9-27-2023

Transmission Power based Congestion Control using Q-Learning algorithm in Vehicular Ad Hoc Networks (VANET)

Pooja Chandrasekharan
University of Windsor

Transmission Power based Congestion Control using Q-Learning algorithm in Vehicular Ad Hoc Networks (VANET)

by

Pooja Chandrasekharan

A Thesis
Submitted to the Faculty of Graduate Studies
through the School of Computer Science
in Partial Fulfillment of the Requirements for
the Degree of Master of Science
at the University of Windsor

Windsor, Ontario, Canada

2023

Transmission Power based Congestion Control using Q-Learning algorithm in Vehicular Ad Hoc Networks (VANET)

by

Pooja Chandrasekharan

APPROVED BY:

A. Sarker
Department of Mathematics and Statistics

S. Khan
School of Computer Science

A. Jaekel, Advisor
School of Computer Science

September 7, 2023

DECLARATION OF CO-AUTHORSHIP

I. Co-Authorship

I hereby declare that this thesis incorporates material that is the result of research conducted under the supervision of Dr. Arunita Jaekel. In all cases, the key ideas, primary contribution, experimental designs, data analysis, and interpretation were performed by the author, and the contribution of the co-author was primarily through providing feedback and the proofreading of the published manuscripts.

I am aware of the University of Windsor Senate Policy on Authorship, and I certify that I have properly acknowledged the contribution of other researchers to my thesis and have obtained written permission from each of the co-author(s) to include the above material(s) in my thesis. I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my work.

II. General

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution. I understand that my thesis may be made electronically available to the public.

ABSTRACT

To enhance road safety, Vehicular ad hoc networks (VANETs), an emerging wireless technology used for vehicle-to-vehicle and vehicle-to-infrastructure communication, are essential components to reduce road accidents and traffic congestion in Intelligent Transportation Systems (ITS). It also provides additional services to vehicles and their users. However, vehicles must balance awareness and congestion control in a dynamic environment to efficiently transmit basic safety messages (BSMs) and event-driven warnings. The limited channel capacity makes the reliable delivery of BSMs a challenging problem for VANETs. This paper aims to optimize the performance of VANETs by effectively managing channel load and reducing congestion by maintaining the channel busy ratio (CBR) near the threshold value of 0.6. This is resolved using a transmission power-based congestion control algorithm that employs a Markov decision process (MDP) and solves it using a Q-Learning algorithm. The algorithm uses varying transmission power levels to lower the channel busy ratio while maintaining high awareness for surrounding vehicles. According to simulation results for various traffic scenarios, the suggested technique chooses a suitable transmission power depending on the present channel circumstances to achieve a balance between awareness and bandwidth usage. The findings show that the proposed strategy reliably maintained the channel load at or near the stipulated level without surpassing it for both low and high traffic densities.

DEDICATION

I dedicate this thesis to my parents for their unwavering support, patience, and love, without which this journey would have been impossible. You were my rock in times of uncertainty, my guiding star in times of darkness, and my cheerleader in times of achievement.

To my supervisor Dr. Arunita Jaekel, for their invaluable guidance and insights, shaping my intellect and always challenging me to aim higher. Their wisdom and tireless dedication have inspired me beyond measure.

To AK, for their constant encouragement, shared laughter, and for standing by me through the many ups and downs. Your companionship made this journey less daunting and more rewarding.

Lastly, to all the curious minds that seek knowledge and truth in the world. May this work contribute to the vast ocean of understanding and inspire others as I have been inspired.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to all those who provided me with the possibility to complete this thesis. Special gratitude is given to my supervisor Dr. Arunita Jaekel, whose contribution in stimulating suggestions and encouragement helped me to coordinate my project, especially in writing this thesis.

I am heartily thankful to my internal reader, Dr. Shafaq Khan, and my external reader, Dr. Animesh Sarker, whose invaluable feedback, expertise, and thorough reading of the manuscript added substantially to the improvement of the final version of this thesis. Your insightful critique and suggestions were invaluable.

I wish to express my sincere gratitude to Ph.D. student Xiaofeng Liu and Ben St. Amour, whose unwavering assistance and thoughtful guidance were instrumental in my research process. Their depth of understanding, patience, and willingness to help at all times were truly indispensable.

My sincere thanks go to my parents and friends, who have stood by me throughout this journey. Their moral support, understanding, and patience during the completion of this thesis are greatly appreciated and have been a source of continuous motivation for me.

I would also like to acknowledge the entire Department of Computer Science for providing a conducive environment, resources, and support throughout my studies. I am particularly grateful for the assistance given by Christine Weisener.

Lastly, my profound gratitude goes to all the authors and researchers whose studies have contributed to the foundation of this work. Your exhaustive efforts and groundbreaking research were the pillars upon which this thesis was built.

Ultimately, I acknowledge that the research journey is as important as the destination, and I am grateful for every step that led me here.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

VANET	Vehicular ad-hoc network
MANET	Mobile Ad-Hoc Network
RSU	Road-side Unit
OBU	On-board Unit
DSRC	Dedicated short-range communication
FCC	Federal Communication Commission
C-V2X	Cellular Vehicle to everything
WAVE	Wireless Access in Vehicular Environment
V2V	Vehicle-to-Vehicle
V2I	Vehicle-to-Infrastructure
V2P	Vehicle-to-Pedestrian
V2X	Vehicle-to-Everything
BSM	Basic safety message
ITS	Intelligent Transportation system
MAC	Medium Access Control
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CBR	Channel Busy Ration

IPD	Inter-Packet Delay
BER	Beacon Error Rate
RL	Reinforcement Learning
MDP	Markov Decision Process
VD	Vehicle Density
SUMO	Simulation of Urban MObility
OMNeT++	Objective Modular Network Testbed
Veins	Vehicles in Simulation
QBTPCC	Q-Learning Based Transmission Power Congestion Control
BACVT	Balancing Awareness and Congestion with Variable Tx Power
MDPRP	MDP-based transmission policies (Rate and Power)
AI	Artificial Intelligence
ML	Machine Learning

CHAPTER 1

Introduction

1.1 Vehicular ad-hoc networks

For more than 80 years, ensuring road traffic safety has remained an ongoing concern. Since motor vehicles are the predominant mode of transportation, their safety holds immense significance. Despite the global pandemic in 2020, the projected number of fatalities in motor vehicle crashes reached 38,680, the highest recorded since 2007 [1]. Alongside focusing on driver behavior and attitudes, enhancing vehicle safety through inter-vehicular communication is a vital factor to address.

A vehicular Ad-Hoc Network (VANET) is a wireless ad-hoc network that facilitates vehicle communication [2]. It is a subset of Mobile Ad Hoc Networks (MANETs) that employ vehicles, whether in motion or stationary, as nodes within a mobile network connected through a wireless network infrastructure [2]. In VANETs, communication between nodes is typically accomplished using the North American Dedicated Short-Range Communication (DSRC) standard, which utilizes the IEEE 802.11p VANET standard for wireless communication. [3]

In 1999, the Federal Communication Commission (FCC) of the United States allocated Dedicated Short-Range Communication (DSRC), a licensed spectrum of 75MHz in 5.9 GHz frequency bandwidth for communication between vehicles and road-side units [4]. DSRC, which stands for Dedicated Short-Range Communication, is a service utilized for communication over short to medium distances. It offers high-speed data transfer while minimizing latency. To facilitate communication within vehicular networks and establish standards for DSRC, the Wireless Access in Vehicular Environment (WAVE) protocol was introduced. WAVE, based on the IEEE 802.11p standard, enables communication in vehicular networks and supports the use of DSRC [5]. However, DSRC has limitations when transmitting large volumes of data and accessing the Internet of Vehicles. A new standard called Cellular-V2X (C-V2X) has been introduced to address these limitations. C-V2X, short for the cellular vehicle to everything, leverages cellular technology to enable connectivity between vehicles, road-side units, central authorities, and cloud-based services [6].

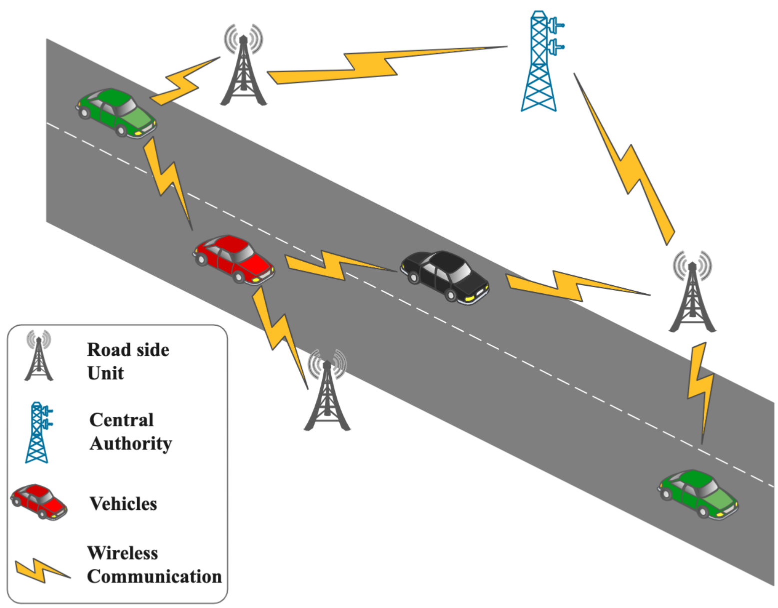

Figure 1.1: An example of a Vehicular ad-hoc network. [7]

The 5.9 GHz band has been specifically allocated 75 MHz of spectrum for vehicle-to-vehicle (V2V) communication in Vehicular Ad-Hoc Networks (VANETs) using dedicated short-range communications/wireless access for the vehicular environment (DSRC/WAVE) [4]. An additional 30 MHz spectrum is also reserved for cellular vehicle-to-everything (C-V2X) communication. However, this type of communication introduces the possibility of simultaneous transmissions, which can lead to packet collisions and diminish the reliability of the communication [8]. As the number of vehicles increases, broadcasting Basic Safety Messages (BSMs) can easily result in congestion on this single channel, leading to lower reception probabilities and reduced transmission ranges [3].

1.2 Motivation

The primary objective of this research is to enhance the performance of Vehicular Ad-Hoc Networks (VANETs) by implementing effective channel load management techniques and minimizing congestion. This is accomplished by maintaining the Channel Busy Ratio (CBR) close to a predetermined threshold value.

The research endeavors to strike a balance in VANETs, where the paramount importance lies in establishing awareness among vehicles. In VANETs, vehicles must communicate and collaborate with each other to facilitate efficient and safe transportation [9]. By addressing the challenges associated with channel load and congestion, this study aims to optimize the overall performance of VANETs, thereby enhancing the effectiveness of communication and cooperation among vehicles.

1.3 Problem Statement

Vehicular Ad-Hoc Networks (VANETs) are fast emerging as a transformative technology, capable of significantly improving communication efficiency and safety among vehicles on the move [10]. Nonetheless, the inherently dynamic and rapidly growing VANET environment presents formidable challenges, notably in channel congestion control [10]. Maintaining robust control over channel load and congestion is key to sustaining reliable and prompt communication among vehicles [10].

Present congestion control strategies in VANETs often fall short of reconciling the dual needs of maintaining high vehicle awareness while preserving a stable channel congestion level [11]. This inadequacy leads to many complications, including escalating packet loss, lagging transmission, and a downward spiral in overall network performance. Further complicating matters is that traditional congestion control methodologies developed for wired or stationary wireless networks may not directly translate into effective solutions for VANETs due to their unique operational characteristic [11].

Consequently, there is a pressing need to devise a novel solution tailored to address channel congestion control in the dynamic VANET environment [12]. This research endeavors to develop an optimized congestion control algorithm, utilizing innovative strategies such as transmission power adjustment and reinforcement learning techniques like Q-learning. This dynamic algorithm will adaptively regulate the transmission power of vehicles, managing channel load and maintaining a stable congestion level, ensuring efficient and reliable communication within the VANET [12].

The central goal of this thesis is to formulate a congestion control algorithm that balances high vehicle awareness and effective congestion control in VANETs. This innovative algorithm aims to optimize this delicate balance, promising efficient data transmission, reduced packet loss, and minimized transmission delays. The proposed solution, therefore, holds the potential to significantly enhance VANET performance and pave the way toward the realization of secure and dependable vehicular communication systems.

By pioneering an innovative approach to tackle channel congestion control in VANETs, this research aspires to contribute significantly to the body of knowledge in this field and provide practical insights for the future design and deployment of efficient congestion control mechanisms in real-world VANET scenarios.

1.4 Solution Outline

This research focuses on designing a congestion control algorithm for Vehicular Ad-Hoc Networks (VANETs) that utilizes Q-learning and transmission power adjustment. The algorithm balances the need for high vehicle awareness and maintains a stable channel congestion level.

Achieving this balance is crucial for the efficient and safe operation of VANETs.

Vehicle awareness refers to a vehicle's ability to have up-to-date knowledge about its surrounding environment, typically obtained through exchanging information with other vehicles or infrastructure. Maintaining a stable channel congestion level is essential to prevent data traffic overload, which can result in packet loss and increased transmission delays, thereby affecting the overall performance of vehicular networks.

The proposed algorithm aims to optimize the trade-off between awareness and congestion control in VANETs by employing Q-learning and adjusting transmission power. This research seeks to develop an effective solution that enables vehicles to maintain a high level of awareness while simultaneously ensuring a stable channel congestion level, ultimately enhancing the efficiency and safety of vehicular networks.

1.4.1 Contributions

The contribution of this research is summarized as follows:

- **Development of a Novel Congestion Control Algorithm:** The foremost contribution of this research is the development of a novel congestion control algorithm that utilizes dynamic transmission power adjustments and reinforcement learning techniques. This innovative algorithm effectively addresses the unique congestion control challenges in VANETs, offering a customized solution beyond traditional methodologies.

- **Enhancement of Vehicle Awareness and Congestion Control Balance:** By dynamically regulating transmission power and using reinforcement learning, the proposed algorithm balances high vehicle awareness and stable channel congestion. This contributes to more efficient data transmission, reduced packet loss, and reduced transmission delays in VANETs.

- **Provision of Practical Insights for Future VANET Development:** The findings and insights gained from this research provide valuable information for the future design and implementation of VANETs. They could guide the development of new congestion control mechanisms and help optimize existing systems.

- **Contribution to Academic Literature:** This research adds to the growing body of academic literature on congestion control in VANETs. It presents a novel approach to a longstanding problem, offering fresh perspectives that can spur further studies and technological advancements in this field.

  In essence, this research contributes significantly both to the academic community and the practical implementation of VANETs. It introduces a unique approach to congestion control that holds promise for enhancing vehicular communication systems' performance and safety.

To approach the complex problem of channel congestion in VANETs, our research introduces a framework that leverages Reinforcement Learning (RL) methods to solve the Markov Decision Process (MDP), focusing particularly on discrete action and state spaces [9]. At the heart of this framework is the Q-learning algorithm, a model-free RL method that aids in determining an optimal action-selection policy for a given MDP.

The training data for our Q-learning algorithm is drawn directly from a simulated dynamic traffic environment, providing a more realistic representation of state transitions. We gather observations of the Channel Busy Ratio (CBR) values under various transmission powers and vehicle densities, feeding these into our learning model to guide decision-making.

In our approach, we define a reward function that integrates both CBR and transmission power. The function is designed to balance the dual objectives of keeping the channel load under a targeted threshold (to manage congestion) and maximizing the transmission power (to maintain high vehicle awareness).

The performance of our approach is validated through extensive simulations. The results demonstrate that the proposed Q-learning-based solution effectively sustains the desired channel load under various dynamic traffic scenarios. Our solution exhibits a lower Beacon Error Rate (BER) than existing methods, underlining its superior performance in managing channel congestion. This result attests to the effectiveness of RL, specifically Q-learning, in addressing the challenges of congestion control in VANETs [9].

1.5 Thesis Organization

This thesis is divided into five main chapters, each detailing a significant stage in the research process. The remaining outline of this thesis is as follows:

Chapter two provides a comprehensive overview of the broad research area and explains the fundamental concepts and terminologies related to the study. It delves into the current research problems and solutions and presents an extensive literature review of crucial papers about the specific research problem. This chapter serves as the foundation of the research, setting the stage for the unique contribution proposed in the following chapter.

Chapter three discusses the proposed approach in detail. It begins with a high-level outline of the approach, highlighting its differences from existing methodologies and elucidating any novel or innovative features. The chapter then provides an in-depth description of the approach, including detailed flowcharts, algorithms, and comprehensive explanations of these elements. This chapter underscores the unique contribution of this research to the field of study.

The fourth chapter presents the results of the research. It starts by explaining the validation process of the proposed approach, including a description of the experimental and simulation setup. The chapter outlines how the 'success' of the approach is evaluated, such as by comparing performance with existing techniques or using standard/benchmark test inputs. It then presents the results, including data in tables, figures, and graphs, with detailed discussions to justify the claims and conclusions.

The final chapter summarizes the research findings, discusses the implications of the results, and proposes directions for future research. It revisits the research objectives and explains how the study has achieved them, underscoring the importance and relevance of the research contributions.

CHAPTER 2

Background Review

2.1 Basic Terminology

The following terminologies form the backbone of the concepts and methodologies discussed in this thesis, providing the necessary language for understanding and implementing the proposed approach to channel congestion control in VANETs.

- **Vehicular Ad-Hoc Networks (VANETs):** VANETs are a subclass of mobile ad-hoc networks that allow vehicles to communicate with each other (Vehicle-to-Vehicle, V2V) or with roadside infrastructure (Vehicle-to-Infrastructure, V2I) to provide safety and comfort to passengers [2].

- **Mobile Ad-Hoc Networks (MANETs):** MANETs are wireless networks where all the nodes are mobile and directly connected to each other without a centralized administrator or infrastructure [13].

- **Nodes:** In the context of this research, a node refers to a vehicle equipped with an onboard unit (OBU) that enables it to send, receive, and relay messages.

- **Onboard Unit (OBU):** An OBU is a device equipped in vehicles participating in VANETs that allows the vehicle to communicate wirelessly with other vehicles or infrastructure [14].

- **Channel Load:** This refers to the amount of data the communication channel carries at any given time. High channel load can lead to congestion, resulting in decreased network performance.

- **Channel Congestion:** Channel congestion occurs when the demand for a channel's capacity exceeds its ability to transmit data, leading to packet loss, delay, or blocking new connections.

- **Transmission Power:** In wireless communication, transmission power is the amount of electrical power sent out by a transmitter in the form of a signal.

- **Reinforcement Learning:** Reinforcement learning is an area of machine learning concerned with how software agents should take actions in an environment to maximize some notion of cumulative reward [15].

- **Q-learning:** Q-learning is a model-free reinforcement learning algorithm that seeks to learn the value of being in a given state and taking a specific action under a policy [16].

- **Basic Safety Message (BSM):** In the context of Intelligent Transportation Systems (ITS) and particularly VANETs, a BSM is a packet of data broadcasted periodically by vehicles to provide situational awareness and enhance roadway safety. It is also not event-driven. It typically includes important status information such as vehicle speed, heading, size, and position data. BSMs form a crucial part of the communication protocol in VANETs, providing vital real-time information for safety applications.

- **Channel Busy Ration (CBR):** The CBR is defined as the ratio between the time the channel is detected as busy and the total observation time. CBR is a helpful indicator of the channel load, with higher values indicating higher channel load and vice versa.

2.2 Overview of Research Area

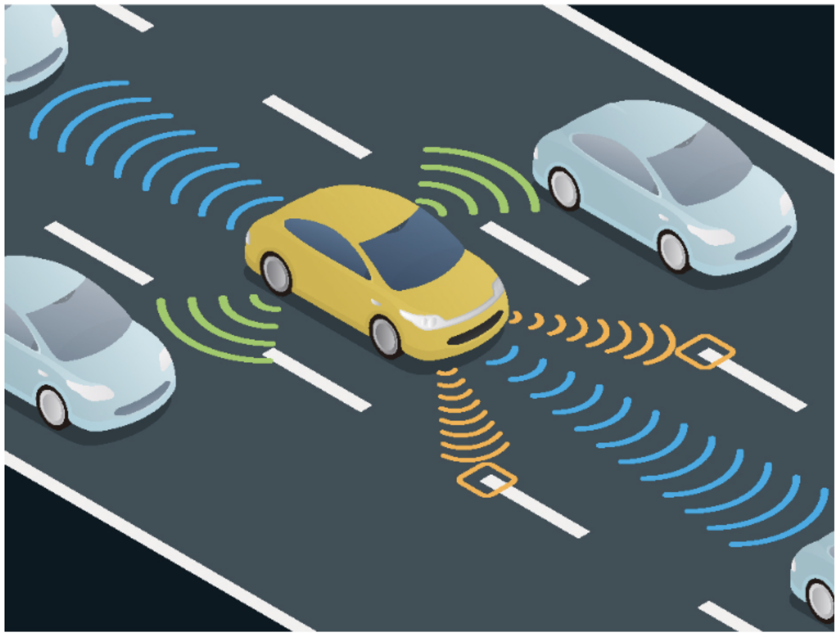

Figure 2.1: Vehicular Ad Hoc Network (VANET). [17]

Vehicular Ad-Hoc Networks (VANETs) are a subclass of Mobile Ad-Hoc Networks (MANETs) that are specially designed for the context of vehicles connected in a wireless manner [18]. Intending to improve road safety and facilitate better transportation services, VANETs stand at the intersection of several research disciplines, including computer science, telecommunications, and traffic engineering.

VANETs consist of vehicles, often referred to as nodes, that form a dynamic network and communicate with each other without the need for a fixed infrastructure [19]. These vehicles have onboard units (OBUs) to transmit and receive messages. The primary purpose of these networks is to facilitate Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communication, allowing for the sharing of valuable information such as traffic conditions, accident warnings, route guidance, and other safety alerts [19].

In VANETs, each vehicle has the ability to act as a wireless router or node, allowing cars approximately 100 to 300 meters from each other to connect and, in turn, create a network with a wide range. As vehicles fall out of the signal range and drop out of the network, other vehicles can join in, connecting vehicles to one another and creating a mobile Internet.

The dynamic and fast-changing nature of VANETs, the high node mobility, and the unique vehicle patterns present various technical challenges, including security, privacy, and the efficient management of channel load and congestion, making it a rich area for research and development [20]. The problem of channel congestion control, in particular, is the focus of this thesis.

2.3 Overview of VANET

2.3.1 VANET Environment

The VANET environment is a highly dynamic and complex communication environment designed to improve road safety and driving comfort [21]. A set of core components defines this environment:

- **Vehicles and their On-Board Units (OBUs):** Vehicles are the primary nodes in the VANET, each equipped with an On-Board Unit. The OBU is a device that enables a vehicle to communicate wirelessly with other vehicles and with infrastructure in its vicinity [22]. The OBU can transmit and receive messages, process data, and execute applications using information from other vehicles or infrastructure [22]. It is crucial in Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications.

- **Road-Side Units (RSUs):** RSUs are fixed infrastructure components that complement the OBUs in vehicles. They can provide various services, such as relaying messages over longer distances, disseminating safety messages or traffic updates, and connecting the VANET with other networks, like the Internet [23]. RSUs enhance the communication capabilities within the VANET, enabling information exchange even when direct V2V communication is not possible.

- **Other Infrastructure:** Besides OBUs and RSUs, other infrastructural components, like traffic signals, can be part of the VANET environment. These components can be equipped with communication capabilities and sensors to gather and disseminate data about traffic, road conditions, weather, and other relevant factors. The data collected by these components can feed into various VANET applications to improve road safety and traffic efficiency [14].

- **Communication Channels:** The VANET environment also includes the communication channels used for data transmission. These channels can become congested due to high vehicle densities or high data transmission rates, a challenge this thesis aims to address.

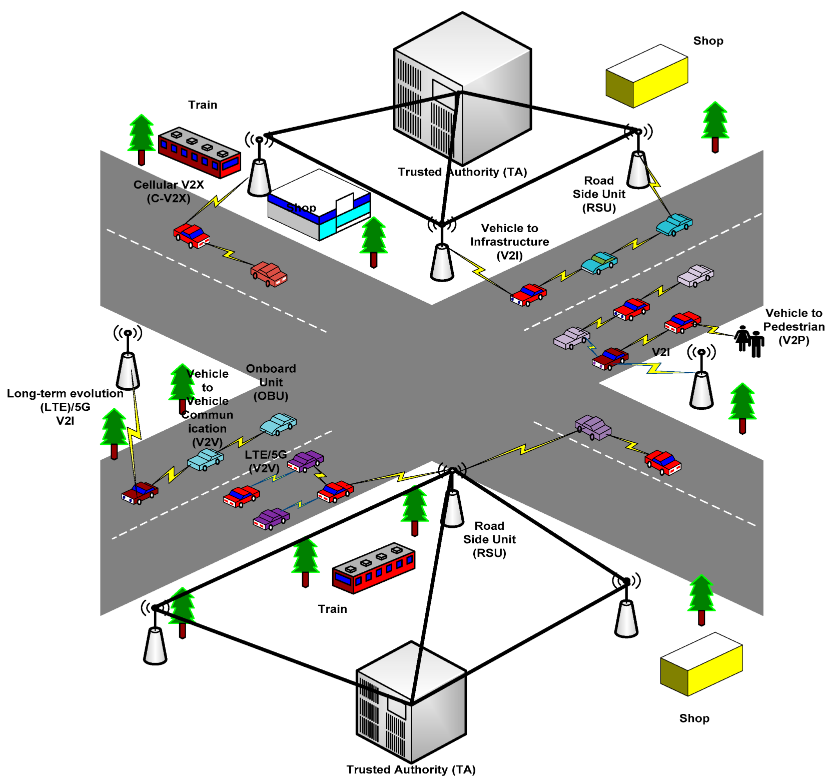

Figure 2.2: VANET Model Diagram. [19]

The VANET environment, while providing significant potential for improving road safety and traffic efficiency, is characterized by high dynamics due to vehicle mobility, varying vehicle densities, and changing environmental conditions [24]. This poses specific challenges for network performance and stability, including the challenge of channel congestion management that this research focuses on.

2.3.2 Types of Communication

VANETs enable direct communication between vehicles through onboard units (OBUs) or with infrastructure nodes such as roadside units (RSUs), thereby facilitating the dissemination of safety-related information. The different types of communication are:

- **In-Vehicle Communication:** This refers to communication within a single vehicle, typically between different devices or components of the vehicle's internal network [25]. For instance, various sensors, control units, and display systems in modern cars often communicate with each other to operate effectively. This type of communication is crucial for in-car infotainment systems, driver assistance features, and overall vehicle functionality.

- **Vehicle-to-Vehicle (V2V) Communication:** As the name suggests, V2V communication involves the direct exchange of information between vehicles [25]. It's a fundamental part of VANETs, enabling vehicles to share data such as speed, direction, position, and other safety-related information. This type of communication can significantly enhance road safety, allowing vehicles to anticipate potential hazards and adjust their actions accordingly.

- **Vehicle-to-Infrastructure (V2I) Communication:** This refers to the communication between vehicles and fixed infrastructure components such as Road-Side Units (RSUs), traffic lights, and traffic control centers. V2I communication can provide vehicles with important information about road conditions, traffic congestion, weather updates, and more [25]. This can enhance both safety and traffic efficiency.

- **Vehicle-to-Everything (V2X) Communication:** This umbrella term encompasses all types of vehicle communication. It includes V2V, V2I, and communications with pedestrians (V2P), cyclists, and the broader network [25]. V2X communication is expected to play a key role in the advent of autonomous driving and smart cities, enabling seamless integration of vehicles into the broader IoT ecosystem.

2.3.3 Congestion Control in VANET

Congestion control in Vehicular Ad-hoc Networks (VANETs) is a critical aspect affecting the performance and reliability of vehicle communication.

The 802.11p wireless protocol is commonly used in VANETs. It employs the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocols to minimize collisions and provide fair access to the communication channel. However, as a random-access protocol, CSMA/CA can lead to simultaneous transmissions, collisions, and network congestion [26].

Additionally, in the 802.11p Medium Access Control (MAC) layer, the request-to-send/clear-to-send (RTS/CTS) handshake is disabled, eliminating a potential means of reducing collisions and congestion.

As the number of vehicles increases, there will be a rise in the broadcasting of Basic Safety Messages (BSMs). This amplification in communication activity can cause congestion on the channel. This can lower the reception probability of messages and decrease transmission ranges.

In the dynamic and rapidly changing environment of VANETs, achieving optimal solutions for congestion and awareness is challenging due to the highly mobile nodes. As vehicular density fluctuates, managing congestion and ensuring the efficient transmission of data becomes complex. When channel congestion occurs, it manifests in multiple ways:

- **Packet Loss:** This refers to data packets sent from one vehicle not being received by another due to congested channels, resulting in loss of information and reduced efficiency of VANET applications [27].

- **Increased Delay:** The time taken for data to be transmitted from one node to another can significantly increase during congestion, especially in high-density situations. This can hinder the performance of VANET applications that rely on timely data transmission [27].
- **Reduced Throughput:** The overall network data capacity or throughput can be reduced during congestion, meaning fewer data can be transmitted over the network. This results in slower and less efficient data exchange [27].

Beyond channel congestion, VANETs face several other challenges, such as ensuring network security, maintaining privacy, and delivering quality service.

Channel congestion, in particular, refers to the state where the wireless communication channel used by vehicles in a VANET becomes saturated due to a high number of concurrent data transmissions. An excessive number of vehicles communicating on the same channel can lead to channel congestion, causing packet loss, increased delay, and reduced throughput. Hence, effective congestion control strategies are essential for the efficient operation of VANETs.

2.4 Overview of Machine Learning

Machine learning is the Artificial Intelligence branch that facilitates machines to perform specific jobs faster and skillfully using statistical learning [28].

Machine Learning is a subfield of Artificial Intelligence (AI) that allows computers to learn patterns and make decisions from data without being explicitly programmed. Essentially, it's about creating algorithms and models that use statistical methods to improve performance over time. Machine Learning can be broadly categorized into three types:

- **Supervised Learning:** This type of learning uses labeled data to train algorithms. In other words, the data input into the algorithm comes with desired output labels [28]. By analyzing the input-output pairs, the algorithm learns to map the input to the correct output. Examples of supervised learning include regression and classification problems.

- **Unsupervised Learning:** Unlike supervised learning, unsupervised learning deals with unlabeled data. The goal is to discover the data's underlying patterns, structures, or hidden relationships [28]. Standard techniques include clustering and dimensionality reduction.

- **Reinforcement Learning:** This type of learning is about making a series of decisions. The machine is trained to make specific decisions by rewarding and penalizing its actions [28]. In the long run, it learns to make the best decision from its experiences. This method is commonly used in robotics, gaming, and navigation.

2.5 Fundamental Concepts

2.5.1 Channel Congestion

In wireless communication networks, channel congestion refers to the scenario where too many packets are contending for transmission over a limited bandwidth wireless channel [29]. As with general network congestion, this can result in packet loss, increased latency, and decreased overall network performance. In dynamic and high-mobility networks like VANETs, channel congestion can be particularly challenging due to changing network topologies and the potentially high number of nodes contending for channel access.

2.5.2 Congestion Control

Congestion control refers to the mechanisms and techniques used in computer networks to avoid or manage congestion, ensuring stable and efficient network performance. Congestion occurs when the network's resource demand exceeds capacity, leading to packet loss, decreased throughput, and increased transmission delays [27]. Congestion control aims to prevent such occurrences by controlling the rate at which packets are sent into the network, often by adjusting it based on network conditions.

Congestion control can be reactive (responding to congestion after it happens by reducing transmission power) or proactive (preventing congestion by controlling the data entry rate into the network). In many modern networks, congestion control is performed by end systems where the sending rate is adjusted based on perceived network congestion.

2.5.3 Transmission Power

Transmission power in wireless communication refers to the power used to send signals from a transmitter. It's a critical parameter because it determines the transmitted signal's range and strength, impacting communication reliability, the network's coverage area, and power consumption.

In VANETs, controlling transmission power is a potential mechanism for managing channel load and congestion. By reducing transmission power, the communication range of a node can be limited, reducing the number of contending nodes for channel access and potentially alleviating congestion. Conversely, increasing transmission power may increase the communication range and improve connectivity but may also lead to increased interference and potential congestion.

2.5.4 Reinforcement Learning

Reinforcement learning (RL) is a type of machine learning where an agent learns to make decisions by taking actions in an environment to achieve maximum cumulative reward [30]. The agent learns from trial and error, receiving rewards or penalties for actions. Over time, it learns the optimal policy, i.e., the best action to take in each state to maximize its total reward over time [30].

RL is particularly useful in scenarios where there's a need to balance immediate rewards with long-term outcomes and where the optimal decision-making policy is not known upfront and must be learned from interaction with the environment.

2.5.5 Q-Learning

Q-Learning is a specific model-free reinforcement learning algorithm that does not require a model of its environment. Q-Learning aims to find an optimal policy by learning a Q-function, which gives a certain action's expected utility (value) in a given state [30].

The Q-function is typically represented as a table, with states as rows, possible actions as columns, and the expected utility of each action in each state as table entries. Q-Learning involves updating these entries over time based on the rewards received from the environment and the agent's exploration of different state-action pairs.

Q-learning is particularly effective in problems where there is a need to balance exploration (trying out new actions to see their effect) and exploitation (choosing the actions known to yield the best results) and where the environment dynamics may be complex and non-deterministic.

2.6 Literature Review

Vehicular Ad Hoc Networks (VANETs) are essential for supporting various applications, including safety-critical ones such as forward collision warnings and traffic signal violation warnings, as well as comfort-oriented applications like weather information systems and restaurant recommendations [31]. Reliable delivery of alert messages and periodic broadcasts of Basic Safety Messages (BSMs) containing crucial information such as vehicle position, speed, and heading is crucial for these safety applications. However, due to the limited channel capacity and high transmission power needed for maintaining awareness, ensuring reliable delivery of BSMs can be challenging in VANETs. This section reviews relevant studies and research contributions on congestion control algorithms for vehicular networks.

One classical approach is the linear message rate integration control (LIMERIC) algorithm [32], which aims to distribute the available channel bandwidth fairly among vehicles by dynamically adjusting the BSM transmission rate in each iteration. LIMERIC adapts the transmission rate using linear feedback to manage congestion. Another approach is the error model-based adaptive rate control (EMBARC) algorithm [33], which improves LIMERIC by preemptively scheduling messages based on a vehicle's movement, thereby reducing instances of large tracking errors. The BSM rate control over IEEE802.11p vehicular networks (BRAEVE) algorithm [34] estimates the number of vehicles to achieve smoother convergence and lower packet error ratio, inter-packet delay (IPD), and tracking error compared to other algorithms.

Another proposed scheme [35] analyzes congestion detection schemes and utilizes a priority model to adjust the transmission rate of beacon messages, employing the search algorithm to control network congestion.

Transmission power, another adjustable parameter for congestion control, is crucial in managing the number of vehicles a BSM can reach. The distributed fair transmit power adjustment for VANETs (D-FPAV) algorithm [36] sets the node transmission power based on the prediction of application-layer traffic and the observed number of vehicles in the surrounding area. Adaptive beacon transmission power (ABTP) [37] adjusts transmission power based on vehicle position prediction error, increasing it for vehicles with large errors and reducing it for vehicles with small errors.

Data rate control is another aspect of congestion control, with studies like the binary rate algorithm [38] dynamically adjusting the data rate based on channel load. The authors in [39] propose an algorithm that directly estimates the appropriate data rate based on the current BSM reception rate (CBR), leading to faster convergence and improved packet delivery ratio.

Another study [40] introduces combined power and message-rate adaptation (CPMRA) and combined power and data-rate adaptation (CPDRA) mechanisms, adapting transmission power and data rate based on the current congestion situation. Channel-aware congestion control (CACC) [41] considers received signal strength (RSS) to diagnose packet loss and determine channel conditions, adjusting data rate and transmission power accordingly.

Liu, Amour, and Jaekel (2023) [9] introduce a congestion control mechanism for V2V communication in VANETs that leverages reinforcement learning. Their approach enables vehicles to learn optimal congestion control policies through interactions with the environment and rewards. By adapting the transmission rate, the proposed approach aims to mitigate congestion and enhance the performance of V2V communication. The authors provide a detailed methodology and evaluate the effectiveness of their approach through simulations. The key contribution of this work lies in the application of reinforcement learning for congestion control in V2V communication. By utilizing reinforcement learning, the proposed mechanism allows vehicles to autonomously learn optimal control policies based on rewards obtained from the environment. The algorithm optimizes congestion control decisions through the iterative learning process to improve communication performance in VANETs.

Aznar-Poveda et al. (2021) [16] propose a novel approach called MDPRP, which leverages Q-learning to control the beaconing rate and transmission power in VANETs jointly. The goal is to achieve a trade-off between communication reliability and resource utilization by dynamically adapting these two parameters. The authors present a detailed methodology and conduct experiments to evaluate the effectiveness of their approach. This work's key contribution lies in applying Q-learning to the joint control of beaconing rate and transmission power. The MDPRP algorithm enables vehicles to autonomously learn optimal policies based on local observations and rewards, facilitating an adaptive control mechanism. By dynamically adjusting the beaconing rate and transmission power, the algorithm aims to improve communication reliability and resource efficiency in VANETs.Compared to related studies, Aznar-Poveda et al.'s approach stands out due to its focus on the joint control of beaconing rate and transmission power using Q-learning. While previous research has explored individual control mechanisms or relied on fixed settings for these parameters, this work presents a holistic approach considering their interaction. This allows for adaptive decision-making that optimizes the trade-off between communication reliability and resource utilization.

Liu et al. (2021) in "Balancing Awareness and Congestion in Vehicular Networks Using Variable Transmission Power" [42] propose a novel approach that dynamically adjusts the transmission power of vehicles to strike a balance between maintaining awareness of surrounding vehicles and mitigating network congestion. This paper addresses balancing awareness and congestion in vehicular networks by utilizing variable transmission power. By adapting the transmission power based on the density of nearby vehicles and the current network congestion level, their algorithm aims to optimize communication reliability and network efficiency. The authors conducted experiments and simulations to evaluate the effectiveness of their approach under various scenarios. The key contribution of this work lies in the variable transmission power scheme, which allows vehicles to adjust their transmission power levels dynamically. By increasing the transmission power in sparser areas with low congestion, vehicles can establish and maintain communication links with nearby vehicles, thus enhancing situational awareness. Conversely, vehicles can lower their transmission power in denser areas or when congestion levels rise to mitigate interference and reduce congestion, thereby improving network efficiency.

The paper titled "Research on adaptive beacon message transmission power in VANETs" [37] addresses the challenge of adapting beacon message transmission power in VANETs to optimize communication performance. Wang et al. (2020) propose an approach that dynamically adjusts the transmission power of beacon messages in VANETs based on the network conditions and communication requirements. Their algorithm balances the trade-off between communication range and energy consumption by adaptively controlling the transmission power. The authors conducted experiments and simulations to evaluate the effectiveness of their approach in various

VANET scenarios. The key contribution of this work lies in the adaptive adjustment of beacon message transmission power. The algorithm optimizes the communication range while minimizing energy consumption by dynamically modifying the transmission power based on vehicle density, traffic conditions, and communication requirements.

Facchina and Jaekel (2020) [43] propose a congestion control scheme that leverages vehicle speed as a key metric to regulate traffic flow and mitigate congestion in vehicular networks. It is a distributed approach, where each vehicle autonomously adjusts its speed based on local congestion observations. This work's key contribution lies in utilizing vehicle speed as a control parameter for congestion control. By adjusting vehicle speed in response to observed congestion levels, the scheme aims to maintain optimal traffic flow and reduce the risk of congestion-related problems, such as packet loss and increased communication delays.

Patil et al. (2019) [44] propose a decentralized congestion control mechanism that dynamically adjusts the transmit data rate of vehicles in VANETs to alleviate congestion. Their algorithm aims to regulate the amount of data transmitted by each vehicle based on the observed network conditions and congestion levels. The key contribution of this work lies in the transmit data rate control mechanism for decentralized congestion control. The algorithm allows each vehicle to autonomously adapt its data transmission behavior by adjusting the data transmission rate based on local congestion observations. This decentralized approach reduces the reliance on centralized control and communication, thereby enhancing the scalability and robustness of the system.

Mohammed et al. (2022) in [45] proposes a decentralized congestion control mechanism for Vehicular Ad Hoc Networks (VANETs). The authors introduce a novel congestion control mechanism that dynamically adjusts the message transmission rate based on the exponential function. The goal is to manage the communication load and mitigate congestion within VANETs effectively. They utilize the exponential function to adaptively adjust the rate at which messages are transmitted in response to the changing network conditions and congestion levels. The proposed approach aims to optimize network performance and ensure efficient information dissemination in VANETs by dynamically modifying the message transmission rate. This approach offers a unique perspective on managing congestion within VANETs.

Math et al. (2015) [46] present a congestion control mechanism for V2V communication that focuses on adjusting the data rate to manage congestion levels in traffic safety scenarios. The authors recognize the criticality of maintaining reliable communication for traffic safety applications and propose a methodology that dynamically adapts the data rate based on observed congestion conditions. The paper analyzes the proposed approach and presents simulation results to demonstrate its effectiveness. The key contribution of this work lies in the data rate-based congestion control mechanism tailored specifically for traffic safety applications in V2V communication. By adjusting the data rate based on observed congestion levels, the proposed approach aims to mitigate congestion and maintain reliable communication links, ensuring a timely and accurate exchange of safety-related information among vehicles.

Subramaniam et al. (2022) [47] introduce the use of traffic density as a crucial parameter in their congestion control method. By leveraging real-time traffic density information, vehicles can dynamically adjust their transmission behavior to adapt to the current traffic conditions and optimize network performance. This approach offers a new perspective on congestion control in VANETs, focusing on the traffic density aspect as a key determinant of congestion. The researchers introduce a congestion-aware message (CAM) for beacon signals in the vehicle environment that utilizes vehicle IDs. The CAM model incorporates the unique automobile IDs into the back-off procedure, weighting the randomized back-off numbers chosen by each vehicle. This results in the generation of car ID-based randomized back-off codes that reduce the risk of collisions caused by identical back-off numbers. In this paper, BSMs as the standard safety messages are not considered.

While the mentioned approaches address various aspects of congestion control in VANETs, each has strengths and limitations. Further research is needed to evaluate their performance under different traffic scenarios and consider their applicability in real-world deployments.

Table 2.1: Comparison Table of Literature Review.

No.	Research Paper	Algorithm	Parameter Used	Approach
1	Liu et al. [9]	QBACC	Beacon Rate	Learning-based
2	Aznar-Poveda et al. [16]	MDPRP	Hybrid	Learning-based
3	Liu et al. [42]	BACVT	Transmission Power	Rule-based
4	Wang et al. [37]	ABTP	Transmission Power	Learning-based
5	Facchina et al. [43]	DACC	Transmission Power	Rule-based
6	Patil et al. [44]	CPMRA	Hybrid	Rule-based
7	Math et al. [46]	DR-DCC	Date Rate	Rule-based
8	Willis et al. [48]	DCC	Transmission Power	Learning-based
9	Bansal et al. [32]	LIMERIC	Beacon Rate	Rule-based
10	Bansal et al. [33]	EMBARC	Beacon Rate	Rule-based
11	Torrent-Moreno et al. [36]	D-FPAV	Transmission Power	Rule-based
12	Cho et al. [41]	CACC	Hybrid	Rule-based
13	**Proposed Method**	QBTPCC	Transmission Power	Learning-based

2.7 Discussion of Current Research Problems

Through the literature reviews, several research problems and solutions have been identified regarding congestion control in VANETs. These findings shed light on the field's current state and highlight areas for further investigation.

The rapid growth in the number of vehicles equipped with communication devices has led to increased congestion in the communication channels. This results in decreased performance and reliability of the network and is a major research problem in VANETs. Current solutions propose adaptive transmission power and rate control, dynamic spectrum access, and priority-based message dissemination. However, these solutions often struggle to balance the need for high vehicle awareness and congestion control.

One significant challenge is finding the right balance between maintaining awareness through reliable message delivery and managing congestion effectively. Researchers have explored various approaches as transmission power and rate directly impact awareness and congestion. For instance, an RL-based Q-learning approach balances transmission rate and power trade-offs. This approach utilizes a Markov decision process (MDP) and Q-learning techniques to optimize the transmission parameters. Other studies, such as CPMRA and CPDRA, dynamically adapt transmission power and data rate to balance awareness and congestion based on the current congestion situation.

Another research problem is determining the optimal selection of transmission parameters, such as transmission rate, power, and data rate. Different algorithms and approaches have been proposed to address this problem. For instance, the EMBARC algorithm improves the linear message rate integration control (LIMERIC) algorithm by preemptively scheduling messages based on vehicle movement. The BRAEVE algorithm estimates the number of vehicles to achieve smoother convergence and lower packet error ratio, inter-packet delay (IPD), and tracking error. These approaches demonstrate efforts to optimize parameter selection for congestion control in VANETs.

Reinforcement learning (RL) has emerged as a promising technique for congestion control in VANETs. RL techniques specifically applied the on-policy rule with function approximation to solve the message rate control problem. Their approach formulates the problem as a Markov decision process (MDP) and enables vehicles to learn optimal control policies. Similarly, an RL-based method was proposed that directly estimates the appropriate data rate based on the current BSM reception rate (CBR). These RL-based solutions highlight the potential of RL in addressing congestion control challenges in VANETs.

Several studies emphasize the need for adaptive and dynamic approaches that can effectively respond to changing network conditions and traffic scenarios. For example, D-FPAV sets the node transmission power based on the prediction of application-layer traffic and observed vehicle density. ABTP adjusts transmission power based on vehicle position prediction error. These adaptive approaches demonstrate the importance of dynamically adjusting transmission parameters to manage real-time congestion.

Despite the advancements in congestion control for VANETs, several research challenges remain. These challenges include the scalability of the proposed solutions, accurate prediction of congestion levels, handling diverse traffic scenarios, and considering the impact of network topology. Furthermore, evaluating the proposed approaches through comprehensive simulations and real-world experiments is crucial for assessing their performance and practicality.

The proposed solution leverages Machine Learning (specifically, Q-learning) to dynamically adjust the transmission power of vehicles based on real-time network conditions. By doing so, it aims to effectively manage channel load, maintain high vehicle awareness, and improve overall network performance.

CHAPTER 3

Proposed Approach

3.1 Introduction

As Vehicular Ad Hoc Networks (VANETs) continue to evolve, enabling real-time communication between vehicles and infrastructure, the efficient and reliable exchange of information has become paramount. One of the most pressing challenges that impede the realization of the full potential of VANETs is channel congestion. The sporadic nature of vehicular traffic and the demanding need for high-frequency communication compound this problem, resulting in packet loss, increased latency, and, ultimately, a compromise in vehicle safety and efficiency [49].

The conventional mechanisms for congestion control in VANETs often fail to strike a delicate balance between high vehicle awareness and channel congestion levels. Traditional approaches, borrowed from wired and wireless networks, are ill-suited for the dynamic, high-mobility environment characteristic of VANETs. There is a crucial need for an adaptive congestion control solution explicitly tailored to the unique attributes of VANETs.

Motivated by this pressing need, this thesis presents a novel transmission power-based congestion control algorithm utilizing Q-learning, a form of reinforcement learning, to address the challenges above in VANETs. The primary objective of this algorithm is to dynamically adjust the transmission power of the vehicle's onboard units based on the network conditions, effectively balancing the requirements for high vehicle awareness with the need to maintain stable channel congestion levels.

Using Q-learning empowers the algorithm to learn from the environment and make informed decisions on transmission power adjustments based on historical data and real-time network conditions. This adaptive approach not only enhances network performance but also ensures the reliability and timeliness of communication, which are crucial for the safety and efficiency of vehicular networks.

In the following sections, this chapter will delve into the specifics of the proposed approach, articulating the algorithmic design and detailing how Q-learning is employed to manage transmission power for congestion control in VANETs dynamically.

3.2 High-level Outline of Proposed Approach

3.2.1 Algorithm Design Framework

Regarding safety considerations, it is imperative to recognize that vehicles closer to the Ego Vehicle (EV) significantly impact its safety. The rationale is that the time taken to approach a nearby vehicle is considerably shorter compared to vehicles farther away, allowing for a narrower margin for corrective actions. Consequently, ensuring that safety messages are promptly relayed to nearby vehicles is paramount. The actual outreach of Basic Safety Messages (BSMs) from a vehicle, denoted as the transmission range (Tx range), is contingent upon the transmission power (TxPower) used. When a vehicle falls within the transmission range of the EV, it is expected to receive the BSM effectively, assuming the absence of channel congestion or other interference. As the transmission power is augmented, there is a corresponding increase in the transmission range. Other factors, such as obstacles or the number of vehicles in the vicinity, may also influence the range.

Table 3.1 illustrates the variation in the number of BSMs received between two vehicles with respect to the distance separating them.

In deriving the data in Table 3.1, we simulated a scenario involving two static vehicles placed at varying distances, engaging in periodic BSM exchanges. A rudi-

Table 3.1: Maximum Transmission Range.

Distance (m)	Sent BSMs	10 Hz 2 mW		10 Hz 10 mW		10 Hz 20 mW	
		Received BSMs	Lost BSMs	Received BSMs	Lost BSMs	Received BSMs	Lost BSMs
100	90	90	0	90	0	90	0
200	90	0	90	90	0	90	0
300	90	0	90	90	0	90	0
400	90	0	90	70	20	90	0
500	90	0	90	0	90	89	1
600	90	0	90	0	90	13	77
700	90	0	90	0	90	0	90
800	90	0	90	0	90	0	90
900	90	0	90	0	90	0	90
1000	90	0	90	0	90	0	90

mentary free-space path loss model, devoid of obstacles consideration, was employed as an approximation of the transmission range. With a TxPower of 20 mW, it was observed that most BSMs could be conveyed effectively up to 500 m, and nearly all BSMs fail to reach when the distance between the vehicles exceeds 600 m. A reduced TxPower of 2 mW significantly shortens the Tx range. In such a scenario, most BSMs are successfully received when the vehicles are not more than 100 m apart, and as the distance increases, the loss of BSMs escalates.

It is important to note that the figures in Table 3.1 depict an idealized scenario with no other vehicles on the road segment. In real-world scenarios with higher vehicle density, the effective transmission range will likely be diminished further due to inter-vehicle interference.

In essence, utilizing a high transmission power (TxPower) can convey Basic Safety Messages (BSMs) to vehicles that are far away, but this also has the potential to cause increased interference and channel congestion. The core approach of this work revolves around giving precedence to the delivery of BSMs to vehicles in close proximity, which can be achieved by employing a lower TxPower rather than consistently using the maximum 20 mW TxPower. While retaining some degree of awareness for distant vehicles is favorable, the update frequency can be less frequent than that for vehicles in closer proximity. Modulating the TxPower makes sustaining elevated awareness levels for nearby vehicles feasible. It concurrently enables occasional BSMs to reach distant vehicles by selectively employing a higher TxPower for certain BSMs. This approach ensures that the awareness level concerning emergency vehicles remains optimal for nearby vehicles while simultaneously alleviating channel congestion since fewer BSMs are transmitted at a high TxPower.

3.2.2 Design of Elements

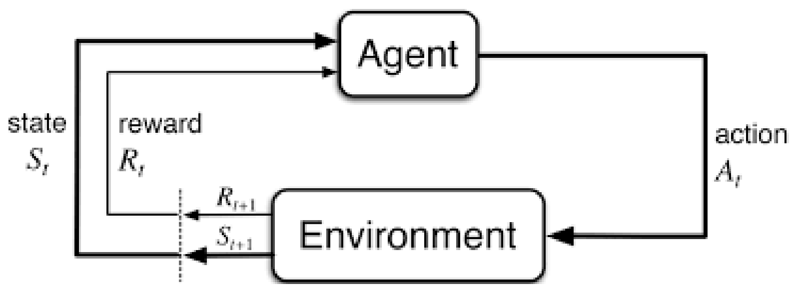

Figure 3.1: Reinforcement Learning (RL) Components. [50]

The problem of making decisions is structured using an MDP framework. The elements listed below form the Reinforcement Learning (RL) framework, which is utilized to solve the MDP in the context of V2V congestion control [9]:

- **The Agent:** The learning agent must have the ability to recognize its environment's state and execute actions that have the potential to modify it. In this context, the vehicle assumes the agent's role and decides the actions to be executed [15].

- **The Environment:** This refers to the volatile world in which the agent performs and communicates. The agent can communicate with the environment and alter it through its actions. However, it cannot control the environment's fundamental rules or behavior [9]. In the Vehicular Ad Hoc Networks (VANETs) context, the environment includes the wireless channels and the surrounding vehicles. Uncertainties arise from various factors, like the dynamic nature of traffic, including fluctuations in vehicle speeds and density.

- **The State (or Observation):** The state in the V2V communication problem aggregates data that reflects the present conditions. This data encompasses aspects like the status of the wireless channel (e.g., CBR, BER, IPD) and vehicle density, defined by the number of neighboring vehicles.

  In our case, we define the state space as a 2-tuple, including the CBR and vehicle density, denoted as

$$
\begin{aligned}
S &= (CBR, VD), \\
&\quad CBR \in \mathbb{R}^{+}, \quad 0 \leq CBR \leq 1, \\
&\quad \text{and} \quad VD \in \mathbb{N}, \quad 1 \leq VD \leq \text{maxVD}.
\end{aligned}
$$

  CBR is a number between 0 and 1, representing the channel busy ratio. Vehicle density means the number of vehicles within a 100m radius, with a maximum value defined as maxVD (set to 50 in this study). Each vehicle's density has ten transmission power levels, so the total state space comprises 500 unique states. The vehicle chooses a new transmission power from the ten available options in each state. It updates its state based on the data received from nearby vehicles via Basic Safety Messages (BSMs).

- **The Action:** Actions are the tools through which the agent communicates with and exerts influence over its environment. The most common actions in the VANET application layer include setting the transmission power, message transmission rate, or data rate of the messages to be transmitted. In this thesis, only transmission power is considered to maintain simplicity, and we use ten discrete powers ranging from 2 to 20 mW.

- **The Goal:** The primary aim is to make optimal choices for actions corresponding to each state to maximize the total rewards. It is essential for the agent (in this case, the vehicle) to have a clearly defined objective, such as minimizing congestion or improving awareness. This study aims to maximize the rewards gained from actions that ensure the Channel Busy Ratio (CBR) stays below 0.6.

- **The Reward:** Rewards are scalar quantities that gauge the efficacy of an action executed by an agent. The agent utilizes the rewards furnished by the environment in response to each action to learn and progressively refine its behavior [15]. Within the scope of V2V communication, the reward is computed based on the environmental observations and the objectives of the vehicle. A reward function, tailored to align with the intended learning goals, calculates rewards. In our proposed methodology, we aim to keep the Channel Busy Ratio (CBR) below a predetermined threshold, denoted as η, while concurrently optimizing the number of Basic Safety Messages (BSMs) transmitted. To this end, we have articulated the reward function as follows:

$$R(CBR, TxPower) = \begin{cases} TxPower \cdot CBR \cdot -1 & \text{if } \eta < CBR \\ TxPower \cdot CBR \cdot 1 & \text{if } \eta \geq CBR \end{cases} \tag{1}$$

  An action that leads to the CBR surpassing η is assigned a negative reward, which can catalyze the learning trajectory [15]. A reward is deliberately diminished for exceedingly low, unfavorable transmission power. This study adopted $\eta = 0.6$ as the desired channel occupancy. Depending on variant learning goals, this value can be adjusted, or an alternate reward function can be deployed.

3.2.2.1 The Reward Function

In V2V communication, the reward function is a crucial component that quantifies the efficacy of an agent's (vehicle) action. In our approach, the reward function takes into account the Channel Busy Ratio (CBR), the Transmission Power (TxPower), and a predefined threshold for CBR, denoted as η. Below, we elaborate on these elements and explain how the reward function is structured:

- **Channel Busy Ratio (CBR):** CBR is a metric that reflects the occupancy of the communication channel. It ranges from 0 to 1, where a value close to 1 denotes that the track is nearly saturated, and a value close to 0 suggests that the channel is primarily unoccupied.

- **Transmission Power (TxPower):** This refers to the power level at which the vehicle transmits messages. A higher transmission power typically correlates with an extended communication range.

- **Threshold η:** η serves as the acceptable limit for CBR. Maintaining CBR below this threshold is essential to prevent the communication channel from becoming overly congested.

  The reward function can be described as follows:

  - **Positive Reward:** When CBR is lesser than or equal to the threshold η, it is positive. In this scenario, the reward is directly proportional to TxPower and CBR. This implies that higher transmission power and channel utilization (without overloading) are favorable.

  - **Negative Reward (Penalty):** Conversely, when CBR surpasses the threshold η, it is negative. This gives rise to a penalty, as the channel is overloaded. The penalty motivates the learning agent to select actions that alleviate channel congestion.

  - **TxPower $\cdot$ CBR:** This term signifies that the reward is directly proportional to the CBR and TxPower. The higher these values, the higher the reward.

This reward function motivates the agent to keep the Channel Busy Ratio under the designated threshold while efficiently utilizing the channel by preferring higher transmission powers when the channel is not excessively loaded. The reward function aims to optimize the transmission power based on current network conditions. If a vehicle's CBR falls below η, it might decide to increase its

TxPower (or make other adjustments) to try and improve CBR. On the other hand, if a vehicle's CBR is consistently above η, it will reduce its TxPower to reduce congestion. The function dynamically adjusts transmission behaviors in a VANET setting, rewarding vehicles or nodes for maintaining their CBR around a desired threshold η while considering the power used for transmission.

3.2.3 Adaptive Congestion Control Using RL Techniques

In the Vehicular Ad Hoc Networks (VANETs) domain, congestion control is imperative for ensuring secure and efficient communication over the constrained bandwidth of wireless channels. The primary goal of congestion control is to mitigate the effects of channel congestion. A delicate equilibrium must be maintained between focusing on congestion control and awareness control, which becomes exceedingly complex in environments with dynamic mobility patterns. The ability to make informed decisions in varied scenarios is paramount to the efficacy of congestion control. For instance, amplifying the transmission power might be advantageous to achieve an extended transmission range in scenarios where vehicle density is relatively low.

Conversely, it becomes necessary to calibrate the transmission power per the prevailing conditions in high-density scenarios. The decision-making process is compounded by numerous variables, including, but not limited to, vehicle density, channel congestion, and packet delay. These factors render identifying an optimal transmission power through conventional methodologies a daunting challenge, especially when certain constraints may be at odds.

RL is a machine learning paradigm where agents learn how to behave in an environment by performing certain actions and receiving rewards or penalties in return [51]. The agent's goal is to learn the optimal policy that will result in the maximum cumulative reward over time.

It is essential to recognize that in VANETs, the decision-making process should be predicated exclusively on the immediate state of affairs. This is due to the Markov property. A Markov Decision Process (MDP) is a mathematical model used in decision theory and reinforcement learning to describe environments in which an agent can interact. States, actions, transition probabilities between states, and rewards characterize an MDP. The key property of an MDP is the Markov property, which states that the future state depends only on the current state and action, not on the sequence of states and actions that preceded it [52].

$$P[St+1|St] = P[St+1|S_1, \ldots, St] \tag{2}$$

Equation (2) elucidates that the state at the subsequent time step, t + 1, exclusively relies on the current state at time step t. This characteristic enables us to model the problem as a Markov Decision Process (MDP). Reinforcement Learning (RL) is a productive framework for deriving solutions [15]. The fundamental learning mechanism of an RL cycle, particularly in the context of Vehicle-to-Vehicle (V2V) communication, is depicted in Figure 3.2.

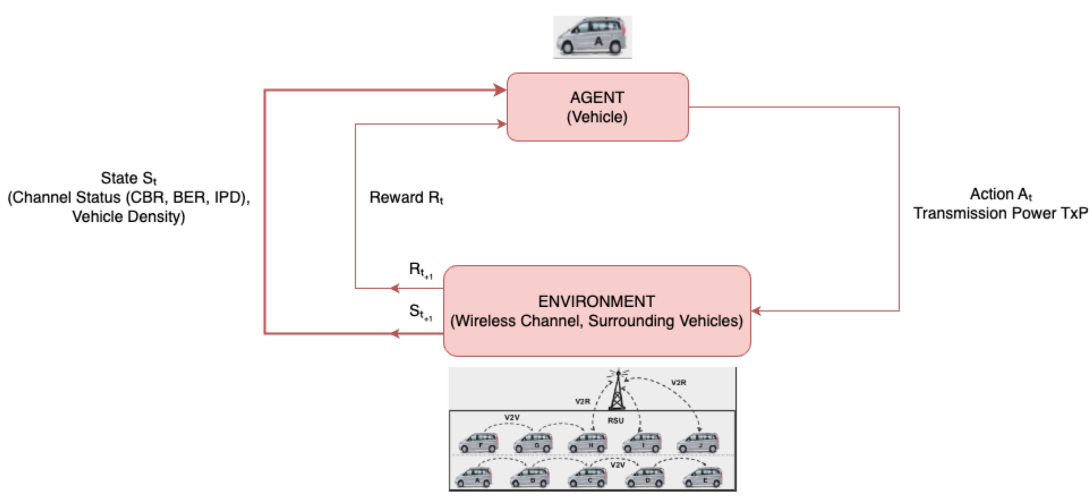

Figure 3.2: Typical RL Cycle in V2V communication. [9]

1. **Initialization:**
   - Define the state space encompassing channel status and vehicle density.
   - Define the action space, transmission power, TxP.
   - Initialize a Q-table with arbitrary values or zeros.
2. **Exploration:**
   - Based on the current state S and some exploration strategy (like ε-greedy), the vehicle selects an action A. Exploration can be diminished over time to prioritize exploiting the learned values.
3. **Interaction with the Environment:**
   - The vehicle takes action A in state S, affecting the V2V communication environment.
4. **Observation:**
   - After taking the action, the vehicle transitions to a new state S' and receives a reward R.
5. **Update Q-values:**
   - The Q-value for the taken action in the original state is updated using the Q-learning update rule:
6. **Loop:**
   - Return to the exploration step with updated Q-values.
7. **Policy Extraction:**
   - A deterministic policy can be derived from the Q-values. For any state S, the best action A^* is the one with the highest Q-value.

The overarching objective for the vehicle is to maximize the cumulative reward garnered over time. This entails maximizing the immediate and aggregate rewards accumulated over the long term.

If we designate the sequence of rewards received post time step t as Rt+1, Rt+2, Rt+3, ..., the maximized reward, referred to as the return Gt, can be computed as the sum of the rewards received at each time step up to the terminal state.

However, it is important to factor in the time value of rewards in this computation. This is achieved by introducing a discount factor, γ, which gauges the present value of future rewards. Specifically, a reward procured k time steps into the future has ascribed a value of $\gamma^{(k-1)}$ times the original reward. This discounting ensures that immediate rewards are accorded a higher weight relative to distant rewards and also circumvents the possibility of the return being infinite in non-terminating environments.

The return, G_t, can thus be mathematically represented as [53]:

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \ldots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \tag{3}$$

The Equation (3) represents the cumulative discounted reward, often referred to as the "return," at time step t. Let's break down the components of this Equation.

- G_t: This denotes the return or the total accumulated reward starting from time step t. It measures how good it is for an agent to be in a particular state, considering the future rewards. In Reinforcement Learning, the goal is usually to maximize the expected return.
- $R_{t+1}, R_{t+2}, R_{t+3}, \ldots$: These represent the rewards that the agent receives at each successive time step. R_{t+1} is the reward received one step into the future, R_{t+2} two steps, and so on.
- γ: The discount factor is a number between 0 and 1. The discount factor determines the present value of future rewards - a reward received k time steps in the future is worth γ^{k-1} times what it would be worth if it were received immediately. If γ is close to 0, the agent will only care about

rewards in the near future, whereas if γ is close to 1, the agent will consider rewards far into the future.

- $\sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$: This is the summation notation which sums up all the discounted rewards from time $t+1$ to infinity. In practice, we often work in environments with finite time steps, so the sum doesn't go to infinity until the end of the episode.

The entire Equation says that the return G_t at time step t is the sum of all the agent's future rewards. Still, each future reward is discounted by a factor of γ raised to the power corresponding to how far the reward is received into the future. This mathematical expression is fundamental in Reinforcement Learning and is known as the discounted return.

3.2.4 The RL Structure for Managing Congestion and Awareness

This thesis emphasizes the role of Reinforcement Learning (RL) in determining the best transmission power to manage traffic congestion in Vehicle-to-Vehicle (V2V) interactions. This decision-making mechanism is framed within the Markov Decision Process (MDP). Here, we explore the intricacies of crafting the MDP, bearing in mind these factors:

- **Defined state and action realms:** At the application level in VANETs, it's inferred that every vehicle has a restricted set of potential moves or actions available in any given situation. Given that the state realm is also restricted, Q-learning emerges as a suitable approach for this challenge.
- **Spotting vehicles in proximity:** A vehicle discerns the number of nearby vehicles by interacting with its surroundings, flagged by the Basic Safety Messages (BSMs) it receives from these vehicles.
- **Inferences drawn from trials:** The conclusions are deduced from test actions executed by the vehicle.

– **Independent decision-making:** Every vehicle makes its own choices based on observations. The sole data exchange between vehicles is restricted to BSMs.

3.3 Proposed Approach

3.3.1 Q-Learning Based Transmission Power Congestion Control (QBTPCC)

Q-learning is a model-free reinforcement learning (RL) algorithm used to find the optimal action-selection policy for a given finite Markov decision process. It helps agents learn how to choose optimal actions that yield the most reward over time, even when they don't know anything about their environment beforehand. The "Q" in Q-learning stands for "quality." Q-values represent the expected future reward of an action taken in a given state. These values are stored in a table called the Q-table.

This thesis uses a Q-learning algorithm with the data for training directly obtained from a simulated dynamic traffic environment. Observing the CBR values with different transmission power and vehicle densities makes the vehicle's state transitions more realistic. We define a reward function combining CBR and transmission power to maintain the channel load under a target threshold with the maximum transmission power possible for congestion control. The proposed congestion control algorithm, leveraging Q-learning, unfolds in a two-tiered approach as described below:

3.3.1.1 Initial Phase

The Q-learning algorithm is executed through simulation-derived observation data, resulting in a Q-table. This table showcases the best policies for each state, as presented in Algorithm 1.

Algorithm 1 summarizes the proposed Q-Learning Based Transmission Power Congestion Control (QBTPCC) methodology. The algorithm commences by setting all the entries in the Q-table to 0; this table encompasses all possible action-state combinations. Subsequently, at every time instance t, the vehicle picks an action a_t, observes the environment, receives a reward r_t, transitions to a new state s_{t+1}, and utilizes Equation (4) to update the Q-value, $Q(s_t, a_t)$ [9]:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left(R(s_t, a_t) + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right) \tag{4}$$

In Equation (4):

- α denotes the learning rate, which falls in the range $0 < \alpha \leq 1$.
- γ symbolizes the discount factor, also lying in the range $0 < \gamma \leq 1$.
- $Q(s_t, a_t)$ is the present estimate of $Q(s, a)$.
- $\max_a Q(s_{t+1}, a)$ approximates the optimal future Q-value.
- $R(s_t, a_t)$ signifies the reward garnered by the agent when action a is performed in state s at instance t.

The components and functionality of the Q-Learning Based Transmission Power Congestion Control (QBTPCC) Algorithm:

1. **Q-table Initialization:** The algorithm initializes a Q-table with all entries set to 0. The Q-table is essentially a matrix where each row represents a state, and each column represents an action. The entries of the Q-table, denoted $Q(s, a)$, store the expected future rewards for taking action a in the state s.
2. **Action Selection (a_t):** At each time step t, the vehicle (agent) selects an action, which could be based on a policy such as ϵ-greedy (choosing the best action with probability $1 - \epsilon$ or a random action with probability ϵ).

3. **Observe Environment and Receive reward (R_t):** After executing the action, the agent observes the new state of the environment and receives a reward R_t. The reward is typically a scalar feedback signal that tells the agent how well it is doing concerning the goal it wants to achieve.

4. **Transition to a New State (s_{t+1}):** Based on the action taken, the agent transitions from its current state s_t to a new state s_{t+1}.

5. **Q-value Update:** The Q-value $Q(s_t, a_t)$ is updated based on the difference between the estimated future rewards and the current Q-value. The learning rate α determines how much the Q-value should be updated, whereas the discount factor γ defines the importance of future rewards.

6. α (**Learning Rate**): The learning rate determines to what extent newly acquired information overrides old information. A value of 0 makes the agent not learn anything, while a value of 1 makes the agent consider only the most recent data. Typically, it's set between these extremes.

7. γ (**Discount Factor**): The discount factor determines how much the agent cares about rewards in the distant future relative to those in the immediate future. If γ is close to 0, the agent will be short-sighted and only consider current rewards. If γ is close to 1, the agent will focus more on the long-term rewards.

The Q-learning update rule is the Equation (4). It essentially states that the new Q-value blends the old and new information learned by taking action.

The algorithm continues to learn as it interacts with the environment. Over time, it converges to an optimal policy where the Q-values represent the expected future rewards for each action in each state, allowing the agent to make informed decisions that maximize its rewards.

Algorithm 1 Q-Learning Based Transmission Power Congestion Control (QBTPCC)

Input: Step size $\alpha \in (0, 1]$, small $\varepsilon > 0$, total episodes = 80,000
Output: Updated Q-table representing the value of each state-action combination

Setup state space $\mathcal{S}$ with each transmission power level as a distinct state
Define actions $\mathcal{A}(s)$ for each state s as the ten available transmission powers, TxP, regardless of s's value
Create Q-table, Q, and set $Q(s, a) = 0$ for all pairs (s, a) where $s \in \mathcal{S}$ and $a \in \mathcal{A}(s)$
for each episode **do**
  Randomly select an initial state s from $\mathcal{S}$
  for each step in the episode **do**
    Determine action a for the current state s using the ε-greedy strategy
    Calculate the reward via Equation (1). Use Equation (6) for CBR computation, incorporating vehicle density
    Refine the Q-value $Q(s, a)$ with Equation (4) using the computed reward
    Execute action a transitioning to the next state
    Update current state s to the newly arrived state
  end for
end for

In Algorithm 1, the Q-table is generated systematically, incorporating simulation data based on different traffic models. The traffic models represent various scenarios by altering vehicle densities in a 100 m radius, with densities ranging from 0 to a maximum value of maxVD, which is set at 50 vehicles.

For each traffic model, simulations were conducted utilizing a predefined action space pertaining to varying transmission power levels, spanning from 2 mW up to 20 mW. This simulation exercise aids in obtaining observation data, specifically the Channel Busy Ratio (CBR), for every state-action pair.

This observation data was then utilized to generate functions for curves that best fit the data. Essentially, these curves indicate the relationship between the average transmission power employed by vehicles and the average CBR experienced within the network. This relation is pivotal for understanding how the transmission power affects the occupancy of the communication channel under varying traffic conditions.

Having established the curves, we created Equation (5), which estimates the CBR for each transmission power and vehicle density combination. This estimation is based on the trends observed in the data.

$$estCBR = f(VD, TP) \tag{5}$$

estCBR is the estimated Channel Busy Ratio, an estimate of the channel occupancy. VD is the current vehicle density within a 100m radius. TP is the estimated average transmission power utilized by the surrounding vehicles. f(VD, TP) is a function that estimates CBR based on the best-fit curves generated earlier.

Notably, a cap of 0.92 was implemented as the maximum return value for the estimated CBR [9]. This cap ensures that the estimated CBR does not exceed 1 for densities exceeding the tested range. This is critical because the empirical data indicated a negligible change in CBR at high densities, and exceeding one would not be meaningful as CBR is generally represented as a fraction between 0 and 1.

Subsequently, the Q-learning algorithm is deployed with the observation data and Equation (5) to generate the Q-table systematically. This table is formulated for each vehicle density and transmission power combination, accounting for the entire state space.

The generated Q-table is central to the Q-learning algorithm. It stores the expected future rewards of taking certain actions (transmission powers) in specific states (vehicle densities). Through repeated interactions with the environment and constant updates to the Q-table, the learning algorithm endeavors to find an optimal policy that minimizes channel congestion while considering the varying conditions of the vehicular network.

By utilizing these simulations and observations to create the Q-table, Algorithm 1 is better equipped to make informed decisions on transmission power levels, ultimately aiming for efficient communication in vehicular networks, even in environments with varying vehicle densities.

$$\text{estCBR}(VD, TP) = \begin{cases} 0.0137(VD) + 0.0911 & \text{if TP} = 2 \\ 0.0156(VD) + 0.129 & \text{if TP} = 4 \\ 0.0163(VD) + 0.1502 & \text{if TP} = 6 \\ 0.0165(VD) + 0.1553 & \text{if TP} = 8 \\ 0.0165(VD) + 0.1553 & \text{if TP} = 10 \\ 0.0165(VD) + 0.155 & \text{if TP} = 12 \\ 0.0164(VD) + 0.1559 & \text{if TP} = 14 \\ 0.0165(VD) + 0.1556 & \text{if TP} = 16 \\ 0.0165(VD) + 0.1557 & \text{if TP} = 18 \\ 0.0165(VD) + 0.1552 & \text{if TP} = 20 \end{cases} \tag{6}$$

Utilizing the Equation (6), the Q-learning algorithm predicts the Channel Busy Ratio (CBR) for each possible pairing of vehicle density and transmission power. This prediction is integral to the learning process as it estimates how congested the communication channel will likely be for different combinations. This enables the Q-learning algorithm to discern which combinations are most conducive to efficient communication.

In the Q-table, each row signifies a unique state characterized by a specific combination of vehicle density (within the range of 0 to 50) and estimated average transmission power (spanning from 2 to 20) used by nearby vehicles. The columns of the Q-table represent different transmission power levels that can be adopted by the vehicle in question.

For instance, if the vehicle density is 20 and the average transmission power of neighboring vehicles is 6, then the corresponding row in the Q-table would contain the Q-values associated with taking different transmission power levels under this state.

By calculating the CBR for all combinations and updating the Q-values accordingly, the algorithm learns which actions (i.e., transmission power levels) will likely yield each state's highest rewards (or the most efficient communication) for each state.

As the Q-table gets populated, it becomes an invaluable resource for decision-making. When the vehicle needs to choose a transmission power level, it can refer to the Q-table for its state (based on vehicle density and neighboring transmission power) and select the action with the highest Q-value. This means that the vehicle is choosing the transmission power level that, according to the learning process, is expected to yield the best results regarding communication efficiency.

In Algorithm 1, the construction and utilization of the Q-table are central to the learning process. This table aids in storing and updating the expected rewards for taking various actions, in this case, transmission power levels, under different states characterized by vehicle density and average transmission power of neighboring vehicles.

Initially, steps 1 and 2 in the algorithm define the state space and action space, respectively. The state space is the set of all possible combinations of vehicle density (ranging from 0 to 50) and the estimated average transmission power used by surrounding vehicles. The action space comprises the transmission power levels that can be selected by the vehicle.

Step 3 involves initializing the Q-table, where each cell corresponds to a specific state-action pair and is initially set to 0. Given the absence of a terminal state in this problem, the algorithm instead relies on a predefined number of episodes for the learning process. In this study, the algorithm runs for 80,000 episodes.

Steps 4 to 13 are executed throughout each episode to update the Q-table continuously. This involves selecting actions and observing the environment to gauge their impact. The algorithm employs a mixed strategy of exploiting the best-known actions according to the current Q-table (optimal policy) and occasionally selecting random actions to explore potentially better strategies. This combination ensures a balance between exploitation and exploration, which is fundamental in reinforcement learning for discovering optimal policies.

The parameters γ (gamma) and α (alpha) represent the discount factor and learning rate, respectively. In this context, γ is set to 0.9, indicating a considerable emphasis on future rewards, while α is set to 0.01, representing a cautious learning rate in updating the Q-values.

As the algorithm iterates over the episodes, it improves the Q-table. After 80,000 episodes, it is observed that the differences between consecutive Q-tables become negligible, signifying convergence of the algorithm. At this juncture, the Q-table is deemed to adequately represent the expected future rewards of actions under different states and is saved to a file for subsequent utilization in Phase 2.

The complexity of this algorithm is O(n), as mentioned in [15], with n representing the number of unique state-action pairs. This linear complexity is typical when dealing with Q-learning algorithms and indicates that the computational cost increases linearly with the number of state-action pairs. This should be considered, especially in real-world implementations, to ensure the algorithm remains computationally feasible.

3.3.1.2 Q-Table

The Q-table is a 2D list (matrix). This table has rows corresponding to each state and columns corresponding to each action. In our Q-table, each state corresponds to the environment with vehicle density, the estimated average transmission power of surrounding vehicles, and the corresponding estimated CBR, and each action corresponds to a transmission power that the vehicle can use.

In Table 3.2, each "Q" is a numerical value that represents the learned value of taking a particular action (column) from a particular state (row). For instance, the value in the cell corresponding to a state with TxPower of 2, VD of 25, and CBR of 0.4336 represents the Q-value of taking an action with TxPower levels of 2, 10, or 20 when the system is in that state.

Table 3.2: Q-table.

TxPower	**VD**	**CBR**	**TxPower**		
			2	**10**	**20**
2	25	0.4336	Q	Q	Q
10	25	0.5678	Q	Q	Q
20	25	0.6455	Q	Q	Q

1. **State Representation:** Each row in the Q-table represents a specific state of the system, described by the combination of `TxPower`, `VD`, and `CBR`.
2. **Actions:** The actions the agent can take in any given state seem to be changing the `TxPower` to 2, 10, or 20.
3. **Q-values:** The Q-values in Table 3.2 represent the expected future reward of taking a particular `TxPower` action from a given state.

Initially, in Table 3.3, all Q-values are set to 0, indicating the agent has no prior knowledge about which actions are best. For each episode, the function starts in a random state, and either takes action with the maximum Q-value (exploit) or a random action (explore) based on the epsilon. The Q-values are then updated using the Q-learning formula. As the agent interacts with its environment and learns from its experiences, these Q-values will be updated to reflect the agent's understanding of which actions yield the highest expected future reward.

Table 3.3: Initial Q-table.

TxPower	**VD**	**CBR**	**TxPower**		
			2	**10**	**20**
2	25	0.4336	0	0	0
10	25	0.5678	0	0	0
20	25	0.6455	0	0	0

The agent typically starts its journey from a randomly selected state. Let's assume it started from the state with 'TxPower' of 10, 'VD' of 25, and 'CBR' of 0.5678. The agent decides on an action based on an exploration-exploitation strategy (like ε-greedy). In the early stages, it's more likely to explore, which means choosing a random action.

Let's say it chose the action 'TxPower' 10. The agent interacts with the environment using the selected action ('TxPower' 10). As a result, it receives a reward from the environment. Let's say this reward was quite positive, thus encouraging the agent to favor this action in this state. Using the Q-learning update rule 4, the agent adjusts its Q-value for the selected action in the given state. Based on the formula 4, the Q-value for the state-action pair ('TxPower' 10, 'TxPower' 10) was updated to 25.

After the action, the agent transitions to a new state. Let's say it moves to the state with 'TxPower' 20, 'VD' 25, and 'CBR' 0.6455. The agent repeats the process of action selection, environment interaction, and Q-value update until the end of the episode. For instance, from the state with 'TxPower' 20, 'VD' 25, and 'CBR' 0.6455, it might have chosen action 'TxPower' 20 and received a moderate reward, leading to an updated Q-value of 10 for that state-action pair.

Once the episode concludes, the agent has updated its Q-values based on the experience of that episode, resulting in the updated Q-table we see in Table 3.4.

Table 3.4: Q-table after the first episode.

TxPower	**VD**	**CBR**	**TxPower**		
			2	**10**	**20**
2	25	0.4336	0	0	0
10	25	0.5678	0	25	0
20	25	0.6455	0	0	100

Following our initial Q-learning episode as presented in Table 3.4, the Q-table experienced updates in its values, reflecting the agent's learned experiences and expected future rewards. Table 3.5 depicts the Q-values after additional episodes of learning. It demonstrates the agent's evolving understanding of its environment based on its interactions and received rewards.

Starting again, let's consider the agent found itself in the state with 'TxPower' of 2, 'VD' of 25, and 'CBR' of 0.4336. By now, it's possible that the agent has shifted slightly from pure exploration and is beginning to exploit its past experiences more. Using the ε-greedy strategy, the agent might decide to take action with the highest Q-value (due to exploitation) or a random action (exploration).

Suppose the agent opts for the action 'TxPower' 20. Upon taking this action, the environment provides feedback in the form of a reward. Given that the Q-value corresponding to this state-action pair (‘TxPower’ 2, ‘TxPower’ 20) has been updated to 86, it’s safe to infer that this feedback was significant and positive. The Q-learning update rule 4 aids the agent in updating this value.

Transitioning further, the agent might land in the state with ‘TxPower’ 10, ‘VD’ 25, and ‘CBR’ 0.5678. With a mix of its exploration-exploitation strategy, it could decide to continue with the action ‘TxPower’ 20. This choice might be influenced by past positive experiences or merely the drive to explore unknown state-action pairs. The environment once again provides feedback - the Q-value for this state-action combination (‘TxPower’ 10, ‘TxPower’ 20) gets updated to 93, indicating a strong positive reward.

Lastly, in the state with ‘TxPower’ 20, ‘VD’ 25, and ‘CBR’ 0.6455, the agent might heavily exploit its past knowledge, choosing the action ‘TxPower’ 2 due to its high Q-value of 99. This choice represents an action that has, in the agent’s past experience, led to favorable outcomes in similar states. The significant Q-value suggests a series of positive feedbacks when this action was taken in this state.

Table 3.5: Partial Q-table.

TxPower	**VD**	**CBR**	**TxPower**		
			2	**10**	**20**
2	25	0.4336	29	27	86
10	25	0.5678	24	25	93
20	25	0.6455	99	64	100

3.3.1.3 Application Phase

In this tier, the vehicle harnesses the Q-table produced during the initial phase to ascertain the transmission power for beacon dispatch. A detailed explanation can be found in Algorithm 2.

In Phase 2, delineated in Algorithm 2, a vehicle employs the optimal policy ascertained in Phase 1 to make informed choices regarding its transmission power. This phase is essential for practical implementation as it enables the vehicle to dynamically adapt its transmission power based on the prevailing traffic conditions, thus ensuring efficient Vehicle-to-Vehicle (V2V) communication.

Algorithm 2 QBTPCC Policy Execution in OMNeT++ Framework

1: Fetch current CBR as $curCBR$
2: Acquire current vehicle density as $curVD$
3: Adjust $curVD$: $curVD = \min(curVD, maxVD)$
4: Set default index: $index = 1$
5: Initialize maximum Q-value tracker: $maxVal = -\infty$
6: Start with a default level: $level = 1$
7: Set TxPower array: $TxPower = [2, 4, 6, 8, 10, 12, 14, 16, 18, 20]$
8: **for** each power level in $TxPower$ **do**
9:     **if** Estimated CBR with current $curVD$ and power level using Equation (6) is less than or equal to $curCBR$ **then**
10:         Update $index$ with current power level index
11:         **break**
12:     **end if**
13: **end for**
14: **for** each row entry from 0 to 9 in the Q-table corresponding to $curVD$ and $index$ **do**
15:     Fetch Q-value for the entry as $qVal$
16:     **if** $qVal$ surpasses $maxVal$ **then**
17:         Set $maxVal$ as $qVal$
18:         Update $level$ with current entry index
19:     **end if**
20: **end for**
21: Determine optimal transmission power: $bestTxPower = TxP[level]$
22: Transmit beacon at power $bestTxPower$

In steps 1 and 2, the vehicle conducts environmental sensing to gauge the Channel Busy Ratio (CBR) and the density of vehicles within its vicinity. If the observed vehicle density exceeds the predetermined maximum value, denoted by maxVD, it is set to this maximum value (step 3). This ensures the algorithm operates within the bounds established during the initial phase.

In steps 5, 6, and 7, three variables, index, maxVal, and level are initialized. maxVal is set to a considerably low value to guarantee that higher values will be encountered during the subsequent exploration of the Q-table.

Between steps 8 and 13, the algorithm considers neighboring vehicles' various average transmission power levels. Using these values alongside the detected vehicle density, it consults Equation (6) to discern which combination results in the lowest estimated CBR that is still greater than or equal to the current CBR. Essentially, this part of the algorithm seeks to identify an optimal state that matches the current traffic conditions and minimizes channel congestion without underutilizing it.

From steps 14 to 20, the vehicle selects the transmission power that is deemed to be the most suitable according to the Q-table. The Q-table is structured such that each row corresponds to a particular state defined by vehicle density, the estimated average transmission power of neighboring vehicles, and the estimated CBR. The columns represent different transmission power levels that the vehicle may adopt. The transmission power associated with the highest Q-value within a particular row is selected as this indicates the optimal policy for that state. This transmission power is expected to yield the best balance between communication efficiency and channel congestion for the given traffic conditions.

This process is repeated as the vehicle moves and the environment changes, allowing for dynamic adaptation of transmission power levels for efficient V2V communication in various traffic scenarios.

In steps 14 through 20, the vehicle undergoes a systematic procedure to identify the optimal transmission power for its current state. This is achieved by examining each possible transmission power and selecting the one with the highest associated Q-value from the Q-table.

Step 16 involves the vehicle accessing the relevant row in the Q-table that corresponds to its current state. The state encapsulates the prevailing traffic conditions, such as vehicle density and channel busy ratio (CBR).

Following this, between steps 16 and 19, the algorithm undertakes a comparative analysis. For each possible transmission power, it evaluates the corresponding Q-value and compares it with the maximum Q-value encountered up to that point. If a higher Q-value is found, the maximum value is updated. This process continues iteratively for each transmission power option.

Upon completing this iterative comparison, in step 21, the vehicle selects the transmission power associated with the highest Q-value. This is the optimal action the vehicle should take according to the Q-learning algorithm, and it is expected to be the most effective choice in balancing communication efficiency and minimizing channel congestion for the given traffic conditions.

It is important to note that the number of iterations is fixed in Algorithm 2, implying that the algorithm performs a constant number of operations irrespective of the input size. As a result, the time complexity of this algorithm is O(1), indicating that it executes in constant time. This is highly beneficial in a vehicular communication environment, as it enables the vehicle to rapidly determine the best action, which is crucial for real-time decision-making.

CHAPTER 4

Simulation Result

This chapter delves into the intricacies of the results procured from the application of the Q-Learning Based Transmission Power Congestion Control (QBTPCC) method. The relevance and reliability of these results are heightened by the use of the Vehicles in Network Simulation (Veins) framework [54]. This framework offers a robust platform, simulating real-life vehicular communication settings which capture a wide array of variables such as traffic densities, vehicle speeds, and environmental dynamics.

In our study on vehicular communications, we primarily examined the impact of our QBTPCC on crucial metrics like the Channel Busy Ratio (CBR), indicating network congestion and transmission efficiency. Besides CBR, we evaluated several other performance indicators relevant to vehicular ad-hoc networks (VANETs). The results are presented broadly before delving into a detailed analysis of parameter interactions.

By analyzing the simulation data, we provide insights into the relative efficacy of the QBTPCC approach under different conditions, underscoring its advantages and potential areas of improvement. Additionally, we compare our model's performance against conventional models to establish its practical applicability and superiority further.

4.1 Experimental Setup

In the subsequent section, we assess the QBTPCC methodology utilizing the Vehicles in Network Simulation (Veins) framework [54]. This framework incorporates the essential modules of IEEE 802.11p and IEEE 1609 protocols, enabling the testing of V2V networks. Veins act as a bridge between the widely recognized network simulation software OMNeT++ (Objective Modular Network Testbed in C++) [55] and the urban mobility simulator SUMO (Simulation of Urban Mobility) [56].

Figure 4.1: VANET Simulation Tools.

The purpose of this assessment was to gauge the performance of our Q-learning approach. To accomplish this, we simulated a 20-kilometer-long highway with four lanes, two for each direction. However, we only focused on a central 4-kilometer segment of the highway to remove inaccuracies from vehicles entering or exiting the simulation.

The simulation incorporated 500 and 800 vehicles, with their velocities randomly set between 80 to 130 kilometers per hour. This setup allowed us to generate dynamic traffic flows with fluctuating vehicle densities. The two-vehicle quantities were selected to exemplify low and high-traffic density scenarios. The total simulation duration was 100 seconds.

4.1.1 Simulation Parameters

The parameters utilized for our performance evaluation are collated and presented in Table 4.1. This table provides a comprehensive overview of the specific configurations employed for this study.

Table 4.1: Configuration Parameters.

Name	**Value**
Transmission Power	Variable: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20 mW
Beacon Rate	2.5 Hz
BSM Size	512 Bytes
Data Rate	6 Mbps
High Vehicle Density	800 Vehicles
Low Vehicle Density	500 Vehicles
Highway Length/Lanes	4 km
Number of Lanes	4
Simulation Time	100 s

The effectiveness of our QBTPCC method is assessed through a comparative analysis with other existing techniques, all of which are examined under the dynamic traffic model as described earlier. The comparative study aims to illuminate the differences in performance, highlighting the advantages conferred by our QBTPCC approach.

4.2 Simulation Result

We conduct a comparative performance analysis using several established algorithms. Each algorithm is scrutinized under two different traffic scenarios – one with low-density traffic comprising 500 vehicles and another with high-density traffic comprising 800 vehicles. This comprehensive assessment allows us to fully gauge the strengths and weaknesses of each algorithm under varying conditions. To evaluate the proposed algorithm, we have compared its performance with the following existing techniques in terms of key performance indicators such as CBR, BER, and IPD.

1. 2 mW: All BSMs transmitted with 2 mW power.
2. 10 mW: All BSMs transmitted with 10 mW power.
3. 20 mW: All BSMs transmitted with 20 mW power.
4. BACVT8_2: 8 BSMs with Tx power = Tx_{n}, 2 BSMs with Tx power = Tx_{f}. [42]
5. BACVT5_5: 5 BSMs with Tx power = Tx_{n}, 5 BSMs with Tx power = Tx_{f}. [42]
6. MDPRP: An RL-based congestion control algorithm [16]. Results for 800 vehicles were not available for MDPRP, so we have only shown comparisons for 500 vehicles.

We note that it is possible to use the BACVT algorithms with different values of Txf and Txn [42]. The transmission power for the far and near ranges used in BACVT are set to Txf = 20 mW and Txn = 5 mW, respectively. For comparing the MDPRP method, we use the simulation result published in [9] for 500 vehicles.

4.2.1 Average Channel Busy Ratio

Channel Busy Ratio (CBR) is a fundamental performance metric in evaluating vehicular communications systems, providing a dynamic measure of channel congestion level. By definition, CBR is the ratio of the time duration in which the channel is sensed as busy (actively transmitting or receiving data) to the total observation time [9]. Consequently, it serves as a reliable indicator of the overall load on the channel: a higher CBR value typically correlates with a more loaded and potentially congested channel. In comparison, a lower value suggests a less loaded channel.

To allow a comprehensive observation of the CBR changes during the simulation process, each vehicle in the network is configured to calculate the prevailing CBR prior to each Basic Safety Message (BSM) dispatch. This calculation is based on the channel's status as detected by the vehicle, taking into account both the channel's current activity and the overall traffic density in the network.

However, given the potentially high variability of CBR values across time and across different vehicles, gauging an overall understanding of channel usage from these individual CBR calculations can be challenging. To overcome this, we introduce an additional metric referred to as Average CBR.

The Average CBR metric calculates the mean CBR across all vehicles in the network at regular intervals - specifically, every 5 seconds. This approach ensures that we capture the cumulative channel usage across the entire vehicular network rather than isolated usage from individual vehicles. This metric thus provides a more holistic and informative perspective on the overall network performance, simplifying the analysis of CBR trends over the course of the simulation.

Figures 4.2 and 4.3 present a comparative analysis of the average Channel Busy Ratio (CBR) across six different methodologies, employing both 500 and 800 vehicles, respectively. The 20 mW transmission power consistently registers the highest average CBR across all evaluated scenarios, leading to substantial channel congestion. On the other hand, the 2 mW transmission power produces a lower average CBR compared to 10 mW, which is not an optimal choice as it sends a lesser number of packets for a low CBR value, and the awareness is also lower for 2 mW which is discussed in section 4.2.3.

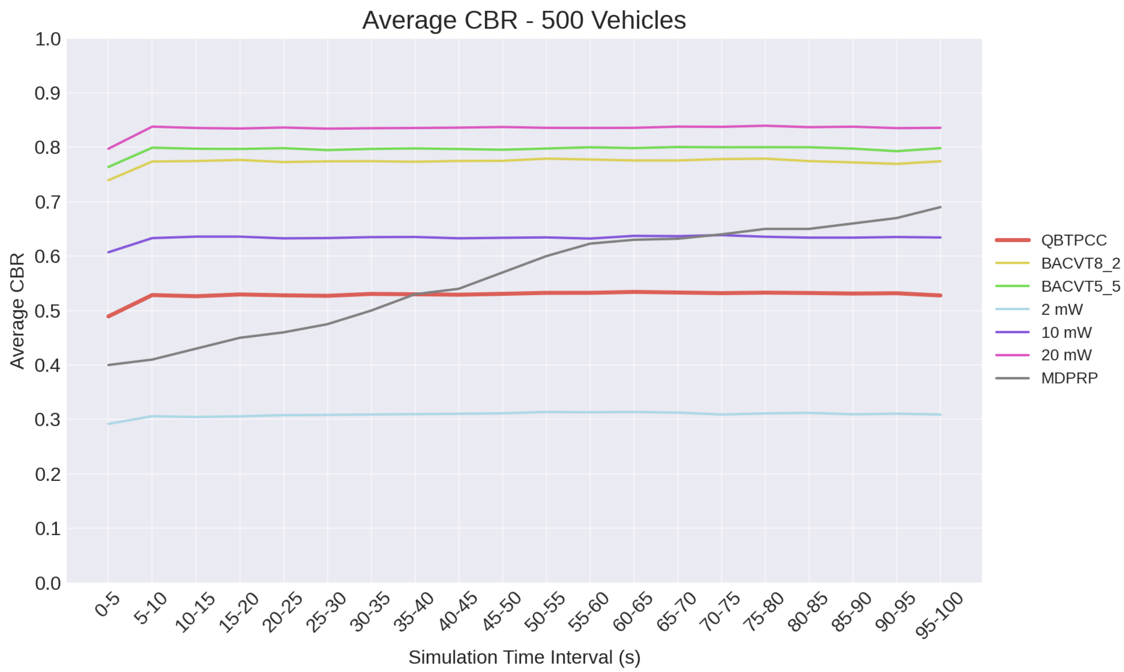

Figure 4.2: Average CBR for 500 vehicles.

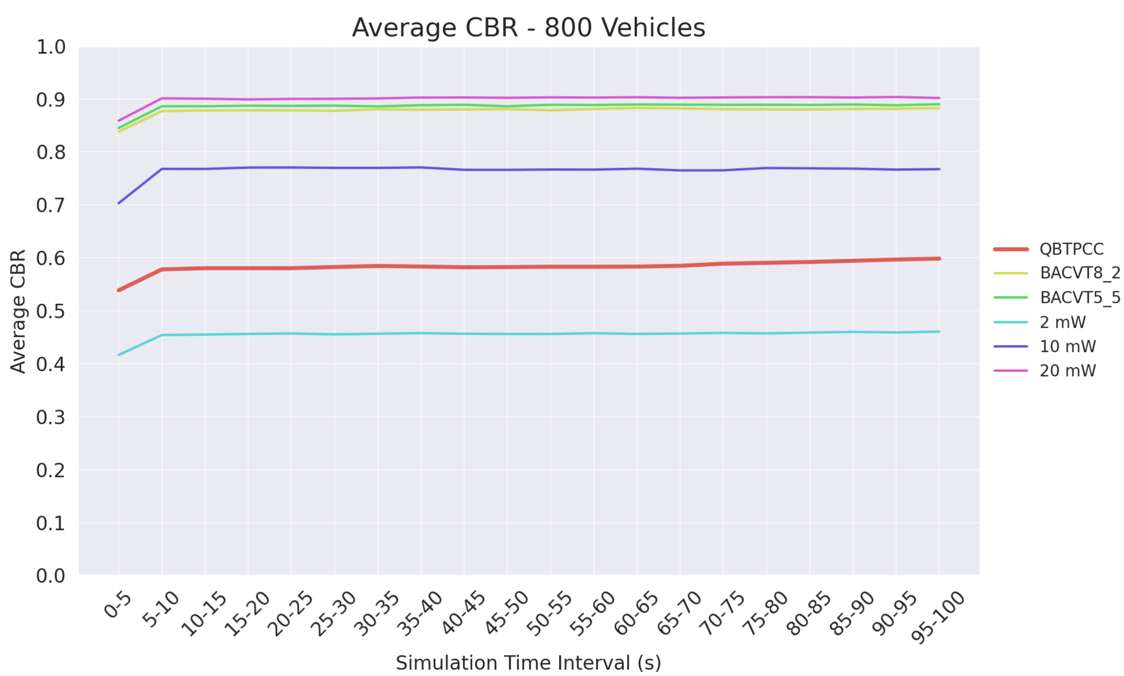

Figure 4.3: Average CBR for 800 vehicles.

Both graphs reveal that BACVT5_5 displays a higher average Channel Busy Ratio (CBR) in comparison to BACVT8_2, especially when contrasted with our proposed method, QBTPCC. A noteworthy observation in both traffic paradigms is the stability exhibited by QBTPCC; its average CBR values uniformly remain beneath the threshold of 0.6, a criterion established within the reward function.

In this comparison, we also compare our method to the MDPRP approach discussed in the paper [9]. In contrast to QBTPCC, MDPRP exhibits a higher CBR value, reflecting a distinction between the performance characteristics of these two approaches.

The results presented in Figures 4.2 and 4.3 highlight the effectiveness of the QBTPCC approach compared to the other methodologies tested. Not only does QBTPCC demonstrate a lower average Channel Busy Ratio (CBR) in both scenarios of 500 and 800 vehicles, but it also exhibits remarkable stability, maintaining CBR values consistently beneath the threshold of 0.6. This stability and efficient channel load management set QBTPCC apart from other methods, making it a compelling choice for handling dynamic traffic scenarios.

4.2.2 Beacon Error Rate

The Beacon Error Rate (BER) is a crucial performance metric in vehicular communication systems that provides a quantifiable measure of transmission reliability. Specifically, it represents the proportion of Basic Safety Messages (BSMs) disseminated throughout the simulation that fails to reach their intended recipients, indicating transmission loss or failure.

Mathematically, BER is computed as the ratio of the total number of lost BSMs (L) to the total number of sent BSMs (S) in the simulation, i.e., BER = L/S. The lost BSMs include all messages that, for various reasons such as congestion, interference, or signal loss, are not successfully received by another vehicle in the network.

A lower BER implies a more reliable and efficient transmission system as it signifies fewer lost messages and, thus, more successful communication attempts. In contrast, a higher BER may indicate problems in the communication channel, such as high congestion or excessive interference. Therefore, by minimizing the BER, we aim to maximize the effectiveness of the V2V communication, enhancing safety and operational efficiency in the vehicular network.

Thus, the BER serves as an important benchmark for assessing the performance of our QBTPCC approach, allowing us to gauge its reliability and efficiency in handling safety-critical BSMs in different traffic density scenarios.

The Beacon Error Rate (BER), a metric that quantifies the proportion of sent packets that were lost for 500 and 800 Vehicles, is illustrated in Figures 4.4 and 4.5. Among the methodologies examined, QBTPCC consistently demonstrated low BER values in 500 and 800 vehicles scenarios. Conversely, the MDPRP approach registered elevated BER values when compared to QBTPCC in 500 vehicles.

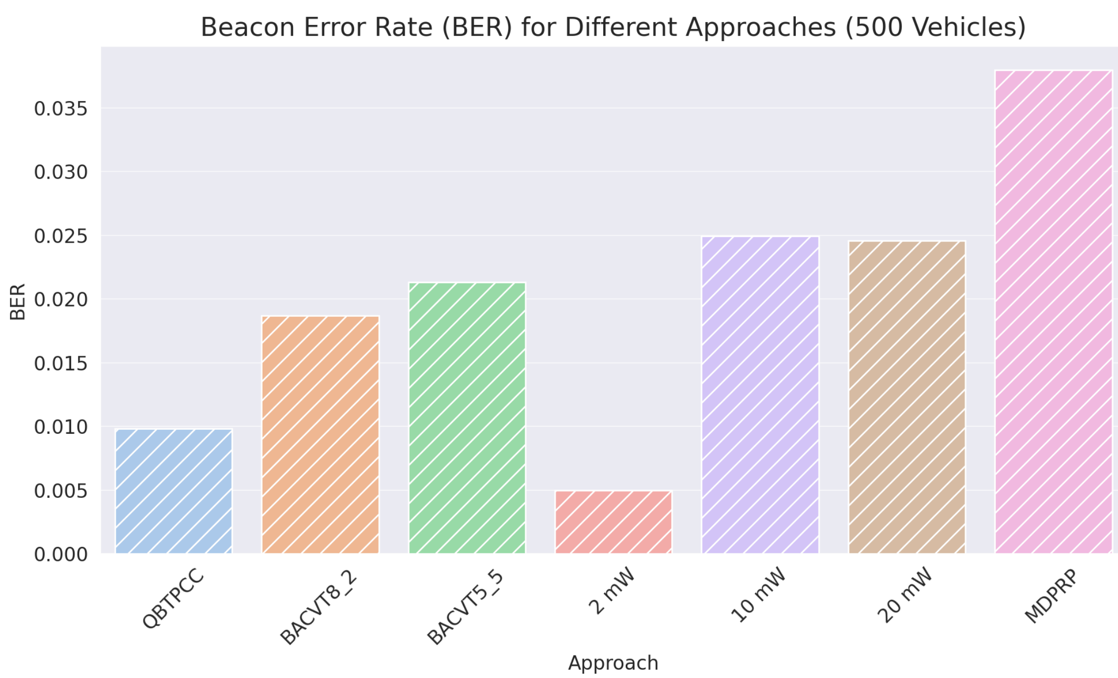

Figure 4.4: BER for 500 vehicles.

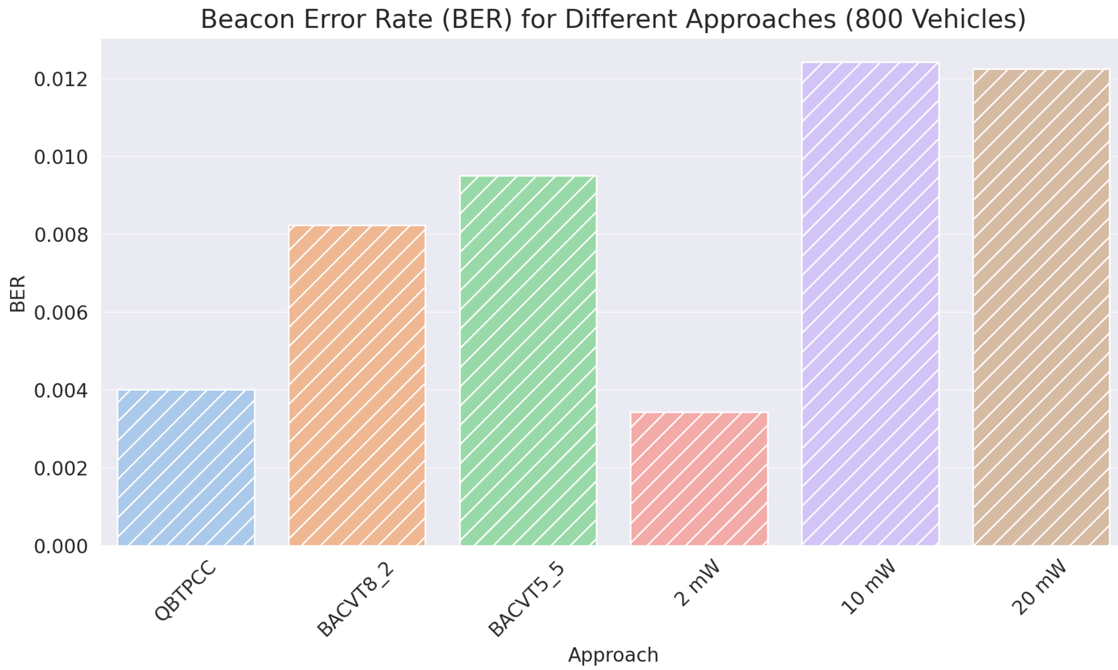

Figure 4.5: BER for 800 vehicles.

The comparative analysis with the BACVT algorithms further enriches this study. All BACVT methods recorded a lower BER compared to the 10 mW and 20 mW transmission power due to their reduced susceptibility to packet collisions but a higher value than QBTPCC. In line with expectations, the 2 mW power level exhibited the lowest BER, a consequence of the diminished transmission power curtailing the likelihood of packet collisions even further. This detailed understanding highlights how transmission power and network efficiency are connected, and it confirms that the QBTPCC approach holds great potential for improving vehicular communications.

In synthesis, the effectiveness of QBTPCC in managing vehicular communication is evident. By consistently maintaining the channel load below the designated threshold of 0.6, as stipulated in the reward function, QBTPCC plays a vital role in ensuring efficient data flow within congested traffic environments. Adhering closely to the threshold significantly reduces packet loss probability, leading to enhanced overall network performance, as evidenced by the observed metrics for BER.

4.2.3 Inter-Packet Delay

The Inter-Packet Delay (IPD) is a pivotal metric in evaluating the performance of a vehicular network, defined as the average delay between the successful receipt of consecutive Basic Safety Messages (BSMs) from one vehicle to another. We denote 'V' as the set encompassing all vehicles involved in the simulation.

For a specific pair of vehicles, denoted as (s, r), where s and r belong to the set V and s is distinct from r, we calculate IPD as follows. Let's say $M_{s,r}$ is the set of all successfully received BSMs for this vehicle pair, and $t_{s,r,i}$ represents the reception time of the ith BSM in the set $M_{s,r}$ from vehicle s at vehicle r. Then, the average IPD across the entire simulation, inclusive of all vehicles, is given by the formula [42]:

$$\text{AverageIPD} = \frac{\sum_{s\in V}\sum_{r\in V, s\neq r}\sum_{i\in M_{s,r}}(t_{s,r,i} - t_{s,r,i-1})}{\sum_{s\in V}\sum_{r\in V, s\neq r}|M_{s,r}|} \tag{1}$$

This formula implies that we take the sum of differences in reception times for all consecutive BSMs, divided by the total count of successfully received BSMs across all vehicle pairs in the simulation.

Furthermore, to obtain a more granular analysis of IPD, we can calculate it within a specific range. For a particular vehicle u in the set V, and a specific distance range $(d_1 < d \leq d_2)$, we define V_{u,d_1,d_2} as the set of all vehicles within distance d from vehicle u. Subsequently, we modify the Average IPD formula as follows [42]:

$$\text{AverageIPD}_{d_1,d_2} = \frac{\sum_{s\in V}\sum_{r\in V_{s,d_1,d_2}}\sum_{i\in M_{s,r}}(t_{s,r,i} - t_{s,r,i-1})}{\sum_{s\in V}\sum_{r\in V_{s,d_1,d_2}}|M_{s,r}|} \tag{2}$$

This formula enables us to calculate the average IPD for vehicles within a specific distance range from a given vehicle. This parameter provides valuable insight into the network's performance in varying proximity scenarios, ultimately augmenting the depth of our analysis.

The successful implementation of safety applications in VANETs requires low IPD values. Suppose the delay between the successful reception of two consecutive BSMs is too high. In that case, it may lead to outdated or irrelevant information being received by the vehicles, which in turn may lead to inefficient decision-making and reduced overall safety.

As demonstrated in Figures 4.6 and 4.7, our simulation results indicate that the Inter-Packet Delay (IPD) performance of all examined algorithms is remarkably consistent for messages originating from vehicles within a 280 m radius. Beyond this range, BACVT8_2 shows the highest delay due to its strategy of transmitting a larger number of Basic Safety Messages (BSMs) at a lower transmission power. Nevertheless, IPD performance improves significantly when the high-to-low transmission power ratio is adjusted, as demonstrated by BACVT5_5.

As anticipated, the 20 mW configuration showcases the least IPD values for larger distances, a result of all BSMs being transmitted at this power level. It's worth noting that, despite its admirable performance in terms of Beacon Error Rate (BER), the 2 mW configuration falls short when considering IPD for vehicles beyond the 240 m threshold. In both 500 and 800 vehicle densities, this scheme ceases to receive BSMs beyond a certain distance. Moreover, while the 10 mW configuration demonstrates a lower IPD, it also stops receiving packets beyond a specified range, albeit outperforming the 2 mW configuration.

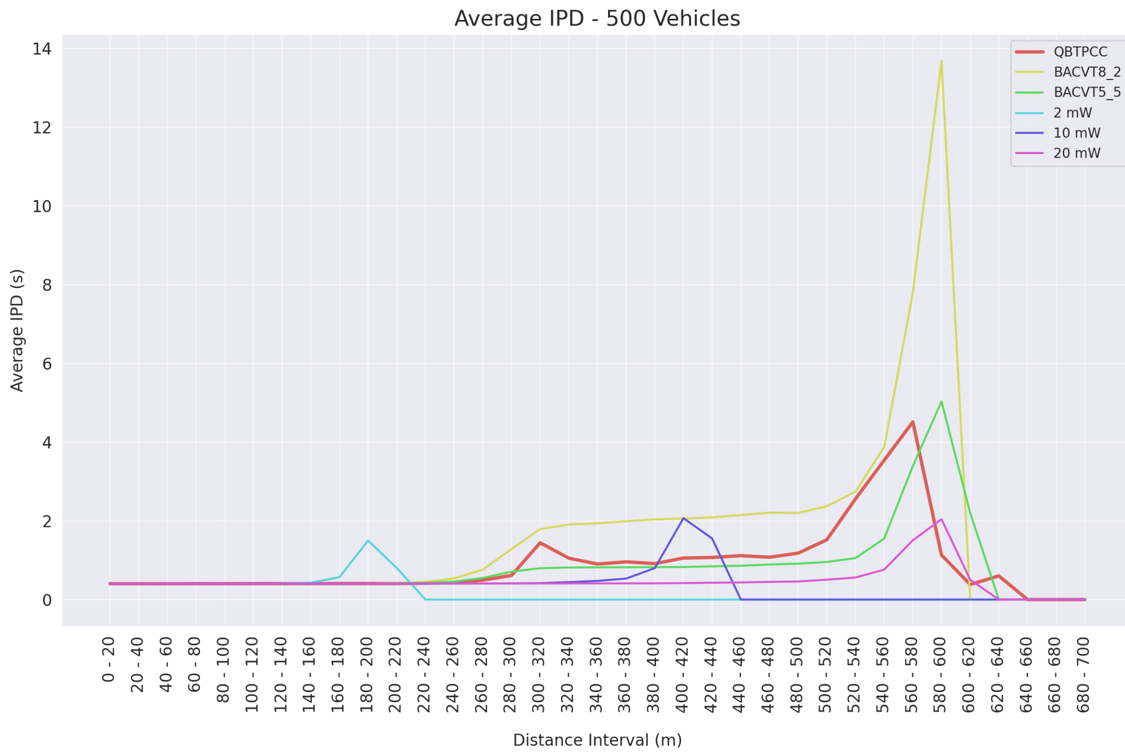

Figure 4.6: IPD for 500 Vehicles.

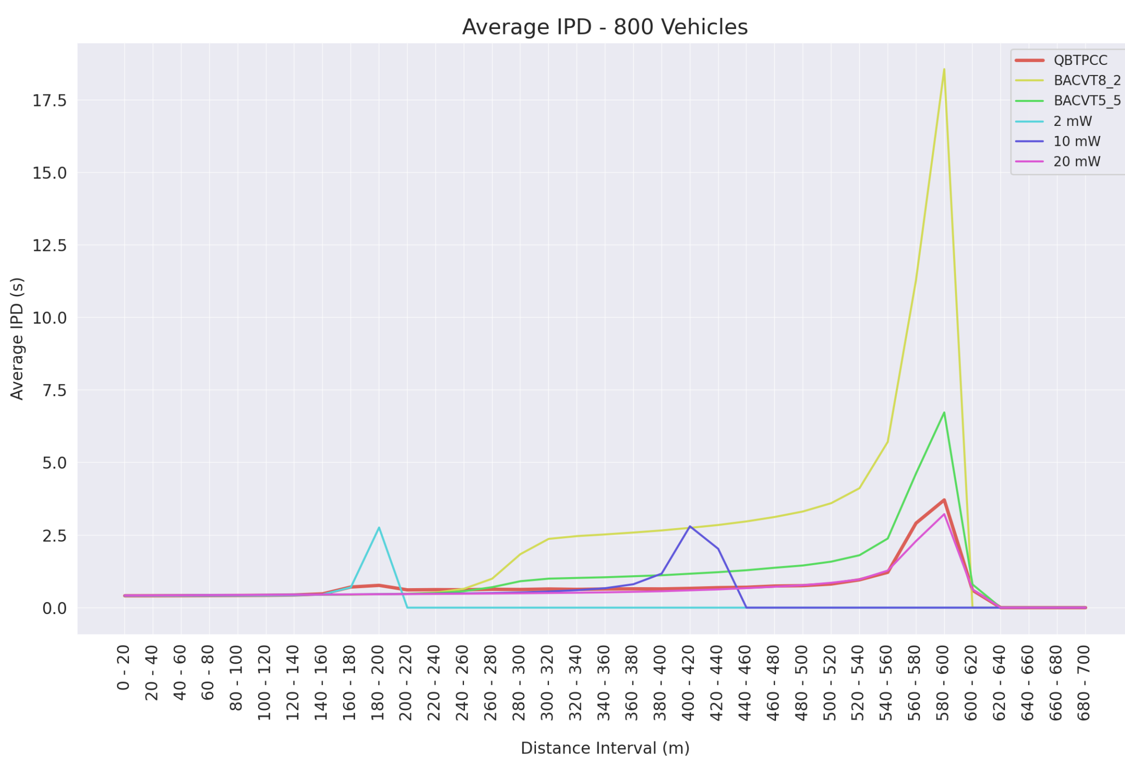

Figure 4.7: IPD for 800 Vehicles.

Notably, our proposed method, QBTPCC, attains an IPD metric marginally above that of the 20 mW configuration, making it an intriguing finding. Across both vehicular densities, QBTPCC consistently maintains a low IPD and continues to receive packets across the entire distance, demonstrating the effectiveness of our approach.

CHAPTER 5

Conclusion and Future Work

5.1 Conclusion

In the realm of Vehicular Ad-hoc Networks (VANETs), the effective management of the limited 10 MHz channel capacity is a significant challenge, particularly with the escalating congestion as vehicular density increases. This scenario poses a threat to the bandwidth's ability to ensure the reliable and timely delivery of crucial safety messages.

In response to this issue, this study presents a pioneering solution: the Q-learning Based Transmission Power Control for Channel Congestion (QBTPCC). Leveraging the capabilities of Reinforcement Learning (RL), specifically Q-learning, QBTPCC strategically manages the transmission power for Basic Safety Messages (BSMs), balancing the Channel Busy Ratio (CBR) beneath a predetermined threshold while optimizing situational awareness in various traffic conditions.

Evaluations of QBTPCC, based on robust simulations and benchmarking against existing strategies using constant transmission power and another RL-based methodology, consistently demonstrated superior performance. QBTPCC effectively maintained channel load at or near the set level, irrespective of whether the traffic densities were low or high. This consistent performance, keeping the CBR within the range of 0.5 to 0.6 and preserving high-quality situational awareness (as reflected in IPD metrics), is indicative of QBTPCC's capacity to navigate the complex balance between awareness and congestion. This capability is crucial for efficient and safe communication in VANETs.

Our research provides a substantial contribution towards the evolution of intelligent transportation systems and paves the way towards more efficient, safer, and technologically advanced road networks. However, recognizing the potential complexities of real-world scenarios, we suggest that future research should involve comprehensive field tests. These tests would further validate and refine our model, ensuring it can adapt to the wide range of challenges inherent in real-world situations.

In conclusion, our study offers an innovative and effective approach for handling transmission power in VANETs, showcasing the tremendous potential of machine learning techniques in addressing dynamic and complex issues. We hope this work acts as a launchpad for further research and development in the field, pushing the limits of vehicular communication networks and contributing significantly to the broader ambitions of smart city development.

5.2 Future Work

While our study has laid a solid foundation in the application of Q-learning for transmission power control in vehicular ad hoc networks, there are several directions in which this research can be extended and refined.

We can enhance QBTPCC by designing a comprehensive reward function that takes into account additional metrics. We are also exploring the possibility of adjusting multiple parameters, transmission power, and data rate, in addition to the transmission power, to create a more robust Q-table.

Multi-agent Learning: In our current work, we have focused on the perspective of individual vehicles. Future research can explore multi-agent learning, where multiple vehicles work together and learn collectively. This could lead to more coordinated and effective decisions in the network, potentially further improving network performance.

Advanced Machine Learning Techniques: While Q-learning has provided promising results, there are numerous other machine-learning techniques that could be explored in this context. Deep reinforcement learning, for instance, could potentially capture more complex patterns and dependencies in the data, providing even more effective transmission power control.

Security Concerns: Security issues become even more important as we move towards a more connected vehicular network. Future research could look into integrating security measures into our approach to protect the integrity and confidentiality of the data transmitted in the network.

By building upon the work completed in this thesis and addressing these opportunities for future research, we can further improve the performance and robustness of VANETs, contributing to safer and more efficient road transport.

REFERENCES

[1] National Highway Traffic Safety Administration et al. *Early estimate of motor vehicle traffic fatalities in 2020. Washington DC: United States Department of Transportation*. 2022.

[2] M. S. Sheikh and J. Liang. "A Comprehensive Survey on VANET Security Services in Traffic Management System". In: *Wireless Communications and Mobile Computing* 2019 (2019), pp. 1–25. DOI: 10.1155/2019/2423915.

[3] K.S. Bilstrup, E. Uhlemann, and E.G. Strom. "Scalability issues of the MAC methods STDMA and CSMA of IEEE 802.11p when used in VANETs". In: *Proceedings of the 2010 IEEE International Conference on Communications Workshops*. Cape Town, South Africa, May 2010, pp. 1–5.

[4] John B Kenney. "Dedicated short-range communications (DSRC) standards in the United States". In: *Proceedings of the IEEE* 99.7 (2011), pp. 1162–1182.

[5] Ghassan MT Abdalla, Mosa Ali Abu-Rgheff, and Sidi Mohammed Senouci. "Current trends in vehicular ad hoc networks". In: *Ubiquitous Computing and Communication Journal* (2007), pp. 1–9.

[6] Federico Poli. "Vehicular communications: from DSRC to Cellular V2X". PhD thesis. Politecnico di Torino, 2018.

[7] Aekta Sharma and Arunita Jaekel. "Machine learning based misbehaviour detection in VANET using consecutive BSM approach". In: *IEEE Open Journal of Vehicular Technology* 3 (2021), pp. 1–14.

[8] J. Eggerton. *FCC to Split Up 5.9 GHz*. 2019. URL: https://www.nexttv.com/news/fcc-to-split-up-5-9-ghz.

[9] Xiaofeng Liu, Ben St. Amour, and Arunita Jaekel. "A Reinforcement Learning-Based Congestion Control Approach for V2V Communication in VANET". In: *Applied Sciences* 13.6 (2023). ISSN: 2076-3417. DOI: 10.3390/app13063640. URL: https://www.mdpi.com/2076-3417/13/6/3640.

[10] Gursimrat Singh. *Cross layer design in vehicular ad-hoc networks (VANETs): issues, approaches and solutions*. Tech. rep. Technical Report, 2021.

[11] Anirudh Paranjothi, Mohammad S. Khan, and Sherali Zeadally. "A survey on congestion detection and control in connected vehicles". In: *Ad Hoc Networks* 108 (2020), p. 102277. ISSN: 1570-8705. DOI: https://doi.org/10.1016/j.adhoc.2020.102277. URL: https://www.sciencedirect.com/science/article/pii/S1570870520306387.

[12] Manisha Chahal et al. "A Survey on software-defined networking in vehicular ad hoc networks: Challenges, applications and use cases". In: *Sustainable Cities and Society* 35 (2017), pp. 830–840. ISSN: 2210-6707. DOI: https://doi.org/10.1016/j.scs.2017.07.007. URL: https://www.sciencedirect.com/science/article/pii/S221067071730361X.

[13] Hongmei Deng, Wei Li, and Dharma P Agrawal. "Routing security in wireless ad hoc networks". In: *IEEE Communications magazine* 40.10 (2002), pp. 70–75.

[14] Saif Al-Sultan et al. "A comprehensive survey on vehicular Ad Hoc network". In: *Journal of Network and Computer Applications* 37 (2014), pp. 380–392. ISSN: 1084-8045. DOI: https://doi.org/10.1016/j.

jnca.2013.02.036. URL: https://www.sciencedirect.com/science/article/pii/S108480451300074X.

[15] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[16] Juan Aznar-Poveda et al. "MDPRP: A Q-learning approach for the joint control of beaconing rate and transmission power in VANETs". In: *IEEE Access* 9 (2021), pp. 10166–10178.

[17] Apoorva Jindal. *Self-Driving Car in Gym Environment Using Reinforcement Learning*. Medium. 2018. URL: https://medium.com/@apoorva.jindal.99/self-driving-car-in-gym-environment-using-reinforcement-learning-4b6a6ab81bd3.

[18] Mahmood A. Al-Shareeda et al. "Survey of Authentication and Privacy Schemes in Vehicular ad hoc Networks". In: *IEEE Sensors Journal* 21.2 (2021), pp. 2422–2433. DOI: 10.1109/JSEN.2020.3021731.

[19] Muhammad Sameer Sheikh, Jun Liang, and Wensong Wang. "A survey of security services, attacks, and applications for vehicular ad hoc networks (vanets)". In: *Sensors* 19.16 (2019), p. 3589.

[20] "Medium access control protocols for safety applications in Vehicular Ad-Hoc Network: A classification and comprehensive survey". In: *Vehicular Communications* 2.4 (2015), pp. 223–237. ISSN: 2214-2096. DOI: https://doi.org/10.1016/j.vehcom.2015.10.001. URL: https://www.sciencedirect.com/science/article/pii/S2214209615000546.

[21] Mohamed Nidhal Mejri, Jalel Ben-Othman, and Mohamed Hamdi. "Survey on VANET security challenges and possible cryptographic solutions". In: *Vehicular Communications* 1.2 (2014), pp. 53–66. ISSN: 2214-2096. DOI: https://doi.org/10.1016/j.vehcom.2014.05.001. URL: https://www.sciencedirect.com/science/article/pii/S2214209614000187.

[22] N Ganeshkumar and Sanjay Kumar. "Obu (on-board unit) wireless devices in vanet (s) for effective communication—A review". In: *Computational Methods and Data Engineering: Proceedings of ICMDE 2020, Volume 2* (2020), pp. 191–202.

[23] Amit Dua, Neeraj Kumar, and Seema Bawa. "A systematic review on routing protocols for Vehicular Ad Hoc Networks". In: *Vehicular Communications* 1.1 (2014), pp. 33–52. ISSN: 2214-2096. DOI: `https://doi.org/10.1016/j.vehcom.2014.01.001`. URL: `https://www.sciencedirect.com/science/article/pii/S2214209614000059`.

[24] "Data communication in VANETs: Protocols, applications and challenges". In: *Ad Hoc Networks* 44 (2016), pp. 90–103. ISSN: 1570-8705. DOI: `https://doi.org/10.1016/j.adhoc.2016.02.017`. URL: `https://www.sciencedirect.com/science/article/pii/S1570870516300580`.

[25] Felipe Domingos Da Cunha et al. "Data communication in VANETs: a survey, challenges and applications". In: (2014).

[26] Mohamed Hadded et al. "TDMA-Based MAC Protocols for Vehicular Ad Hoc Networks: A Survey, Qualitative Analysis, and Open Research Issues". In: *IEEE Communications Surveys Tutorials* 17.4 (2015), pp. 2461–2492. DOI: `10.1109/COMST.2015.2440374`.

[27] Kashif Naseer Qureshi et al. "A dynamic congestion control scheme for safety applications in vehicular ad hoc networks". In: *Computers & Electrical Engineering* 72 (2018), pp. 774–788.

[28] Mohssen Mohammed, Muhammad Badruddin Khan, and Eihab Bashier Mohammed Bashier. *Machine learning: algorithms and applications*. Crc Press, 2016.

[29] Xiaofeng Liu and Arunita Jaekel. "Congestion control in V2V safety communication: Problem, analysis, approaches". In: *Electronics* 8.5 (2019), p. 540.

[30] Kurt Driessens. "Relational reinforcement learning". In: *Multi-Agent Systems and Applications: 9th ECCAI Advanced Course, ACAI 2001 and Agent Link's 3rd European Agent Systems Summer School, EASSS 2001 Prague, Czech Republic, July 2–13, 2001 Selected Tutorial Papers 9* (2001), pp. 271–280.

[31] Amit Kumar Goyal et al. "Systematic Study of VANET: Applications, Challenges, Threats, Attacks, Schemes and Issues in Research". In: *Green Computing in Network Security* (2022), pp. 33–52.

[32] Gaurav Bansal, John B Kenney, and Charles E Rohrs. "LIMERIC: A linear adaptive message rate algorithm for DSRC congestion control". In: *IEEE Transactions on Vehicular Technology* 62.9 (2013), pp. 4182–4197.

[33] Gaurav Bansal et al. "EMBARC: Error model based adaptive rate control for vehicle-to-vehicle communications". In: *Proceeding of the tenth ACM international workshop on Vehicular inter-networking, systems, and applications*. 2013, pp. 41–50.

[34] Kazumine Ogura, Jiro Katto, and Mineo Takai. "BRAEVE: Stable and adaptive BSM rate control over IEEE802. 11p vehicular networks". In: *2013 IEEE 10th Consumer Communications and Networking Conference (CCNC)*. IEEE. 2013, pp. 745–748.

[35] Swati Sharma and Manisha Panjeta. "Optimization transmit rate-based decentralized congestion control scheme in vehicular ad hoc networks". In: *AIP Conference Proceedings*. Vol. 2555. 1. AIP Publishing LLC. 2022, p. 030006.

[36] Mare Torrent-Moreno, Paolo Santi, and Hannes Hartenstein. "Distributed fair transmit power adjustment for vehicular ad hoc networks". In: *2006 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks*. Vol. 2. IEEE. 2006, pp. 479–488.

[37] M. Wang et al. "Research on adaptive beacon message transmission power in VANETs". In: *Journal of Ambient Intelligence and Humanized Computing* 13 (2022), pp. 1307–1319. DOI: 10.1007/s12652-020-02575-x.

[38] Seungnam Yang, Hyogon Kim, and Seungho Kuk. "Less is more: need to simplify ETSI distributed congestion control algorithm". In: *Electronics Letters* 50.4 (2014), pp. 279–281.

[39] Srihari Jayachandran and Arunita Jaekel. "Adaptive Data Rate Based Congestion Control in Vehicular Ad Hoc Networks (VANET)". In: *International Conference on Ad Hoc Networks*. Springer. 2021, pp. 144–157.

[40] M Deeksha et al. "Multistate active combined power and message/data rate adaptive decentralized congestion control mechanisms for vehicular ad hoc networks". In: *Journal of Physics: Conference Series*. Vol. 2161. 1. IOP Publishing. 2022, p. 012018.

[41] Byeong-Moon Cho, Min-Seong Jang, and Kyung-Joon Park. "Channel-aware congestion control in vehicular cyber-physical systems". In: *IEEE Access* 8 (2020), pp. 73193–73203.

[42] Xiaofeng Liu, Ben St. Amour, and Arunita Jaekel. "Balancing Awareness and Congestion in Vehicular Networks Using Variable Transmission Power". In: *Electronics* 10.16 (2021). ISSN: 2079-9292. DOI: 10.3390/electronics10161902. URL: https://www.mdpi.com/2079-9292/10/16/1902.

[43] Caitlin Facchina and Arunita Jaekel. "Speed based distributed congestion control scheme for vehicular networks". In: *2020 IEEE Symposium on Computers and Communications (ISCC)*. IEEE. 2020, pp. 1–4.

[44] Ashish Patil et al. "Transmit Data Rate Control Based Decentralized Congestion Control Mechanism for VANETs". In: *Proceedings of the 2019 International Conference on Data Science and Communication*. 2019, pp. 1–5. DOI: 10.1109/IconDSC.2019.8816902.

[45] Noor Mohammed and Rasim Azeez Kadhim. "Congestion Control in VANETs based on Message Rate Adaptation by the Exponential Function". In: *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. IEEE. 2022, pp. 24–28.

[46] Chetan Belagal Math et al. "Data Rate based Congestion Control in V2V communication for traffic safety applications". In: *2015 IEEE Symposium on Communications and Vehicular Technology in the Benelux (SCVT)*. IEEE. 2015, pp. 1–6.

[47] Mahendrakumar Subramaniam et al. "A traffic density-based congestion control method for VANETs". In: *Wireless Communications and Mobile Computing* 2022 (2022).

[48] Jordan T. Willis, Arunita Jaekel, and Ikjot Saini. "Decentralized congestion control algorithm for vehicular networks using oscillating transmission power". In: *2017 Wireless Telecommunications Symposium (WTS)*. 2017, pp. 1–5. DOI: 10.1109/WTS.2017.7943531.

[49] Richard Gilles Engoulou et al. "VANET security surveys". In: *Computer Communications* 44 (2014), pp. 1–13.

[50] Tommaso Fedullo et al. "Adaptive lorawan transmission exploiting reinforcement learning: The industrial case". In: *2021 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4. 0&IoT)*. IEEE. 2021, pp. 671–676.

[51] Marco A Wiering and Martijn Van Otterlo. "Reinforcement learning". In: *Adaptation, learning, and optimization* 12.3 (2012), p. 729.

[52] David Schön Myers, Lisa Wallin, and Petter Wikström. "An introduction to Markov chains and their applications within finance". In: *MVE220 Financial Risk: Reading Project* 26 (2017).

[53] David Silver. "Lecture 2: Markov decision processes". In: *UCL. Retrieved from www0. cs. ucl. ac. uk/staff/d. silver/web/Teaching files/MDP. pdf* (2015).

[54] Christoph Sommer, Reinhard German, and Falko Dressler. "Bidirectionally coupled network and road traffic simulation for improved IVC analysis". In: *IEEE Transactions on mobile computing* 10.1 (2010), pp. 3–15.

[55] András Varga and Rudolf Hornig. "An overview of the OMNeT++ simulation environment". In: *1st International ICST Conference on Simulation Tools and Techniques for Communications, Networks and Systems*. 2010.

[56] Michael Behrisch et al. "SUMO–simulation of urban mobility: an overview". In: *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation*. ThinkMind. 2011.

VITA AUCTORIS

NAME:	Pooja Chandrasekharan
PLACE OF BIRTH:	Nagpur, Maharashtra, India
EDUCATION:	Bachelor of Technology in Computer Science, Government Model Engineering College, Kochi, Kerala, India, (2020)
	M.Sc. Computer Science, University of Windsor, Windsor, Ontario, Canada, (2023)