed algorithms have highlighted that the introduction of intelligence appears to be an essential requirement in order to mitigate the mobility-related concern, therefore, provide a reliable and robust radio access network for a diverse range of use cases especially in rapidly changing environments.

RL is one of the most popular and powerful approaches in both wireless networks and other artificial domains due to its self-organizing and self-co-ordination abilities. Moreover, it eliminates the need for manual intervention that is potentially challenging and time-consuming to keep up with dynamic varying traffic load and network topologies. However, an inherent disadvantage of RL algorithms is the need for a longer exploration phase, a characteristic that is undesirable in real-time applications. To overcome this case-based reinforcement learning user association approach was proposed in Chapter 6. The scheme combined case-based transfer learning with heuristically accelerated RL to improve the reliability and adaptability of the developed RL based user association algorithms in Chapter 5. The CBRL algorithm was able to improve the initial as well as steady-state performance. Similar to other investigated RL schemes, CBRL algorithms were also able to restructure their current policy to learn reliable solution in a rapidly changing ultra-dense small cell vehicular environment such that the varying demand for data coverage and capacity by the users on the move are met.

A more detailed chapter-by-chapter discussion of the original contributions of this thesis towards the traffic dynamics and adaptability of intelligent user association in dynamic environments is given in the following subsection.

7.1.1. Original Contributions

Ultra-dense Small Cell Vehicular Network Simulator

In order to evaluate the empirical properties of the proposed user association approaches, a model that could accurately represent an ultra-dense small cell vehicular network was absolutely important. Considerable effort has been made to develop a sufficiently complex event-based system level simulator. It is an integration of a dynamic vehicular network, a highway, and an ultra-dense small cell network. The model considers relevant aspects of LTE simulation, such as multi-cell environments with uplink flows, user mobility, handover procedures, cell planning, scheduling, interference calculation, and QoS management in a dynamic environment. Moreover, the developed simulator enables the analysis of the influence of user mobility on network performance across different mobility levels. Chapter 3 presents the modelling techniques used to develop the simulation model and key performance metrics utilized for assessment of these policies. The developed simulation model allows system and user experience to be modelled, empirically assess the performance and reliability of the proposed algorithms and serves as the firm foundation for the main research problem.

Performance Metric based Three-Step Sequence Approach

In Chapter 4, a conventional, performance metric based user association algorithm that follows a three-step sequence to shortlist, select and choose an appropriate base station for association was proposed. The scheme effectively determines a robust solution on a diverse range of vehicle speed. The designed approach uses parameters such as vehicle speed, the direction of vehicle flow, desired performance metric and geographical locations of eNBs and vehicles to identify an appropriate eNB for the association. The results provided insight into the influence of different performance indicators such as spectrum efficiency, network load and dwell time per cell on the network performance and the number of handovers at different traffic conditions that relates to the mobility-related concerns. The simulation results obtained demonstrated the equivalence between vehicular traffic and radio communication in an ultra-dense small cell environment that was subsequently quantified and exploited to develop more sophisticated user association schemes.

Variable Learning Rate and Reward Function Approach

In Chapter 5, few RL based user association schemes, using the Q-learning framework, have been proposed to enhance the network performance and user experience. Among them, the conventional Q-Learning approach learnt a solution which was able to provide the guaranteed QoS but failed to assure that the learnt solution is the global maximum. However, the WoLF-QL approach was developed by employing a variable learning rate to learn the best solution. The scheme guided the learning agent to learn a solution slowly in case of winning and quickly in case of unsuccessful trials. The algorithm performed better than the conventional Q-Learning; however, it converged to a sub-optimal solution. On the contrary, motivated by the high-risk-high-return approach, the VR-QL algorithm was proposed for user association. The scheme learnt the maximum association range by varying the reward function while the learning rate remains constant. The conventional QL, WoLF-QL and the VR-QL algorithm learnt an action value purely through trial-and-error iterations and were promising techniques to achieve a single parameter optimization, i.e., the schemes either learnt an action value that delivered a guaranteed network QoS or learnt the global maxima at the expense of an extremely degraded network performance.

Variable Reward Quality-Aware Q-learning Approach

Later in Chapter 5, the variable reward quality-aware Q-learning (VR-QAQL) algorithm was proposed with an aim to achieve multiple parameter optimization. VR-QAQL followed *learn-execute-evaluate-formulate-improve* principle to learn the best action value for user association. The algorithm recursively performed policy evaluation and policy improvement processes, using the VR-QL and the real-time control feedback approach, enabling the learning agent to learn a solution that might have not been proven to be promising only in the early simulation phase. The strong correlation between the policy evaluation and policy improvement process in the VR-QAQL algorithm outperform the conventional max-RSS user association approach, the real-time controlled feedback approach, as well as other developed RL approaches, by learning the best action value whilst delivering guaranteed network QoS across all vehicle speeds in the considered vehicular scenario.

Case-Based Reinforcement Learning Approach

The novel case-based reinforcement learning (CBRL) technique proposed in Chapter 6 is an effective user association approach in ultra-dense dynamic environments. The techniques combined the heuristically accelerated reinforcement learning and case-based transfer learning using a variable reward function and variable learning rate respectively. The large-scale system level simulations of a highway network with a rapidly changing temporal-spatial vehicular distribution show that augmenting classical distributed Q-learning with case-based transfer learning and variable reward function improves the reliability of the learnt solution significantly as well as learns the maximum action value at different network conditions. The developed case-based reinforcement learning algorithms outperformed all the other user association techniques investigated by exploiting the strong correlation between the classical Q-learning, heuristically-accelerated reinforcement learning, transfer learning and variable reward function, thus, achieving a consistent improvement in system QoS at all vehicle speed.

7.1.2. Hypothesis Revisited

The research has been guided by the following hypothesis as stated in Chapter 1.

"Appropriate exploitation of heuristic information through reinforcement learning for user association in ultra-dense dynamic vehicular environments can significantly reduce the handover rate whilst delivering a guaranteed network quality of service."

The hypothesis has indicated the need for designing learning-based user association schemes to facilitate self-organization and self-management in ultra-dense dynamic vehicular environments. In the context of the above hypothesis, the key contributions of this thesis can be summarized as follows:

- The three-step sequence user association approach proposed in Chapter 4, demonstrate the influence of individual performance metric on user experience and network performance in the considered dynamic vehicular, ultra-dense environment.

- The WoLF-QL approach proposed in Chapter 5 uses the variable learning rate policy to learn a reliable action, thus, increasing the adaptability of the conventional RL based user association approach. The persistent exploration encourages the learning agent to continuously reconstruct its current policy such that it does not converge to local maxima.

- Similarly, the VR-QL approach, also proposed in Chapter 5, uses a variable reward function to learn an appropriate action value. The results of VR-QL and real-time control feedback approach motivated us to combine both techniques to form the VR-QAQL user association approach. The VR-QAQL algorithm employs the high risk and high reward policy to learn the maximum action-value, during the policy evaluation and policy improvement to improve the robustness of the learnt solution.

- Chapter 6 proposes a case-based reinforcement learning user association approach. The scheme employs the merits of transfer learning and heuristic acceleration on a VR-QAQL and WoLF-QAQL framework to enable the learning agent to rationally evaluate and improve the current policy, thus, adapt to the rapidly changing environment.

These contributions have been empirically evaluated in this thesis. The results show dramatic improvements in the adaptability of the RL based user association methods to vehicular networks, thus, proving the hypothesis of this research.

7.2 Recommendation for Future Work

This section provides recommendations for further work that includes extensions of the ideas and potential application on the areas explored in this thesis.

Application of CBRL to Urban Intersection Scenario

The research work presented in Chapter 6 demonstrates the effectiveness of case-based reinforcement learning user association specifically on a highway scenario, where the vehicles move at a constant speed. Due to the variation in mobility pattern of users at different vehicle speed, especially in a more complex dynamic scenario such as in urban intersection with free-flow traffic and restricted-flow traffic more sophisticated CBRL algorithms will be required to provide more advanced reliable communication and stable QoS. It is therefore recommended to first investigate the feasibility and reliability of the proposed CBRL algorithm, as well as other investigated RL approaches on these scenarios before more advanced CBRL algorithms for these scenarios, are investigated. Some of the machine learning schemes such as density estimation, state-action-reward-state-action (SARSA), semi-supervised learning, active learning, decision tree, and decision forest may be used to develop sophisticated handoff strategies that may further enhance the stability and reliability of dynamic user association in vehicular networks. Moreover, the results obtained in chapter 6 may be used as the baseline results for comparison with more advanced user association algorithms.

Effect of Exploration on Agent's Performance

In the considered dynamic vehicular scenario, due to the node mobility, frequent changes in the network topology and radio network performance are experienced. The learning strategies developed uses epsilon-greedy exploration, with a persistent ϵ value, to constantly simulate exploration, yet select an action value that performs the best. The learning agent selects a greedy action with probability 1-ϵ, while uses an exploratory approach with the probability of ϵ. A persistent exploration thus updates the current learning policy depending on the current environmental conditions. However, the disadvantage of using persistent learning is that the learning agent constantly explores other actions even when it has converged which may lead to learning a sub-optimal action. Therefore, to understand the impact of exploration function on learning agent's performance in such uncertain environment, different exploration strategies such as random exploration, greedy exploration, ϵ-greedy exploration, decaying ϵ-greedy exploration, and softmax exploration should be investigated. The results obtained may be compared in terms of reliability, adaptability and convergence to provide a better understanding of the impact of exploration on learning optimal policy in dynamic scenarios.

Employing Inverse Reinforcement Learning to Learn Reward Function

The learning algorithms investigated in the presented thesis used a constant reward value to evaluate an action and learn an appropriate solution. In the case of traditional Q-Learning and the WoLF-QL, the reward value is a constant; ($r = \pm 1$). However, in the case of VR-QL approach, the VR-QAQL approach, and HA-VR-QAQL approach, the reward function depends on the action chosen, i.e., an action is rewarded with a positive reward equal to the action value if a trail is successful else it is punished; ($r = \pm k$). The simulation results assisted to understand the impact of reward function on learning performance in a vehicular environment. However, the empirical evaluation was not able to identify an optimal reward function that would generate a straightforward desirable behaviour. An emerging technique to extract or optimize a reward function based on an observed optimal behaviour is the Inverse Reinforcement Learning (IRL). In the state-of-the-art Q-Learning, the purpose is to learn an optimal policy by mapping states to action such that the overall reward is maximized. However, IRL intends to recover the reward value for each action learnt and use the values to generate a desirable behaviour. The simulation results obtained in this thesis may serve as heuristic information to identify the reward function using the IRL approach.

Traffic-aware Cell Management

One of the essential parameters to maximize the coverage and capacity provisioning to the moving users and to improve the QoS on the access link is load balancing. It relates to network load management, i.e., equalizing the offered network traffic between neighbouring or overlaid cells. Chapter 4 discusses the influence of different performance metrics based user association approaches on the network performance and user experience. However, the later chapters of the thesis predominantly focus towards the application of RL based approaches to mitigate the challenges due to user mobility in the ultra-dense small cell environment. Therefore, it is highly recommended to first explore traditional load balancing techniques in vehicular networks and then develop reinforcement learning based load balancing approaches in vehicular environments that enhance the network capacity and performance. Moreover, machine learning enabled load balancing and load unbalancing approaches may also assist towards adaptive base station switching on an off using traffic load as heuristic information, resulting in the development of adaptive green frameworks.

Exploiting Clustering Techniques for User Association

Another possible direction to extend the research work to mitigate the uncertainty in larger scenarios due to high node mobility and higher node-density is to group vehicles with similar properties. Clustering is a popular machine learning technique to classify or group data with similar properties. In the context of vehicular networks, clustering of vehicles could be performed based on predefined parameters such as vehicle density, velocity, the signal strength received, the direction of flow and/or geographical locations. The research should begin by grouping vehicle users using different clustering schemes such that the cluster structure is maintained without any increase in the communication overhead. The authors of [193] provides an extensive survey of different clustering algorithms for wireless networks. The research should then focus to develop a machine learning based horizontal and vertical handoff strategies for efficient communications. Further, the proposed algorithms should be assessed at different vehicle speed and vehicle density to analyze their adaptability and reliability in dynamic environments.

Glossary

BBU Baseband Unit

BS Base Station

CA Carrier Aggregation

CAPEX Capital Expenditure

CBR Case Based Reasoning

CBTL Case Based Transfer Learning

CB Coordinated Beamforming

CoMP Coordinated Multipoint

CS Coordinated Scheduling

C-RAN Cloud-RAN

CRE Cell Range Expansion

DP Dual Connectivity

DP Dynamic Programming

eNB Evolved Node B

FTP File Transfer Protocol

HARL Heuristically Accelerated Reinforcement Learning

HA-VR-QAQL Heuristically Accelerated Variable Reward Quality Aware Q-Learning

HA-WoLF-QAQL Heuristically Accelerated Win-or-Learn Fast Quality Aware Q-Learning

HeTNet Heterogeneous Network

KPI Key Performance Indicator

LOS Line-of-Sight

LTE Long Term Evolution

MDP Markov Decision Process

MIMO Multiple Input Multiple Output

NLOS Non-Line-of-Sight

MARL Multi-agent Reinforcement Learning

MS Mobile Station

OPEX Operational Expenditure

QoE Quality of Experience

QoS Quality of Service

QL Q-Learning

RAN Radio Access Network

RAT Radio Access Technology

RL Reinforcement Learning

RRH Remote Radio Head

RSS Radio Signal Strength

RSRQ Reference Signal Received Quality

RSRP Reference Signal Received Power

RSSI Reference Signal Strength Indicator

SARSA State action reward state action

SINR Signal-to-Interference plus Noise

SNR Signal-to-Noise Ratio

TL Transfer Learning

UDN Ultra dense Network

UE User Equipment

V2I Vehicle to Infrastructure

V2V Vehicle to Vehicle

VNI Visual Network Index

VR-QL Variable Reward Q-Learning

VR-QAQL Variable Reward Quality Aware Q-Learning

WoLF Win-or-Learn-Fast

WoLF-QL Win-or-Learn-Fast Q-Learning

References

[1] V. Cisco Mobile, “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 20162021 White Paper,” ed: Accessed on Apr, 2017.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, “What will 5G be?,” IEEE Journal on Selected Areas in Communications, vol. 32, no. 6, pp. 1065-1082, 2014.

[3] P. Demestichas, A. Georgakopoulos, D. Karvounas, K. Tsagkaris, V. Stavroulaki, J. Lu, C. Xiong, J. Yao, “5G on the horizon: Key Challenges for the Radio-Access Network,” Vehicular Technology Magazine, IEEE, vol. 8, no. 3, pp. 47-53, 2013.

[4] M. Agiwal, A. Roy, and N. Saxena, “Next generation 5G wireless networks: A Comprehensive Survey,” IEEE Communications Surveys & Tutorials, vol. 18, no. 3, pp. 1617-1655, 2016.

[5] E. Hossain and M. Hasan, “5G cellular: Key enabling technologies and research challenges,” IEEE Instrumentation & Measurement Magazine, vol. 18, no. 3, pp. 11-21, 2015.

[6] X. Lin, R. K. Ganti, P. J. Fleming, and J. G. Andrews, “Towards understanding the fundamentals of mobility in cellular networks,” IEEE Transactions on Wireless Communications, vol. 12, no. 4, pp. 1686-1698, 2013.

[7] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K. K. Wong, R. Schober and L. Hanzo, “User association in 5G networks: A survey and an outlook,” IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1018-1044, 2016.

[8] C. Chi, X. Cai, R. Hao, and F. Liu, “Modeling and analysis of handover algorithms,” in Global Telecommunications Conference, 2007. GLOBECOM’07. IEEE, 2007, pp. 4473-4477: IEEE.

[9] C. Guo, Z. Guo, Q. Zhang, and W. Zhu, “A seamless and proactive end-to-end mobility solution for roaming across heterogeneous wireless networks,” IEEE

Journal on Selected Areas in Communications, vol. 22, no. 5, pp. 834-848, 2004.

[10] Q. Kuang, J. Belschner, Z. Bleicher, H. Droste, and J. Speidel, “A measurement-based study of handover improvement through range expansion and interference coordination,” Wireless Communications and Mobile Computing, vol. 15, no. 14, pp. 1784-1798, 2015.

[11] G. P. Pollini, “Trends in handover design,” IEEE Communications Magazine, vol. 34, no. 3, pp. 82-90, 1996.

[12] N. D. Tripathi, J. H. Reed, and H. F. VanLandinoham, “Handoff in cellular systems,” IEEE Personal Communications, vol. 5, no. 6, pp. 26-37, 1998.

[13] M. Gudmundson, “Analysis of handover algorithms (microcellular radio),” in Vehicular Technology Conference, 1991. Gateway to the Future Technology in Motion., 41st IEEE, 1991, pp. 537-542: IEEE.

[14] G. Gódor, Z. Jakó, Á. Knapp, and S. Imre, “A survey of handover management in LTE-based multi-tier femtocell networks: Requirements, challenges and solutions,” Computer Networks, vol. 76, pp. 17-41, 2015.

[15] P. Valente Klaine, M. A. Imran, O. Onireti, and R. D. Souza, “A survey of machine learning techniques applied to self organizing cellular networks,” IEEE Communications Surveys and Tutorials, 2017.

[16] A. Imran, A. Zoha, and A. Abu-Dayya, “Challenges in 5G: how to empower SON with big data for enabling 5G,” IEEE Network, vol. 28, no. 6, pp. 27-33, 2014.

[17] O. G. Aliu, A. Imran, M. A. Imran, and B. Evans, “A survey of self organisation in future cellular networks,” IEEE Communications Surveys & Tutorials, vol. 15, no. 1, pp. 336-361, 2013.

[18] V. N. I. Cisco, “Global Mobile Data Traffic Forecast Update, 20152020 White Paper,” Document ID, vol. 958959758, 2016.

[19] M. Döttling, W. Mohr, and A. Osseiran, “Radio technologies and concepts for IMT-advanced. John Wiley & Sons, 2009.

[20] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu, “5G network capacity: Key elements and technologies,” IEEE Vehicular Technology Magazine, vol. 9, no. 1, pp. 71-78, 2014.

[21] D. López-Prez, I. Güven, G. De La Roche, M. Kountouris, T. Q. S. Quek, and J. Zhang, “Enhanced intercell interference coordination challenges in heterogeneous networks,” IEEE Wireless Communications, vol. 18, no. 3, pp. 22-30, 2011, Art. no. 5876497.

[22] 5G Mobile Research Lab, “Prospective scenarios in the 5G wireless networks”. Available: http://5gmobile.fel.cvut.cz/activities/, last accessed October 2018.

[23] K. Ullah, “On the use of opportunistic vehicular communication for roadside services advertisement and discovery,” Universidade de So Paulo.

[24] P. Kela, J. Puttonen, N. Kolehmainen, T. Ristaniemi, T. Henttonen, and M. Moisio, “Dynamic packet scheduling performance in UTRA long term evolution downlink,” in Wireless Pervasive Computing, 2008. ISWPC 2008. 3rd International Symposium on, 2008, pp. 308-313: IEEE.

[25] D. Astely, E. Dahlman, G. Fodor, S. Parkvall, and J. Sachs, “LTE release 12 and beyond [accepted from open call],” IEEE Communications Magazine, vol. 51, no. 7, pp. 154-160, 2013.

[26] M. Kamel, W. Hamouda, and A. Youssef, “Ultra-dense networks: A survey,” IEEE Communications Surveys & Tutorials, vol. 18, no. 4, pp. 2522-2545, 2016.

[27] M. Ding, D. López-Pérez, G. Mao, P. Wang, and Z. Lin, “Will the area spectral efficiency monotonically grow as small cells go dense?,” in Global Communications Conference (GLOBECOM), 2015 IEEE, 2015, pp. 1-7: IEEE.

[28] J. Deissner and G. P. Fettweis, “A study on hierarchical cellular structures with inter-layer reuse in an enhanced GSM radio network,” in Proc. Conf. Mobile Multimedia Communications (MoMuC99), 1999, pp. 243-251.

[29] J. G. Andrews, X. Zhang, G. D. Durgin, and A. K. Gupta, “Are we approaching the fundamental limits of wireless network densification?,” IEEE Communica-

tions Magazine, vol. 54, no. 10, pp. 184-190, 2016.

[30] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," IEEE Journal on Selected Areas in Communications, vol. 30, no. 3, pp. 497-508, 2012.

[31] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," IEEE Wireless Communications, vol. 18, no. 3, 2011.

[32] M. Peng, Y. Li, Z. Zhao, and C. Wang, "System architecture and key technologies for 5G heterogeneous cloud radio access networks," IEEE Network, vol. 29, no. 2, pp. 6-14, 2015.

[33] R. Jain and C. So-In, "System-level modeling of IEEE 802.16 E mobile WiMAX networks: key issues," IEEE Wireless Communications, vol. 15, no. 5, 2008.

[34] D. Liu, Y. Chen, K. K. Chai, and T. Zhang, "Optimal user association for delay-power tradeoffs in HetNets with hybrid energy sources," in Personal, Indoor, and Mobile Radio Communication (PIMRC), 2014 IEEE 25th Annual International Symposium on, 2014, pp. 1857-1861: IEEE.

[35] H. Zhu, S. Wang, and D. Chen, "Energy-efficient user association for heterogenous cloud cellular networks," in GLOBECOM Workshops, 2012, pp. 273-278.

[36] D. Bethanabhotla, O. Y. Bursalioglu, H. C. Papadopoulos, and G. Caire, "Optimal user-cell association for massive MIMO wireless networks," IEEE Transactions on Wireless Communications, vol. 15, no. 3, pp. 1835-1850, 2016.

[37] H. S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis," IEEE Transactions on Wireless Communications, vol. 11, no. 10, pp. 3484-3495, 2012.

[38] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," IEEE Transactions on Wireless Communications, vol. 12, no. 6, pp. 2706-2716, 2013.

[39] 3GPP, “Mobility enhancements in heterogeneous networks,”tr 36.839V11.0.0, Sep. 2012.

[40] K. Zhu, D. Niyato, P. Wang, E. Hossain, and D. In Kim, “Mobility and handoff management in vehicular networks: a survey,” Wireless Communications and Mobile Computing, vol. 11, no. 4, pp. 459-476, 2011.

[41] A. J. Ransom, “Handoff considerations in microcellular systems planning,” in Personal, Indoor and Mobile Radio Communications, 1995. PIMRC’95. Wireless: Merging onto the Information Superhighway, Sixth IEEE International Symposium on, 1995, vol. 2, pp. 804-808: IEEE.

[42] O. Grimlund and B. Gudmundson, “Handoff strategies in microcellular systems,” in Vehicular Technology Conference, 1991. Gateway to the Future Technology in Motion, 41st IEEE, 1991, pp. 505-510: IEEE.

[43] S. Chia, “The control of handover initiation in microcells,” in Vehicular Technology Conference, 1991. Gateway to the Future Technology in Motion, 41st IEEE, 1991, pp. 531-536: IEEE.

[44] S. Nanda, “Teletraffic models for urban and suburban microcells: Cell sizes and handoff rates,” in Wireless Communications: Springer, 1993, pp. 157-175.

[45] S. Sadr and R. S. Adve, “Handoff rate and coverage analysis in multi-tier heterogeneous networks,” IEEE Transactions on Wireless Communications, vol. 14, no. 5, pp. 2626-2638, 2015.

[46] J. G. Andrews, “Seven ways that HetNets are a cellular paradigm shift,” IEEE Communications Magazine, vol. 51, no. 3, pp. 136-144, 2013.

[47] 3GPP. Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Mobility enhancements in heterogeneous networks (3GPP TR 36.839 version 11.1.0 Release 11), December 2010.

[48] K. Okino, T. Nakayama, C. Yamazaki, H. Sato, and Y. Kusano, “Pico cell range expansion with interference mitigation toward LTE-Advanced heterogeneous networks,” in Communications Workshops (ICC), 2011 IEEE International Conference on, 2011, pp. 1-5: IEEE.

[49] T. Kudo and T. Ohtsuki, "Cell range expansion using distributed Q-learning in heterogeneous networks," EURASIP Journal on Wireless Communications and Networking, vol. 2013, no. 1, p. 61, 2013.

[50] F. Zheng, M. Wu, and H. Lu, "Coordinated multi-point transmission and reception for LTE-advanced," in Wireless Communications, Networking and Mobile Computing, 2009. WiCom'09. 5th International Conference on, 2009, pp. 1-4: IEEE.

[51] A. Barbieri, P. Gaal, S. Geirhofer, T. Ji, D. Malladi, Y. Wei, and F. Xue, "Coordinated downlink multi-point communications in heterogeneous cellular networks," in Information Theory and Applications Workshop (ITA), 2012, pp. 7-16: IEEE.

[52] S. Bassoy, H. Farooq, M. A. Imran, and A. Imran, "Coordinated multi-point clustering schemes: A survey," IEEE Communications Surveys & Tutorials, vol. 19, no. 2, pp. 743-764, 2017.

[53] P. Shang, L. Zhang, M. You, Y. Yang, and Q. Zhang, "Performance of uplink joint reception CoMP with antenna selection for reducing complexity in LTE-A systems," in Wireless Communications and Networking Conference (WCNC), IEEE, 2015, pp. 977-982: IEEE.

[54] 3GPP, "Study on small cell enhancements for EUTRA and E-UTRAN Higher layer aspects (3gpp tr 36.842), 2013.

[55] 3GPP, "3gpp work items on small cells v0.0.4, 2014.

[56] WGC, White Paper in Wireless Research Forum.

[57] S. C. Jha, K. Sivanesan, R. Vannithamby, and A. T. Koc, "Dual connectivity in LTE small cell networks," in Globecom Workshops (GC Wkshps), 2014, pp. 1205-1210: IEEE.

[58] Z. Wu, W. Xie, F. Yang, and Q. Bi, "User association in heterogeneous network with dual connectivity and constrained backhaul," China Communications, vol. 13, no. 2, pp. 11-20, 2016.

[59] J. Lun and D. Grace, “Cognitive green backhaul deployments for future 5G networks,” in Cognitive Cellular Systems (CCS), 2014 1st International Workshop on, 2014, pp. 1-5: IEEE.

[60] H. Hartenstein and K. P. Laberteaux, “A tutorial survey on vehicular ad hoc networks,” IEEE Communications Magazine, vol. 46, no. 6, p. 164, 2008.

[61] J. Härri, F. Filali, and C. Bonnet, “Mobility models for vehicular ad hoc networks: a survey and taxonomy,” IEEE Communications Surveys & Tutorials, vol. 11, no. 4, 2009.

[62] G. Yan, J. Lin, D. Rawat, and J. C. Enyart, “The role of network and mobility simulators in evaluating vehicular networks,” in Intelligent Computing and Information Science: Springer, 2011, pp. 706-712.

[63] G. Piro, L. A. Grieco, G. Boggia, F. Capozzi, and P. Camarda, “Simulating LTE cellular systems: An open-source framework,” IEEE Transactions on Vehicular Technology, vol. 60, no. 2, pp. 498-513, 2011.

[64] J. Blumenstein, J. C. Ikuno, J. Prokopec, and M. Rupp, “Simulating the long term evolution uplink physical layer,” in ELMAR, 2011 Proceedings, 2011, pp. 141-144: IEEE.

[65] Aricent, LTE layer 1, LTE baseband/PHY library. [Online]. Available: http://www.aricent.com/Expertise/LTE.aspx, last accessed October 2018.

[66] J. J. Sánchez, G. Gómez, D. Morales-Jiménez, and J. T. Entrambasaguas, “Performance evaluation of OFDMA wireless systems using WM-SIM platform,” in Proceedings of the 4th ACM International Workshop on Mobility Management and Wireless Access, 2006, pp. 131-134: ACM.

[67] C. Mehlfhrer, M. Wrulich, J. C. Ikuno, D. Bosanska, and M. Rupp, “Simulating the long term evolution physical layer,” in Signal Processing Conference, 2009 17th European, 2009, pp. 1471-1478: IEEE.

[68] J. C. Ikuno, M. Wrulich, and M. Rupp, “System level simulation of LTE networks,” in Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st, 2010, pp. 1-5: IEEE.

[69] T. Issariyakul and E. Hossain, “Introduction to Network Simulator 2 (NS2),” in Introduction to Network Simulator NS2: Springer, 2012, pp. 21-40.

[70] QualNet Network Simulator, “Scalable network technologies,” Inc.[Online]. Available:web.scalable-networks.com/qualnet-network-simulator-software, 2011, last accessed October 2018.

[71] O. Ghica, “SIDnet-SWANS manual,” Evanston, IL, USA: Northwestern University, 2010.

[72] M. Antoniotti and A. Göllü, “SHIFT and SMART-AHS: A Language for Hybrid System Engineering Modeling and Simulation,” in DSL, 1997, vol. 97, pp. 14-14.

[73] A. Halati, H. Lieu, and S. Walker, “CORSIM-corridor traffic simulation model,” in Traffic Congestion and Traffic Safety in the 21st Century: Challenges, Innovations, and Opportunities Urban Transportation Division, ASCE; Highway Division, ASCE; Federal Highway Administration, USDOT; and National Highway Traffic Safety Administration, USDOT, 1997.

[74] J. Härri, F. Filali, C. Bonnet, and M. Fiore, “VanetMobiSim: generating realistic mobility patterns for VANETs,” in Proceedings of the 3rd International Workshop on Vehicular Ad hoc Networks, 2006, pp. 96-97: ACM.

[75] M. Fellendorf, “VISSIM: A microscopic simulation tool to evaluate actuated signal control including bus priority,” in 64th Institute of Transportation Engineers Annual Meeting, 1994, pp. 1-9: Springer.

[76] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, “Sumo-simulation of urban mobility,” in The Third International Conference on Advances in System Simulation (SIMUL 2011), Barcelona, Spain, 2011, vol. 42.

[77] R. Bishop, “A Survey of intelligent Vehicle Application,” in Proceedings of IEEE Intelligent Vehicles Symposium, 2000.

[78] A. Girard, J. Sousa, and K. Hedrick, “An Overview of Emerging Results in Networked Multi-Vehicle Systems (I),” in IEEE Conference on Decision and Control, 2001, vol. 2, pp. 1485-1490: IEEE; 1998.

[79] Y. Toor, P. Muhlethaler, and A. Laouiti, "Vehicle ad hoc networks: Applications and related technical issues," IEEE Communications Surveys & Tutorials, vol. 10, no. 3, 2008.

[80] P. Papadimitratos, A. De La Fortelle, K. Evenssen, R. Brignolo, and S. Cosenza, "Vehicular communication systems: Enabling technologies, applications, and future outlook on intelligent transportation," IEEE Communications Magazine, vol. 47, no. 11, 2009.

[81] C2C-CC, "Car to Car Communication Consortium Manifesto: Overview of the C2C-CC System, Car to Car Communication Consortium, Tech. Rep. Version 1.1, 2007.

[82] T. ETSI, "Intelligent transport systems (ITS); vehicular communications; basic set of applications; definitions," Tech. Rep. ETSI TR 102 6382009.

[83] G. Karagiannis, O. Altintas, E. Ekici, G. Heijenk, B. Jarupan, K. Lin, and T. Weil, "Vehicular networking: A survey and tutorial on requirements, architectures, challenges, standards and solutions," IEEE Communications Surveys & Tutorials, vol. 13, no. 4, pp. 584-616, 2011.

[84] M. Fiore, J. Härri, F. Filali, and C. Bonnet, "Understanding vehicular mobility in network simulation," in Mobile Adhoc and Sensor Systems, 2007. MASS 2007. IEEE International Conference on, 2007, pp. 1-6: IEEE.

[85] M. Fiore, "Vehicular mobility and network simulation," in Handbook on Vehicular Networks, S. Olariu and MC Weigle Eds., Taylor and Francis, 2008.

[86] U. Models, "UDel Models For Simulation of Urban Mobile Wireless Networks," ed: ed, 2007.

[87] Q. Zheng, X. Hong, and J. Liu, "An agenda based mobility model," in Simulation Symposium, 2006. 39th Annual, 2006, pp. 8 pp.-195: IEEE.

[88] M. Smith, G. Duncan, and S. Druitt, "PARAMICS: microscopic traffic simulation for congestion management," 1995.

[89] L. Neumann and S. Deymann, “TRANSSIM-NODE-a simulation tool for logistics nodes,” in Industrial Simulation Conference. Lyon: The European Simulation Society, 2008, pp. 283-287.

[90] S. Krauß, “Microscopic modeling of traffic flow: Investigation of collision free vehicle dynamics,” Universitat zu Koln., 1998.

[91] C. F. Daganzo, “The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory,” Transportation Research Part B: Methodological, vol. 28, no. 4, pp. 269-287, 1994.

[92] C. F. Daganzo, “The cell transmission model, part II: network traffic,” Transportation Research Part B: Methodological, vol. 29, no. 2, pp. 79-93, 1995.

[93] S. P. Hoogendoorn and P. H. Bovy, “State-of-the-art of vehicular traffic flow modelling,” Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, vol. 215, no. 4, pp. 283-303, 2001.

[94] Bononi L, Di Felice M, Bertini M, Croci E., “Parallel and distributed simulation of wireless vehicular ad hoc networks,” Proceedings of the 9th ACM international symposium on Modeling analysis and simulation of wireless and mobile systems, Oct 7, pp. 28-35, ACM, 2006.

[95] Traffic and Network Simulator, http://lca.epfl.ch/projects/trans/ last accessed Feburary 2019

[96] Anda J, LeBrun J, Ghosal D, Chuah CN, Zhang M., “ VGrid: vehicular adhoc networking and computing grid for intelligent traffic control”, Proceedings of IEEE 61st Vehicular Technology Conference, May 30, Vol. 5, pp. 2905-2909, IEEE 2005.

[97] A. Abadi, T. Rajabioun, and P. A. Ioannou, “Traffic Flow Prediction for Road Transportation Networks With Limited Traffic Data,” IEEE Transaction on Intelligent Transportation Systems, vol. 16, no. 2, pp. 653-662, 2015.

[98] C. Ameixieira, A. Cardote, F. Neves, R. Meireles, S. Sargento, L. Coelho, J. Afonso, B. Areias, E. Mota, R. Costa, and R. Matos, “Harbornet: a real-world

testbed for vehicular networks," IEEE Communications Magazine, vol. 52, no. 9, pp. 108-114, 2014.

[99] F. Hagenauer, F. Dressler, and C. Sommer, "Poster: A simulator for heterogeneous vehicular networks," in Vehicular Networking Conference (VNC), 2014 IEEE, 2014, pp. 185-186: IEEE.

[100] R. S. Sutton, A. G. Barto, and F. Bach, Reinforcement learning: An introduction. MIT press, 1998.

[101] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," Journal of Artificial Intelligence Research, vol. 4, pp. 237-285, 1996.

[102] A. Gosavi, "Reinforcement learning: A tutorial survey and recent advances," INFORMS Journal on Computing, vol. 21, no. 2, pp. 178-192, 2009.

[103] R. E. Bellman and S. E. Dreyfus, Applied dynamic programming. Princeton university press, 2015.

[104] G. A. Rummery and M. Niranjan, On-line Q-learning using connectionist systems. University of Cambridge, Department of Engineering Cambridge, England, 1994.

[105] C. J. C. H. Watkins, "Learning from delayed rewards," King's College, Cambridge, 1989.

[106] C. Claus and C. Boutilier, "The dynamics of reinforcement learning in cooperative multiagent systems," AAAI/IAAI, vol. 1998, pp. 746-752, 1998.

[107] S. Kapetanakis and D. Kudenko, "Reinforcement learning of coordination in cooperative multi-agent systems," AAAI/IAAI, vol. 2002, pp. 326-331, 2002.

[108] M. Tan, "Multi-agent reinforcement learning: Independent vs. cooperative agents," in Proceedings of the Tenth International Conference on Machine Learning, 1993, pp. 330-337.

[109] Y. Shoham, R. Powers, and T. Grenager, "Multi-agent reinforcement learning: a critical survey," Technical report, Stanford University 2003.

[110] M. L. Littman, “Markov games as a framework for multi-agent reinforcement learning,” in Machine Learning Proceedings 1994: Elsevier, 1994, pp. 157-163.

[111] L. Busoniu, R. Babuška, and B. De Schutter, “Multi-agent reinforcement learning: A survey,” in Control, Automation, Robotics and Vision, 2006. ICARCV’06. 9th International Conference on, 2006, pp. 1-6: IEEE.

[112] L. Busoniu, R. Babuška, and B. De Schutter, “Multi-agent reinforcement learning: An overview,” Innovations in Multi-Agent Systems and Applications - 1, vol. 310, pp. 183-221, 2010.

[113] E. Maskin, “Nash equilibrium and welfare optimality,” The Review of Economic Studies, vol. 66, no. 1, pp. 23-38, 1999.

[114] M. Bowling and M. Veloso, “Multiagent learning using a variable learning rate,” Artificial Intelligence, vol. 136, no. 2, pp. 215-250, 2002.

[115] M. Lauer and M. Riedmiller, “An algorithm for distributed reinforcement learning in cooperative multi-agent systems,” in Proceedings of the Seventeenth International Conference on Machine Learning, 2000: Citeseer.

[116] H. Ye and G. Y. Li, “Deep reinforcement learning for resource allocation in V2V communications,” arXiv preprint arXiv:1711.00968, 2017.

[117] I. Arel, C. Liu, T. Urbanik, and A. Kohls, “Reinforcement learning-based multi-agent system for network traffic signal control,” IET Intelligent Transport Systems, vol. 4, no. 2, pp. 128-135, 2010.

[118] S. J. Pan and Q. Yang, “A survey on transfer learning,” IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, 2010.

[119] M. E. Taylor and P. Stone, “Transfer learning for reinforcement learning domains: A survey,” Journal of Machine Learning Research, vol. 10, no. Jul, pp. 1633-1685, 2009.

[120] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” Journal of Big Data, vol. 3, no. 1, p. 9, 2016.

[121] Q. Zhao and D. Grace, "Transfer learning for QoS aware topology management in energy efficient 5G cognitive radio networks," in 5G for Ubiquitous Connectivity (5GU), 2014 1st International Conference on, 2014, pp. 152-157: IEEE.

[122] Y. Han, D. Grace, and P. Mitchell, "Energy efficient topology management for beyond next generation mobile broadband systems," in Wireless Communication Systems (ISWCS), 2012 International Symposium on, 2012, pp. 331-335: IEEE.

[123] K. Gomez, K. Sithamparanathan, M. Vidal, V. Boussemart, R. Ramos, R. Hermenier, T. Rasheed, L. Goratti, D. Grace, Q. Zhao, and Y. Han, "Aerial base stations with opportunistic links for next generation emergency communications," IEEE Communications Magazine, pp. 31-39, 2016.

[124] E. Baştuğ, M. Bennis, and M. Debbah, "A transfer learning approach for cache-enabled wireless networks," in Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2015 13th International Symposium on, 2015, pp. 161-166: IEEE.

[125] A. Galindo-Serrano, L. Giupponi, and G. Auer, "Distributed Learning in Multiuser OFDMA Femtocell Networks," in VTC Spring, 2011, pp. 1-6.

[126] Q. Zhao, D. Grace, A. Vilhar, and T. Javornik, "Using k-means clustering with transfer and Q learning for spectrum, load and energy optimization in opportunistic mobile broadband networks," in Wireless Communication Systems (ISWCS), 2015 International Symposium on, 2015, pp. 116-120: IEEE.

[127] S. Zeadally, R. Hunt, Y. S. Chen, A. Irwin, and A. Hassan, "Vehicular ad hoc networks (VANETS): status, results, and challenges," Telecommunication Systems, vol. 50, no. 4, pp. 217-241, 2012.

[128] BuNGee, Broadband Radio Access Networks (BRAN); Very high capacity density BWA networks; System architecture, economic model and derivation of technical requirements, ETSI TR 101 534 V1.1.1, www.ict-bungee.eu.

[129] R. Lamm, B. Psarianos, and T. Mailaender, Highway design and traffic safety engineering handbook. New York : McGraw-Hill, 1999.

[130] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/341513/pdfmanforstreets.pdf, last accessed June 2017.

[131] 3GPP. Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA physical layer aspects (3GPP TR 36.814 version 9.0.0 Release 9). December 2010.

[132] P. Kyösti, J. Meinilä, L. Hentilä, X. Zhao, T. Jämsä, C. Schneider, M. Narandzić, M. Milojević, A. Hong, J. Ylitalo, V. Holappa, M. Alatossava, R. Bultitude, Y. de Jong, and T. Rautiainen. IST-4-027756 WINNER II Deliverable D1.1.2: WINNER II channel models. 2008.

[133] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, Stochastic geometry and its applications. John Wiley & Sons, 2013.

[134] 3GPP. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) system scenarios (3GPP TR 36.952 version 11.0.0 Release 11). December 2012.

[135] 3GPP. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (3GPP TS 36.213 version 11.5.0 Release 11). December 2013.

[136] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell association and interference coordination in heterogeneous LTE-A cellular networks," IEEE Journal on Selected Areas in Communications, vol. 28, no. 9, pp. 1479-1489, 2010, Art. no. 5646339.

[137] N. W. Sung, N. T. Pham, H. Yoon, S. Lee, and W. J. Hwang, "Base station association schemes to reduce unnecessary handovers using location awareness in femtocell networks," Wireless Networks, vol. 19, no. 5, pp. 741-753, 2013.

[138] C. S. Chiu and C. C. Lin, "Comparative downlink shared channel evaluation of WCDMA release 99 and HSDPA," in Networking, Sensing and Control, 2004 IEEE International Conference on, 2004, vol. 2, pp. 1165-1170: IEEE.

[139] 3GPP. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) procedures in idle mode (3GPP TS 36.304 version 9.1.0 Re-

lease 9). December 2010.

[140] S. Russell, and P. Norvig, Artificial Intelligence - A modern approach, Prentice-Hall, Englewood Cliffs, vol. 25, p. 27, 1995.

[141] M. Haenggi, “Mean interference in hard-core wireless networks,” IEEE Communications Letters, vol. 15, no. 8, pp. 792-794, 2011.

[142] J. Lim and D. Hong, “Management of neighbor cell lists and physical cell identifiers in self-organizing heterogeneous networks,” Journal of Communications and Networks, vol. 13, no. 4, pp. 367-376, 2011.

[143] Z. Becvar, P. Mach, and M. Vondra, “Optimization of SINR-based neighbor cell list for networks with small cells,” in Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on, 2013, pp. 2346-2351: IEEE.

[144] F. Pantisano, M. Bennis, W. Saad, S. Valentin, M. Debbah, and A. Zappone, “Proactive user association in wireless small cell networks via collaborative filtering,” in Signals, Systems and Computers, 2013 Asilomar Conference on, 2013, pp. 1601-1605: IEEE.

[145] P. Coucheney, E. Hyon, and J. -M. Kelif, “Mobile association problem in heterogenous wireless networks with mobility,” in Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on, 2013, pp. 3129-3133: IEEE.

[146] W. Saad, Z. Han, R. Zheng, M. Debbah, and H. V. Poor, “A college admissions game for uplink user association in wireless small cell networks,” in INFOCOM, 2014 Proceedings IEEE, 2014, pp. 1096-1104: IEEE.

[147] J. G. Andrews, S. Singh, Q. Ye, X. Lin, and H. S. Dhillon, “An overview of load balancing in HetNets: Old myths and open problems,” IEEE Wireless Communications, vol. 21, no. 2, pp. 18-25, 2014.

[148] D. Bethanabhotla, O. Y. Bursalioglu, H. C. Papadopoulos, and G. Caire, “User association and load balancing for cellular massive MIMO,” in Information Theory and Applications Workshop (ITA), 2014, pp. 1-10: IEEE.

[149] A. AlAmmouri, H. ElSawy, and M. -S. Alouini, “Load-aware modeling for uplink cellular networks in a multi-channel environment,” in Personal, Indoor, and Mobile Radio Communication (PIMRC), 2014 IEEE 25th Annual International Symposium on, 2014, pp. 1591-1596: IEEE.

[150] N. Namvar, W. Saad, B. Maham, and S. Valentin, “A context-aware matching game for user association in wireless small cell networks,” in Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, 2014, pp. 439-443: IEEE.

[151] F. Zhou, B. Wang, W. Wang, Y. Zhang, M. Lei, and L. Jiang, “Analysis of Spectral Efficiency and Management of Cross-Tier Interference in Femtocell Network,” in Vehicular Technology Conference (VTC Fall), 2013 IEEE 78th, 2013, pp. 1-6: IEEE.

[152] E. Demarchou, C. Psomas, and I. Krikidis, “Mobility management in ultra-dense networks: Handover skipping techniques,” IEEE Access, vol. 6, pp. 11921-11930, 2018.

[153] E. Ndashimye, N. I. Sarkar, and S. K. Ray, “A mobility-aware network selection method for vehicle-to-infrastructure communication over LTE-A multi-tier networks,” in Information Networking (ICOIN), 2017 International Conference on, 2017, pp. 315-320: IEEE.

[154] K. Okada, D. K. Park, And S. Yoshimoto, “A dynamic channel assignment strategy using information on speed and moving direction for micro cellular systems,” IEICE Transactions on Communications, vol. 79, no. 3, pp. 279-288, 1996.

[155] B. Li, H. Zhang, P. Hao, and J. Li, “Sojourn time estimation-based small cell selection in Ultra-Dense Networks,” in Personal, Indoor, and Mobile Radio Communications (PIMRC), 2017 IEEE 28th Annual International Symposium on, 2017, pp. 1-5: IEEE.

[156] T. Liu, S. Zhou, and Z. N. Tsinghua, “Mobility-aware coded-caching scheme for small cell network,” in Communications (ICC), 2017 IEEE International Conference on, 2017, pp. 1-6: IEEE.

[157] C. H. Lee and Z. S. Syu, "Handover analysis of macro-assisted small cell networks," in Internet of Things (iThings), 2014 IEEE International Conference on, and Green Computing and Communications (GreenCom), IEEE and Cyber, Physical and Social Computing (CPSCom), IEEE, 2014, pp. 604-609: IEEE.

[158] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavasi, C. Patel and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," Communications Magazine, IEEE, vol. 52, no. 2, pp. 82-89, 2014.

[159] A. Rath and S. Panwar, "Fast handover in cellular networks with femtocells," Proc. 2012 IEEE ICC, pp. 2752-2757, 0000.

[160] K. Doya, "Reinforcement learning in continuous time and space," Neural Computation, vol. 12, no. 1, pp. 219-245, 2000.

[161] Y. I. Kim, K. J. Lee, and Y. O. Chin, "Analysis of multi-level threshold handoff algorithm," in Global Telecommunications Conference, 1996. GLOBECOM'96.'Communications: The Key to Global Prosperity, 1996, vol. 2, pp. 1141-1145: IEEE.

[162] M. D. Austin and G. L. Stuber, "Velocity adaptive handoff algorithms for microcellular systems," IEEE Transactions on Vehicular Technology, vol. 43, no. 3, pp. 549-561, 1994.

[163] X. Ge, H. Cheng, G. Mao, Y. Yang, and S. Tu, "Vehicular communications for 5G cooperative small-cell networks," IEEE Transactions on Vehicular Technology, vol. 65, no. 10, pp. 7882-7894, 2016.

[164] L. Li, Y. Xu, B. H. Soong, and L. Ma, "A reinforcement sensor embedded vertical handoff controller for vehicular heterogeneous wireless networks," Sensors, vol. 13, no. 11, pp. 15026-15047, 2013.

[165] H. Van Hasselt and M. A. Wiering, "Reinforcement learning in continuous action spaces," in Approximate Dynamic Programming and Reinforcement Learning, 2007. ADPRL 2007. IEEE International Symposium on, 2007, pp. 272-279: IEEE.

[166] Z. Li, C. Wang, and C. J. Jiang, "User association for load balancing in vehicular networks: An online reinforcement learning approach," IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 8, pp. 2217-2228, 2017.

[167] R. Bellman, Dynamic Programming. Courier Corporation, 2013.

[168] R. S. Sutton, "Integrated modeling and control based on reinforcement learning and dynamic programming," in Advances in Neural Information Processing Systems, 1991, pp. 471-478.

[169] C. J. Watkins and P. Dayan, "Q-learning," Machine Learning, vol. 8, no. 3-4, pp. 279-292, 1992.

[170] T. Jiang, Q. Zhao, D. Grace, A. G. Burr, and T. Clarke, "Single-state Q-learning for self-organised radio resource management in dual-hop, 5G high capacity density networks," Transactions on Emerging Telecommunications Technologies, 2016.

[171] N. Morozs, D. Grace, and T. Clarke, "Case-based reinforcement learning for cognitive spectrum assignment in cellular networks with dynamic topologies," in Military Communications and Information Systems Conference (MCC), 2013, pp. 1-6: IEEE.

[172] K. Katahira, "The relation between reinforcement learning parameters and the influence of reinforcement history on choice behavior," Journal of Mathematical Psychology, vol. 66, pp. 59-69, 2015.

[173] M. Bowling, "Convergence and no-regret in multiagent learning," in Advances in Neural Information Processing Systems, 2005, pp. 209-216.

[174] M. Bowling and M. Veloso, "Rational and convergent learning in stochastic games," in International Joint Conference on Artificial Intelligence, 2001, vol. 17, no. 1, pp. 1021-1026: Lawrence Erlbaum Associates Ltd.

[175] N. Morozs, T. Clarke, D. Grace, and Q. Zhao, "Distributed Q-learning based dynamic spectrum management in cognitive cellular systems: Choosing the right learning rate," in Computers and Communication (ISCC), 2014 IEEE Symposium on, 2014, pp. 1-6: IEEE.

[176] N. Mehta, S. Natarajan, P. Tadepalli, and A. Fern, “Transfer in variable-reward hierarchical reinforcement learning,” Machine Learning, vol. 73, no. 3, p. 289, 2008.

[177] N. Morozs, T. Clarke, and D. Grace, “Distributed heuristically accelerated Q-learning for robust cognitive spectrum management in LTE cellular systems,” IEEE Transactions on Mobile Computing, vol. 15, no. 4, pp. 817-825, 2016.

[178] S. Kapoor, D. Grace, and T. Clarke, “A base station selection scheme for handover in a mobility-aware ultra-dense small cell urban vehicular environment,” in Personal, Indoor, and Mobile Radio Communications (PIMRC), 2017 IEEE 28th Annual International Symposium on, 2017, pp. 1-5: IEEE.

[179] D. Grace, “Distributed dynamic channel assignment for the wireless environment,” PhD Thesis, University of York, 1999.

[180] R. A. Bianchi, M. F. Martins, C. H. Ribeiro, and A. H. Costa, “Heuristically-accelerated multiagent reinforcement learning,” IEEE Transactions on Cybernetics, vol. 44, no. 2, pp. 252-265, 2014.

[181] R. A. Bianchi, C. H. Ribeiro, and A. H. Costa, “Heuristically Accelerated Q-Learning: a new approach to speed up Reinforcement Learning,” in Brazilian Symposium on Artificial Intelligence, 2004, pp. 245-254: Springer.

[182] A. Walid, A. Kobbane, A. Mabrouk, E. Sabir, T. Taleb, and M. El Koutbi, “Group vertical handoff management in heterogeneous networks,” Wireless Communications and Mobile Computing, vol. 16, no. 10, pp. 1256-1270, 2016.

[183] S. Singh, M. Kearns, and Y. Mansour, “Nash convergence of gradient dynamics in general-sum games,” in Proceedings of the Sixteenth conference on Uncertainty in Artificial Intelligence, 2000, pp. 541-548: Morgan Kaufmann Publishers Inc.

[184] M. Nazir, M. Bennis, K. Ghaboosi, A. B. MacKenzie, and M. Latva-aho, “Learning based mechanisms for interference mitigation in self-organized femtocell networks,” in Signals, Systems and Computers (ASILOMAR), 2010 Con-

ference Record of the 44th Asilomar Conference on, 2010, pp. 1886-1890: IEEE.

[185] C. Drummond, "Accelerating reinforcement learning by composing solutions of automatically identified subtasks," Journal of Artificial Intelligence Research, vol. 16, pp. 59-104, 2002.

[186] L. M. Gambardella and M. Dorigo, "Ant-Q: A reinforcement learning approach to the traveling salesman problem," in Machine Learning Proceedings 1995: Elsevier, 1995, pp. 252-260.

[187] R. A. Bianchi, L. A. Celiberto Jr, P. E. Santos, J. P. Matsuura, and R. L. de Mantaras, "Transferring knowledge as heuristics in reinforcement learning: A case-based approach," Artificial Intelligence, vol. 226, pp. 102-121, 2015.

[188] R. A. Bianchi and R. L. de Mantaras, "Case-Based Multiagent Reinforcement Learning: Cases as Heuristics for Selection of Actions," in ECAI, 2010, pp. 355-360.

[189] L. Torrey and J. Shavlik, Transfer Learning, Handbook of Research on Machine Learning Applications, 2009.

[190] Q. Zhao, "Intelligent Radio Resource Management for Mobile Broadband Networks," University of York, 2013.

[191] T. Gabel and M. Riedmiller, "Multi-agent case-based reasoning for cooperative reinforcement learners," in European Conference on Case-Based Reasoning, 2006, pp. 32-46: Springer.

[192] C. Dhahri and T. Ohtsuki, "Learning-based cell selection method for femtocell networks," in Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th, 2012, pp. 1-5: IEEE.

[193] A. A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," Computer Communications, vol. 30, no. 14-15, pp. 2826-2841, 2007.