VEHICLE ASSISTED ENERGY-EFFICIENT DATA DISSEMINATION FRAMEWORK

A THESIS SUBMITTED TO AUCKLAND UNIVERSITY OF TECHNOLOGY
IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Salman Naseer
School of Engineering, Computer and Mathematical Sciences

May 2021

Copyright

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning.

Signature of candidate

Abstract

One of the main issues of maintaining a smart city (SC) is to handle a huge volume of data generated from various data sources and to transport them to the central control units for further processing and analysis. These data sources incur not only high running cost of core networks but also results in overburdening of such networks. The high energy utilization and carbon discharges from these core networks posing serious environmental issues. Therefore, in this thesis, to overcome the issues of high energy consumption (EC) and carbon emission (CE) of congested core networks, an energy-efficient vehicle assisted data dissemination (VADD) framework to tackle big data traffic is being proposed. In the investigation, different delay tolerant data forwarding case scenarios are designed and analysed by analytical modelling and data-driven approaches. To illustrate the efficiency of our proposed framework, we present mathematical models and data forwarding schemes for Auckland City (New Zealand) and Beijing City (China) case scenarios to transfer delay tolerant data from various data sources to a control centre.

One of the motivations for designing a sustainable data dissemination framework is to efficiently utilize the vehicular mobility to offload data from traditional networks. However, the main contributions of this thesis is highlighted below.

First, we explore the potential of VADD framework for data dissemination, by considering the annual average daily traffic of vehicles in Auckland city. Second, we analyse the network coverage and capacity of the proposed framework by considering

the displacement of vehicles in Beijing City.

Third, we propose a mathematical model for data transmission, numerically analyse the data forwarding strategies for the Auckland case by considering the vehicle mobility with Poisson distribution and with different parameters (e.g effective distance, deliver probability, and vehicle density). Fourth, we use the data-driven approach to transmit an updated software code to a massive number of smart sensors in a SC by our proposed framework. For this purpose an optimal code forwarding scheme (OCFS) is presented, to select the optimal vehicle and optimal roadside unit (RSU) for code transmission.

Finally, we develop an EC model for our proposed framework and use different optimization techniques to prove that our proposed framework is energy efficient, as compared to traditional networks in terms of EC and CE.

List of Publications

Journal Publications

- **Naseer, Salman**, William Liu, and Nurul I. Sarkar. "Energy-Efficient Massive Data Dissemination through Vehicle Mobility in Smart Cities." SI: Vehicular Sensor Networks: Applications, Advances and Challenges, MDPI: **Sensors** 19, no. 21 (2019): 4735. **Impact Factor: 3.275**

Book Chapters

- **Naseer, Salman**, William Liu, Nurul I. Sarkar, Peter Han Joo Chong, Edmund Lai, Maode Ma, Rangarao Venkatesha Prasad. "A sustainable marriage of telcos and transp in the era of big data: Are we ready?." In book: Smart Grid and Innovative Frontiers in Telecommunications , pp. 210-219. Springer,2018.

Conference Publications

- **Naseer, Salman**, William Liu, Nurul I. Sarkar, P. H. J. Chong, E. Lai, and R. Venkatesha Parsad, "A sustainable vehicular based energy efficient data dissemination approach," in proceedings of the 27^{th} International Telecommunication Networks and Applications Conference (ITNAC), Melbourne, Australia, November 2017.

- **Naseer, Salman**, William Liu, Nurul I. Sarkar, P. H. J. Chong, E. Lai, and R. Venkatesha Parsad, "A sustainable marriage of telcos and transp in the Era of big data: Are we ready?". in proceedings of the 3^{rd} EAI International conference SMARTGIFT, Auckland, New Zealand, April 23-24, 2020.

Articles under review

- **Naseer, Salman**, William Liu, and Nurul I. Sarkar, Muhammad, Shafiq " Analysing the potential of Intelligent Sensing System of Vehicles in smart cities." MDPI: **Sustainability**, SI: "Machine Learning, Data Mining, and IoT Applications in Smart and Sustainable Networks", **Impact Factor: 2.58**

- **Naseer, Salman**, W. Liu, N. I Sarkar, Muhammad, Shafiq. "Sustainable data dissemination in Smart Cities by destination trajectory prediction of taxis" In progress, targeted journal: Computer, Materials and Continua, **Impact Factor: 4.89**

Acknowledgements

I thank the Almighty for empowering me with the opportunity to complete this challenging assignment. I would like to express to my primary supervisor my sincere gratitude, Dr William Liu, philosopher, and mentor, for his patience, inspiration, and tremendous knowledge, for the continuous help of my PhD studies and related research. Throughout the study and writing of this thesis, his advice helped me. For my PhD studies, I could not have imagined getting a better mentor.

I would like to thank Associate Prof. Dr Nurul I Sarkar, in addition to my primary supervisor, for his insightful comments and motivation, and also for his difficult questioning, which inspired me to extend my study to look at different perspectives.

My grateful appreciation also goes to my alma mater University of the Punjab, Lahore, Pakistan, which provided me with fully-funded scholarships under the faculty development program to finance my three years of PhD study. Without it, I can't fully dedicate myself to the PhD research journey. Thankful to Auckland University of Technology (AUT), Auckland, New Zealand for providing me excellent supervision team and research opportunities.

My gratitude to the people who taught me the value of education, Babu Khan, Molvi Muhamad Amin, Mubashar Atiq Nutt, all teachers, Dada G Muhammad Din (Late), Nana G Noor Muhammad (Late), Chacha G Habbib ur Rahman (Late), Ch. Muhammad Saleem (Late), my uncle Muhammad Anwar (Late), Muhammad Ishaq (Late), Riaz ud Din (Late), Saeed Zafar (Late).

I wish to express my gratitude to my father Naseer ud Din Adil, my mother Parveen Akhter, and my siblings for their patience, love, and encouragement. My wonderful children Irha, Abdul Rahman, and Eshaal for coming into my life and giving me beautiful joy and moral support to become a stronger and fully-fledged father.

Last but not the least, my wife Mrs Iqra Zafar, for all love, caring, and understanding through this difficult time, and I thanked her for the wonderful support during my MPhil and PhD studies. Thanks for managing my family in my absence, specifically in this difficult situation of Covid-19. Thanks for supporting me spiritually throughout the writing of this thesis and my life in general.

Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

3G	Third Generation
3GPP	Third Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
AADT	Annual Average Daily Traffic
ADTN	Assisted Delay Tolerant Network
AP	Access Point
APN	Access Packet Name
AWS	Amazon Web Server
BI	Business Insider
BS	Base Station
CAPEX	Capital Expenditure
CBD	Central Business District
CC	Central Controller
CE	Carbon Emission
CFCR	Code Forwarding Coverage Rate
CFTD	Code Forwarding Time Duration
CN	Cellular Network
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CSI	Channel State Information
CTP	Cabernet Transport Protocol
CU	Control Unit
CUE	Cellular User Equipment
D2D	Device to Device
DC	Data Center
DM	Data Mule
DNN	Deep Neural Network
DPI	Deep Packet Inspection
DS	Data Source

DSL	Digital Subscriber Line
DSRC	Dedicated Short Range Communication
DTN	Delay Tolerant Network
EBRP	Energy-Efficient and Buffer-Aware Routing Protocol
EC	Energy Consumption
FTTH	Fiber To The Home
FTTN	Fibre To The Node
GPS	Global Positioning System
GSM	Global System for Mobile Communications
HAN	Home Area Network
HART	Highway Addressable Remote Transducer Protocol
HSPA	High Speed Packet Access
ICT	Information Communication Technology
IEA	International Energy Agency
IEEE	Institute of Electrical and Electronic Engineering
IoT	Internet of Thing
IP	Internet Protocol
ISM	Industrial Scientific and Medical Band
ITS	Intelligent Transport System
I-WLAN	Industrial Wireless LAN
LAN	Local Area Network
LTE	Long Term Evolution
MANET	Mobile Adhoc Network
MBS	Mobile Base Station
MIMO	Multiple Input Multiple Output
MN	Mobile Node
MoSoNet	Mobile Social Network
MU	Mobile User
MULE	Mobile Ubiquitous LAN Extensions
NCP	Network Control Protocol
NP	Non-Deterministic Polynomial
OCFS	Optimal Code Forwarding Scheme
OPEX	Operational Expense
OppNets	Opportunistic Networks
OSI	Open System Interconnection Model
P2P	Peer to Peer
PL	Path Loss
PLC	Power Line Communication
PRoPHET	Probabilistic Routing Protocol uses History of Encounters/Transitivity
QoS	Quality of Service
RNC	Radio Network Controller
RNN	Recurrent Neural Network
RSU	Roadside Unit
SaS	Statistical Analysis System

SC	Smart City
SCTP	Stream Control Transmission Protocol
SG	Smart Grid
SIM	Subscriber Identification Module
SIPTO	Standard Selected IP Traffic Offload
SNMP	Simple Network Management Protocol
SS	Smart Sensor
TCP	Transmission Control Protocol
TWh	Terawatt-hour
UE	User Equipment
V2I	Vehicle to Infrastructure
V2V	Vehicle to Vehicle
V2X	Vehicle to Everything
VADD	Vehicular Assisted Data Dissemination
VANET	Vehicular Adhoc Network
VDSL	Very high-speed Digital Subscriber Line
VDTN	Vehicular Delay Tolerant Network
VPN	Virtual Private Network
VUE	Vehicular User Equipment
WCDMA	Wideband Code-Division Multiple Access
Wi-Fi	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
CUE	Cellular User Equipment
D2D	Device to Device
DC	Data Center
DM	Data Mule
DNN	Deep Neural Network
DPI	Deep Packet Inspection

Chapter 1

Introduction

In this thesis, vehicle mobility and a huge volume of big data migration are addressed. The key idea is to utilize the numerous vehicular trips to reduce energy consumption and load on the congested networks (such as cellular networks and the internet). This proposed approach will provide a complimentary network with traditional wired and wireless network-based data transmission. By offloading the high volume of data, using the concept of the delay-tolerant network to the moving vehicles for the data delivery in a heterogeneous network environment. We are living in heterogeneous network environments, especially in city areas where we have 3/4G cellular network coverage, Wi-Fi, and other network connectivity. However, achieving quality of services (QoS), increase spectral efficiency with the low operational expense (OPEX) and energy consumption in heterogeneous networks can be challenging tasks. The strength of this proposed research in this thesis is the optimization of data aggregation from various data sources for smart cities.

In this chapter, the background and motivation behind this work are discussed in Section 1.1. The research questions are formulated in Section 1.2. The thesis main contributions are presented in Section 1.3. The research methodology adopted in this thesis is discussed in Section 1.4. Finally, the thesis structure is presented in Section

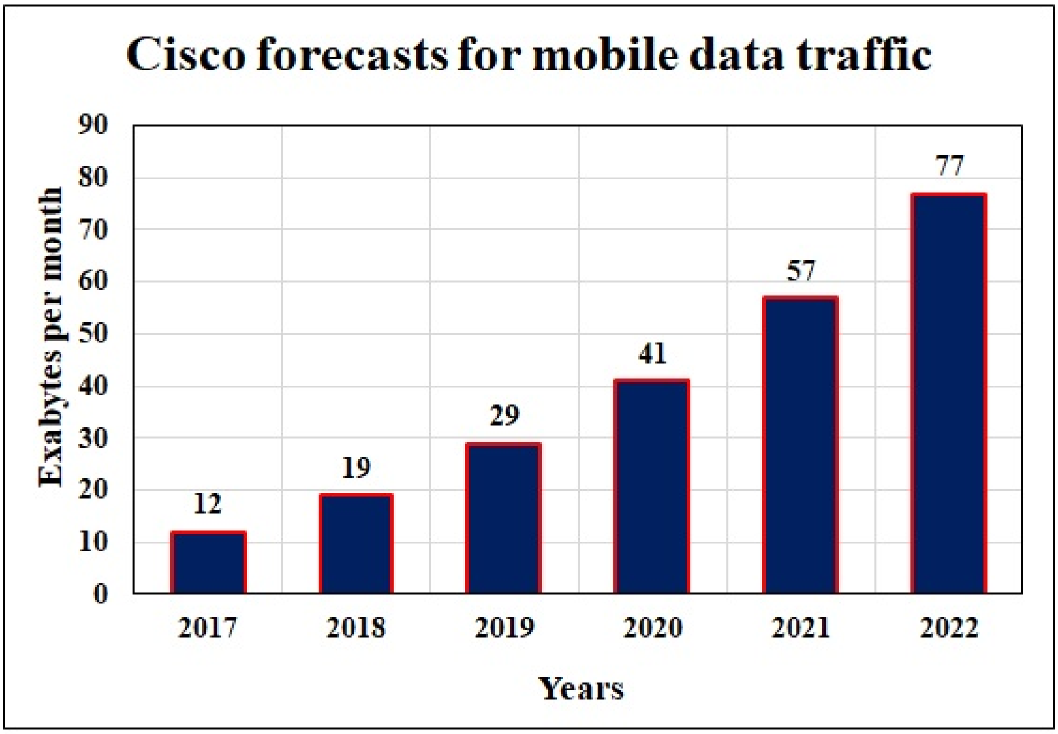


Figure 1.1: Cisco Forecasts of Mobile Data Traffic Exabyte/ Month by 2022.

1.5.

1.1 Background

The reduction of unnecessary energy consumption is becoming a major concern in wired and wireless networking because of the potential economic benefits and its expected environmental impact. These issues usually referred to as “green networking”, relate to embedding energy-awareness in the architecture and product design, device usage, and the protocols of the networks. The main concern is to maximize the performance of communication systems while producing sufficient information with minimum energy cost and controlling carbon emissions. To address the problem of energy consumption and network congestion, we propose a new data dissemination approach to offload data from the wired/wireless network to public vehicles on the road. The public

Table 1.1: Data rates comparisons

	3G	4G	5G
Development	2004-05	2006-10	2020
Bandwidth	2 mbps	200 mbps	>1 gbps
Latency	100-500 milliseconds	20-30 milliseconds	<10 milliseconds
Average Speed	144 kbps	25 mbps	200-400 mbps

transport vehicles are the moving agents or moving routers in this new communication paradigm, and they enable an alternative communication option, in addition to the current infrastructure based communications to reduce the energy consumption and carbon footprint. With the constantly increasing energy consumption in transmissions of explosive growing data, the more energy-efficient and sustainable solutions are necessary to be incorporated into the future wired and wireless networks. Figure 1.1 [4], shows Cisco's forecast in 2017 for the mobile traffic Exabyte per month till 2022, and the projected data traffic almost sevenfold in 5 years. The total Internet traffic has experienced dramatic growth in the past two decades and is still growing very fast.

Table 1.1 compares 3G, 4G, and 5G, where the data rates of 5G are five times faster than 4G, whereas the mobile data growth is sevenfold in 2021 [5]. Hence, data offloading is a promising solution to meet this exponential growth of mobile data.

Figures 1.3 and 1.4 [1] show that ICT consumption will be a big ratio (51%) of global energy consumption in 2030. The global electricity supply by all types of equipment is defined in figure 1.2 with the best, expected, and worst-case scenarios. The electricity demand of communication technology is increasing and will be 30,715 TWh of a worst-case scenario in 2030. In summary, figure 1.4 presents the percentage share of communication technology in global EC. Depending on different scenarios, EC in 2010 is 8% to 14%, in 2020 6% to 21% and in 2030 8% to 51%, respectively. Therefore, EC is a critical issue that needs to be addressed.

Figure 1.5 shows three digital laws, namely, Kryder's law, Moore law, and Neilson's

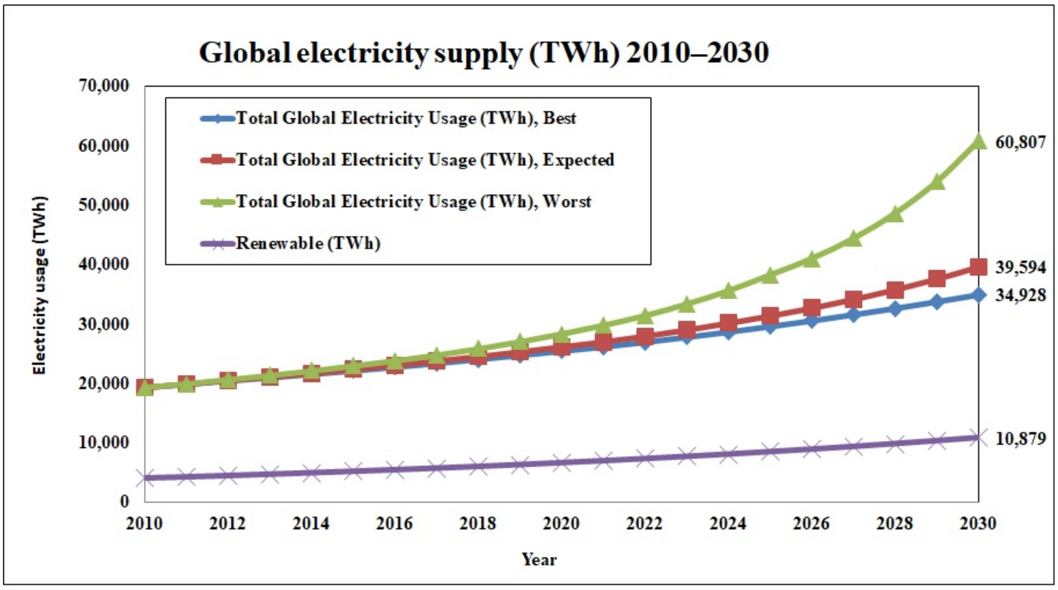


Figure 1.2: Global electricity supply 2010–2030 [1]

law [6] state new products come into the market with each passing year with new technology. The basic idea of Kryder's law is to double the storage capacity every 12 months. Moore's law is somewhat like Kryder, but works on the processing speed of chips, which get doubled every 18 months. Moving forward to Neilson's law, which estimates that bandwidth doubles every twenty-one months, so this last component of digital experience lags behind both storage and processing speed. These three laws clearly explain that, whatever new network technology comes into the market, the available data (in online storage) is never being fully accessed by the new network technology, and also the end-users. There will always be a gap between the available bandwidth and the available data/information storage online [6]. The need for big data will never be satisfied with the Internet. Whenever there is a conflict between modern technology and users' desires, human desires always win, they will keep advancing Internet technology. This non-stopping data need leads to more and more energy consumption, which further leads to carbon emission that harms the environment more adversely. Therefore, we want to introduce a new option of utilizing the existing

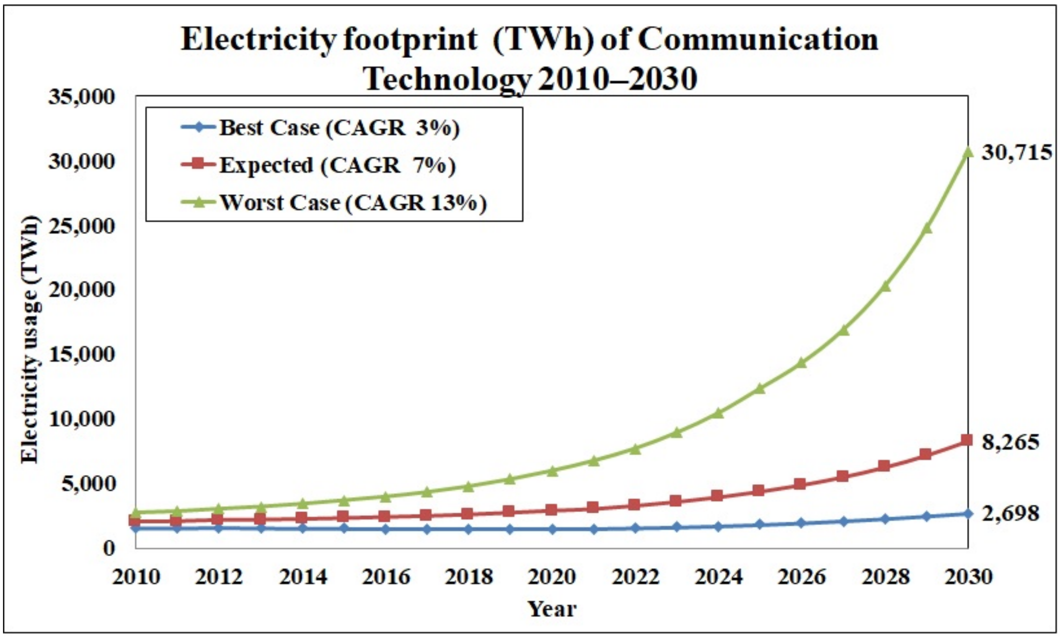


Figure 1.3: Electricity demand of communication technology 2010–2030 [1]

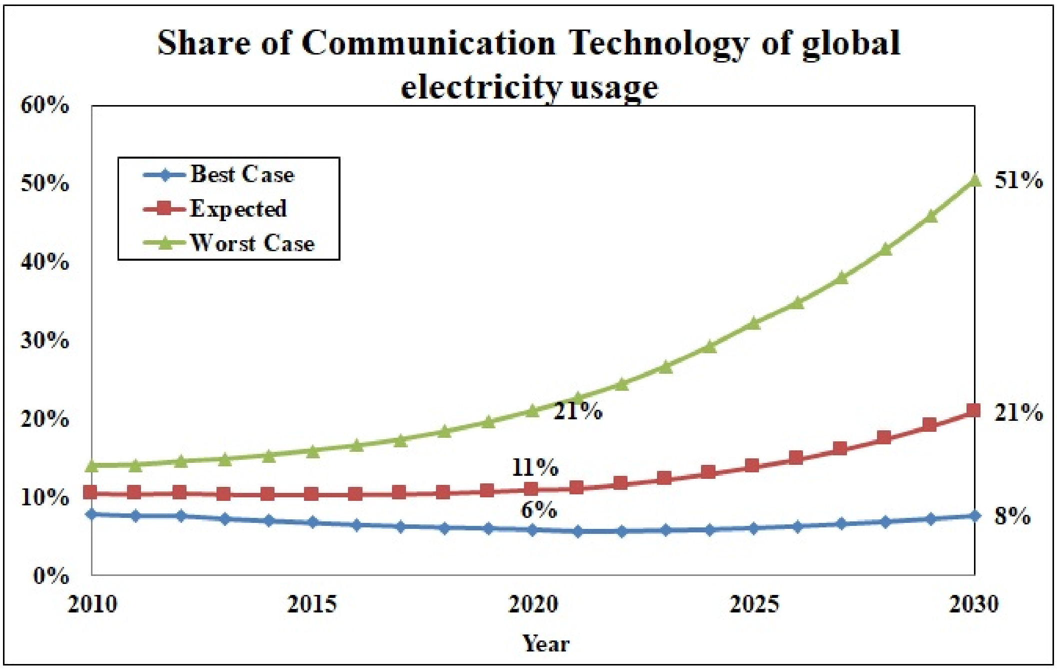


Figure 1.4: Share of ICT in global electricity usage 2010–2030 [1].

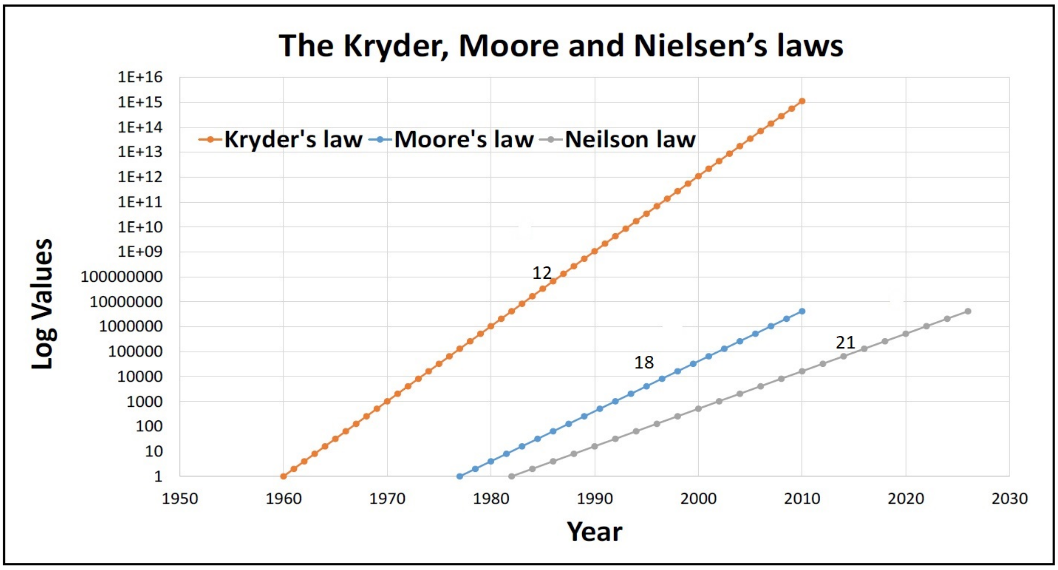


Figure 1.5: The Kryder, Moore, and Nielsen's laws.

vehicular networks to address this problem. Our solution can have a possible way-out through decoupling the energy consumption from the data size and its transmission distance; some time might need to sacrifice the delay of data delivery.

Many network researchers have devoted their efforts to improving the resource efficiency of the ICT systems, but all are still dedicating in the telecommunications and Internet domains themselves. Several attempts [7][8] in developing efficient, sustainable, and integrated (wired/wireless) networks have been made. The opportunistic network is one of the techniques to overcome this problem while disseminating data by store carry and the forward manner by connecting mobile devices. The key idea of the Opportunistic Network is that the mobility of nodes may help to deliver messages, whether it is connected or disconnected asynchronously in time [9].

Recently, there is fast-growing literature on energy efficiency and QoS. Nevertheless, our work will focus on both perspectives simultaneously in a heterogeneous network. The proposed research emphasis on data dissemination techniques, which will utilize existing vehicles on the road to make an energy-efficient network.

This research aims to propose a novel data dissemination framework suitable for heterogeneous networks, in which data will be offloaded from the congested networks towards the existing vehicular networks. To overcome the problem of core network congestion, energy consumption, and carbon emissions, an energy-efficient data dissemination framework is proposed in this thesis. The idea is to utilize the existing vehicular infrastructure to pick up and deliver big data from designated pick-up points and data centers in a smart city.

To recognize SC, a huge number of smart sensors (SSs) and different data sources (DSs) are needed to install throughout the city. The sensing requirements of different applications may change from time to time, which leads to software code updates in SSs. Many of the SSs in SC cannot be connected to the data center (DC) clouds directly because of the high cost of infrastructure network and power resources. This is the reason, managing software update code transmission to SSs from DCs in traditional ways is challenging. The same is the case for data collection from various data sources and transfer to data centers.

On the other hand, there exists a displacement of a huge number of mobile vehicles in SC. These vehicles travel on streets and roads almost everywhere and at any time. Many vehicles are equipped with storage, processing, and communication facilities. The motivation behind this work is that we can use the resources of vehicles, and a huge displacement of vehicles as mobile nodes to transmit a big amount of data.

1.2 Research Questions

The main objective of this thesis is to design an energy-efficient framework that will efficiently utilize the existing vehicles on the road for data transmission from various data sources and smart sensors to their designated data centers. To design this framework, the following questions are addressed in this thesis.

Research Question 1: What sustainable data dissemination framework in smart cities can be developed using the existing vehicular network?

Research Question 2: What can be done to quantify the performance gain of the proposed data dissemination framework?

Sub-questions:

2a. How to take advantage of opportunistic communication to improve the network performance in terms of energy efficiency and QoS and to optimize operational costs of the proposed network?

2b. How efficiently, our proposed framework can be utilized to reduce the energy consumption and carbon emissions by using different optimization techniques?

The proposed research focuses on data dissemination techniques, while offloading data from infrastructure to existing vehicle volume on the roads, to obtain a better quality of services (QoS), with less consumption of energy in a heterogeneous network. QoS can be defined in terms of delivery probabilities, delays, and cost. The literature survey concludes that people have been using the traditional way of communication for a long time to transmit the data from data sources to data centers of a smart city. The problem of the interconnection of data sources in SC should be solved by using some other type of networks, rather than the internet, Wi-Fi, 4G, 3G, etc. Under this topic, Vehicular Delay Tolerant Networks VDTN, by leveraging the existing vehicle's routine rides in the city, could be a wise solution to transport data from these data sources to control units.

Explore the potential of the system in terms of data volume transmission, end-to-end delay, network coverage and capacity, and how this proposed system is better than the traditional ways of communication in terms of energy consumption.

The thesis main contributions are highlighted next.

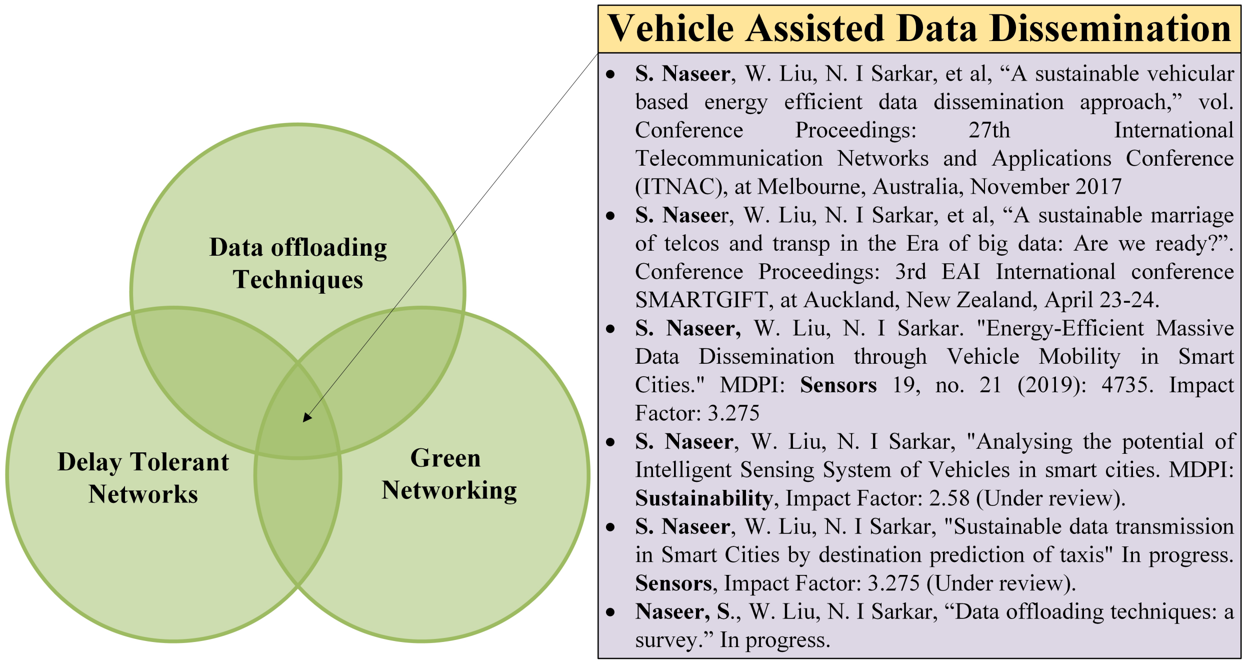


Figure 1.6: Research Gap

1.3 Contributions

We have widely explored the literature on green networks, delay tolerant networks, and data offloading techniques. A research gap has been identified, which is shown in figure 1.6. A need for an energy-efficient alternate network has been identified as a research gap, the network that can efficiently utilize the existing vehicles on the road to transfer big data from different data sources to data centers in smart cities. This finding has become the motivation for the following contributions in this thesis.

Contribution 1: We propose a framework called the Vehicle Assisted Data Dissemination (VADD) and explore its potential by using the daily vehicle count on the roads of Auckland city and real taxi traces of Beijing city. By inspiring the work of Baron et al. [10], Marincic et al. [11], and Wang et al. [12], we use the existing vehicles on the road for data transmission from various data sources to designated data clouds.

We develop a mathematical model and use the daily vehicle count on the roads of Auckland city for data dissemination and use Monte Carlo simulation to compare the end-to-end delay of our proposed network with the traditional network. The delay of our proposed network is affected by the vehicle count on the road, vehicle speed, and the distance between source and destination on the road.

Next, we use the data-driven approach and Grid Clustering to analyse the network coverage and capacity of our proposed VADD framework. For this purpose, we apply the various algorithms on the real taxi traces of Beijing city and find how vehicle density affects the performance of our proposed system. Finally, a case of code forwarding has been evaluated by using direct communication, and by using roadside units RSUs on the road. All of the findings show that our system outperforms with higher delay-tolerant intervals. This contribution is reported in Chapter 3.

Contribution 2: We developed a mathematical model for data transmission by our proposed framework VADD and proposed a data forwarding scheme. Validate its effectiveness by using the annual average daily traffic of Auckland city, New Zealand.

We updated the mathematical model and found the effective distance of a vehicle from a SS or RSU, and calculated the time duration that a vehicle remains in the wireless coverage of a device. We use the different densities and speeds of vehicles and calculate the degree of data offloading by using the Poisson arrival of vehicles. Finally, use the Monte Carlo simulation to compare the end-to-end delay of our proposed system with the traditional networks. This contribution is reported in Chapter 4.

Contribution 3: We propose the optimal code forwarding scheme (OCFS) in SC by using a data-driven approach. Validate its efficacy by using the displacement of vehicles in Beijing city, China.

We usually use smart sensors in smart cities. Their working requirements may change from time to time, and we need to update their software code for their task adjustment. To transfer this software update code to smart sensors from data centers, we propose a code transmission scheme. To verify the effectiveness of our proposed scheme we use the displacement of vehicles in Beijing City. Optimal vehicle selection and optimal RSU selection have been proposed in this scheme to solve the NP-hard multi-objective optimization problem of minimum set coverage. Finally, compare our results with the random and the fixed deployment of RSUs. The results show that our proposed scheme got the highest coverage and reduced the delay in this code transmission scheme. This contribution is reported in Chapter 4.

Contribution 4: Save the energy cost and minimize the carbon emission by using different optimization techniques.

Our proposed system is better than the traditional networks in terms of delay as discussed earlier. Here we discuss the betterment of VADD in terms of energy consumption and carbon emissions. We developed mathematical models for VADD and traditional networks to calculate the energy consumption for data transmission. A case scenario is evaluated by Monte Carlo simulation and our system consumes less energy to transfer big data between source and destination. In 2nd case, a problem is formulated by using a min-cost flow problem. Our solution selects the optimal routes to save energy costs and reduce carbon emissions. In 3rd case, we developed an algorithm to select the optimal transmission mode, and a problem is formulated by using traditional multi-commodity flow problems to minimize the energy cost. Our algorithm selects optimal mode according to different data demands to save energy costs and reduce carbon emissions. This contribution is reported in Chapter 5.

These four contributions address all the research questions outlined in Section 1.2. The research methodology adopted in this thesis is discussed next.

1.4 Research Methodology

The main objective of this study is to design a framework for sustainable data dissemination in smart cities by using the vehicular network. To achieve this research objective, analytical models, simulations, and data-driven approaches have been used to evaluate the performance of the proposed VADD framework.

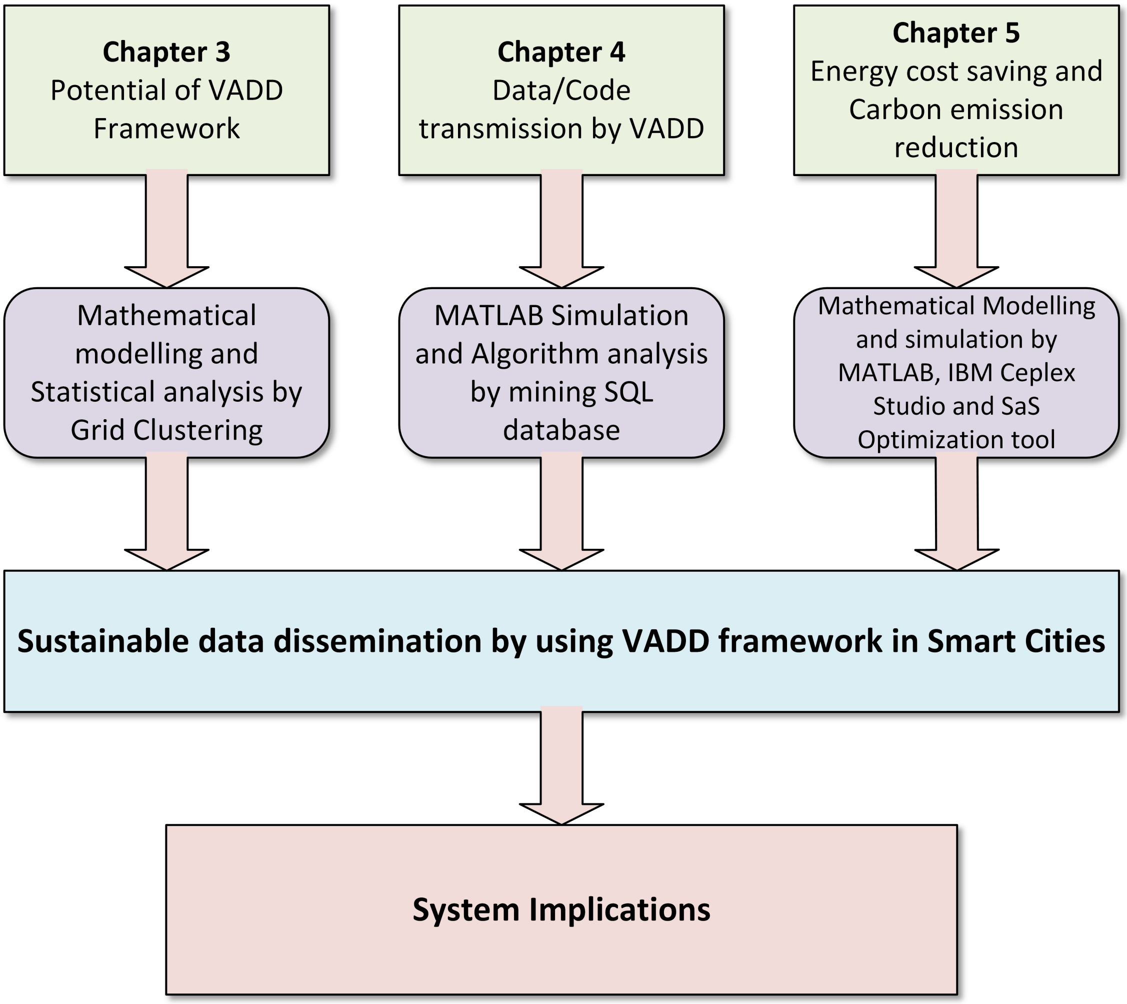


Figure 1.7: Block diagram of the adopted methodology.

The methodology approach adopted in this study is illustrated in Fig. 1.7. First, we explore the potential of Vehicle Assisted Data Dissemination (VADD) framework in terms of delay, network coverage, and capacity (Chapter 3). Second, we design data forwarding schemes for proposed VADD framework by considering the mobility of vehicles (Chapter 4). Third, a software update code forwarding scheme is proposed by considering the displacement of vehicles and their trajectories history (Chapter 4). Finally, we explore the energy cost saving and reduction in carbon emissions by using different optimization techniques (Chapter 5). These four contributions constitute a framework VADD, that can be used for big data transmission in smart cities. It can collect the data from millions of smart sensors and smart devices in a SC, and can transfer this data to designated data centers by saving huge infrastructure costs. Moreover, software code can also be disseminated to smart sensors by using this proposed framework. In addition to saving the infrastructure cost, it also saves energy and reduces the end-to-end delay for data transmission in SC.

Taking into account the need for this study, it is highly costly to evaluate the efficiency of the proposed system in the real-time environment. Thus, to test and evaluate the performance of the proposed VADD framework, this research adopts mathematical modelling, computer simulations and statistical analysis by using the MySql server database. Also, by evaluating different network scenarios, the simulation and analysis approaches enable the generalization of the results. For numerical analysis, we chose the Monte Carlo Simulations [13] by using MATLAB to evaluate the performance of the mathematical concepts of the designed algorithms (Chapter 3 to 5). Monte Carlo experiments are a broad class of computational algorithms that, to obtain numerical results, rely on repeated random sampling. They are also used in physical and mathematical problems and are most useful when other methods are hard or difficult to use.

Moreover, for data-driven approach and statistical analysis, we use the Grid clustering algorithm (Chapter 3 and 4). Divide the whole city into different sizes of cells according to the wireless coverage range of different wireless technologies. Extensive analysis is performed to calculate the network coverage and capacity. Test the performance of code forwarding technique OCFS. Calculate the vehicle density by using MATLAB function Scatter Plot coloured by Kernel Density Estimate to calculate the vehicle density to identify the most important locations to deploy RSUs.

Finally, we used IBM CPLEX Studio and SaS Optimization tools [14][15] (Chapter 5), to minimize the energy cost and reduce carbon emissions of our proposed VADD framework. IBM ILOG CPLEX Optimization Studio is a prescriptive analytics solution that enables rapid development and deployment of decision optimization models using mathematical and constraint programming. By using CPLEX studio we apply the min-cost flow problem for finding the optimal routes to reduce the energy consumption in multi-hop communication. Similarly, SaS is also a powerful array of optimization and simulation tools for identifying actions that will get the best results while operating within resource limitations and other relevant restrictions. We use the SaS optimization tool to transfer multiple data demands by using a multi-commodity flow problem for optimal network mode selection by the proposed VADD framework.

1.5 Thesis Structure

Figure 1.8 presents the overall framework and structure of this study. It shows three parts, namely introductory (Chapters 1 and 2), main contributions (Chapters 3 to 5), and Concluding part (Chapter 6). Before considering the data dissemination framework, it is necessary to establish a solid foundation for these areas. For the study, the introductory chapter (Chapters 1 and 2) provides the basis and background material. Chapter 2, provides detailed literature on delay-tolerant networks, green networks, and data offloading techniques.

The thesis main contributions are discussed in Chapters 3 to 5, which mainly focus on the promotion of the dissemination of data in Smart Cities through the vehicle network. In Chapter 3, the potential of the proposed VADD is explored in terms of big data dissemination, network coverage, and capacity. Whereas in Chapter 4, we propose data and code forwarding schemes for VADD. Benefits based on energy consumptions and carbon emissions by our proposed VADD framework are discussed in Chapter 5.

Finally, the thesis is concluded in Chapter 6, all the contributions and the findings in this thesis are discussed in detail. The future work and system implications of the proposed VADD framework are also discussed in this chapter.

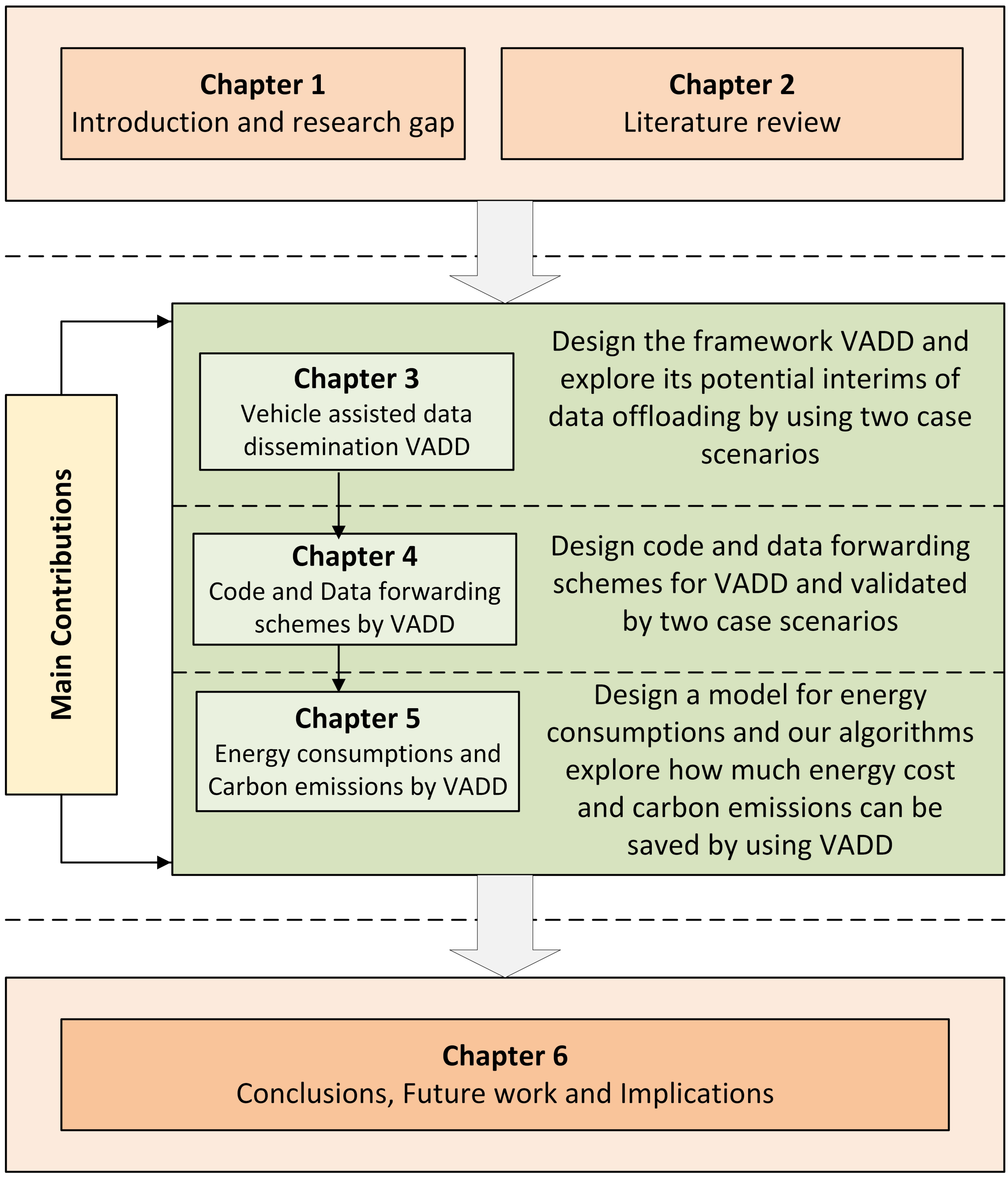


Figure 1.8: Thesis Structure

Chapter 2

Litrature Review

2.1 Introduction

In chapter 1, we briefly explained the research gap. In this chapter, detailed literature regarding green networking, delay-tolerant networking, and techniques of data offloading are reviewed, so to find the reason why the data offloading by using the vehicular network is required to fulfil the future demands of big data transmission. The literature about the green network is presented in Section 2.2. What are delay-tolerant networks and how they can be used to offload the data, are discussed in Section 2.3. In Section 2.4, we describe what is data offloading and different techniques of data offloading with respect to the energy spectrum band have been presented. Section 2.5, discusses the research methodology adopted in this thesis. Finally, this chapter is summarized in Section 2.6.

2.2 Green Networking

Green Networking is defined as the selection of energy-efficient networking technologies and better utilization of resources whenever possible. It has been estimated that

3% of the world's yearly electrical energy consumption and 2% of CO_2 emissions are caused by information and communication technology (ICT) infrastructure [16]. Moreover, it is estimated that ICT energy consumption [16] is rising at 15–20 per cent per year. Specifically, 57% of the energy consumption of the ICT business goes to users and network devices in mobile and remote networks [4], The rapid development of energy consumption by the user and network devices has created major issues [17]; many efforts are being done by the researcher for sustaining Quality of Services, throughput, adaptability Devices and their infrastructure are arranged in an execution way to get good QoS and better utilization of resources [18]. The trade-off between execution and energy utilization should be exactly exploited. The connection between energy and execution as indicated by [19]. Efforts are done for energy efficiency by Implementing Virtualization, server consolidation, upgrading older products to new energy-efficient products.

2.2.1 Green networks

There is a very broad area to validate the concept of Green Networking, which is defined in following terms :(i) Adaptive Link Rate, (ii) Interface proxying, (iii) Energy-aware infrastructures and (iv) Energy-aware applications.

1. Adaptive Link Rate We know that, in case of Ethernet link, energy consumption is based upon its utilization. The adaptive link rate has designed in such a way that there will be a low utilization of online resources and consumption will be less. This method can be used in two different ways: a) switching links to low energy consumption mode or sleeping mode. b) Rate switching, which means that reduction of line rate, according to low utilization.

2. Interface Proxying: The main aim of interface proxying is to delegate traffic from the main-board CPU to low power consumption of the Network Interface Card

[20]. Such data can be offloaded from one system to other to utilize the resources and less consumption of energy.

3. Energy-Aware Infrastructure: This section categorizes into energy-aware architecture and routing schemes [21]. In Energy-aware architecture, a new approach suggests amending structure over existing infrastructure (e.g. Grid 5000 is a good example used by many researchers). Energy-Aware routing aims to cumulate traffic over other network devices and allows other connected devices to be in OFF mode for re-routing strategies.

4. Energy-Aware Applications: Applications are divided into 2 categories a) User-level Application b) Kernel-level network stack. In user-level applications [22], a strategy is used to test peers whose status is not known. Peers present their energy state and therefore energy efficient. The peer does not wake up idle peers if not necessary. In a kernel-level network, many applications get shared by others. They store data and does not send immediately, which helps to reduce energy consumption.

All over the world, various techniques are used for green network terminology. The Japanese government also demonstrates additional interest by introducing Green IT for power efficiency. Different applications range from lower energy usage, low modern observing of large-scale energy compelled environmental monitoring. In network topology, energy is consumed at scanning, transmission, and data acquisition. All current approaches to green networking promise high energy savings at the cost of reducing network performance. Since network performance indexes have been so far the only yardstick for operators and manufacturers, the proposal of such trade-off may appear odd and hard to accept. Thus, the current challenge of a large part of researchers involved in the area of the green network is to find specific solutions/mechanisms, working with a negligible impact on the network level performance.

2.2.2 Green Technology and Next-Generation Networks

To better discuss the main issues in each field, we describe the most representative research contributions in different sub-sections, each related to a different network type or part which is given

1. Wired Access network: Wired access networks are a very critical scenario for reducing the overall carbon footprint of telecommunication infrastructures since they represent a large share of network energy requirements. Current green networking technologies [23] in this area include two approaches: a) Re-engineering Approach b) Dynamic Adoption Approach, the re-engineering of current technologies means by replacing copper-based technologies with fiber ones to the maximum possible extent and as well as the design of power scaling mechanisms. Based on a model of a typical operator network, [24] compared the power consumption of different broadband access technologies and architectures, especially DSL, FTTN+ VDSL and FTTH (Fiber To The Home). Even though VDSL power management improves performance, the authors stated that there is still a clear advantage of FTTH for energy efficiency. In Dynamic adoption approaches, they consider the trade-off between dynamic spectrum management and the rate maximization of digital subscriber lines according to a given transmit power constraint of modems

2. Wireless Cellular Network: Similarly, to the wired access network, also research approaches in wireless/cellular network infrastructures are mainly devoted to re-engineering current device platforms (e.g., access points, base stations, etc.), and to include support for the dynamic adaptation of network resources (and then of power requirements) to the actual traffic loads. In re-engineering approaches, specific research activities have been performed to reduce the energy wastes of fixed infrastructures of cellular networks (e.g., Base stations [25]) and WLAN access points [26], with more detail in [27]). Moreover, dynamic approaches include interoperability with legacy protocols and devices Another factor to consider is that wireless networks mostly operate at significantly lower data rates than wired networks [28]. Correspondingly, the need for buffer capacity can be lower and acceptable latencies can be higher [29].

3. Network Router and Switches: As far as an energy-efficient router and switch architectures are concerned, re-engineering approaches mainly focus on how reducing the complexity of internal architectures and using more efficient hardware technologies, Power-aware network investigates potential saving in design and routing [30, 31]. The author use core and edge routers to investigate and ultimately results show that power consumption can vary as much as an order of magnitude and there are substantial opportunities to control it. Various dynamic approaches have introduced to reduce the carbon footprint of network equipment by following two main approaches, namely power-efficient design and power-saving design, respectively [32, 33, 34, 35].

4. Network Topology Control:

Power saving mechanisms based on network and topology control is currently originated on the extension of traffic engineering and routing criteria to use [36]. Researchers propose to reduce the network energy requirements by switching off unused links and nodes. Their main concern is to move traffic flows among network nodes to find the minimum number of network resources (i.e., links and nodes), guaranteeing the best trade-off between end-to-end network performance and overall power consumption [37]. Many dynamic adoption approaches have been proposed to switch off some portions of the core network, while still guaranteeing full connectivity and maximum link utilization.

5. Green Ethernet: Green Ethernet is a technology at the link layer, and it is highly widespread through various standard versions, at both customer and network sides [38, 39] . Ethernet interfaces diffusion have attracted the attention of several researchers, since even a marginal reduction in the power requirements of Ethernet interfaces may potentially lead to saving a huge amount of energy [40]. Many re-engineering and dynamic adoption approaches are working on energy-efficient solutions.

6. SNMP Green Extensions: The power state includes all supported power management capabilities, current settings, total and currently active, inactive, and sleep times, and statistics on wakeup and sleep events. With knowledge of the power state of network devices, a network manager could remotely audit the energy consumption of IT equipment and make changes to power management settings [41].

7. End-users and Applications: Many End user's applications have been proposed in terms of green networking. Christensen et al. introduced the NCP to maintain the network presence of sleeping hosts. Recently, they evolved the NCP architecture to handle the P2P traffic of common file-sharing applications [22, 42, 43]. In more detail, while focuses on greening the behaviour of Bit-torrent clients, the authors proposed a prototype Gnutella like P2P power management [44]. Authors in [38, 45], demonstrated that a complex trade-off exists between balancing the complexity of the proxy, the amount of energy saved and the sophistication of idle-time functionality.

2.2.3 Optimization Technique

Various techniques have been followed in the wireless sensor network for optimization [46]. Research on energy-efficient networking has been going on for several years . Different energy-related issues have been altogether examined covering an extensive variety of subjects (routing, cross-layer plans, scope conventions, range portion, media access control conventions, resource allocation, scheduling, and so forth. Some of the techniques are discussed below.

1. Energy Management in network connecting devices- routers and switches: Over the last decade, various kinds of router and switches are in use for a high-speed network. These routers and switches are designed with increasing capacity and performance, which also increased power usage. In [47], power density issues were considered and started focusing on designing aspects. Energy management depends upon: a) exploiting idle state, by putting switch/router on/off mode when no use b) clocking hardware at low speed by just trade-off between power and performance [32, 48].

2. Energy-Efficient Communication Protocol and networking Approach: There are various publications on the energy-efficient protocol for the wireless Adhoc network, wireless sensor network, multihop wireless networks and delay-tolerant network in the literature. Many researchers work to improve the performance of wired/wireless, and heterogeneous networks [49, 50, 51]. Energy consumption associated with the execution of transport protocol is one of an important issue to explore. They compared many protocols with new amendments to make it energy efficient [52, 53]. In [54], an energy-efficient TCP quick time-out technique was proposed for wireless local area networks improve the efficiency of energy consumption by five times better than before. In case of Delay Tolerant Communication protocol, Energy Efficient Epidemic Protocol also considered performance and energy efficiency while communication. With the growth of the Internet (including wired networks) and the emergence of wireless networks such as wireless sensor networks, multi-hop networks, mesh networks, the various energy-related publication has been released. All of them works on improving the commodity-based network.

3. Network Selection and Hand-Off: Hand-Off Mechanism is to be used one of the strategies used for green networking. This strategy keeps connection active while moving from one network to another. Homogeneous network hand-off is to be called horizontal hand-off and on the contrary vertical hand-off, the term is used in the heterogeneous network. This strategy consumes lots of energy and therefore network selection is being done using interfaces rather than keeping it alive all the time. Many researchers worked for making energy-efficient network selection in homogeneous and heterogeneous networks. Researchers are still working on this aspect to make energy efficiency with good performance factors [55, 56, 57, 58].

4. Reducing Energy Consumption in Base Station/ Cellular Network: It is exposed in [59], that carbon emission is due to energy consumed at the base station and mobile terminals. The trade-off between base station density and energy consumption is being analyzed by many researchers [28, 60, 61]. These papers categorized how much energy is consumed at a base station, which further gives direction to move towards wireless communication for energy efficiency. Power consumption at the base station is being predictable by all its constituent parts and core radio devices. While considering energy metrics, improvements in energy efficiency proceed in base station hardware, which is still struggling. Consequently, the proposed research determined to use fewer base stations and moving towards mobile nodes/DTN approach to forwarding data to the final destination. The important technique for reducing energy consumption in the base station is to improve the hardware energy efficiency of transceiver and power amplifier. So Far, a different type of linearization method has been used [29]. While considering the software feature of base stations, one approach to energy efficiency is by keeping mobile network off during less traffic time. Designing of a base station with an aim of power-efficient using IP and Radio on Fiber technologies were proposed [18]. Several approaches such as traffic requirements, channel information and load balancing are proposed by Cellular Network. The ICT sector (including laptops, personal computers, telecommunication networks, data centres and computer networks, mobile devices) keeps on developing on a worldwide scale and the force will increment in coming years as more ICT equipments are being sent and utilized as a part of different segments of society.

In 2015 [62], authors introduced an energy-efficient technique named as EBRP: An energy-efficient and buffer-aware routing protocol. They put their attention on an adaptive clustering algorithm which can achieve the optimal community division and extract the community attributes of nodes. They consider sociality of the node and dynamic changes of the attributes of nodes while selecting the best relay nodes to forward the messages. These attributes define the energy and buffer storage for messages. They evaluated the performance of their strategy with three real-life mobility traces

A study in [63], has classified various technologies to truncate energy consumption in wireless communication. New invention about smart-phones and their features incorporate new architecture, new network management, and forwarding techniques. Therefore, it is important to make a shift from traditional spectral efficiency towards Energy Efficient designing and routing strategies. The future communication network is to be designed to offload traffic to other network devices while using comparative fewer resources. Many researchers worked on energy efficiency techniques in wireless or wired infrastructure with changes in hardware requirements or routing strategies. This work proposes a new solution with guaranteed delivery and using less consumption of energy in a heterogeneous network. So far, we introduced Delay Tolerant Network with a tendency to store, carry and forward. We use DTN, where wired and the wireless network can't approach or if there are a loose time of requirements. The DTNs and it's routing protocols are discussed next.

2.3 Delay Tolerant Network

Delay tolerant networks (DTN) is case type of Opportunistic Networks (OppNets), which is developing as a critical method for correspondence between the remote heterogeneous devices because of their convenience, adaptability of operation and calculation [64]. All of the devices used for communication have increased their storage capacities, processing, and their communication technologies, Moreover, this communication capacity has enabled a novel class of applications extending from mobile networks to Vehicular Networks [65]. These applications work on the principle of forwarding messages one device to another and known as Opportunistic Network. DTNs [66] are inadequate versatile networks, where an end-to-end path may not exist. The main principle is to store, carry and forward in the path towards the final destination. DTNs are often susceptible to the problem of high delay. Many applications are merging on the concept of DTN, for example, crisis management, wildlife monitoring, transport engineering, packet-switched network and disaster management etc. Mostly the traditional ad-hoc networks and internet routing protocols don't function admirably in the event of OppNets where there is never a settled and dependable way amongst sender and recipient because of node's mobility, network allotments, node failure and other features which is reliable embedded in a dynamic wireless environment [67]. The communication in DTN depends on the contact opportunity between the nodes that emerge because of their versatility, and the store-convey and forward strategies. The message is gone to the intermediate nodes, which takes it closer and nearer to the destination lastly to the destination itself. On the off chance that the intermediate node does not locate an appropriate location which guarantees to take the message nearer to the destination, it keeps the duplicate message in its buffer until it finds a reasonable node to pass messages or finds the destination node itself [68]. Such type of networks leads to the long delay of messages. Recently, many applications are emerging with the same concept of DTN. This concept is being advanced from Mobile Ad-hoc Network (MANET) but the basic difference between both is MANET nodes use TCP/IP for communication and DTN uses application-layer bundle (Transport and Network Layer) and uses the concept of store-carry and forward with other peers [66, 69]. It may take a long time to find an appropriate relay node, so need to use message carriers to store messages in the buffer for a long time. Many Opportunistic Routing Algorithm has defined in terms of forwarding messages to the destination [68].

2.3.1 Routing protocols in DTN

1. Epidemic Routing: It is a forwarding protocol which works on the concept of spreading a message like a disease [70]. A node starts spreading the message to all its neighbouring nodes and keeps on repetition until the message reaches on destination. It consumes huge resources and provides optimal routing performance in terms of delivery ratio while minimizing delay. In this algorithm, the author considered a case of 45 nodes in the random way-point model and their traffic patterns. They place an upper bound on message hop count and buffer space according to a single node. In such a way, increment bounds on all these parameters and applications just increase the probability that the message will be successfully delivered whether they have to consume more resources. Though, random pair-wise message dissemination among all nodes give ensure of message delivery. This routing protocol involves two steps:

a. Exchange of summary vectors: Every node contains an index of messages stored in its buffer, which include unique message-ids.

b. Exchange of messages: One node computes set of messages carried by other nodes. Same is being done by other nodes as well. If such computed node is not empty, it requests its peer to transmit a message with its corresponding ID's.

While considering Epidemic as a benchmark, many researchers have given their hypothesis as a new invention. Zhang et al. [71], worked on performance modelling of Epidemic Router with many other schemes by using a model. Due to the scattering feature of Epidemic Routing protocol, it is being used in medical science. Matsuda and Taking [72], used this protocol wherein a relay node duplicate message to another hand-off node with likelihood, and q is the probability of the source node to replicate a message of the relay node. With the variation of , the behaviour of Epidemic routing protocol changes subsequently [73].

2. Direct Delivery:

In this protocol source node does not pass the message to the satisfactory nodes, yet rather keeps it with it until it comes directly in contact with the destination node. This plan is straightforward, simple to send and, uses least data transmission and system assets for message exchange following every message is transmitted at most once to the destination node. Then again, there may be a long defers for message conveyance either for the situation the source never meets the destination or there may not be a quick contact between the source and the destination, however, a path exists through the centre of the road nodes for the message passing. On the off chance that the source hub comes up short then, the message will be lost as there is stand out duplicate accessible in the system. In this plan's likelihood of conveyance is poor along these lines, it is not best for the circumstance where high conveyance likelihood is needed [74].

3. Spray and Wait: This algorithm is divided into two parts a) Spray Phase: where many numbers of copies are generated from the source and spread towards distinct nodes and they use direct transmission towards the destination. B) Wait Phase: If the destination does not get discovered, each of distinct nodes uses the concept of direct transmission to reach to the final destination. Initially, it uses the concept if Epidemic routing, but much better than Epidemic in terms of generating low contention specifically when high traffic load flows and highly scalable despite making any change in network size and node density. This strategy outperforms all existing algorithms in terms of transmission and delivery delays and finally got an optimal solution which is very scalable [75].

4. Single Copy Algorithm: Spyropoulous et al. [75], defined single-copy based protocols for an Opportunistic network where nodes forward a message only it knows the destination and transmits a single copy of the message once. Here, utility function defines the usability of nodes according to the number of meetings between all numbers of nodes and with their last encountered information. As soon as, high utility node is revealed, they use the approach of seeking and focus and finally reaches the destination.

Nodes take local forwarding decision based upon their connectivity and prediction of future connectivity details. A combination of simple random policy and utility-based policy followed to reach the final destination.

5. PRoPHET: PRoPHET (Probabilistic Routing Protocol uses History of Encounters and Transitivity) uses non-random mobility and contact pattern with a probabilistic metric called delivery probability. It considers the history of previous contacts which is not to be considered as random and quite possible to identify mobility patterns. This model is based upon the prediction of the probability of each node for all known destinations. The predictability of delivery between A and B is denoted as $P_{A,B}$ and its range is defined as "$0 \leq P_{A,B} \leq 1$" [76].

2.3.2 From DTN to VDTN

Vehicular Ad-hoc Networks (VANETs) have been an important research topic for many years [77]. It is an extension of Mobile Ad hoc Networks (MANETs) to vehicle systems. VANET has many special characteristics e.g. Predictable mobility, High mobility, Variable topology in time and place, partitioned network etc. VDTNs extend VANETs with DTN capabilities to support long disruptions in network connectivity. VDTN has some unique characteristics borrowed from VANET and DTN. In VDTN, some nodes (vehicles) store and carry network data while waiting for opportunities to forward it. A Vehicle which is also used to carry the data towards the destination, which is unlike the other non-VDTN approaches for VANETs is to be called MULEs (Mobile Ubiquitous LAN Extensions). Mules are mobile nodes that pick data in one place and drop it in another place. Mules are added to the network to extend coverage and/or the number of communication opportunities. Stationary relay nodes are fixed devices with store-and-forward capabilities that are located at road intersections. The proposed research is using the vehicle as a data carrier to transport it from one place to another in terms of energy efficiency. To implement our idea, data offloading can help to better understand the usage of vehicles in the transportation of data.

Grociz et al. [78] used vehicle carrier for big data transfer. Due to increased internet traffic in recent years, they decided to exploit the existing worldwide road infrastructure as an offloading channel to help the legacy Internet assuage its traffic. Motivated by the need for technical flexibility and cost-effective scalability, large companies, organizations, universities, and governmental agencies constantly move their data and applications within and between data centres to balance workloads, handle replication, and consolidate resources. They put their efforts in using conventional vehicles for offloading scheme, cost reduction, and capacity improvement. Assisted-DTN architectures involve various data carriers or forwarders ranging from buses to aeroplanes to compensate for the lack of continuous connectivity by bridging disconnected nodes. They justified that vehicle help to reduce traffic and cost with less delay.

2.3.3 Vehicular Transport Network

Vehicular DTN is based on the concept of DTN. It handles non-real-time applications at low cost and unreliable conditions. Vehicles can be used as a data carrier between terminal nodes either in rural areas or in emergency scenarios. VDTN follows the Open System Interconnection reference model. This protocol works on two layers of the OSI model: Physical layer and data link layer for functioning and data management. Data plane aggregates data into bundles and transports it to the destination. At the source node, Bundle aggregation layer aggregates packets and then destination address gets convert into VDTN terminal node address. Many techniques can be used for data assembly. Therefore, this strategy can help in the vehicle to vehicle communication [79].

Kashihara et al. [80], proposed a data offloading technique to offload data, particularly to public transport vehicles by using high-speed short-range wireless communication. The motivation was to reduce traffic congestion because of the high volume data. Firstly, the user checks all the requirements that are there any offloading spot, then fast and stable data transport. In uploading data, the user drops a large packet to data port at any offloading spot or directly to the storage system of any bus. This drop decision is also depending upon delay time. If the delay time is acceptable then the only user drops it off otherwise not. After uploading, the bus carries data to the internet through a high-speed connection. In downloading, a user request to the server to deliver particular data to get it at any bus stop. If it is available, they will be delivered immediately otherwise the vehicle will carry it from the terminal and deliver it to the data port.

Michele et al. [81], explored BUSNET algorithm that achieves effective routing in a bus environment. It considers routing at a bus-line-level instead of bus level. They do not consider paths among buses, but paths among bus lines. The nodes are no more individual buses, but buses on a specific bus line. They use route contact oracle where every bus is a mobile node that can store, carry and forward packets of information. A node can transmit a packet to a neighbour node by establishing an ad hoc connection if it is in the communication range; nodes also have a buffer to store intermediate packets, until the correct destination is reachable.

ALARMS [82], is one of message scheduling approach which uses message ferries to forward messages to achieve good QoS. Ferry nodes move along pre-defined routes to exchange message with gateways of each node of the route and also pass to the gateway nodes look-ahead routing information about when it will arrive at each gateway node on the route in the next two rounds. In such a way, gateway node estimates the delivery delay of each message and schedule the message to be delivered by some other ferry to deliver it soon. They go for message prioritization and utilizes a scheduled ferry to deliver it.

The proposed research will introduce a vehicle as a part of communication if there are a loose time of requirements. Moreover, there are many rural ways where the wired/wireless network can't approach. Then, these vehicles can help to deliver messages with some delay. However, we can't ignore our traditional way of combination. In some cases, we can even lose our messages with DTN approach and we don't want to compromise with our quality. Therefore, we will use the heterogeneous network to make communication more effective and energy-efficient.

2.4 Mobile Data Offloading

The use of complementary network technologies to distribute data initially aimed at cellular networks is called mobile data offloading. Offloading decreases the amount of data transported on the cellular bands, releasing the bandwidth resources for other users in congested networks. It is often used in circumstances where local cell reception can be low, enabling the user with better coverage to connect via wired networks.

We divide the data offloading techniques into three approaches based on the use of frequency spectrum band resources. Some of the approaches use the licenced frequency and pay cost while some other use licenced free frequency. On the other hand, some techniques get the advantages of both bands licenced and unlicensed frequency spectrum are called hybrid techniques. These categories are explained as follow.

2.4.1 Data Offloading by Unlicensed Frequency Band

Industrial Scientific and Medical ISM band also called unlicensed or licence free frequency in the energy spectrum band as shown in figure 2.1. Any equipment having a suitable transmitter and receiver can use this band without having any reservation in this frequency range. Nevertheless, to use this frequency range a device must have to

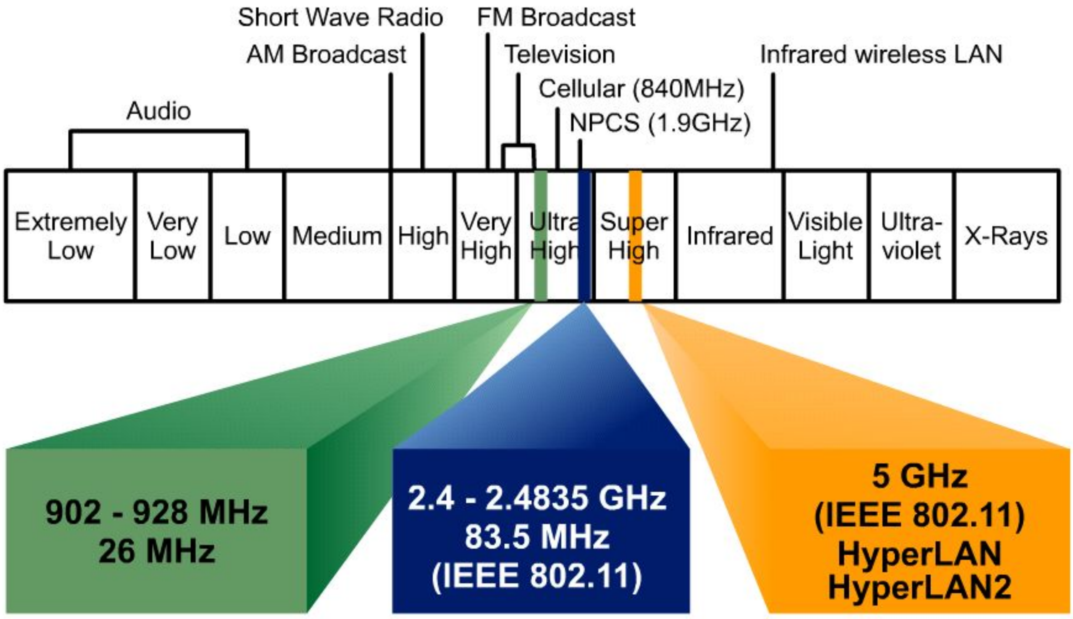


Figure 2.1: Energy Spectrum Band [2].

follow some certain rules and regulations like restrictions on maximum signal power for data transmission [83]. IEEE 802.15 standards like Wireless HART, Bluetooth, and ZigBee uses a frequency range of 2.4 GHz also some WLAN standards IEEE 802.11 family uses 2.4 and 5.8 GHz frequency range of the ISM band.

During the peak hours when a lot of users are using the cellular networks, we use these bands to offload data of congested and costly links of cellular networks to reduce the load on these links. Recently, the integration of Wifi and Cellular networks got the attention of researchers for cellular mobile data offloading.

1. WiFi Offloading

Wireless Fidelity is abbreviated as WiFi, IEEE 802.11 is a WiFi standard provides wireless connectivity. The main use of this technology is to access wireless broadband services in an indoor environment. WiFi delivers higher throughput with limited mobility and coverage as compared to other traditional wireless communication technologies like Long Term Evolution, Universal Mobile Telecommunications services (UMTS), WiMax, and High Speed Packet Access (HSPA). Currently, WiFi is gaining popularity for its paradigm shift towards its city-wide/outdoor ubiquity availability

Several reasons show that WiFi is an efficient solution for mobile data offloading. Firstly, we can see that WiFi is available in many indoor areas like residential buildings, train stations, shopping malls, airports and different private and public locations. Secondly, it seems feasible the deployment of WiFi at an outdoor area by the operators. Also, because of advancements in technology, most of the end-users devices like laptop, smartphones and tablets contain WiFi interface. Last but not the least, energy efficiency, less cost and sharing the load from mobile networks are some of the key benefits of WiFi offloading.

There are different ways to categorize WiFi offloading. Delayed and non-delayed WiFi offloading, are explained in [84], based on data traffic nature and the delay requirements that can be tolerated by the users in the delivery of data. Whenever a device has some delay constraints on data delivery and has priority to choose WiFi interface rather than the cellular interface, it offloads its data by using non-delayed or on-the-spot offloading. In this strategy, data is offloaded to WiFi network until the user remains under the coverage of WiFi area and when the user crosses this coverage then data flow continues on cellular links to meet the delay requirements. On the other hand, in delayed data offloading, when a user can tolerate on some delays in data delivery, then data transmission remain stopped while user's device comes out of the signal range of WiFi network and resume again when it re-enters in the signal coverage area of WiFi. Significance of this shallowness can be evident, when there are long deadline delay requirements and delayed offloading is used to reduce the burden on the cellular links[85].

The experimental results in [86], showed that 65% of the cellular traffic can be offloaded by using non-delayed offloading and it can be improved up to 82.1% if we use the delayed offloading strategy. For both operator and subscriber, delayed offloading is more desirable in an economical point of view. Results in [86], indicate that it rises the revenue of operator from 21% to 152% and from the user point of view it reduces cost and energy from 73% to 319% in comparison of different scenarios of non-offloading.

For operators, depending upon the integration of cellular and WiFi Network there are three main categories unmanaged data offloading, managed data offloading and integrated data offloading [85].

a. Unmanaged data offloading or Network bypass: In this case whenever there is WiFi coverage area the user secession is transferred from cellular network to WiFi network and it bypasses core cellular network for data-oriented services. Whereas voice-oriented services remain to be continued on the cellular core network. It is a simple and attractive approach because we don't need to apply any modification to network infrastructure, but there exist some drawbacks in this strategy. First, when a user is in the WiFi coverage area then the operator loses its control over the user. Second, the subscribed contents (Corporate VPN, Blackberry, ringtones etc.), can't be delivered by the network operator and it leads to revenue loss for the network operator. Regardless of its drawbacks, it can be adopted as candidate offloading approach because of its easy deployment. This technique is also useful from the user point of view because of user control over data delivery. It is very similar to switching on the user's device WiFi interface whenever it is in a WiFi coverage area. The operator can implement this offloading scheme very simply by installing an application program on the user's device that automatically switches on the Wi-Fi interface whenever it detects the WiFi coverage area.

b. Managed data offloading: To get some control over the subscriber, different operator uses the managed data offloading approach. In this approach, a secession aware gateway is placed intelligently by the operator to monitor the subscriber's secession under the coverage of WiFi on its way to the internet. Full integration of WiFi providers and the cellular network is not required in this approach and the operator can get control over the end-user, again it can't manage the delivery of subscribed contents.

c. Integrated data offloading: To get full control over end-users and to transmit subscribed contents under the coverage of WiFi network, the operator uses integrated data offloading approach. In this technique, data flow can be managed by a bridge between these two networks and it can be accomplished by integrating WiFi and cellular networks. There are two approaches to integrate WiFi and cellular networks tight coupling and loose coupling. In loose coupling, no cooperation is required between these networks and both are independent in this integration. In this strategy, both networks are connected indirectly by some other IP network like the internet. Service is provided by the concept of roaming between these two networks. Whereas in tightly coupled architecture both networks share the common core and WiFi network is also managed by the cellular network operator. Most of the network function like billing, resource management and vertical hand-overs are managed and controlled centrally. For a cellular network operator, 3GPP I-WLAN standard describes the offloading scenario and basic principles of integrating both networks [87]. It defines a solution on how to transfer data from the cellular core network and end-device with the help of Wi-Fi access network. The main idea is, establish a tunnel between a dedicated I-WLAN server and end-device in the core network, to get access over public internet or operator's subscribed contents.

3. Vehicular Wi-Fi Data Offloading

Moving Vehicles on the road and roadside APs of WiFi network can be used as a practical solution for data cellular data offloading, called vehicular Wi-Fi offloading. The research in this direction has an objective to enhance the cellular data offloading performance, particularly for delay-tolerant and non-interactive applications.

For non-vehicular users different offloading strategies mainly emphasis on offloading performance and availability of next approaching WiFi AP because a user wants to have a stable communication link with one WiFi AP. This thing is not applicable to vehicular WiFi communication. Since vehicles have fast speed and in a short period, they may meet more APs having different quality of communication links and may have more opportunities to transmit data. Therefore offloading strategy in these scenarios should consider the prediction of forthcoming WiFi APs to get the advantages of more data transfer opportunities. Delay requirements of an application would have significant impacts on mobile data offloading schemes. Different non-interactive or non-real-time applications, such as regular sensing data transmission to control unit, bulk data transmission from smart meters to smart girds and email attachments are usually throughput-sensitive and can compromise on delays of certain time limits

Literature shows the different proposal of vehicular WiFi offloading strategies [88][89]. Wiffer is proposed as an offloading strategy in [88], which determine when to defer data transmission of an application for WiFi connectivity rather than using communication over the cellular network. Wiffer calculates the delay requirements of each application and predicts the calculations of throughput potential of WiFi, based on the history of the vehicle's route. Based on these calculations it performs the data offloading.

To study the performance and availability of 3G and WiFi networks an experiment is conducted, by placing WiFi APs at more than half locations of the city, results showed that 20% of cellular traffic can be reduced by data offloading through drive-thru WiFi, also due to mobility of vehicles the temporal availability of WiFi is low(12% of the time). Wiffer handles delayed offloading and for delay-sensitive data applications, it performs fast switching to the cellular network. Vehicular users's (VU's) preferences or binary names or port number information are used to determine the delay tolerance of an application.

The effective throughput of WiFi can be estimated or predicted by calculating the number of contacted APs in history. This prediction is based on the duration of inter-contact time of each encounter and it can be same in future as in history average. Based on these predictions WiFi offloading is performed if $W > S.c$, where S is Data size need to be transferred within the given delay constraints, c is conservative quotient that is used to manage the trade-off between application completion time and offloading effectiveness and W is effective throughput of WiFi-based on prediction. For real-time or delay-sensitive applications, Wiffer uses fast switching from WiFi to cellular network when WiFi fails to transmit data packet within a predefined threshold time

Motivated by the experimental results of [88] which showed that mobility and connectivity with AP of the moving vehicles can be predicted, that's why authors in [90], proposed a data pre-fetching strategy. In this approach data is pre-fetched and cached in APs along the predicted route of the vehicle, transmit these data contents to the vehicles when they are in their wireless range. This exploits the vehicular WiFi offloading because vehicle-to-AP uses point to point communication having higher bandwidth than backhaul bandwidth. Moreover, vehicle-to-AP communication can use some special type of transport layer protocols like cabernet transport protocol (CTP) having less porn to wireless losses than TCP [91]. Data pre-fetching is based on mobility prediction model of vehicle, to manage the mobility prediction errors, data is redundantly pre-fetched along with the predicted APs on the route. Nevertheless, in recent years backhaul capacity of WiFi is boosted greatly, that's why the benefit of data pre-fetching might be reduced.

To enhance the throughput of Vehicular WiFi offloading strategy is also proposed at the transport layer [90]. Stream Control Transmission Protocol (SCTP) is presented for data offloading of cellular networks via WiFi and enhance the user's benefit. STCP uses both WiFi and cellular interface at the same time if necessary, and after each time interval, it schedule data packets for each interface. Experimental results showed that SCTP offload data traffic from 63% to 81% and it validates its effectiveness.

4. WiFi Deployment

Data offloading also depends upon the density and deployment strategy of WiFi APs. It has been shown that by deploying 10 APs in the area of per square kilometre, the average throughput per user can be increased up to 300% and this gain is proportional to the AP Density [92]. In other words, we can say that if we increase the number of APs in a geographical area then we can get better performance of cellular data offloading. On the other hand, dense deployment of APs may increase the operational cast and expenditures. Hence there is a trade-off between offloading performance and AP deployment cost and it should be examined properly. WiFi deployment strategy plays an important role to get batter throughput. Many factors should be considered to make a good strategy to deploy APs, like vehicle mobility, communication environment, population density and mobile data usage etc.

In [85], the performance of mobile data offloading is investigated against WiFi deployment. This study identifies the APs having least sessions and user connection times and gradually eliminate those APs to get the performance with different densities of APs. It is identified that even 80% of the WiFi APs are eliminated, whereas the offloading performance is dropped only 10% to 20%, which shows that only a few and well-deployed APs can contribute to batter performance of data offloading with less cost. WiFi deployment with respect to location popularity is presented in [93] and [94]. Location popularity means locations visited by the high frequency of users. Experiments are conducted in an area of 313.83 km2 with the deployment of several hundred APs at prime locations and results show that half of the cellular data can be offloaded efficiently on WiFi networks. It is also identified that most visited places are not definite sources of high data traffic generation. Therefore, by identifying the frequency of data traffic requests, another WiFi APs deployment approach is proposed and evaluated in [95].

Three strategies of WiFi deployment, namely, uniform random, traffic-centric and outage-centric are proposed and evaluated in [92]. In outage-centric and traffic-centric APs are placed by considering the outage and locations having the highest data traffic requests respectively, whereas in Uniform Random APs are deployed uniformly and randomly and deployment metrics are not considered. Results showed that by deploying 10 APs per km^2, in traffic centric average throughput gain was 300%, the outage is reduced 14% in the indoor environment by using the outage-centric approach.

2.4.2 Data offloading by Licensed Frequency band

For indoor (e.g., in office or home) cellular communication, femtocells are used as small base stations. These BSs are connected to provider's core network via some broadband network (e.g, DSL, digital subscriber line) and it helps to service provider to provide its services in the indoor coverage area, like the areas where service is unavailable or limited. Femtocells are more attractive solutions for operators because they offer improvements in capacity and coverage to the macro cell, especially indoors. This concept is applicable to all cellular standards e.g. GSM, World-Wide Interoperability for Microwave Access (WiMAX), wideband code-division multiple access (WCDMA) and LTE. Femtocell introduced effective ways of reducing the data traffic load over the microcell. They improve the experience of mobile users on macrocell BS by freeing the capacity and on the other hand users on the femtocell get improved performance due to having batter radio resources.

Mobile data offloading by using femtocells is also efficient because of different reasons explained in [96]. First, data usage occurs mainly in an indoor environment (office or home). In [97][98], it is explained that data usage occurrence is 26% in office and it is 55% at home. It shows that the operator can offload heavy data traffic via femtocells. Second, femtocells services are deployed and managed by the operator and it provides seamless service experience to end-user. Third, femtocell installation is easy as compared to macrocell BS deployments which involve a huge cost of backhaul, infrastructure and site acquisition etc.

In femtocell network, data traffic flows from users equipment to femtocell's small

BS by using air interface, which is connected to some broadband connection, then from this broadband internet to the provider's core network or/and other destinations of internet. User's equipment (UE) is automatically connected to small BS whenever a user comes in the range of femtocell frees the macrocell BS resources and traffic flows through the users broadband connection. Use of femtocell not only offload the traffic from NodeB (in 3G) or eNodeB (in 4G LTE) but also frees resources of a radio network controller(RNC), which leads towards the further load reduction of the macrocell.

In [99], a new standard selected IP traffic offload (SIPTO) is explained, in which operator offload a certain type of data traffic at a node that is close to the location of end-user equipment. Two types of approaches are explained in this standard, deep packet inspection (DPI) based and access point name (APN) based. In the implementation of SIPTO operator can offload the data traffic from the core network by the direct traffic flow from the femtocell to the internet.

1. Small-Cell-Based Offloading Small cells operate in the licensed band of the energy spectrum and they are access points having low power as compare to macrocell [100]. These cells provide batter cellular coverage, improve capacity and application services for home and enterprise users at metropolitan and rural public spaces. There are three types of cells based on signal transmission power and coverage area in term of radius as shown in table 2.1. Small cell-based data traffic offloading have two approaches: (1) User-oriented traffic offloading, (2) Network oriented offloading

Table 2.1: Comparison between cell types

Characteristic	Femto	Pico	Micro	Macro
Power	20mW	200mW - 2W	2W	greater than 2W
Coverage	Hot Spot	Hot Spot	Wide Area	Wide Area
Configuration	Automatic	Automatic or Manual	Automatic or Manual	Manual
Location	Indoor	Indoor or out door	Out Door	Outdoor
Density	Large Number	Large Number	Small Number	Small Number
Service Location	Designed for data	Designed for Data	Designed for Voice	Designed for voice

a. User-oriented traffic offloading: It investigates the performance enhancement of user equipment through femtocell traffic offloading. Mobile user (MU) can get benefit in term of radio resource utilization, data cost and energy consumption. Due to non-ideal backhaul as well as intermittent and limited small cell coverage, data offloading might cause some additional delay. Hence, it is critical to comprehend under which condition there will be the benefit of data traffic offloading and how it can be adjusted for a trade-off between the benefits from data traffic offloading and consequential disadvantages.

In [101], Cheung et al. find the data transmission delay for offloading data in a cell and optimize the data offloading problem concerning MU's data cost and resultant delay performance in file transfer. In [102][103] the authors suggested different proposals, to give incentives to users and exploit delay-tolerance for data traffic offloading in a small cell to reduce the burden on macrocell. In [104], the author proposed a cost-aware data offloading scheme and considered the budget of MU's data usage, formulated and optimized the offloading problem concerning throughput and delay trade-off.

b. Network-oriented Traffic Offloading: This strategy emphases on the study of network performance improvement by traffic offloading using a small cell. Strong coupling exists among cellular users when data offloading occurs in small cells because of intercell interference and limited radio resources. For example, when data offloading occur on the same band in nearby cells then severe intercell interference may occur. Moreover, severe collision and interference occur when several mobile users offload a huge amount of data in the same cell. That's why to measure offloading performance gain, we must consider the overall network performance enhancement by data traffic offloading using small cells.

In [105], authors measured the interference, during traffic offloading among different cells. They evaluated the overall network performance by formulating utility maximization problem aiming to distribute traffic offloading in different cells. In another research

[106], Chen et al. consider the energy efficiency of the overall network by considering the interference in small cells and defined an optimal on-off controlled strategy for data traffic offloading. In [107], authors measured the co-channel interference in the same cell among different mobile users by considering and maximizing the overall network performance and defined an optimal solution of the user-association problem. A social welfare optimization problem is defined and formulated by Iosifidis et al.[108] that distribute the demands of data traffic offloading in different cells by considering the limited resources of the overall network.

2. Dynamic cell creation by using vehicles

Since small mobile cells have been proposed to improve coverage and capacity for onboard mobile users, these cells could be combined with other data transmission technologies to provide an effective data traffic offloading solution for public transit vehicles.

The number of users using their mobile devices onboard public transit vehicles via mobile networks is rising exponentially [4, 109]. Due to Path Loss (PL), shadowing, and Doppler shift effects resulting from vehicle speed and distance to MBSs, these users can experience poor signal reception and low bandwidth. It is becoming very difficult for operators to boost onboard cellular coverage and the ability to accommodate mobile users. Mobile traffic from groups of moving users also influences the efficiency of the mobile base stations [109]. Recent research has considered the deployment of small cells, including buses and street cars (a.k.a, small mobile cells), in public transport vehicles [110, 111]. Mobile small cells are deployed to maximize coverage and the ability to improve the QoS of vehicle users.

In [112] a framework is proposed, the goal is to offload data from users, that is intended to be transmitted using WiFi via macrocells. Besides, mobile small base stations are fitted with WiFi transmitters mounted on the vehicle's rooftop to provide mobile small base station with backhaul connectivity by connecting to urban WiFi access points (APs) (which are commonly used and already cover many urban cities [113]). Therefore, to relieve overburdened macrocells, mobile data traffic is routed to the cellular network operators via WiFi. Also, to make the offloading process more accurate, they integrated WiFi coverage maps and the profile service history of users in the proposed offloading system. By inspiring this work, we also find the most visited locations by vehicles and identify the locations for roadside unit installation as offloading points.

2.4.3 Data Offloading using Hybrid Frequency Band

In previous sections, mobile data offloading techniques have been categorized based on energy spectrum utilization, in which spectrum resources are utilized in term of licenced or unlicensed bands. In this section, we will explain the categories that use both licenced and unlicensed bands.

1. Device-to-Device Communication

To address the challenges of cellular data crises, a promising solution, D2D device-to-device communication has been suggested for next-generation cellular networks. In D2D communication, mobile devices in the wireless range can communicate with each other without traversing through the base station of the cellular network or the cellular backhaul. By exploiting the vicinity of mobile devices and direct communication, D2D data transmission can enhance the throughput, reduce the communication delay of mobile devices and it can enhance overall spectrum utilization, network throughput and performance [114]. Because of the different advantages of D2D communication, proximate communication is appropriate for many user cases and it can introduce different peer-to-peer and location-based services and applications. In [115], different services, use cases and applications of D2D communication are introduced in different categories, like content dissemination, machine to machine communication, peer to

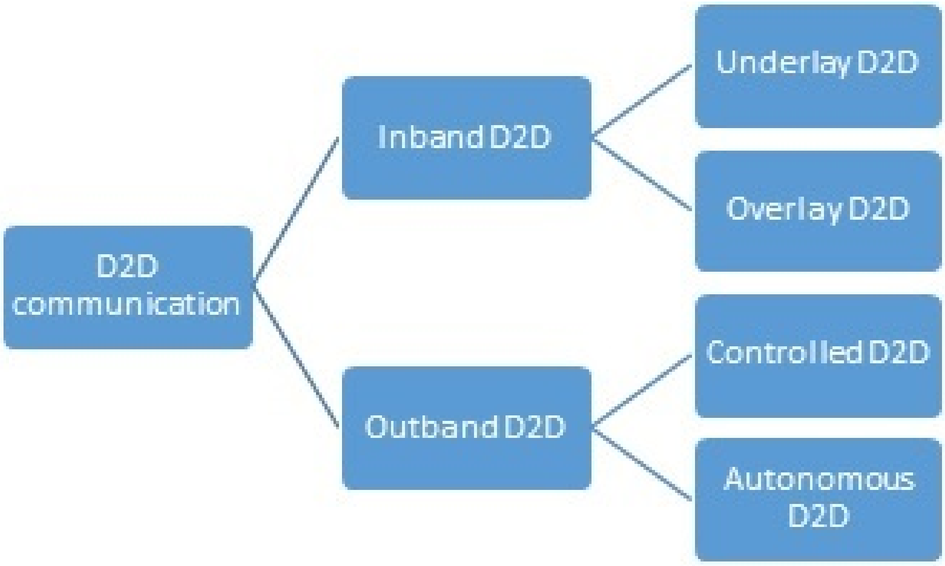


Figure 2.2: D2D communication classification.

peer, multi-casting, and so forth for cellular data offloading

Sound literature of D2D transmission is available, in this section we discuss the research issues of D2D communication and some existing solutions. In [116], multi-hop communication was proposed by using D2D communication in cellular networks. D2D transmission is analogous to Adhoc networks, the main difference is it often includes a cellular network for controlling. A common classification of D2D communication is explained in literature [115, 117] and shown in the figure 2.2. For spectrum utilization, D2D communication is divided into two strategies, 1) Inband D2D communication and 2) Outband D2D communication. In 1^{st} strategy, same energy spectrum band as of cellular communication is used for D2D communication, whereas in 2^{nd} strategy, to avoid the signal interference from cellular networks, some other bands like WiFi, Blue, Tooth, WiFi Direct etc. are used[118]. It is to be noted that, in outband D2D transmission, eNodeB BS can have a different level of control over D2D transmissions such as connection establishment and neighbour discovery, this control can be managed by control channels of cellular networks. Otherwise, D2D data transmissions work in self-governing mode. Based on spectrum resource utilization, Inband D2D data transmissions can be further characterized into two categories. If same spectrum resources as of cellular communications are used for D2D communication then it is called underlay D2D[119] data transmission and if resources are reserved used for D2D communication from cellular spectrum band then it is called overlay D2D transmission[120].

2. Opportunistic Communication

When several users request for same contents in a given area, service provider governs the communication as an in-charge and transferred the requested contents to the base station, and base station forwarded these contents to those users. These requested contents can be delivered to users by using three different approaches. In 1^{st} case, the base station forwards the requested contents directly to each user. In 2^{nd} case some of the target users are selected by the base station, it forwards the data to the target users, and they disseminate the data among their peer users. Whereas in the 3^{rd} case, target users may store the requested contents and delivered to peer users when they are in their wireless range. If the selected users don't meet the timelines to transfer data, then cellular base station are responsible for deriving data to each of data undelivered-subscriber. The latter case is referred to as an opportunistic communication in [93, 121].

Opportunistic communication also reduces load on the cellular base station that's why it can also be counted as a data offloading scheme. This approach can enhance the performance and throughput of each subscriber because channel conditions between users are much better as compared to channel conditions between user and base station. It is to be noted that, opportunistic communications can be useful in case of delay-tolerant data delivery (e.g video and audio files etc., sensor produced data to upload at control units and scientific experiments related data delivery [93] because total end-to-end delay may increase from service providers to subscribers. One of the parameters, in opportunistic communication to select the target users, who are responsible to disseminate the data among their peers is the influence of their social behaviour. They may prefer to use the battery of their devices for their purposes rather than consuming battery for the benefit of others and this is called social selfishness [121]. Users may discard the data after using it because of buffer space limitation in their devices, this is another example of the effect of social behaviour on data dissemination in opportunistic networks. Target and peer user selection have main importance in opportunistic data offloading, these are explained here in more detail.

a. Target user selection: In content delivery networks there are several numbers of the distributed server instead of a single server. These distributed servers store the more popular contents in their caches which are the most periodically requested contents by the users from the main server. Now, it is not necessary for all users to get data from the main server, they can get data from their nearby distributed server, hence the main server can remain uncongested. The same idea is implemented in opportunistic data communication in which target users cache the most popular contents in their cache and behave like distributed servers as in content delivery networks.

In [122], a solution for target user selection is formulated and named as Mobile Social Network (MoSoNet). The users who already received the data from content delivering servers, or stored the contents that are requested by some other users, can be selected as a target user and can disseminate data to the other peer users who are selected as receivers in the opportunistic social network. Several systems like PeopleNet and WhozThat or many applications like CenceMe and Micro-Blog are used for MoSoNet. For example, Micro-Blog is used for querying and sharing the contents in the devices. Whenever a mobile user requests some specific contents from the main server, the server may provide the contents itself to the user or it redirects the user to some other selected target user having requested contents. The status of the target user and its location is sent periodically to the main server.

In opportunistic networks, it is an important question that who is the most suitable user as a target user among all users in the network. In [121], one of the criteria is the social importance of a user and it is defined as a number of peer users connected to him/her. However, this not sufficient for selecting a user as a target user because it is possible that peer users are not interested in the contents that are stored by that user. Hence, another perimeter is defined as the number of interested peer users in the network for the selection of the target user. Therefore, it is a trade-off between these two factors for decision making over the selection of the target user.

In another point of view, the problem of target user selection can also be formulated as an optimization problem having an objective of maximizing the number of peer users (number of content-delivered users) in a required time period so that load on the base station can be reduced. Different methods are defined in [122]. For an already defined number of selected users, in an approach of the greedy algorithm, a set of target users are selected that maximizes the number of peer users or content-delivered users. As it can be considered that, the greedy algorithm doesn't consider the location/time correlation and statistical consistency of user's daily presence. On the other hand in the Heuristic algorithm, this consistency is considered at an initial stage of target user selection. For example, consider a company, all of its employee remain present at its location at almost the same time during each working day of the week. In such a scenario, the heuristic algorithm selects an initial target user set same as before instead of defining the new set every day. Finally, the initial target user selection set is picked up completely in a random way from all users. Simulation results of [122] showed that all these selection algorithms mitigated the traffic load on the cellular network as compared to the scenario where data is delivered to each user by base stations themselves.

b. Peer user selection: Peer users are those that accept the data contents directly from the selected target users. Data contents are transferred from the base station to selected target users, and they can transfer these contents to the peer users in their wireless range via Bluetooth or WiFi, by using some applications like Haggle [121].

They can also disseminate the contents to their peers via some cellular resources in their social friend networks like Twitter or Facebook as mentioned before in Mobile Social Network (MoSoNet) [122].

3. Opportunistic and D2D Communication in vehicular networks

Due to the advantages of opportunistic and D2D technology, it is suitable for many vehicular related use cases and can enable peer-to-peer and location-based application and services. For instance, considering the enormous number of connected the smart cars (connected cars in 2021 [3]), software update of smart car can put a noteworthy load on the cellular infrastructure, and cost a lot of money for car owners. Thus, the software update can be downloaded by some selected vehicles and then it can be transferred to other vehicles by opportunistic and D2D(V-D2D) transmission. In this process, target vehicles can be selected by the cellular network by applying some efficient algorithms and allocate appropriate resources to reduce the interference, optimize the performance and fulfil different QoS requirements. In this way, most of the cellular load and data traffic can be offloaded to V-D2D communication, and thus cost, energy and much of the cellular bandwidth can be saved. Furthermore, software update does not require the real-time communication, it can compromise on some delays, in such cases the vehicular delay-tolerant network (VDTN) can be employed to forward the software update package in a store-carry-forward manner which can further save the cost, energy and offload the cellular traffic. Audio/video streaming and gaming are some other type of data services among vehicular users such as a social network in the vehicular proximity [123]. Generally, such types of applications are maintained by WiFi-direct, LTE-direct or DSRC communications, because of long device paring time and collisions they may not satisfy the proper requirements. V-D2D and opportunistic communication can support such type of services in a batter way due to the use of cellular links for control channels. Cellular-control may provide connection setup within short delay time and batter resource allocation.

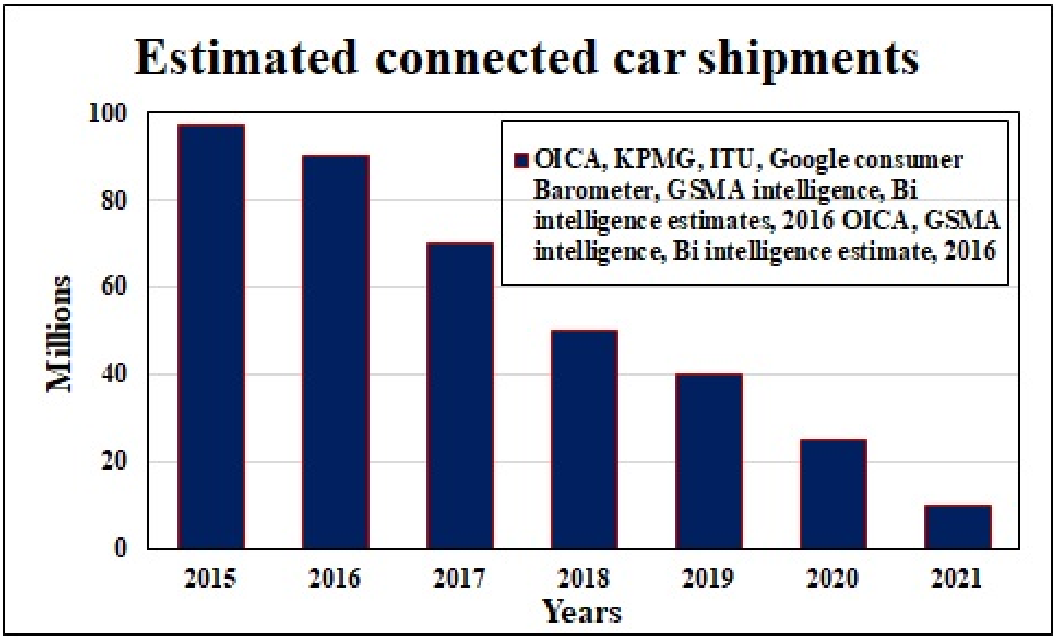


Figure 2.3: Estimated connected car shipments [3].

Numerous research works consider the problems of applying D2D communications in vehicular adhoc networks. In [124], Cheng et al. investigate the validity of D2D communication for intelligent transportation systems (ITS) by applying spatial distribution of high mobility vehicles and on different channel characteristics. The simulation study showed that among all other D2D techniques D2D-underlay got the highest efficiency in spectrum utilization and transmission rate increases with a decrease in distance of vehicles. Moreover, it is observed that with the gradual increase in link density of V-D2D the average spectral efficiency first begins with growth then gradually declines, because when V-D2D link density increases then the network will get more severe interference. In [125], Sun et al. formulated a resource allocation strategy for cellular user equipment's (CUEs) and vehicular user equipment's (VUEs), to exploit the CUEs sum rate while assurance the strict delay reliability and consistency requirements of VUEs' services. To get the optimum performance of V-D2D, Ren et al. in [126] studied power control framework with the joint channel selection, where a sequence of modifications are made to decrease the conditions of full channel state information

(CSI).

2.4.4 Types of data offloading

The work in [84], introduced many offloading techniques in a cellular network to reduce network congestion. These offloading techniques are being categorized as delayed offloading and non-delayed offloading. Table 2.2, describe the summary of these offloading techniques.

1. Non-Delayed Offloading: Non-delayed offloading provides logical solutions as a Wi-Fi hotspot to lessen congestion and cost. Cellular Base stations and D2D approaches use wireless technologies to establish direct communication between devices, and cellular network to device communication. AP Based offloading is a user-driven approach [84].

2. Delayed Offloading: In delayed offloading, content reception can be delayed up to a certain point of time but with good delivery conditions.

Sylvia et al. [127], introduced energy-aware opportunistic mobile data offloading. Their main focus is to decrease the energy consumption while at the same time retaining the application throughput of an opportunistic network. They evaluated the performance of progressive selfishness which combines the merits of two energy-saving mechanisms which includes duty cycling and selfishness. With the use of trace-driven simulation analysis, they evaluate the performance under two modes of nodes cooperation for the progressive selfishness behaviour which helps to save energy up to 85% while sacrificing less than 1% with the distribution of density of node and availability of initial content. However, since mobile nodes are battery-powered, opportunistic networking can only be considered a viable mechanism for offloading data if it delivers high content volumes at a low energy cost.

The authors in [61], have deployed cell base station in addition to the conventional macro base station to boost the network capacity and energy saving. They focused on energy harvesting for on-grid power saving while sustaining Quality of Services (QoS). They optimized the energy-aware traffic offloading scheme, On-Off state of SBS and power control jointly to know the statistical information of renewable energy arrival and amount of offloaded traffic. In this scheme, multiple base stations have been set up in the heterogeneous network to satisfy quality services. Simulation results demonstrate that their scheme can achieve more than 50% power saving gain for typical daily traffic and solar energy profiles, compared with the conventional traffic offloading schemes.

The authors in [128], demonstrated Device-to-Device (D2D) Fogging, which works on the concept of mobile task offloading based on network-assisted D2D collaboration, where mobile users can dynamically and beneficially share the computation and communication resources among each other via the control assistance by the network operators. They proposed an optimization problem to minimize the time-average energy consumption for task executions of all users and taking into account the incentive constraints of preventing the over-exploiting and free-riding behaviours.

Mobile phones and laptops are increasing data traffic every year and this trend is expected to continue further with all new developments. Therefore, Dimatteo et al. ([93], proposes a solution to offload traffic from the cellular network towards metropolitan WIFI access point, which is known as MADNet architecture. While transferring bulk data, this architecture is beneficial even in the case of the sparse WIFI network.

Table 2.2: Overview of offloading schemes

Work	Message Direction	Goal	Offloading Type	Mobile Device Type
MADNet	Infrastructure $\rightarrow$ mobile device	Reduce Infra. load	Delayed	All infra. access, all DTN
Drive-thru Internet	Infrastructure$\rightarrow$ Vehicle	Reduce Infra. Load and more strong connection	Non-Delayed	All vehicle, Wi-Fi access
Wiffler	Infrastructure$\rightarrow$ Mobile device	Prefer Wi-Fi over cellular	Delayed	All Wi-Fi and cellular capable
HRS	Infrastructure$\rightleftarrows$ Mobile device	Offload Infra., prefer DTN	Delayed	Heterogeneous infra. access/adhoc
Kashihara	Wireless nodes $\rightarrow$ Vehicle	Reduce traffic congestion	Delayed	All Wi-Fi and cellular capable
Push-and-Track	Infrastructure$\rightleftarrows$ Mobile device	Reduce Infra. load	Delayed	All infra. access, all DTN

MADnet consists a total of six modules: Connectivity module, a Location module, Protocol module, Naming and Forwarding module, Data Module. All of these modules are responsible for providing connectivity to users according to location, name and then forwarding data. Every node of the system generates data to upload and request content to download on another side. In such a way, at the time of downloading content, a user requests to the nearby base station with all the details (position, speed, and direction) which is known for all Apps position. Base station makes a list of all APs of that route and node consults the particular AP to download the content. This list keeps in the storage of node's system for future direct communication through Bluetooth/Wi-Fi.

In this thesis, we use the huge volume of connected smart vehicles as forecasted in figure **??** [3], and design an energy-efficient data dissemination framework by using the vehicular delay-tolerant network.

2.5 Adopted VADD Performance Evaluation Approach

The primary objective of this thesis is to develop a framework that be used for energy efficient data dissemination from various data sources to data centers. For this data dissemination we use vehicular delay tolerant network. The performance of the proposed framework is measured using QoS parameters, including, end-to-end delay, throughput, energy efficiency, and data loss. This section depicts methods adopted for measuring QoS performance in this thesis.

Taking into account the need for this analysis, evaluating the efficacy of the proposed network in the real-time setting is highly expensive. Thus, by using the real vehicle count and taxi traces in MySql server database, this research adopts mathematical modelling, simulations and statistical analysis to test and evaluate the output of the proposed VADD framework. By evaluating different network scenarios, the simulation and analysis approaches are being used to generalize the research findings.

The Monte Carlo Simulations are used for numerical analysis and MATLAB is used to test the performance of the proposed algorithms' (Chapter 3 to 5). Experiments with Monte Carlo are a wide class of computational algorithms that rely on repeated random sampling to obtain numerical results. They are often used in physical and mathematical problems and are most helpful when it is difficult or impossible to use other approaches [129].

In addition, we use Statistical Information Grid (STING): Grid clustering algorithm (Chapters 3 and 4) for data-driven approach and statistical analysis. STING is the most powerful algorithm used by many researchers for tempo-spatial analysis GPS based data [12]. We divide the entire city into different cell sizes according to the wireless coverage area of various wireless technologies. Extensive analysis is performed to calculate the network coverage and capacity. We use MATLAB MathWorks function “Scatter Plot coloured by Kernel Density Estimate” to calculate the vehicles displacement and density. For RSUs deployment we identify the most important locations with respect to the most visited locations in a selected geographical area. This function use the kernel smoothing function to compute the probability density estimate (PDE) for each point. It uses the PDE has colour for each point and it returns handles to the scatter objects created [130].

Finally, IBM CPLEX Studio and SaS Optimization tools are used[14][15] (Chapter 5), to minimize the energy cost and reduce carbon emissions of our proposed VADD framework. IBM ILOG CPLEX Optimization Studio is a prescriptive analytics solution that enables rapid development and deployment of decision optimization models using mathematical and constraint programming. Using CPLEX studio we apply the min-cost flow problem for finding the optimal routes to reduce the energy consumption in multi-hop communication. Similarly, SaS is also a powerful array of optimization and simulation tools for identifying actions that will get the best results while operating within resource limitations and other relevant restrictions. We use SaS optimization tool to transfer multiple data demands by using a multi-commodity flow problem for optimal network mode selection by the proposed VADD framework.

2.6 Summary

This Chapter reviewed the concept of green networking. How much energy is consumed, and carbon is emitted caused by information and communication technologies. The huge energy usage in ICT illustrates the need for energy-efficient data communication techniques. Different terms related to green networking have been explored and how these concepts have been applied to different types of networks. Then types of different energy optimization techniques explored and how energy can be efficiently utilized for data communication in information and communication technologies. This first part motivates us to design an energy-efficient technique for big data dissemination.

In the second part, we explore the concept of delay tolerant networks and its routing protocols. How the DTN can be used in vehicular delay tolerant networks. This part motivates us, how the existing system of vehicular networks can be used for big data dissemination. In the 3^{rd} part of the literature, we explore the different techniques of data offloading for different bands in the energy spectrum band. Types of data offloading have been explored in the next section. These parts motivate us, why data offloading is necessary to meet the future needs of mobile data demands.

From the literature review of three different areas of research, a research gap has been identified, which is to utilize the potential vehicular network as an alternate network channel for energy-efficient big data dissemination to reduce the burden on the congested networks, as shown in Fig. 1.6. To design an energy-efficient data dissemination framework, we need to investigate the potential of this proposed alternate channel. In the next chapter, we introduce the design of this framework and explore its potential in terms of big data dissemination, network coverage, and network capacity.

Chapter 3

Vehicle Assisted Data Dissemination Framework

In chapter 2, a detailed review of literature has been discussed related to green networking, delay-tolerant networking, and data offloading techniques. A vehicle assisted data dissemination (VADD) is proposed in this Chapter. The main objective of this thesis to support the hypothesis that vehicular networks are a suitable alternate channel for data communication related to delay-tolerant or background traffic in smart cities. One of the main issues for a smart city is the data aggregation from various data sources to the central control units for storage and analysis. While the core networks are winding up progressively over-burden, the cost of keeping up and extending the networks to oblige the big data of these smart cities stays high. To overcome the problems of congested networks, we propose a data dissemination framework. The idea is to utilize the city existing public transport infrastructure to pick up and deliver the big data from designated pick-up points and data centers.

In 1^{st} part of this chapter, we efficiently utilize the vehicle volume on the road and D2D communication of future smart cars with roadside units, under the control of cellular networks, to transport big data from various data sources of SC to their corresponding data centers. Instead of using separate storage media, these smart cars will have their own storage, processing, and radio communication modules. These modules contain Wi-Fi, GPS as well as cellular interfaces. In this part, we propose a framework that offloads the data from congested networks through these vehicles by using their routine rides.

In 2^{nd} part of this chapter, we explore the collection of big data produced by a massive number of smart devices in SC using VADD as an alternate data dissemination channel. By using the opportunistic contacts between the sensors and microscopic movement of vehicles with their GPS location information, it is possible to piggyback accumulated data on moving taxis-cabs for further data delivery at data centers of the corresponding service. Not only it can reduce the load on expensive cellular links but also eliminates the need for new infrastructure deployment.

The findings of this proposed vehicle assisted data dissemination can help network engineers, to configure the network for good coverage and capacity. The main contributions of this chapter are:

- We propose an alternate network for data collection from densely deployed SC's devices, which saves a huge amount of cost for infrastructure deployment.
- We develop a mathematical model and a case of Auckland City is evaluated by using annual average daily traffic (AADT) of Auckland City in our proposed model to reduce the end-to-end delay for big data transmission.
- To explore the potential of vehicular displacement in the entire city, we develop a grid clustering algorithm to divide the whole city in different sizes of rectangular clusters, it reduces the complexity of data analysis to observe network dynamics. A case of Beijing city is evaluated, by applying our proposed algorithms on real taxi traces to observe network coverage and capacity of the proposed system.

- For roadside units (RSU) placement in the smart city, we develop a greedy algorithm to identify the most popular locations and evaluate a case of data dissemination using our proposed framework.

To achieve this, propose VADD framework in Section 3.1. In this chapter, we analyze the potential of our proposed system VADD by considering two cities Auckland, New Zealand, and Beijing, China. In the Auckland case Section 3.2, we use real data of annual average daily traffic (AADT) and our model explores, how much data can be offloaded by using this vehicle count on the roads. The calculations related to an alternate channel for data dissemination are presented in Section 3.3, which can fill a gap between current and future needs of mobile data demands in different cities of New Zealand. In the Beijing case Section 3.4, we use the real traces of taxies to consider the displacement of vehicles around the city, for analysis of coverage, capacity, and data offloading potential of VADD. Finally, the chapter is summarized in Section 3.5.

3.1 Overview of VADD framework

The smart devices and sensors are scattered throughout the smart city. These smart devices are installed in a different type of structures to accumulate the status data for various applications. The wireless coverage of the cellular network is available everywhere in the city, and smart vehicles are moving on the roads. These smart vehicles have GPS, WiFi, and cellular interfaces, also have storage and processing capacity. They can form a network with the base station (BS) by using the cellular interface, with roadside units (RSU) or with other vehicles by using WiFi interface. A central controller (CC) server is installed at the cellular network. This controller will be the manager of overall communication.

In the proposed work various SC devices and sensors generate a huge volume of data in the city. CC selects the cellular network or traditional core network to forward delay-sensitive data. For delay-tolerant data, CC selects cellular network, core network, or proposed vehicular network. When an information source needs to send information to the control unit (CU)/ data center (DS), it sends an information request packet to CC on a cellular control channel. CC will choose the most appropriate alternative, to exchange information utilizing the core network, cellular network or vehicular network. This decision is made by considering the information of vehicle count on the road, delay-tolerant interval, history of the vehicle's trajectory and energy cost. Based on these results, CC selects the suitable network, if infrastructure network is more appropriate for the given set of demands then CC guides the source to send data on path (a) or (b) of Figure 3.1. Otherwise, data can be forwarded by using vehicular transport network on path (c). For each given set of data transfer requirements, CC chooses optimal vehicles by using their trajectory's history and informs the DS to send data by using these particular vehicles. The vehicles participating in this proposed system, update their reputation profile to CC for optimal decision making.

On the receiver side, these vehicles will upload the data whenever they encounter an RSU by using the WiFi interface. The RSUs are connected to the backbone network of the city and send the data to SC CUs. After receiving the data, the RSUs send the acknowledgements to the server. If a particular data packet is not transmitted to the CU before the delay-tolerant indicator expires, then such data will be transferred directly to the CU by using the cellular interface. Moreover, the vehicles in our proposed system are smart, they can sense the data collected from sensors. If the data collected from the sensor is critical and needs to upload urgently then our smart vehicles can upload it by using their cellular links.

Whenever a service provider at CU or data center wants to send data to SC devices e.g software update code, it sends the data to CC. CC will use our proposed system to forward this data from a service provider to smart devices by using smart vehicles.

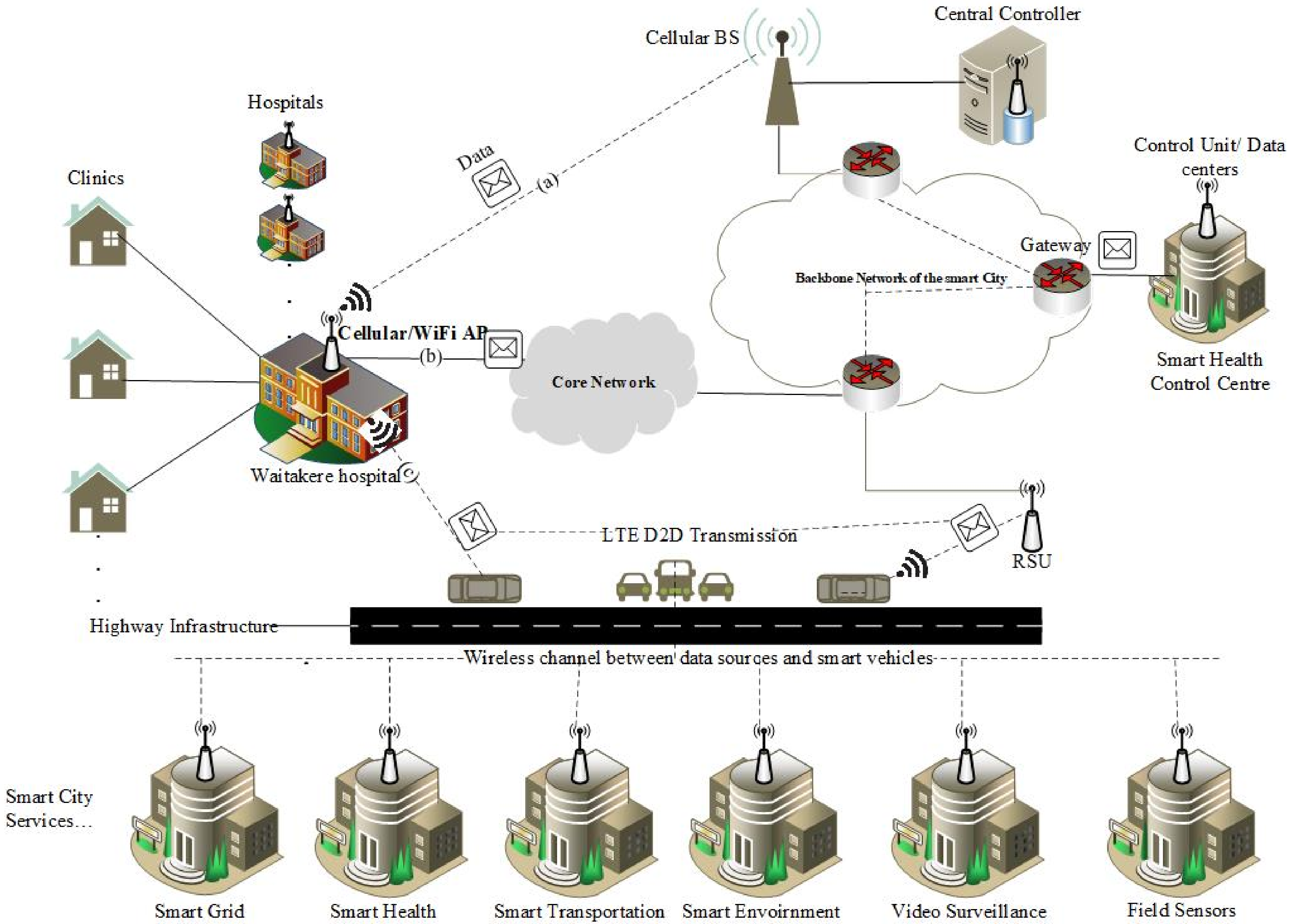


Figure 3.1: VADD system overview

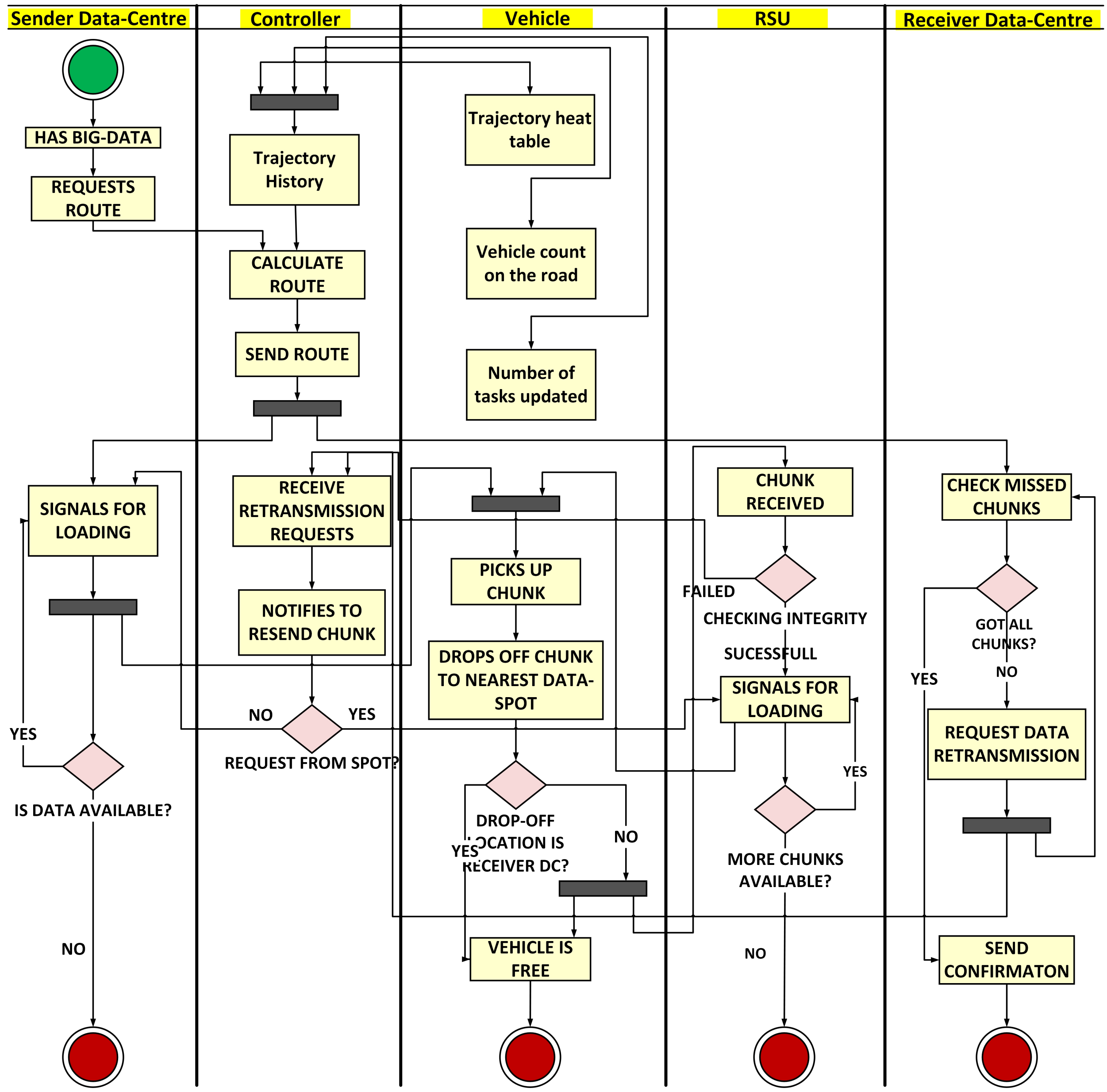


Figure 3.2: Sequence diagram of proposed framework.

The sequence diagram of the proposed system is shown in Figure 3.2. It represents the overall flow of data between different entities.

In the next two sections, we evaluated our proposed framework by using the case of Auckland and Beijing cities.

Table 3.1: Summary of variables used in proposed model

Symbol	Meaning
n	Number of vehicles used to transport the required data
B	Available Bandwidth of internet
V	Vehicle volume on the road
C	System Capacity
D	Delay Tolerant Data,Volume
d	Distance Between,Source and Destination
$\bar{s}$	Average Speed
ρ	Probability of Vehicles to,Participate in Proposed System
k	Number of,offloading points from source to Destination
V_f	Vehicular frequency on the road
T_d	Total Transportation Delay
C_{Veh}	Storage Capacity of Vehicle
T_{D2D}	Time required to transfer data by using D2D Communication.

3.2 Auckland Case

In the Auckland case, we consider daily vehicle count on the roads to explore the potential of our proposed framework. Following are the mathematical models that are used in this work for mathematical calculations and the symbols used in these models are shown in Table 3.1.

3.2.1 Network Delay Model

The time required to transport data between source and destination on the road is based on two aspects. The first, time required to transit a vehicle between source to destination, can be calculated from the distance between two points and the average speed of the vehicle. Second, the time required to pass enough vehicles form the data pick-up point towards the final destination or towards some other offloading point on the way to the destination, it can be calculated by the given data volume, percentage of participating vehicles and storage capacity of each vehicle.

$$T_d = \frac{d_{AB}}{\bar{s}} + data\ loading\ time \tag{3.1}$$

Vehicle volume 'V' is the available vehicles on the road and it is publicly available for each road of Auckland city in traffic count report of Auckland Transport Agency [131]. The storage capacity of the vehicle can be assumed 1 TB or 1/2 TB for calculation. Now the system capacity 'C' can be expressed as follows.

$$C = V \times C_{Veh} \tag{3.2}$$

Suppose there is a probability only 'ρ' drivers are willing to participate in this proposed network based on some incentives and 'V_f' is the vehicular frequency on the road then the bandwidth of each road can be calculated as follows.

$$Bandwidth = V_f \times C_{Veh} \times \rho \tag{3.3}$$

From Eq. (3.1), we can calculate the total delay when no offloading point is in between the source A to destination B.

$$T_d = \frac{d_{AB}}{\bar{s}} + \frac{D}{V_f \times C_{Veh} \times \rho} \tag{3.4}$$

If there are k offloading points in between the source A and destination B, T_{D2D} is the time required at each offloading point to upload and offload the data then total delay can be calculated as follows.

$$T_d = \frac{d_{AB}}{\bar{s}} + \frac{D}{V_f \times C_{Veh} \times \rho} + 2 \times \sum_{i=1}^{k} (T_{D2D_i}) \tag{3.5}$$

And for internet case delay can be calculated by the data volume and available bandwidth as follows.

$$T_{Interent} = \frac{D}{B} \tag{3.6}$$

3.2.2 Performance Analysis

In this section, we compare the proposed system with the internet-based system.

3.2.2.1 Case Scenario

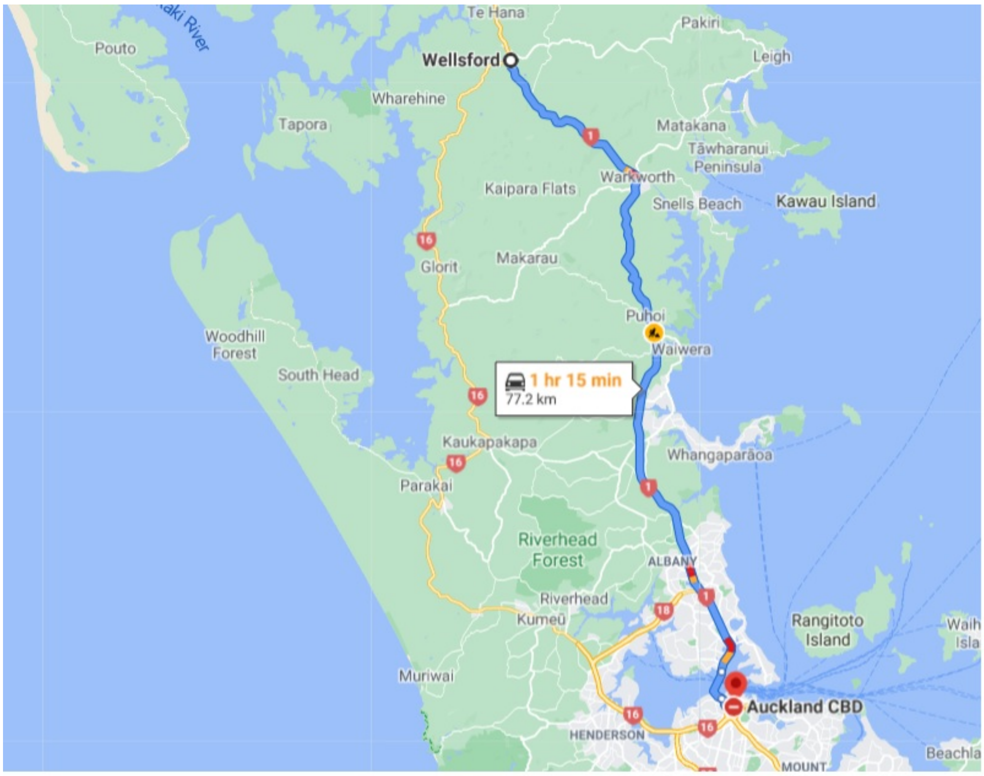


Figure 3.3: Internet based Vs Vehicle based system.

In this section, we compare two cases for data forwarding, internet-based and vehicular based. Suppose a data source at WELLSFORD - Telemetry Site 17 - Nth of Mangawhai Rd with reference ID:01N00336 in Auckland wants to send some data to a destination in Central Business District (CBD) as shown in Figure 3.3. For vehicle case, we get the distance between these two locations is 77.2 Km from google maps

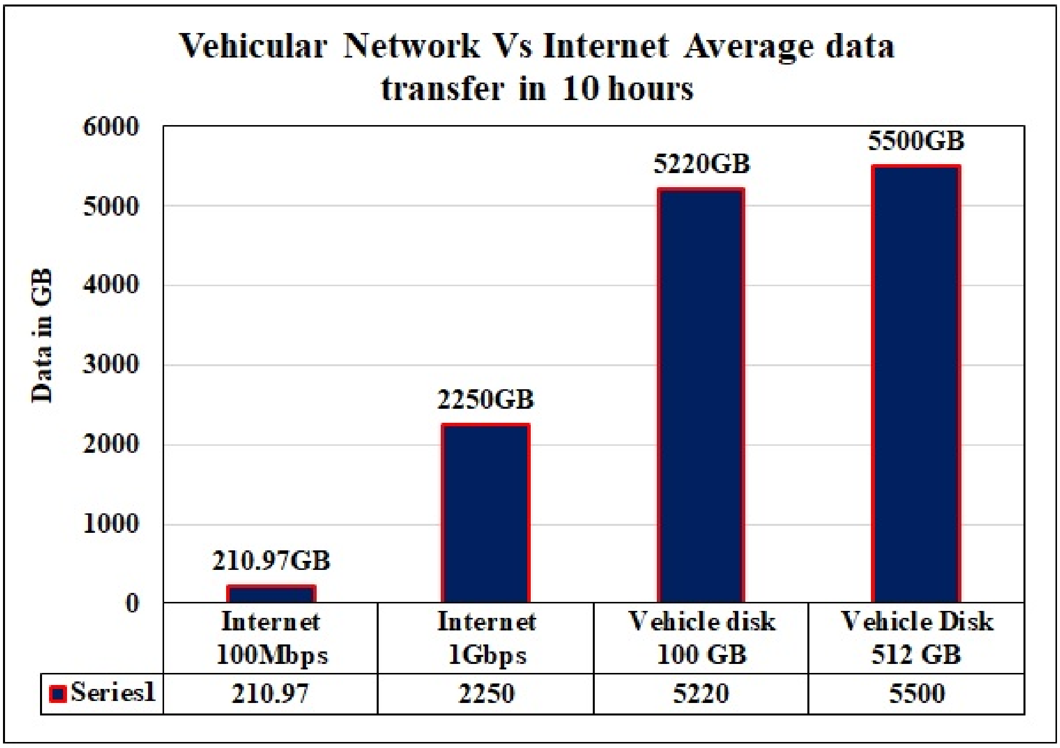


Figure 3.4: Internet based Vs Vehicle based system.

and vehicle volume on the road at the source site is 10055 vehicles per day [131], we consider the average speed of the vehicle is 50 Km/h, each vehicle has a wireless interface having speed 2.5 Gbps for D2D communication [132]. If 20% of the vehicles take part in the proposed network then we can calculate the vehicular delay by using Eq. (3.5). For internet case, data can be forwarded to the destination by using a wired network, cellular network or Wi-Fi network. Core network delay can be calculated by using Eq. (3.6).

We transfer the data between these two points for ten hours and get the results as shown in Fig. 3.4. For internet case, we consider bandwidth 100 Mbps and 1 Gbps, and for a vehicle-based system, we consider the vehicles having disk capacity 100 GB and 512 GB. The results show that our proposed system outperforms than the internet. Vehicular network transfer more than twice data at the same time as compared to the internet has a very high data rate of 1 Gbps.

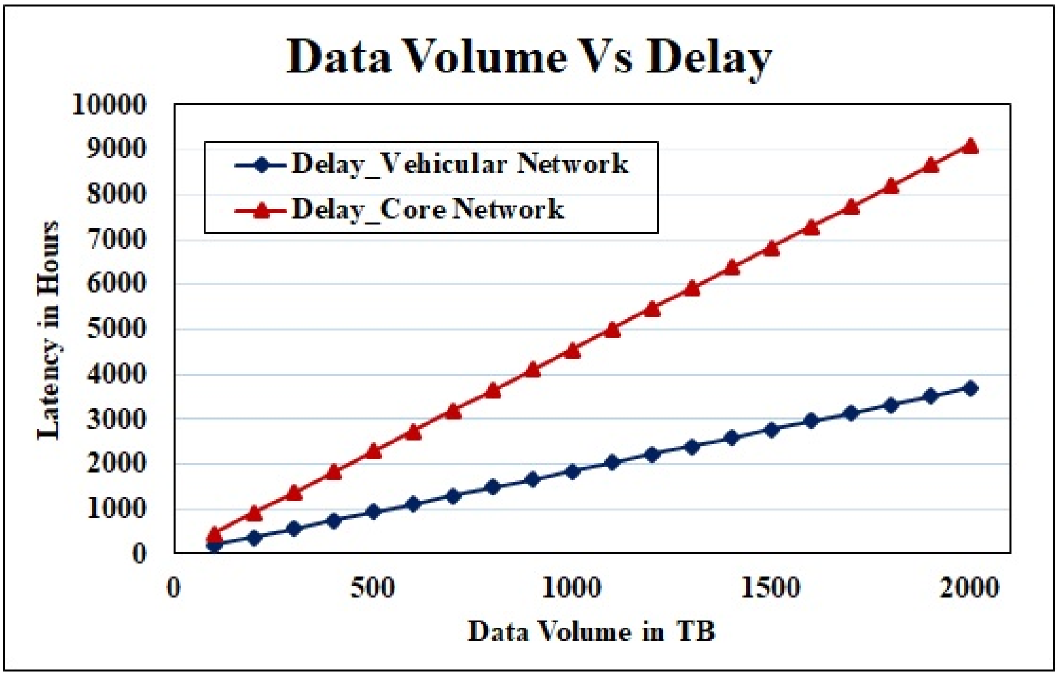
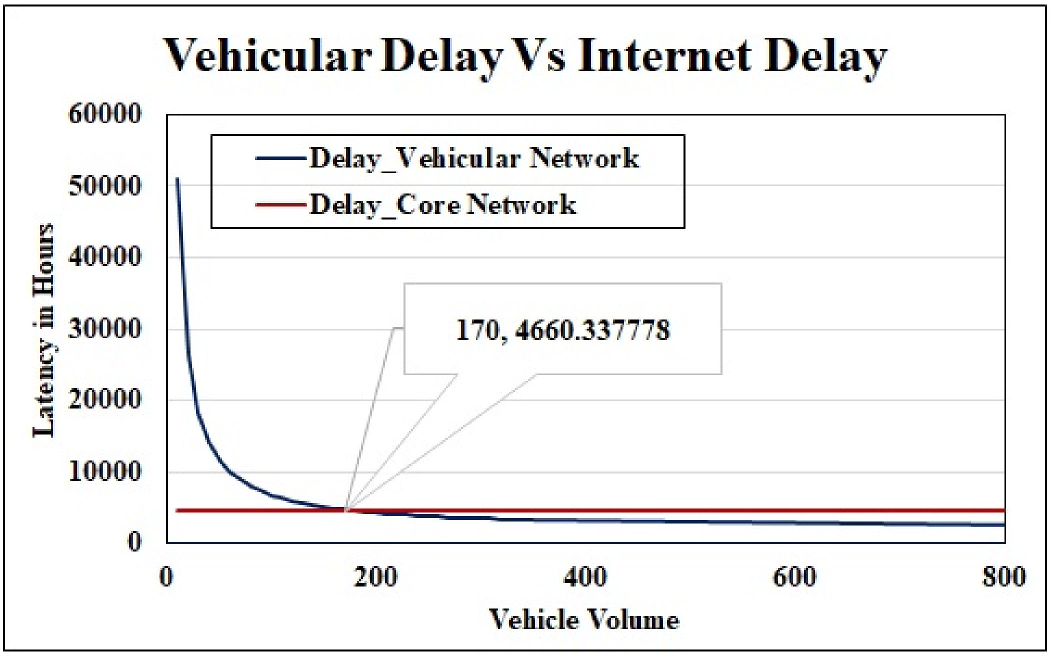


Figure 3.5: Data Volume Vs Delay.

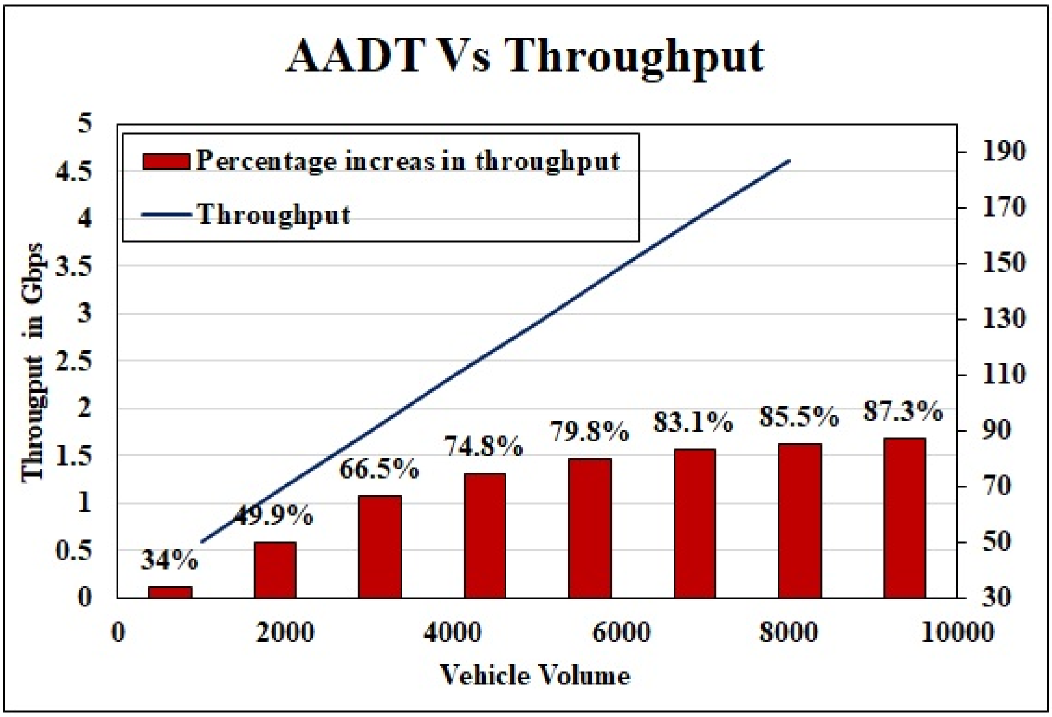
To transfer big data volume we compare internet system having data rate 1 Gbps and vehicular system having disk capacity of 512 GB for each vehicle. In Fig. 3.5, we gradually increase the data volume and examine its effect on data transmission delay. The results show that for big data volume there will be a significant increase in delay for internet case as compared to vehicle case.

Eq. 3.5, demonstrates that the performance of the proposed system depends upon vehicle volume, speed, and distance other than the data volume. Therefore, it is essential to get the effect of these parameters on the system performance.

Vehicle Volume: To get the impact of vehicle volume on the throughput of the proposed system, we transfer 1 PB data from the same source to destination. We vary the vehicle volume from 100 to 850 vehicles per day, by keeping constant the average speed 50 Km/hour, and storage capacity 256 GB for each vehicle. Fig. 3.6(a), shows that at the start latency is very high as compared to internet case and it gradually reduces by increasing the vehicle volume. When the vehicle volume increases more than 170

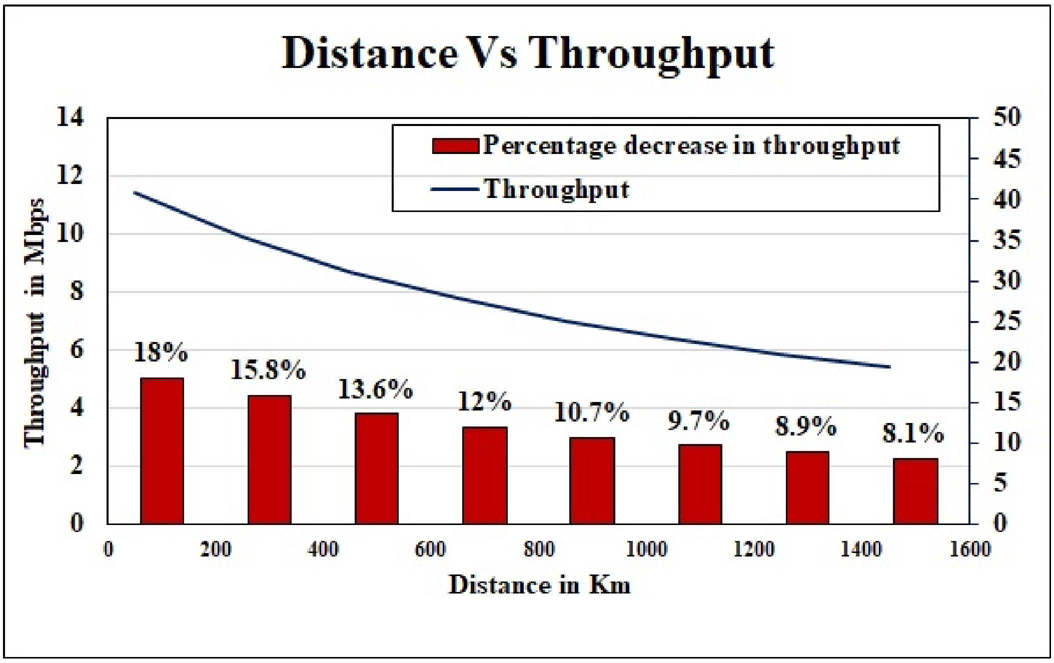


(a) Vehicular network Vs Internet Delay.

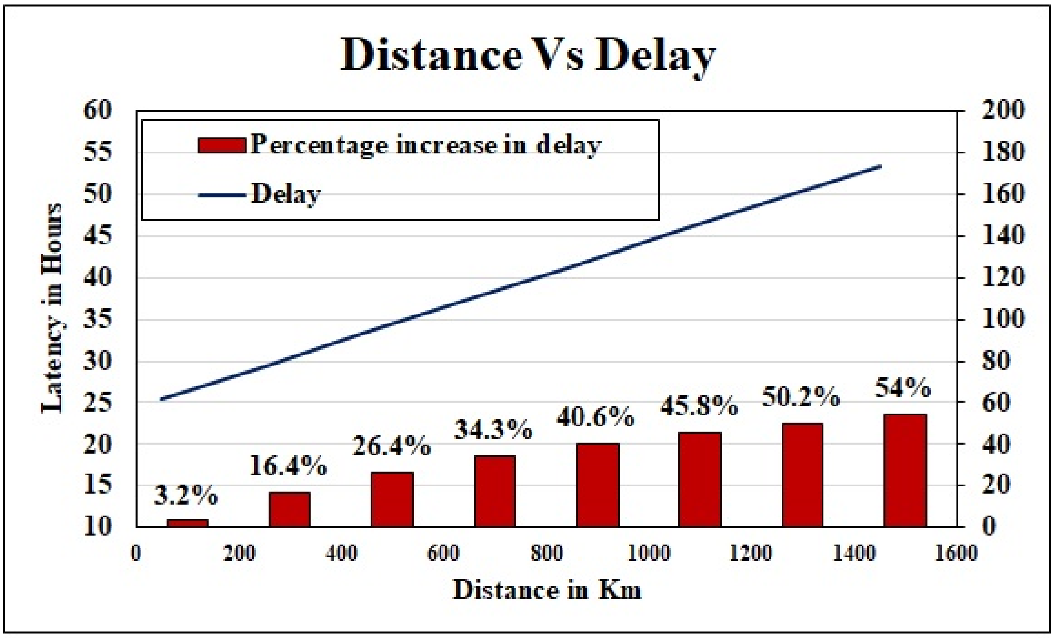


(b) Vehicle count Vs Throughput.

Figure 3.6: Vehicle volume's effect on the performance of the proposed system



(a) Distance Vs Throughput.



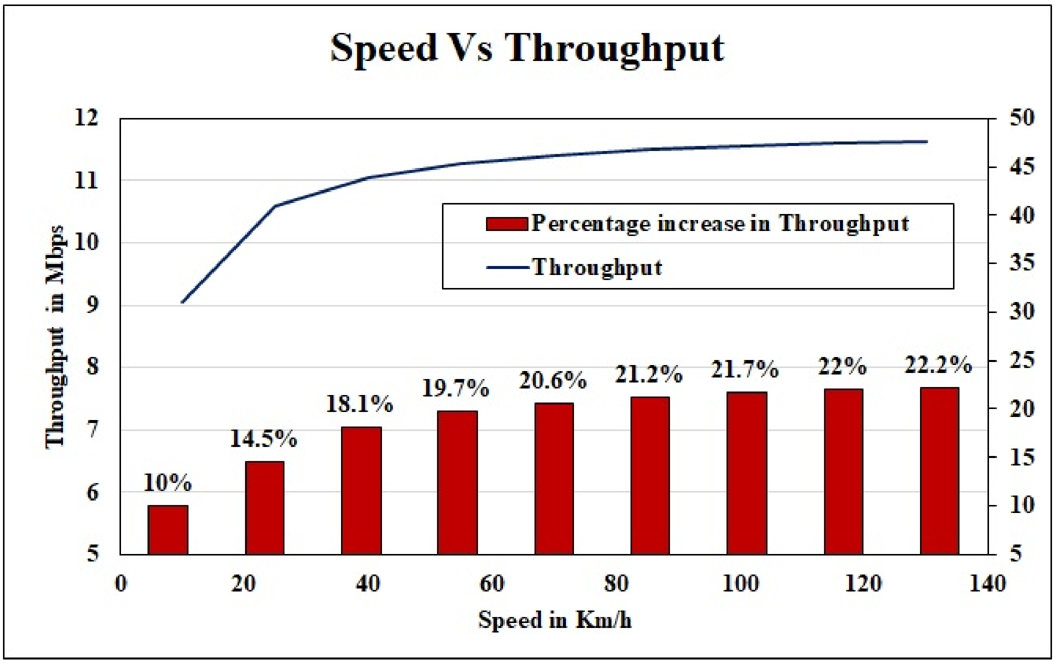
(b) Distance Vs Delay.

Figure 3.7: Effect of distance on the performance of the proposed system

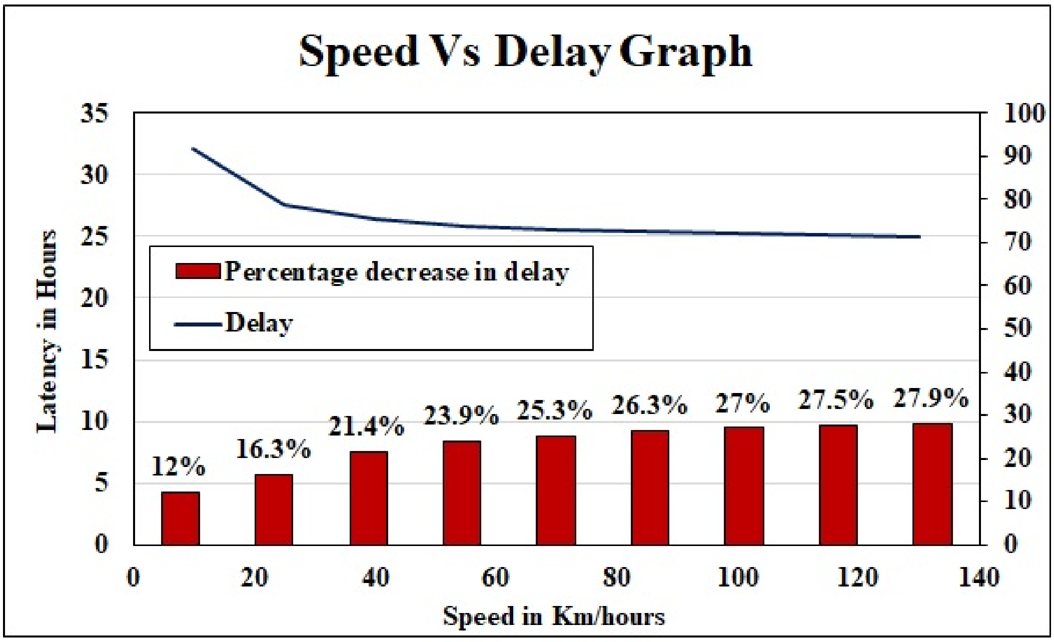
vehicles per day then the latency get decreases in our proposed case as compared to traditional networks. Whereas, Fig. 3.6(b) shows that the throughput of the proposed system is directly proportional to the vehicle count on the road. If more vehicles will take part in the proposed system then there are more chances to transfer big data.

Distance: As the distance between source and destination can vary significantly, we, therefore, tested our system by varying the distance from 50 Km to 1450 Km and by keeping the other parameters constant. The results show that distance is inversely proportional to the throughput of the system, as we increase the distance the throughput of the system decreases as shown in Fig. 3.7(a). On the other hand, Fig. 3.7(b) shows the latency of the proposed system is directly proportional to distance.

Speed: We vary the speed of the vehicle from 10 Km/h to 130 Km/h by keeping the other parameters constant and get the results as shown in Fig. 3.8(a). Results show that speed has a directly proportional impact on the proposed system throughput. Whereas, the latency of the proposed system reduces if we increase the speed, as shown in Fig. 3.8(b).



(a) Speed Vs Throughput.



(b) Speed Vs Delay.

Figure 3.8: Effect of speed on the performance of the proposed system

In next section, how our proposed network bridge the gap between the existing mobile data demands and future mobile data demands of smart cities.

3.3 A Possible Alternate Channel

In this section, we consider the case of smart cities in New Zealand and perform some calculations on the existing vehicle volume in these cities. Here we propose an alternate channel for communication in New Zealand cities by using our proposed VADD.

Like rest of the world, the data demands in New Zealand are increasing each year, as depicted in Figure 3.9. There is an increase of 40% data demand in 2018 and an increase of 54% in the number of fiber connections for faster data communication, to make up 32% of all broadband connections. On the other hand, a significant decrease of 25% is shown on the slow Internet and dial-up connections. More than 70% of all Internet broadband connections with no data limits and cap. These data demands of broadband connections include the use of 281,615,000 GB of data that is equal to 90 million hours of streaming, or 10,700 years, high definition online shows of TV. Mobile phone users consumed 10,089,000 GB data in June 2018, which is up to 56% from 2017 [133]. The use of mobile data has been grown by 600% in the past four years and the expected growth for the next four years is 50% per year [134].

Cisco predicts that a SC with a populace of 1 million could produce 180 million GB data volume for each day or 42.3 ZB/month [135]. If we assume NZ main cities as smart cities, where data sources of different applications of smart cities like smart buildings, smart water, smart public services, smart lighting, smart mobility, smart waste management, smart meters, smart energy, smart sensors, and smart grids etc., will generate a massive volume of data and we need to transmit this big-data to data cloud, data centers or data control units. These expected data demands of 724 PB per day, will exceed the data usage beyond the available limit of existing infrastructure. There will

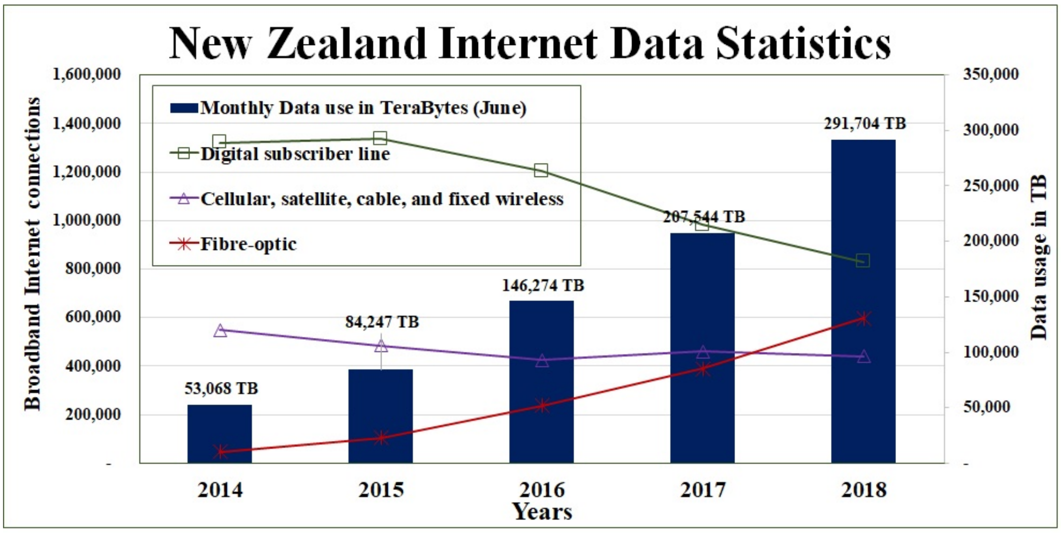


Figure 3.9: Past five years broadband data usage in NZ: Comparison of Fiber optic, Cellular/satellite/ and fixed wireless, DSL, and Monthly data use.

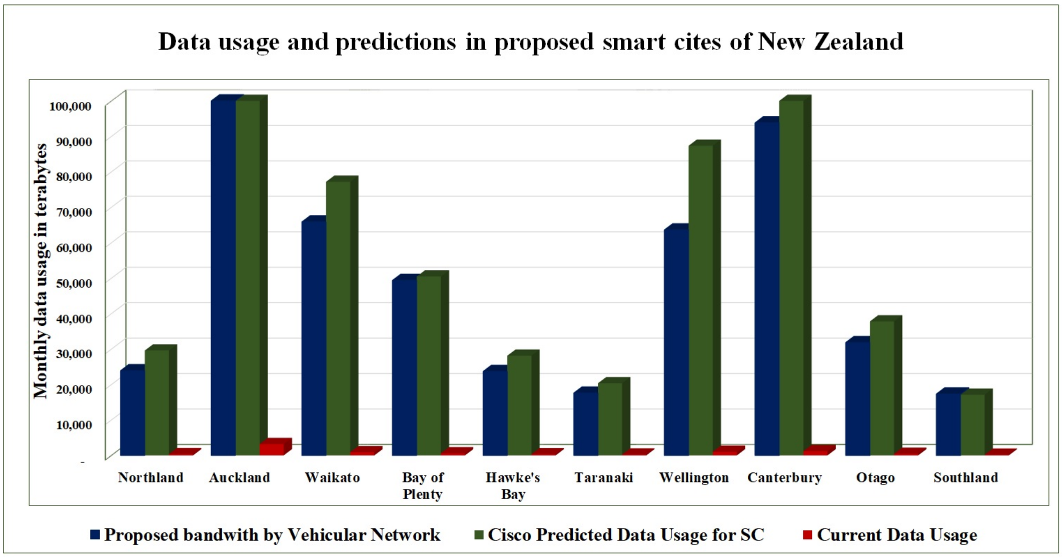


Figure 3.10: Data usage statistics of various NZ cities. Comparison of Cisco predicted, current, and the proposed data usage.

be a massive gap between existing data usage and required data usage of these cities, as shown in Figure 3.10.

Table 3.2: Average percentage increase in per capita vehicles.

City Name	Population	Per capita Vehicle	Number of Vehicles	Yearly %Age Increase
Northland	168,300	0.71	119,493	1.14
Auckland	1,569,900	0.71	111,4629	2.03
Waikato	439,100	0.75	329,325	1.37
Bay of Plenty	287,100	0.86	246,906	1.66
Gisborne	47,400	0.66	31,284	0.62
Hawke's Bay	160,000	0.74	118,400	1.64
Taranaki	115,700	0.76	87,932	0.8
Wellington	496,900	0.64	318,016	1.28
West Coast	32,700	0.90	29,430	1.61
Canterbury	586,400	0.80	469,120	0.25
Otago	215,000	0.74	159,100	1.38
Southland	97,300	0.89	86,597	1.37

We propose a complementary network, the vehicular network as a hybrid network, along with existing infrastructure, and mobile networks. Business Insider (BI), predicted that in 2021, 82% of all vehicles would be shipped as smart and connected vehicles [3]. These smart vehicles will have wireless interfaces, global positioning system, storage capacity, and processing capacity. We assume NZ's main cities as SCs. Ministry of transport NZ predicted that NZ has 0.79 per capita vehicles in 2016. This value keeps on increasing with an average increase of 2.12% in per capita vehicles since the past five years [136] as shown in Table 3.2. Figure 3.10, also predicts the parallel growth in mobile data demands and vehicle volume in NZ. Hence, it is forecasted that this significant volume of connected vehicles can produce huge bandwidth on the road network of NZ's smart cities. By using the data set of the ministry of transport NZ [136, 137] as shown in Table 3.2, our proposed system predicts the vehicular network in SCs of NZ could produce a massive bandwidth up to 607 PB if only 20% smart vehicles in NZ will transfer only one data assignment daily, as shown in Figure 3.10.

3.4 Beijing Case

It sounds exciting, the ubiquitous availability and displacement of taxi cabs can help us in data collection and can move the collected data throughout the smart city. However, we must consider that these taxi-cabs can cooperatively deliver the essential coverage and capacity to fulfil SC requirements. In this regard, we analyse the coverage and capacity of vehicle assisted data dissemination framework, formed by these taxi-cabs by considering their possible contacts in different locations of SC. To evaluate our proposed work we use the real-time taxi traces of Beijing city. At first, we cut the complexity of the proposed network by using a clustering algorithm on big data of taxi traces. In our case, the cluster is a rectangular region from where a vehicle can collect or deliver the data. Moreover, we apply different spatial data mining techniques to filter the data and analyse it for different wireless technologies by using the different sizes of clusters.

To analyse the network coverage and potential we will use the microscopic movement of vehicles based on GPS locations and trajectory traces, for data dissemination. We divide the whole city in smaller grids. Grid is a rectangular region bounded by some longitude and latitude. The length of each grid depends upon the selected wireless technology. In the next section, we describe the model to calculate the effective distance of a vehicle from a wireless device inside a grid. The number of vehicles reported inside a grid form a cluster.

Notations used to calculate effective distance, cluster density, network area coverage and network capacity are given in table 3.3

3.4.1 Effective distance

It is assumed that whenever a vehicle will visit a rectangular grid cluster, it can collect data from all the devices inside that cluster. For this purpose, the size of the grid cluster

Table 3.3: Notations used in this scenario

Symbol	Meaning
d	Effective distance
d_0	Reference distance
n_0	Path loss factor
Ψ	Gaussian random variable
ρ	Vehicle density
r	Length of each side of grid cluster
tid	Taxi id
Lon	Longitude
Lat	Latitude
aid	Cluster area ID
S_i	Size of wireless technology used
t	Unit time interval
m	Total number of clusters in S_i
n	Total number of GPS points in selected area ($25 \times 25Km^2$)
T	Average update time
ST	Stay time in a cluster
θ	Data rate of wireless technology

will be adjusted concerning the selected wireless technology.

To reduce the packet loss, it is essential to keep a vehicle in the appropriate area of a device's wireless coverage region as shown in figure 3.11. We can calculate the path loss $PL(d)$ at a distance d by using path loss formula [138].

$$PL(d)[dB] = PL_F(d_0) + 10n_0 \log{(\frac{d}{d_0})} + \Psi \tag{3.7}$$

Where d_0 is the reference distance where path loss inherits the characteristics of free-space loss PL_F, n_0 is path loss exponent, depending upon the propagation environment. Ψ is the Gaussian random variable.

From Eq. (3.7) we can find the effective distance of a vehicle to wireless device inside a grid.

$$d = d_0 \times 10^{\frac{PL_F(d_0)+\Psi-PL}{10(n_0)}} \tag{3.8}$$

d is the distance the vehicle to a device and effective distance can be $2d$, the diameter of

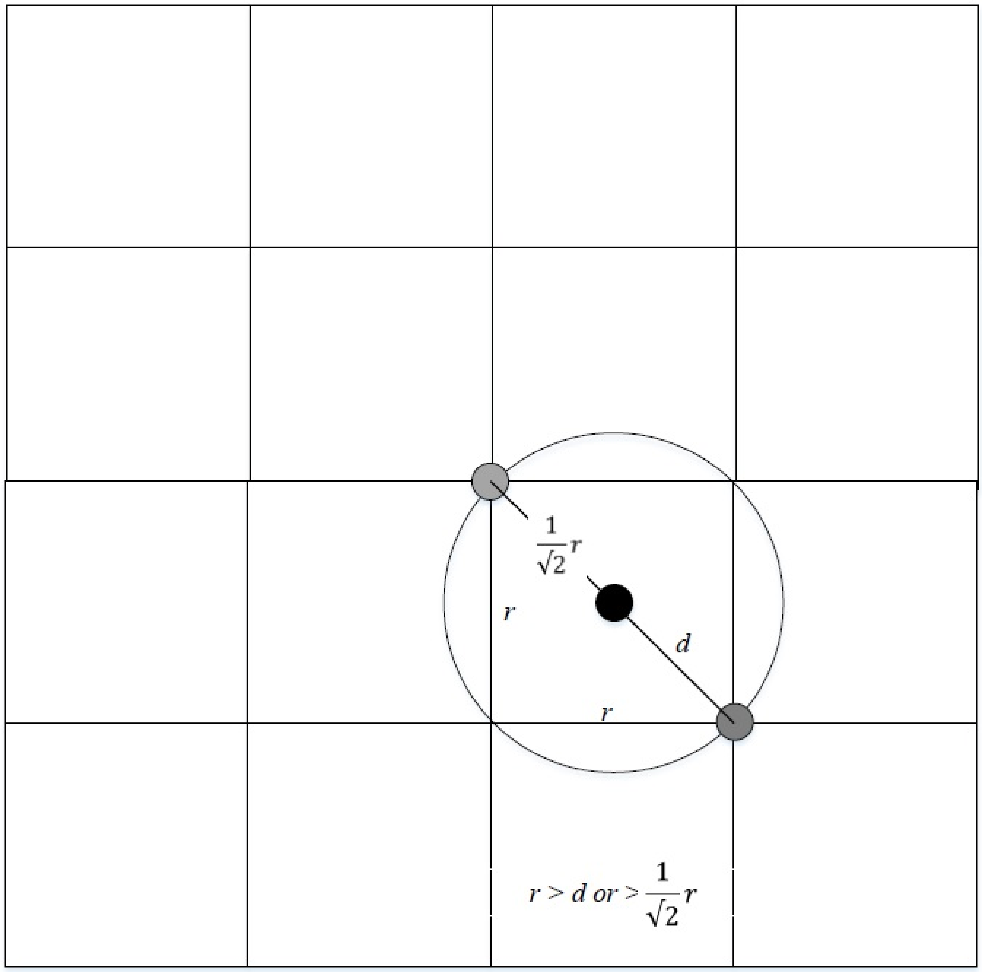


Figure 3.11: Effective distance.

the wireless coverage area of a device. In other words, each vehicle can have a chance to transmit the data packets to a device at a distance from *0 to 2d* meters, as shown in figure 3.11.

To calculate the presence of a vehicle in a given area, we will divide our whole selected area in grids having the length of each side equal to r as shown in figure 3.11. If d is the radius of wireless coverage of a device then maximum length of wireless coverage can be calculated as follow.

$$2d = \sqrt{r^2 + r^2} \tag{3.9}$$

Where r is the length of each side of a grid topology. Eq 3.9 can be written as follow.

$$d = \frac{1}{\sqrt{2}}r \tag{3.10}$$

The length of each side of grid can be calculated as follow.

$$r = \sqrt{2} \times d_0 \times 10^{\frac{PL_F(d_0)+\Psi-PL}{10(n_0)}} \tag{3.11}$$

If r is the length of each side of the grid, then any vehicle at a distance $\frac{1}{\sqrt{2}}\ r$ or double of it is in the wireless range of a device can collect data from that device as shown in the figure 3.11.

In the next section, we describe our algorithm for this division related to different wireless technology.

3.4.2 Grid clustering algorithm

The cost to evaluate all possible taxicabs encounters with geographically stationary sensors is high. In this way, it is important to process, for every taxicab and for all time intervals, the area as characterized by the wireless coverage under consideration. For this situation, as the databases of vehicle trajectories are of noteworthy size, the procedure of complex mining is expensive and tedious. The data set is made out of GPS locations of taxicabs during the whole day's intervals. There are various approaches to dissect huge measures of spatial data, like sampling an array of GPS points, decreasing the investigated area, clustering, measuring the distance between all GPS locations, to cite a few. One of the main objectives of this work is to examine the reasonableness of wireless technology along with its range, that's why we use grid clustering method, it garbs well our proposed study and problem.

10) By inspiring the grid clustering approach for useful data analysis of geospatial

data, we use STING algorithm and reduce the complexity of big data of vehicular traces. [12]. Our grid clustering algorithm 1 will be used for making rectangular and spatial clusters. It will also reduce the complexity of big data analysis. We give taxi ID, date time, location, starting value of longitude and latitude and the set of different wireless technologies coverage range as input to our algorithm 1. Our algorithm will return clusters of different sizes according to given wireless ranges.

Algorithm 1: Grid clustering algorithm (GCA)

Input: Set of vehicle traces Trace={(tid,datetime,lon,lat)}, Set of initial values V={$(lon_s, lat_s, lon_i, lat_i)$}, Set of all sizes S = $\{s_i\}$

Output: Set of vehicle traces with along smaller area ID's $Trace_{Area}$={(tid,datetime,lon,lat,aid)}

$lo \leftarrow lon_s$
$la \leftarrow lat_s$
while $i <= s_i$ **do**
  $la_1 \leftarrow la$
  $la_2 \leftarrow la + lat_i$
  while $j <= s_i$ **do**
    $lo_1 \leftarrow lo$
    $lo_2 \leftarrow lo + lon_i$
    for *new cell values* **do**
      update Trace
      set $aid++$
      where lon between lo_1 and lo_2 and lat between la_1 and la_1
    end
    $lo = lo + lo_i$
  end
  $lo \leftarrow lon_i$
  $la \leftarrow la + la_i$
end
return $Trace_{Area}$

3.4.3 Cluster Density

Density of a grid having size S_i is the number of records mapped or updates reported in that grid until a time t. it can be written as follow.

$$Density(S_i, t) = \mid M(S_i, t) \mid \tag{3.12}$$

The sum of all densities of all clusters m in each time t can be calculated as follow.

$$M_{Density}(t) = \sum_{i=1}^{m} \mid M(S_i, t) \mid \tag{3.13}$$

We use algorithm 2, to estimate vehicle density in each cluster area of size S_i. It is the number of vehicle visits in each cluster. A cluster having more vehicle visits means a high dense cluster, whereas low dense cluster means it has a smaller number of vehicle visits. High dense clusters are more popular areas, can have a high potential for data dissemination and they can be used as a location for RSU installation.

Algorithm 2: Cluster Density

Input: Set of vehicle traces Trace=$\{(tid, datetime, lon, lat)\}$, Set of all sizes S = $\{s_i\}$, Total updates $Count_i$

Output: Set of clusters with vehicle density $Trace_{Area\_density}$ = $\{(aid, count_i)\}$

initialize();
$Trace_{Area} \leftarrow Algorithm\_GCA()$
for $S_j \in S$ **do**
    $Count \leftarrow 0$
    for $\forall aid \in Trace_{Area}$ **do**
        if $tid \in Trace_{Area_i}$ **then**
            $++Count_i$
        else
            do nothing
        end
    end
    $Trace_{Area\_density} \leftarrow \{aid, Count\}$
end
return $M_{Density}(t)$

3.4.4 Network area coverage

Algorithm 3: Area Coverage

Input: Set of vehicle traces Trace=$\{(tid, datetime, lon, lat)\}$, Set of all sizes S = $\{s_i\}$, Area Visits $Count_i$, Set of time intervals T = $\{t_1, t_2, ..., t_n\}$

Output: Percentage coverage of clusters in a given time interval $Coverage(S)$

initialize();

$Trace_{Area} \leftarrow Algorithm\_GCA()$

for $S_j \in S$ **do**

  for $t_i \in T$ **do**

    $Count_i \leftarrow$

    select count (Distinct (aid))

    from $Trace_{Area}$

    where aid is not NULL

    and time = t_i

    $Coverage(S_i) = \frac{\sum Count_i \times 100}{Total_{aid}}$

  end

  $Coverage(S_j) = \frac{\sum Coverage(S_i) \times 100}{Total_{aid}}$

end

return $Coverage(S)$

If vehicles cover k clusters in time t then we can calculate percentage coverage by using equation 3.14.

$$Coverage(S) = \sum_{i=1}^{k} \frac{S_i \times 100}{Total_{aid}} \tag{3.14}$$

We use algorithm 3, to measure the coverage of clusters S_i in our selected area by vehicle displacement in time T. The time T is divided into equal smaller intervals t_i as $T = \{t_1, t_2, ..., t_n\}$. Some areas where vehicles are not allowed to visit have zero coverage.

3.4.5 Network Capacity

Network capacity is the total potential of our proposed system. It is the total data that can be disseminated by considering all possible contacts in each cluster of our selected

Algorithm 4: Network Capacity

Input: Set of vehicle traces Trace=$\{(tid, datetime, lon, lat)\}$, Set of all sizes S = $\{s_i\}$, Total updates $Update_i$, Average updae time T, Set of data rates R = $\{r_i\}$, Stay time ST

Output: Data transfer capacity of the system $Capaciyt(S)$

initialize();
$Trace_{Area} \leftarrow Algorithm\_GCA()$
for $S_j \in S$ **do**
    $Count \leftarrow 0$
    for $aid \in Trace_{Area}$ **do**
        $Update_i \leftarrow$(
        select Number of updates
        from $Trace_{Area}$
        where aid is not NULL
        and time = date and time)
        ST_i = $Update_i \times T$
    end
    ST_j = $\sum ST_i \times T$
    $Capacity(S_j)$ = $\theta \times \sum ST_j$
end
return $Capacity(S)$

area. These contacts can be calculated by the displacement and movement of vehicles in different clusters of size S_i. By getting the total number of possible contacts or updates, we can calculate the total stay time ST of each vehicle in each cluster. Finally, the capacity of the network, with the data rate θ of each selected wireless technology in each cluster of size S_i can be calculated as follows.

$$Capacity(S) = \theta \times T \times \sum_{i=1}^{m} \mid M(S_i, t) \mid \tag{3.15}$$

Where T is the time duration of the average update interval. We can calculate the total amount of data dissemination by using algorithm 4.

3.4.6 RSU placement

RSUs have storage processing and communication capabilities and are directly attached to CC. To transfer data from a service provider to smart sensors we can use RSUs. The service provider can send the data e.g software update for sensors to CC, CC forward this data to RSUs by using some traditional network e.g cellular network. So vehicles can easily get the data from the RSUs and transfer the data to SS. Hence, it is important to deploy the RSUs at important locations. In this work, we consider the most visited locations of a geographical area visited by vehicles and use algorithm 5 for RSUs placement, the greedy approach for most visited locations.

Algorithm 5: Greedy Algorithm for RSU placement

Input: Set of vehicle traces Trace=$\{(tid, datetime, lon, lat)\}$, Set of all sizes S = $\{s_i\}$, $k \leftarrow Number\_of\_RSUs$

Output: Set of most visited clusters $Trace_{RSU\_Locations}$ = $\{(aid, lon, lat)\}$

initialize();
$M_{Density}(t) \leftarrow Algorithm\_ClusterDensity()$
$n_c \leftarrow\mid M(S_i, t) \mid$ // number of clusters in S_i
for *i=1...k* **do**
    for *j=i+1...*n_c **do**
        if $(M_{Density}(t)[j] > M_{Density}(t)[i])$ **then**
            $temp = M_{Density}(t)[j]$
            $M_{Density}(t)[j] = M_{Density}(t)[i]$
            $M_{Density}(t)[i] = temp$
        else
            do nothing
        end
    end
    $Trace_{RSU\_Locations}[i]$ = $M_{Density}(t)[i]$
end
return $Trace_{RSU\_Locations}$

3.4.7 Performance Analysis

To evaluate the coverage, capacity, and performance of our proposed vehicular system for data collection, we selected four subsets of 400 random taxis from Beijing city, having average update time less than 30 seconds. The most important of taxi fleet is that taxicabs can reach to each street at the exact destination/origin of the riders and can provide more excellent coverage. In this section, we will briefly describe the data set, locations with high potential for data transmission, coverage and capacity of the proposed system.

1. Beijing taxi traces overview

Taxi traces of Beijing city are available at Microsoft repository [139], as a T-Drive data set by MSRA. This data set contains the information of 10375 taxis trajectories, having GPS position of taxis in the most populated and congested city of Beijing. The total number of GPS updates in this data set reaches 15 million points, and these taxi trajectories cover a distance up to 9 million kilometres. Figure 3.12 is the one day's visual representation of these traces, the heat map colours represented the different densities of taxis on Monday, February 04, 2008. We draw the heat map by using a MATLAB function Scatter Plot coloured by Kernel Density Estimate to calculate the vehicle density to each GPS location [130]. Figure 3.12, shows that it is difficult to analyse each possible contact in a selected geographical area, that's why we use our grid clustering algorithm to reduce the complexity and calculated possible contacts in each clustered area rather than at each GPS location. The average update time in this data set is 177 seconds; the average distance between the consecutive points is 623 meters. The average speed of the taxis in these traces is 12.67 km/h, which is nearly equal to 7.5 miles/h, that is validated at [140]. This implies that these distinct points have enough data to represent the persistent trajectories of taxicabs [141][142]. Each text file of this repository, name by the ID of the taxi, contains the geographical information of each

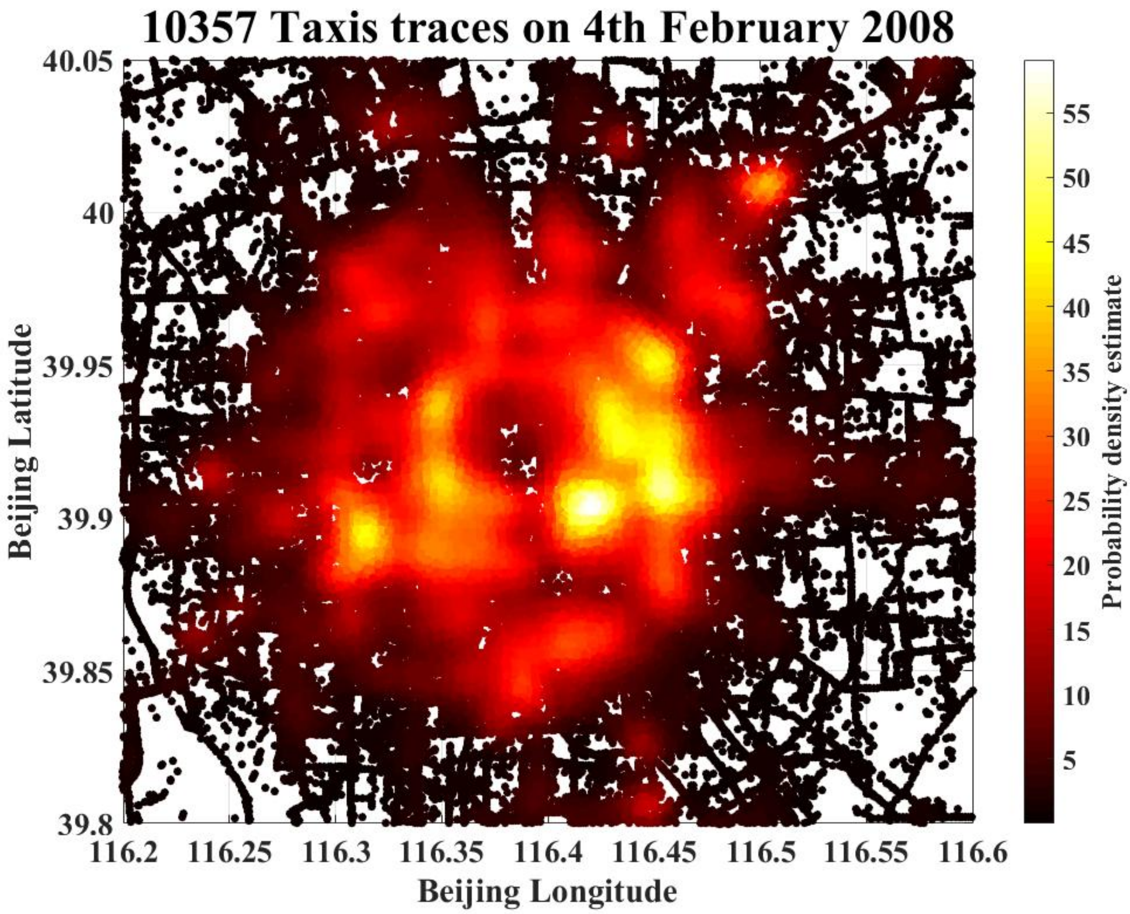


Figure 3.12: Heat map of vehicle density in Beijing selected area of $25 \times 25Km^2$: a visualization of T-Drive Traces data set.

taxicab.

2. Area selection for analysis

For our proposed network performance, coverage and capacity, we have selected an area of 25×25 km^2 having longitude between $116.24°$ E to $116.5335°$ E and latitude between $39.8125°$ N to $40.04°$ N. The initial step of our methodology is to apply grid clustering algorithm on taxi traces to make a graph having every vertex as a geographic area of the selected city called a grid. Every vertex has a weight, that is equivalent to the total number of taxicabs reported inside that grid. After girds creation, the whole area is separated into equivalent measured quadrants. Finally, the GPS locations of

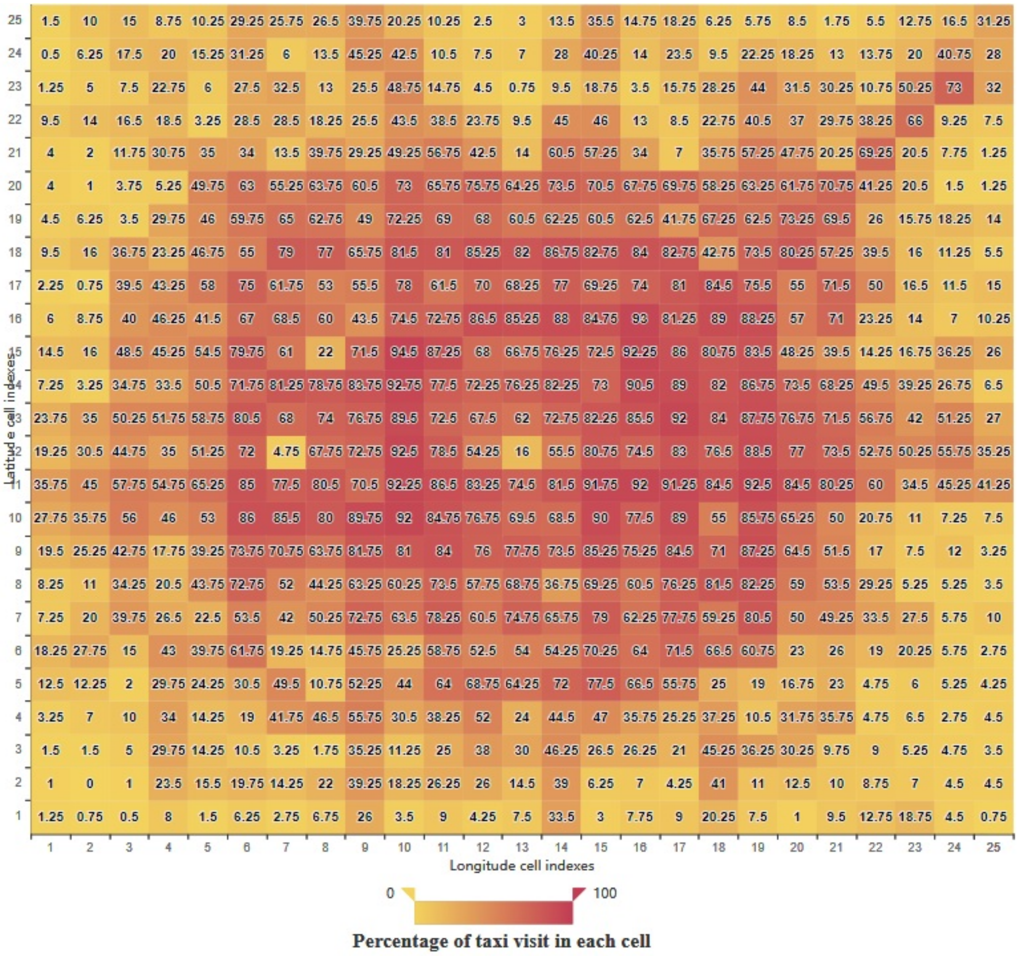


Figure 3.13: Heat table: Vehicle density in each cluster of size $1000 \times 1000m^2$.

each taxi are stored in the database, along with their quadrants (grids), as shown in figure 3.13. A grid is a small region of the city bounded by a square, having a side length r, demarcated by GPS locations of taxis. If n is the total number of GPS points, then during association phase, our algorithm has $O(n)$ complexity, at higher levels this complexity condensed to $O(m)$, where m is the total number of grids in the target region. This is the basic advantage of grid clustering algorithm, it has less complexity because of $m << n$.

It is important to choose the grid size, so that is neither too big nor too small, which influence the performance of our proposed work directly for evaluation of radio ranges of chosen wireless technologies. We analyse the coverage and dynamics of the proposed network by using the four random subsets of 100, 200, 300 and 400 taxi cabs from the above-said data set Beijing Taxi traces, by considering the wireless range of different wireless technologies as shown in table 3.4.

Table 3.4: Cluster size selection with a coverage range of different wireless technologies

ID	Cell Size	Area	Total Cells	Lon. Inc.	Lat. Inc.	Technology
1	1000m^2	25×25	625	0.0174	0.008992	IEEE 802.11p
2	500mm^2	50×50	2500	0.00587	0.004496	IEEE 802.16
3	250mm^2	250×250	10000	0.002935	0.002248	IEEE 802.11n
4	100mm^2	100×100	62500	0.001174	0.008992	IEEE 802.11ay

3. Vehicle Density

To evaluate the network dynamics, it is important to stop at each desired layer of algorithm 1, where the size of the layered grid matches with the radio range of wireless technology. Based on this division, we got a heatmap table of 400 cabs visits in each grid of 1000 m^2 size, as shown in figure 3.13. The colours of the heat table show the vehicle density of cabs in all grids. The cab density varies from yellow grid 0% to the darkest grid of 94.5%. The GPS locations having a great human index, have more cab visits. More vehicle visit's in a grid means more chance to transfer or receive data from sensors to vehicles in that grid. The maximum degree of a grid is 378, the average degree of a grid is 166. The percentage of grids having a degree greater than the average value is 48.16%. Having an average degree of 166 means that the network may be highly connected.

The connectivity of the proposed network can be illustrated by quantifying the displacement of taxi cabs in the city during some specific periods. By analysing GPS locations, it is promising to categorize patterns which are distinctive of Beijing transportation network. By this analysis, we can find the areas or grids where taxicabs stay a longer period of times. These can be taxi stops or highly congested areas. One of the main objective of this work is identifying the locations, with high potential for data transmission. Hence, clusters having high density have a high potential for data

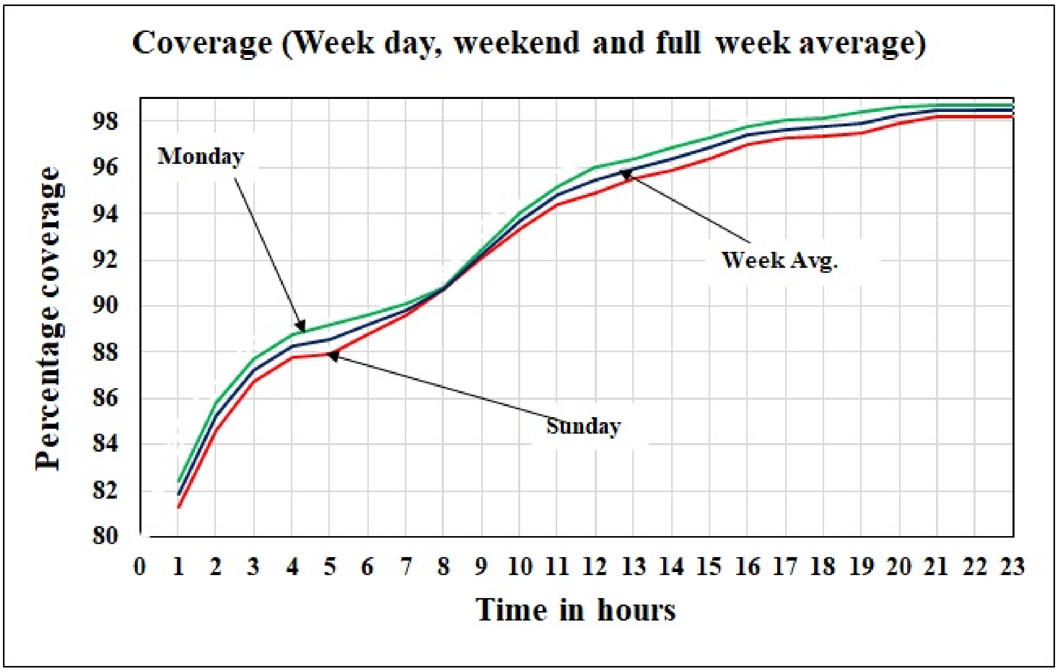


Figure 3.14: Week day, weekend and average week coverage.

dissemination.

4. Network area Coverage

We evaluate the percentage of grid coverage of 1000 m^2 by using a set of 400 random vehicles, for a given interval. Figure 3.14, compare the grid coverage percentage of a working day, a weekend day, and the weekly average. In all three cases, there is a big coverage of the area; almost it is 95% area coverage after a half-day. Some of the grids can never be visited, e.g., the grid with ID 27. So the coverage can never be 100% in this given case.

Figure 3.15 depicts the area coverage in 24 hours by the displacement of 400 random taxi cabs of the given data set. In this experiment, we investigate the coverage concerning different grid sizes varies from 100 m^2 to $1Km^2$. In the case of $1Km^2$ grid size, the coverage of the area varies from 80% to 98.5%. Almost 1.5% of the grids can never be visited by these cabs. By geographical inspection of the selected area, it is found that these grids are in the regions where taxi cabs and vehicle movement is not

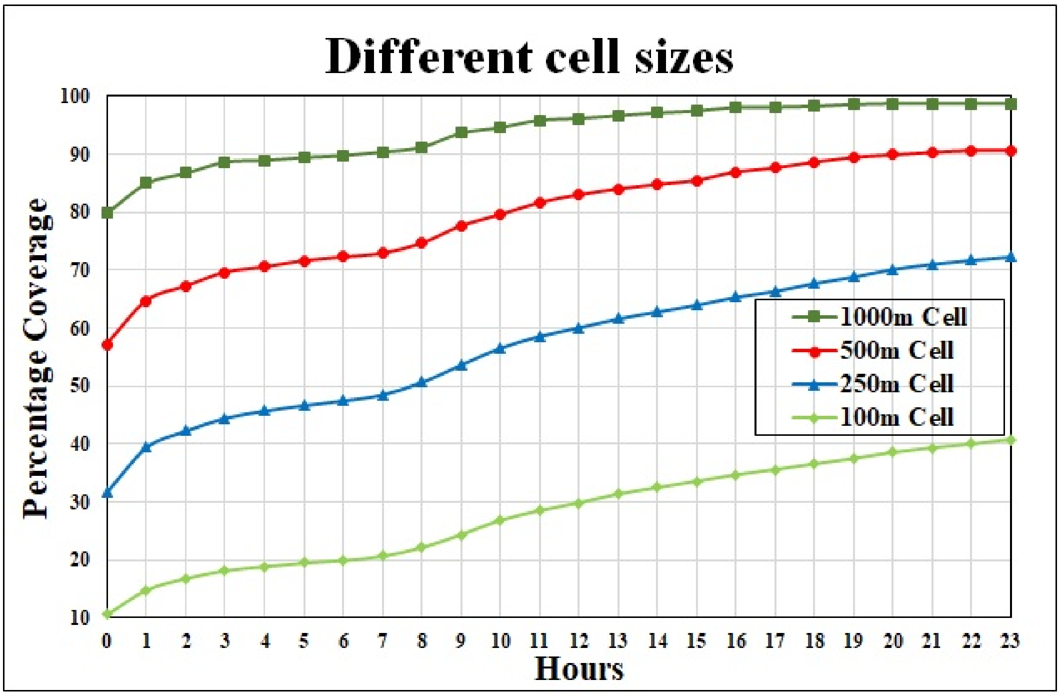


Figure 3.15: Coverage with different cell sizes.

allowed, for example, old cemeteries, big private areas, public gardens, train stations, rivers, hotels, etc. Moreover, when we reduce the size of grids, the coverage of the area also reduces, as shown in the above figure 3.15, with grid sizes $500m^2$, $250m^2$, and $100m^2$. The smallest grid size division $100m^2$ gives the smaller coverage; it varies from 10% to 40% in 24 hours. The average road area in Beijing city is also 26% [143]. In this division, we found that only those grids are covered, which are on the road or near the coverage of the road. All the smaller grids which are away from the road can never send their data directly to the vehicles because they are not in the wireless range of the vehicles having the technology of 100 m^2 wireless range. However, these areas can also be covered if we use the clustering approach of wireless sensor networks. All the sensors, away from the road, will route their accumulated data to their cluster heads periodically. These cluster heads are installed in those grids, which are in the coverage of some road.

Figure 3.16, depicts the coverage of the given area with the different number of

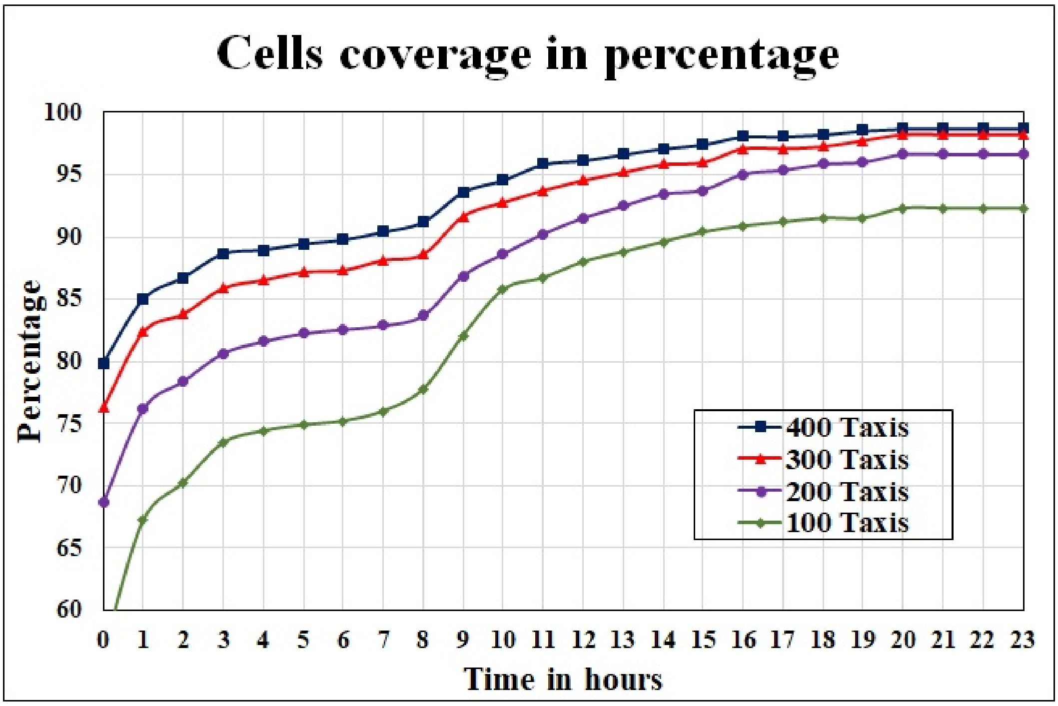


Figure 3.16: Coverage with different taxi cab subsets.

taxi cab sets in 24 hours, on 04 February 2008. This experiment shows that when we increase the number of vehicles, the coverage also goes on increasing. The subset of 100 taxicabs gives coverage from 60% to 92%, of the given area, whereas the bigger set having 400 random cabs give higher coverage from 80% to 98.5% of the given area.

As a last comment, we bring up that we limit our investigation to an urban region of $25 \times 25\ Km^2$ area; one could criticize that this situation won't be relevant in hardly populated regions or rural areas. This objection is valid, without any doubt. Though we think that the locations where more taxicabs move, frequently compare to locations where more information is produced/consumed and at those locations usually the data dissemination is progressively critical. This reflection is true likewise if we consider the different regions of the same city. Furthermore, on the off chance that we consider for example uncommon occasions like domestic carnivals or open exhibitions, we observe that a more noteworthy number of taxis in that area compares to an extended need of data communications among "things"; for example, to report the load levels of trash cans, that they are full more quickly. The outcome is that mobile nodes (cabs) arrangements could likewise give a sort of automatic solution to bring greater capacity where and when it is required.

5. Network Capacity

To measure the data transfer capacity of the network, we assume that each grid has multiple sensors. These sensors can communicate directly with the vehicle inside the grid. Every vehicle can store and forward the data collected by these sensors by making a wireless secession with them inside each grid. Vehicles gathered data from these sensors when they encountered during their routine travel. Finally, they upload to some cloud through some wireless access point called roadside unit (RSU), in a grid. We make no further hypothesis, as far as the applications are concerned, that produce/consume the data. Additionally, by estimating the duration vehicles remain in the radio range of a grid, we can measure the capacity of that grid in term of data transmission and the delay associated with data transmission over the proposed vehicular system.

In conclusion, the estimates of these analyses can be matched with the requirements of different applications that can be supported by the proposed vehicular network. By knowing the requirements of different smart city data applications, for example, the amount of data generated from the smart sensors, the delay-tolerant intervals, it is conceivable to decide the radio technology that can support it. Data transmission capacity is a key component to measure the application requirements of a smart city that can be supported by the proposed vehicular communication. For this purpose, we can associate the total time a taxicab remains in the wireless range of the grid with the time to travel between source and destination grids. That's why we need the number of updates recorded by each vehicle in each grid and the total time a vehicle remain inside each grid.

As each taxicab update its location after some time, and the average update time is 30 seconds for Beijing taxi traces in our selected set [139]. After the formation of grids by our grid clustering algorithm, we got the location updates of taxis in each grid, over the time intervals of 24-hour. When we increase the radio range of technology, then we get a higher number of vehicle contacts inside the grid. In our analysis, the biggest cluster got 12260 number of updates by taxicabs in a day having grid size 1000 m^2. When we decrease the grid size, we get a smaller number of updates. The grid size 500 m^2 receives 9483, 250 m^2 receives 9415, and 100 m^2 receives 9248 number of updates respectively.

Figure 3.17 represents the total time in seconds that all vehicle remains inside a grid, in the time interval of 24-hours, all grids are considered. This sum compares the absolute time that these taxicabs can use to offload information to the system in 24 hours. Each group of vertical bars relates to a different radio range, varying from 100 m^2 to 1000 m^2. A rectangular cluster is created by all taxicabs that update their position inside a grid. In this manner, if a taxicab remains for a longer time in a cluster, it implies that it appears more than once in this grid. Since these clusters are formulated by the number of updates reported by taxicabs, there is an immediate connection to staying time. Besides, there is an association with the number of taxicabs inside a grid, as a taxicab has a limited number of updates in each day, thus a grid cluster having more updates forms the most active area of the city.

To calculate the aggregated data volume that can be collected in an interval of 24-hours, we just need to utilize the multiplying factor, the data rate related to each radio technology and that all taxicabs remain in the wireless coverage area of the grid cluster, introduced in figure 3.17. It is difficult to determine this multiplying factor accurately because there are many other variables involved to be considered for actual calculations. That's why we get help from the literature, gives a few reflections on the throughput of data communication in vehicular systems.

A. Abdelgader et al. [144], define the data rates of IEEE 802.11p, 9, 18, 36, 48, 54 Mbps, by using different modulation techniques, with a wireless range up to

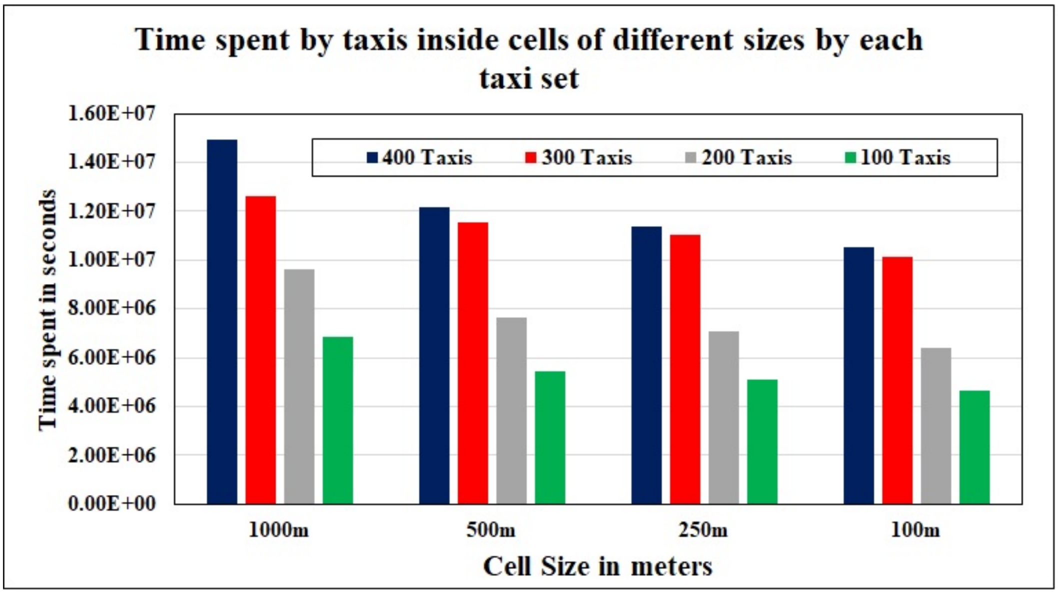


Figure 3.17: Stay time inside cluster of different sizes with each taxi cab subset.

1000 m^2 outdoor and frequency band of 20 MHz bandwidth. Vanhatupa et al. [145], calculated the data rates of IEEE 802.11n 600 Mbps, with frequency range 20MHz to 40 MHz, modulation MIMO-OFDM and wireless range are up to 250 meters outdoor. G. Yasaman et al. [146], define the data rate of IEEE 802.11ay up to 100 Gbps. The data rated defined in [147], 20 Gbps in an outdoor wireless range of 100 m^2, with frequency band 8000 MHz, and OFDM modulation. IEEE 802.16, a metropolitan area network protocol, WiMax (Worldwide Interoperability for Microwave Access) provide the mobile and fixed internet access; it can provide data rate up to 1 Gbps with a frequency band 2 GHz and 11 GHz [148].

The literature shows the great capability of data transfer among vehicles and fixed infrastructures. Each of these studies uses different conditions and equipment which outcomes in a distinctive test-bed. It is difficult to fix the multiplying factor for throughput of data in V2I case, because of different conditions and equipment. Hence, supported by the literature, for SC device to vehicle and the vehicle to the roadside unit, we select IEEE 802.11p for 1000 m^2 cluster, IEEE 802.16 for the 500 m^2 cluster, IEEE 802.11n

Table 3.5: System capacity with various cluster sized and data rates of selected wireless technologies.

Cell Size	Technology	Data rates	Data transfer capacity in PB			
			400 Taxi	300 Taxi	200 Taxi	100 Taxi
1000mm^2	IEEE 802.11p	54	0.133	0.103	0.065	0.036
500mm^2	IEEE 802.16	1000	2.463	1.901	1.204	0.670
250mm^2	IEEE 802.11n	600	1.477	1.141	0.722	0.402
100mm^2	IEEE 802.11ay	20480	51.135	38.966	24.661	13.733

for 250 m^2 cluster and IEEE 802.11ay for 100 m^2 cluster. In this case, the multiplying factors are, 54 Mbps for IEEE 802.11p, 1Gbps for IEEE 802.16, 600Mbps for IEEE 802.11n and 20 Gbps for IEEE 802.11ay, as seen in [144, 145, 146, 147, 148].

All selected sell sizes, technologies, and multiplying factors are shown in table 3.5. By applying the multiplying factor with the total stay time of vehicles in corresponding grids, we can get the potential of the system against each selected set of vehicles. For example, in a grid size of 1000 m^2 and throughput of 54 Mbps, IEEE 802.11p can reach up to 0.133 PB in case of a bigger set of taxicabs. As the size of the grid grows, the number of updates by vehicles in that grid also grow, which implies the greater stay time in that grid, as shown in figure 3.17. But, on the other hand, the table shows that the throughput of selected technologies increases in case of smaller grid sizes, where taxicabs have less stay time. For example, with IEEE 802.16, the capacity reaches up to 2.463 PB. But in this case, it is the most expensive technology and requires a huge infrastructure to implement. IEEE 802.11n provides 1.477 PB capacity with 250 m^2 grid size. With the smallest grid size of 100 m^2, IEEE 802.11ay can reach up to 13.733 PB with the smallest set of taxicabs. These results show that the capacity of the system not only depends upon the vehicles stay time in a cluster but also depends upon the technology selected.

Figure 3.18 explains the stay time of all selected set of taxis in a 1000m cell for the period of each 1-hour during the whole day. It is noted that the time of the day

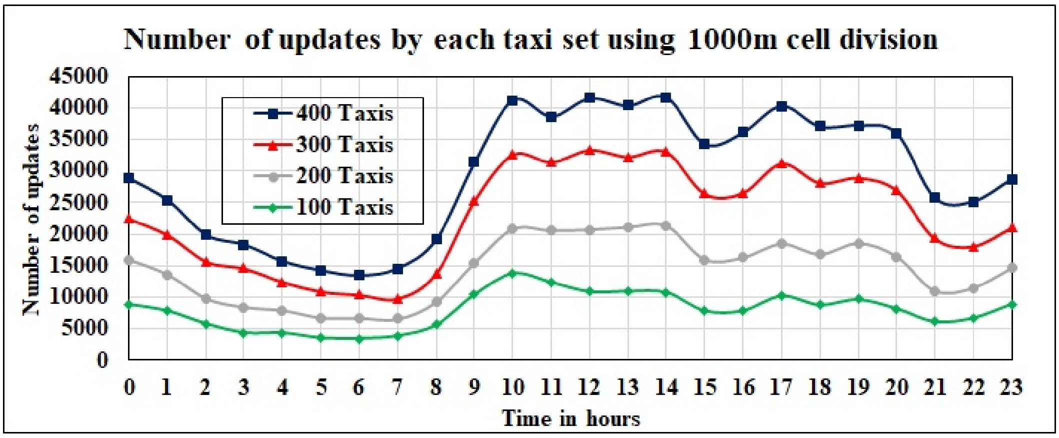


Figure 3.18: Number of updates by each taxi set using 1000 m^2 cell division.

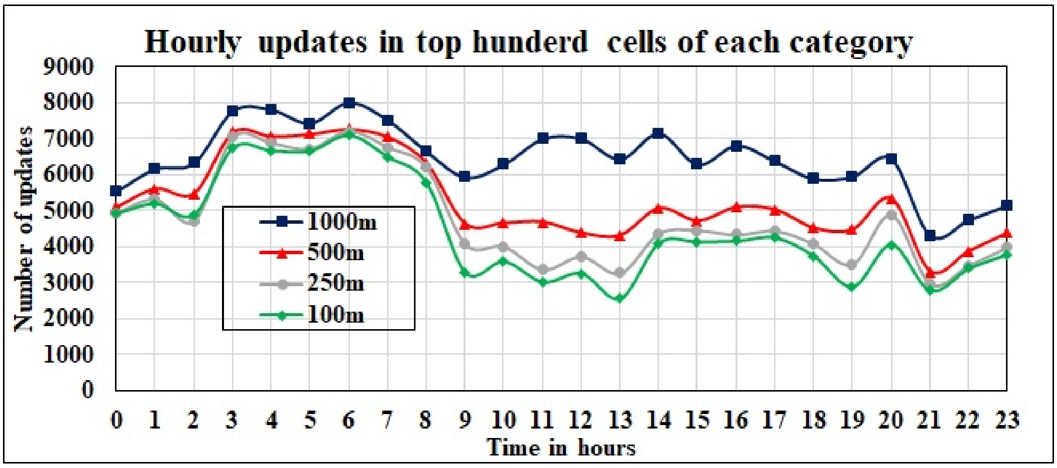


Figure 3.19: Hourly updates in top hundred cells of each category.

impacts the system capacity. During the rush hours, especially in business hours, there is a significant increase in traffic congestion. On those hours, there are more taxicabs to satisfy the need of the people, moving between their homes to the work locations, students heading off to the school, to refer to a couple. Whereas the period from 9 pm to 9 am is less congested leads to less capacity.

On the other hand and interesting factor is shown in figure 3.19, it gives the hourly updates of the biggest set of taxicabs during the whole day, in the top 10 clusters of each size. It shows that the number of updates from taxicabs is highest from 3 am to 8 am. At this time interval, the taxi movement is slow, and they remain in most popular areas of the city, where job ratio remains high. This time interval is more suitable to transfer data from taxicabs to data centers, which they have collected during the whole day. This fact can help us to deploy the RSU network for data centers as well as a suitable time for data offloading at data centers.

3.4.8 Data offloading

To evaluate the data offloading performance of our proposed system, we envision a case, a new software update for SC devices will be available at some service provider at some given time. The service provider needs to forward these updates to all devices. One of the cases is, the service provider will send this data directly to devices. For this purpose, each device should have a SIM card to connect it with the cellular network or it should be directly attached to the internet. This case requires a high cost for infrastructure deployment. In our proposed solutions, the service provider will forward this update to CC. From CC data will be forwarded to SC devices, we consider two cases for the evaluation of our proposed frameworks. The cluster size for this evaluation is considered equal to 1000 m^2 and number of taxi-cabs are 400.

1. Direct communication.

As we assume that the vehicles in a smart city are smart. These vehicles are connected to the internet and can directly receive this software update from CC through the internet. CC selects a set of vehicles and forwards this update to those vehicles. After getting this data, vehicles start the diffusion process and transmit this data to SC devices whenever they are in their wireless range. In this diffusion process, vehicles should be helpful as they use their resources in terms of internet connection, yet this won't impact their trajectory.

2. Using RSUs.

In this case, whenever CC will receive the software update, it will forward this data

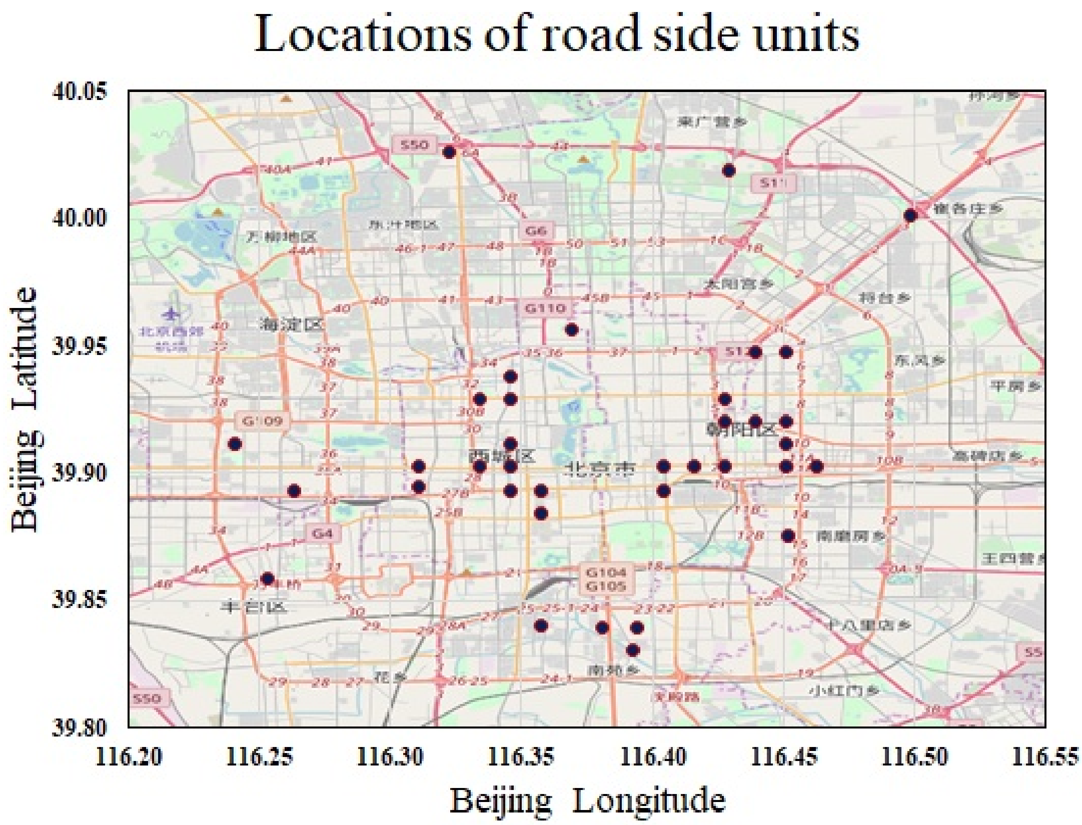


Figure 3.20: Location of road side units.

to RSUs. Now vehicles will pick the data whenever they are in the wireless range of these RSUs and start the diffusion process. After collecting data from RSUs, vehicles transfer data to SC devices, whenever they are in their wireless range. In this case, we apply the greedy algorithm 5 on Beijing city traces, to identify the most popular location of the selected area as shown in figure 3.20. We consider the case of 1000 m^2 grids and identify 36 locations. We assumed that RSUs are installed at these locations and they are directly connected to CC by the internet. In this case, there is no need to share the internet connections of smart vehicles as in the previous case. This process will be slow as compared to direct communication because here a vehicle will first visit any of RSU installed grid clusters then it will start the diffusion process.

In both cases, we assume that for wireless communication we didn't consider any

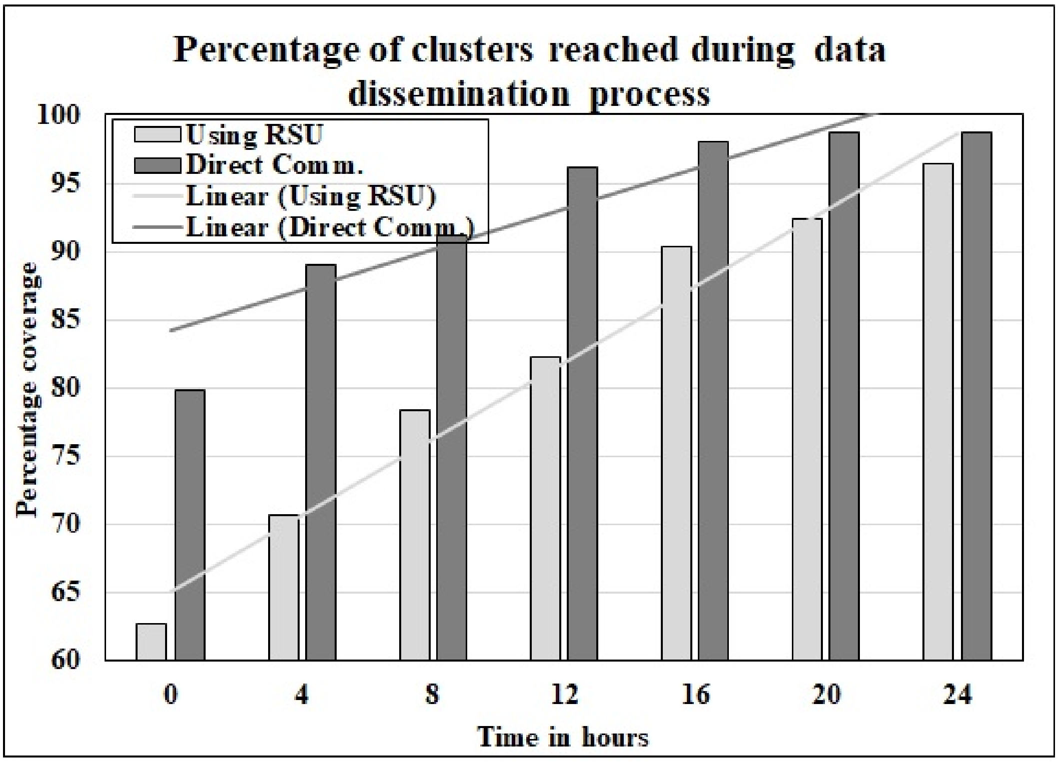


Figure 3.21: Percentage of clusters reached during data diffusion process.

particular propagation model. Also, we didn't consider any pairing delay and remain independent of any wireless or sensor network technology. Figure 3.21, in the latter case, the delay is higher and has a significant difference as compared to direct communication. As the delay-tolerant interval increases the coverage in both cases increases and almost 98% in direct communication and 96% in the latter case when we use the RSUs. When we have a higher delay-tolerant interval then the latter case also provides good coverage and saves the cost of smart vehicles' internet connections.

3.5 Summary

A network architecture for delivering big data in the smart city by using the existing transport infrastructure is proposed in this Chapter. It could best accommodate those massive and delay tolerant-able data transmission demands by utilizing efficiently the existing vehicles' mobility in the urban areas. Moreover, it could release the congestions in the wireless and wired networks.

In 1^{st} part of this chapter, a method of big data aggregation for smart cities using the existing vehicle volume on the road is investigated. For delay tolerant data needs, our proposed method efficiently utilizing the existing vehicles and their energy-efficient wireless links, for big data transmission to reduce end to end delay and infrastructure cost. Our essential and contextual investigations performed in the mathematical evaluation of these case scenarios give solid proof that substantial end-to-end delay and cost savings can be accomplished while as yet ensuring the information delivery. Results obtained from the Auckland case, show that our proposed approach can provide faster transferring the large volume of data by using the existing daily vehicles' mobility, than the conventional transmission network.

In 2^{nd} part of this chapter we used the taxi traces of Beijing city to evaluate our proposed system. It is possible to offload collected data on taxis for further delivering data at data centers in smart cities. We investigate data dissemination potential of the proposed system in terms of network coverage and capacity. The analysis of network capacity and coverage is a function of selected wireless technology range. We evaluate our proposed system by using the microscopic movement of taxi-cabs and their GPS locations in Beijing, China. We apply the area division algorithm on big temporal-spatial data to reduce the complexity of data for analysis of the proposed system. Evaluate the vehicle density in the given area, that is further used to find the potential locations for data offloading. The results propose that even relatively small subsets of taxi fleets, operating in a very large area of Beijing (25×25 Km^2) can achieve more than 90% coverage in 24 hours in case of grid size greater than 500 m^2. Our proposed system can collect a significant amount of data that can be offloaded on to vehicles, for instance, if we assume IEEE 802.11n for communication with vehicles, the system can collect more than 1.4 PB data on daily basis.

The main conclusion is that our proposed system is suitable for delay-tolerant data delivery applications as it can reduce the network load further by sharing the burden of congested networks. The case of code dissemination shows that whenever we have a loose time of requirement then our system outperforms.

This chapter explored the potential of data dissemination in smart cities by using our proposed VADD framework. The data and code forwarding schemes are investigated in the next Chapter.

Chapter 4

Data Forwarding schemes by VADD Framework

4.1 Introduction

In chapter 3, the VADD framework for data dissemination in smart cities to explore the potential in terms of data/code offloading is proposed. We evaluate our proposed framework by two case scenarios of Auckland's vehicle count and Beijing taxi traces. In the Auckland case, a vehicle needs to stop at any point for data offloading. To make our framework more effective we updated our model so that vehicle will pick the data during its movement. In the 1^{st} part of this chapter, we propose a data forwarding scheme for the Auckland case to transfer data from SC devices to the data center. We calculate the effective distance under the coverage of a wireless device. We evaluate the degree of data offloading by using Poisson arrival of vehicles and then calculate end-to-end delay. In the 2^{nd} part of this chapter, we propose a scheme to transfer updated code from CC to SSs. Our proposed algorithms solve the NP-hard problem of the selection of optimal vehicles and optimal RSUs for this transmission. To compare this code transmission scheme we use the random and fixed deployment of RSUs. Finally evaluated this scheme by using real traces of Beijing taxies.

The main contributions in this chapter are listed below..

- We develop a mathematical model to measure the degree of data offloading by considering the Poisson arrival of smart vehicles and RSUs of the road network.
- A case of Auckland City is evaluated by using the annual average daily traffic (AADT) of Auckland City in our proposed model to reduce the end-to-end delay for big data transmission.
- In 2nd part of this chapter, we propose an optimal code forwarding scheme (OCFS) for software update code transmission from end-user to SSs, where optimal mobile nodes and roadside units are selected to solve a multi-objective optimization problem that manages a balance between coverage enhancement and minimize time duration for this transmission.
- Finally, we evaluated the proposed OCFS using real taxi traces of Beijing city. Our results show that OCFS enhances the code transmission coverage rate and reduce the time duration for this dissemination.

Section 4.2, presented a brief background work. Data forwarding delay models are presented in Section 4.3. Data forwarding a data source to the data center is presented in Section 4.4, by considering the daily vehicle count in Auckland city. In Section 4.5, we presented data forwarding from the data center to different sensor locations by considering the Beijing city taxi traces. Finally, the chapter is summarized in Section 4.6.

4.2 Background

In smart cities (SC), a massive number of sensing devices monitor the physical and environmental conditions, create a huge amount of data along with thousands of SC applications. The collection of data from these devices may pose several challenges. The connectivity of these millions of SC's devices to their data centers by using cellular networks or traditional networks may require expensive infrastructure. Our work proposes that an alternate delay-tolerant network exists, to connect such a huge number of smart devices with their data centres, by using smart vehicles, to be specific the transport fleets e.g. taxi cabs.

In future, smart urbanization will improve the satisfaction of their residents with the assistance of different services of data and communication technologies. Similar to other tasks, it includes observing different environmental variables and provide in-time actions based on advance analytics. Because of their very low budgets, smart sensors and smart devices can unquestionably discover their spaces in a widespread scope of SC's applications. For example, collection of temperature and humidity data from vegetation fields along either side of the road, physical structure deformation collection of bridges and road, information about the filling level of all garbage bins in the city for cost optimization in bin collection for city government, storm drains can send warnings when they are obstructed by leaves, surveillance cameras for SC security. These are a few examples where smart sensors can be used as "thing" that can be attached into prevailing objects, to improve their functionality or productivity. However, in this regard we are facing two main challenges: (a) how we can provide power to these sensors; (b) how they can send/receive notifications to/from their administrative bodies.

The energy and cost required for this purpose depending upon the quality of communication requirements. For instance, the real-time communication requires expensive infrastructure over all nodes that could be very expensive. To accomplish these tasks, we need a ubiquitous and reliable communication infrastructure, e.g conventional networks or cellular networks. However, the conventional networks or cellular systems may not be adequate enough to fulfil the requirements on peak load and may fail to provide connectivity and services, at the point when the city needs it most. Moreover, connecting a massive number of smart sensors to a conventional network or some other IP based traditional network may look insignificant technically. It is problematic from both administrative and economic point of view to installing a SIM card on each sensor. It is also unrealistic for the management of sensor gateways or femtocells in the entire city under the same administration.

In this work, we explore the collection of big data produced by a massive number of smart devices in SC using vehicular sensor networks as an alternate data dissemination channel. By using the opportunistic contacts between the sensors and microscopic movement of vehicles with their GPS location information, it is possible to piggyback accumulated data on moving taxis-cabs for further data delivery at data centers of the corresponding service. Not only it can reduce the load on expensive cellular links but also eliminates the need for new infrastructure deployment.

4.3 Network models

The system overview of our proposed framework is given in section 3.1. To calculate the end-to-end delay for data dissemination in smart cities, following are the models, that are used in this case for mathematical calculations. Notations used in the mathematical model are listed in Table 4.1.

To transfer the data from source to destination, following are the two models of end-to-end delay for the transport network and core network.

Table 4.1: Notations used in data forwarding.

Symbol	**Meaning**
d	Effective distance
d_0	Reference distance
n	Path loss factor
Ψ	Gaussian random variable
ρ	Vehicle density
λ	Poisson arrival process parameter
B_v	Data offloading by a vehicle
B_{RSU}	Bandwidth of roadside unit
$\bar{s}$	Average speed of the vehicle
r	Radius of RSU's coverage area
D	Diameter of RSU's coverage area
d_{ac}	Distance between two consecutive RSUs with no wireless coverage
B	Core network bandwidth
D_{Vol}	Data volume
μ	Probability of vehicles to participate
E_{inc}	Incremental Energy cost

4.3.1 Delay Model for Transport Network

Effective distance: We have to identify the suitable region for the vehicle in the wireless coverage area of an RSU. Normally, the packet loss rate depends upon the received power. By increasing the received power we can reduce the packet loss rate. Received power also depends upon the distance from the RSU. Therefore, to calculate the packet loss rate we can use the distance from RSU. Hence, to reduce the packet loss, it is essential to keep a vehicle in the appropriate coverage area of the RSU. We can calculate the path loss $PL(d)$ at a distance d by using path loss formula 3.11.

Delivery probability with vehicle density: The other parameters that affect the effective region of RSU are delay and bandwidth of WiFi is a key factor that can control the delay. RSU has a fixed bandwidth that can be shared among all vehicles inside the coverage area. If vehicle volume on the road is greater, then smaller will be the bandwidth share for vehicles. We assume a fixed number of vehicle m in the coverage area of RSU with bandwidth B_{RSU}. As we know that vehicles that connect to RSU follow Poisson arrival process {f(t), t $\geq$ 0} with parameter $\lambda > 0$ and vehicle volume on the road varies at different intervals of the day [149]. We can convert a day into various intervals, according to vehicle volume on the road, and these intervals follow the Poisson distribution. Suppose that there are n intervals (I_1,I_1 ... I_n) with parameters (λ_1, λ_2 ... λ_n). The probability of vehicles in the coverage area of an RSU at any time t can be calculated as follows.

$$P\{f(t+i) - f(i)\} = e^{-\lambda t}\frac{(\lambda t)^m}{m!} \tag{4.1}$$

Where i is the start time of each interval, $(t + i) \in$ (I_1,I_1 ... I_n), and $\lambda \in$ (λ_1, λ_2 ... λ_n).

The vehicle density in the effective region of an access point depends upon the average speed $\bar{s}$ and Poisson arrival parameter lambda λ. By using different value of lambda (λ_1, λ_2 ... λ_n) in different intervals (I_1,I_1 ... I_n), we calculate the vehicle density ρ in the effective distance d of access point.

$$\rho = \frac{\lambda}{\pi \bar{s} d} \tag{4.2}$$

Suppose B_{RSU} is the bandwidth of RSU and B_v is the shared bandwidth of each vehicle when there are *m* vehicle in the coverage of RSU then

$$m = \frac{B_{RSU}}{B_v} \tag{4.3}$$

Transmission probability Φ, with m number of vehicles can be calculated as follows.

$$\Phi = \sum_{v=1}^{m}(e^{-\lambda t}\frac{(\lambda t)^v}{v!}) \tag{4.4}$$

Degree of data offloading: By using (3.8),(4.1), and (4.4) we can calculate the data transmitted/offloaded by a vehicle to RSU as follows.

$$B_v = B_{RSU} \times \Phi \times \frac{d}{\bar{s}} \tag{4.5}$$

Equation (4.5) can be represented in term of vehicle density (4.2) as follows.

$$B_v = B_{RSU} \times \Phi \times \frac{d}{\frac{\lambda}{\pi \rho d}} \tag{4.6}$$

Where d is the distance of the vehicle from RSU and $\bar{s}$ is the speed of the vehicle.

End-to-end delay: The time T_d required to transfer a big data D_{vol} from source A to destination B depends upon two things. First travel time between A and B T_{AB}, 2nd the data loading time T_L.

$$T_d = T_{AB} + T_L \tag{4.7}$$

If d_{AB} is the distance between these two points and $\bar{s}$ is the average speed then T_{AB} can be calculated as follows.

$$T_{AB} = \frac{d_{AB}}{\bar{s}} \tag{4.8}$$

If N_v is the total number of vehicle required to transfer data D_{vol} then data loading time T_L can be calculated as follows.

$$T_L = \frac{D_{vol}}{N_v} \tag{4.9}$$

N_v depends upon vehicle count V_c on the road, bandwidth share of each vehicle B_v and probability μ of the drivers or vehicles willing to participate in this proposed system, then N_v can be calculated as follows.

$$N_v = V_c \times B_v \times \mu \tag{4.10}$$

By applying the value of N_v, T_{AB} and T_L in (4.7).

$$T_d = \frac{d_{AB}}{\bar{s}} + \frac{D_{vol}}{V_c \times B_v \times \mu} \tag{4.11}$$

By using the value of B_v from (4.5).

$$T_d = \frac{d_{AB}}{\bar{s}} + \frac{D_{Vol}}{V_c \times B_{RSU} \times \Phi \times \frac{d}{\bar{s}} \times \mu} \tag{4.12}$$

4.3.2 Delay Model for Traditional Core Network

For core network case, the total delay depends upon the number of packets and number of intermediate nodes. Suppose L is the length of each packet then the total number of packets N can be calculated as follows.

$$N = \frac{D_{Vol}}{L} \tag{4.13}$$

Let Tp be the processing delay, Tq be the queuing delay, and Tt be the transmission delay and N_{Nodes} is the total number of nodes between sender and receiver then total end to end delay can be calculated as follows.

$$T_{Core} = N_{Nodes} \times 1^{st} PacketDelay + (N-1) \times (T_p + T_q + T_t) \tag{4.14}$$

We assume in case of a dedicated link when a packet arrives at a node; it will find no packet ahead in the queue. Then queuing delay becomes zero. In such case parallel processing will be performed at each node then the final end to end delay can be calculated as follows.

$$T_{Core} = N_{Nodes} \times 1^{st} PacketDelay + (N-1) \times T_t \tag{4.15}$$

If B is the transmission bandwidth then Equation (4.15) can be written as.

$$T_{Core} = N_{Nodes} \times 1^{st} PacketDelay + (N-1) \times \frac{L}{B} \tag{4.16}$$

4.4 Data forwarding

This section describes the numerical analysis of our proposed system with the conventional network system. In first part, we set the value of various parameters and then we present two case scenarios to compare both systems.

4.4.1 Parameters Setting

To evaluate the proposed system model, we consider a straight expressway . Vehicles are moving with some predefined average speed $\bar{s}$ along the straight road. We are assuming that the vehicles are moving only in one direction for simplicity. We calculate the effective distance and data transfer probability in the wireless coverage area of an RSU. We also measure the vehicle density with different speeds of vehicles. Finally, by applying these parameters, we calculate the degree of data offloading by using different speeds and vehicle density.

Effective distance: To calculate the effective distance, we set reference distance $d_0 = 100m$ and carrier $f_c = 2.4$ GHz and apply the different path loss exponent values, according to equation (3.8). The relationship between path loss [dB] and distance is shown in Figure 4.1. Path loss increases with the distance between vehicle and RSU and approaches to 80 dB for 50 to 100 meters. For the remaining analysis we set effective distance d=100 m.

Delivery probability: The fixed bandwidth of RSU is divided among all the vehicles in the wireless coverage area of WiFi. Data transmission probability increases exponentially with different densities of vehicles as shown in Figure 4.2. We vary the Poisson

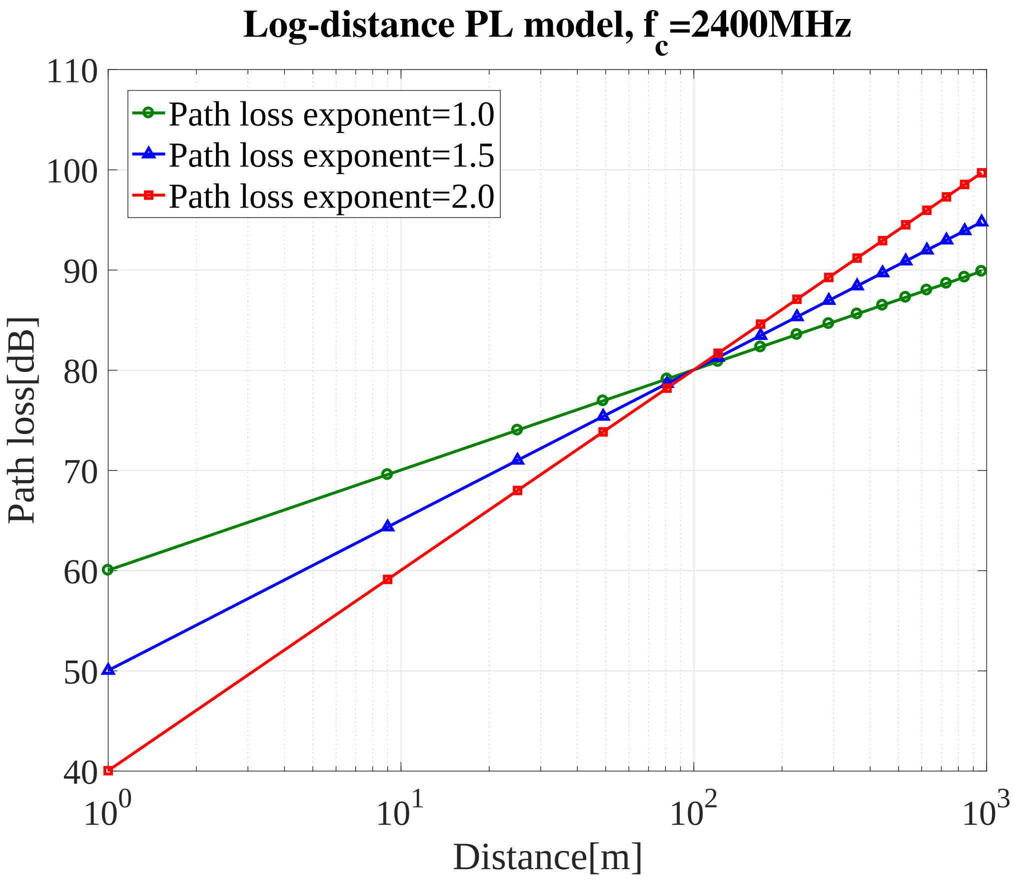


Figure 4.1: RSU Distance vs Path loss.

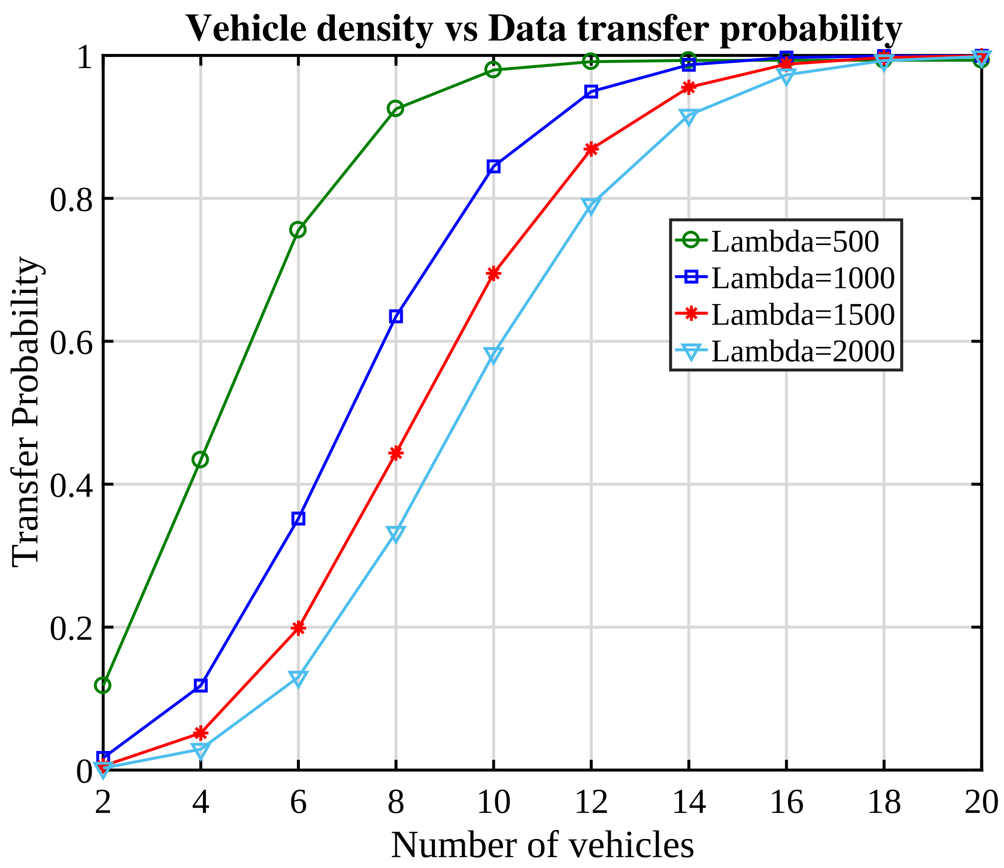


Figure 4.2: Data transfer probability.

arrival parameter λ from 500 to 2000 by using Equation (4.4). The trend shows that the data transfer probability reaches to 0.95 in all cases when the number of vehicles approaches to 15.

Vehicle density: ρ is evaluated with different speeds of vehicles by using effective distance *d* and Poisson arrival parameter λ in Equation (4.2). We vary the value of λ from 500 to 2500 and set effective distance *d = 100*. Vehicle speed is used to get the vehicle density in a specific area as shown in Figure 4.3. If we increase the vehicle speed, then vehicle density decreases exponentially.

Degree of data offloading: By considering the above evaluation, the data offloading can be calculated by using equation (4.5) in the wireless coverage area of an RSU. According to Figure 4.1, we set the effective distance of the vehicle from RSU $d = 100$,

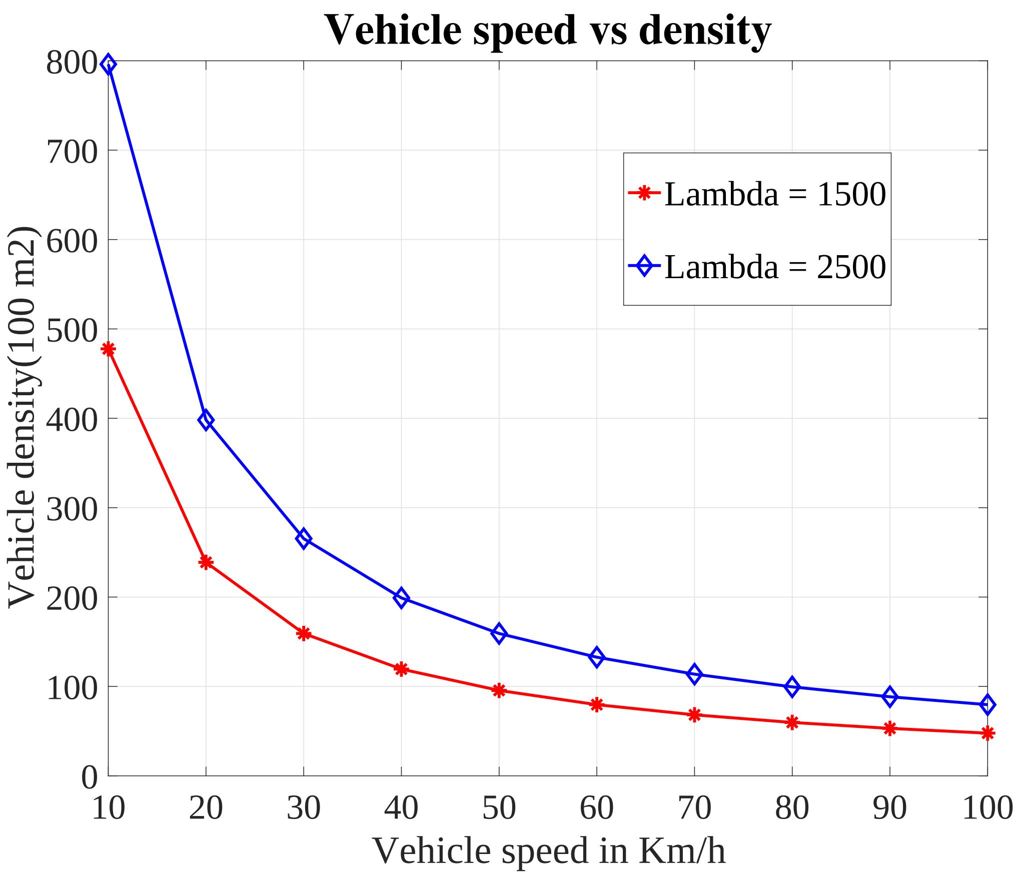


Figure 4.3: Vehicle speed vs density

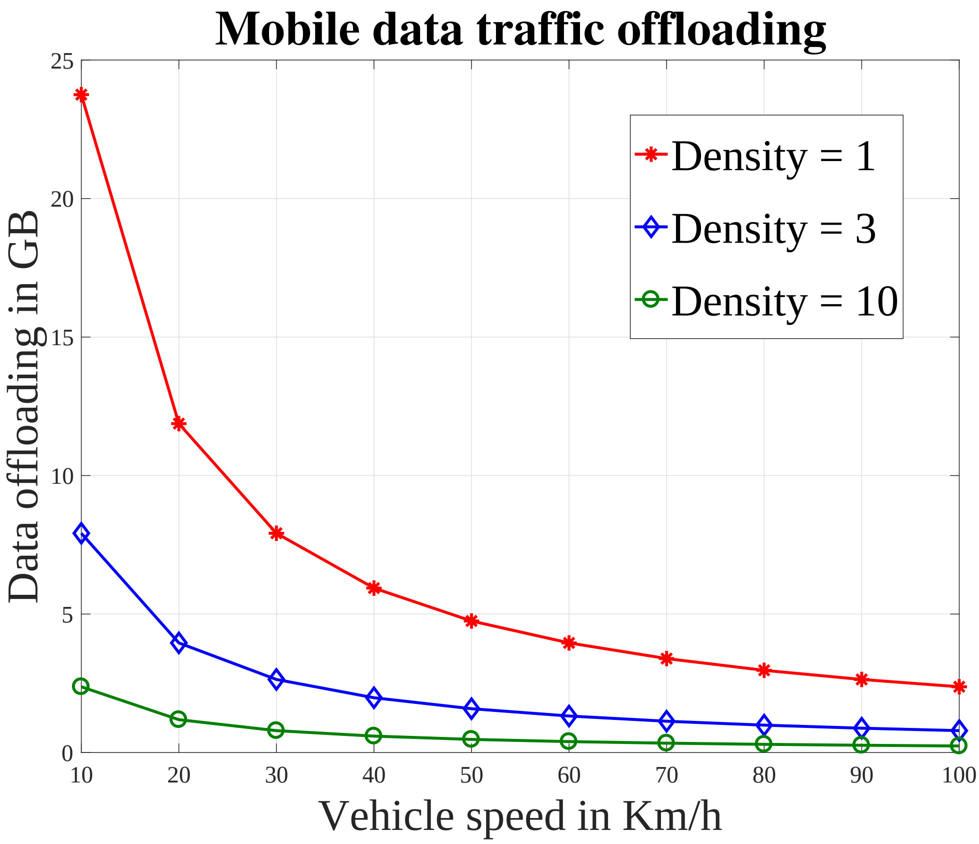


Figure 4.4: Mobile data offloading.

and we set the transmission probability $\rho = 0.95$ as in Figure 4.2. A vehicle can transfer the various amount of data to RSU by using different speeds. The time in which a vehicle stayed connected with an RSU is inversely proportional to vehicle speed. According to Equation (4.5), the data transfer rate varies with the different speeds of vehicle and connection time duration of a vehicle decreases with increase in the speed of the vehicle. Moreover, Figure 4.4, shows that if we increase the vehicle density, i.e., more vehicles want to send the data then the bandwidth will be divided among all vehicles.

In the following two subsections, we present two case scenarios for comparing the performance in terms of delay and EC.

4.4.2 Auckland Case Scenario

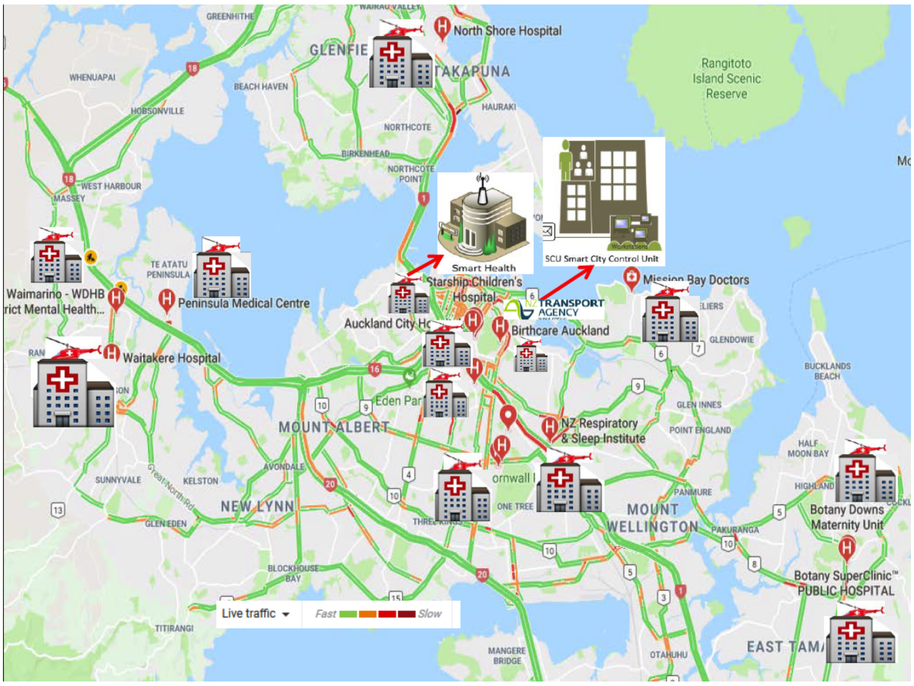


Figure 4.5: Auckland City Case Scenario for Delay Tolerant Study

The proposed case of Auckland City is shown in Figure 4.5. We propose two cases for delay-tolerant information transmission, the core network, and the vehicle transport network. Assume a source at SH16 Royal Rd Off Ramp to Hobsonville Rd/SH18 Off Ramp WB with reference ID:01610016 in Auckland City needs to send some delay-tolerant data D_{Vol} to a destination in the city center. For the vehicular system, we find the distance d between these two points is 23 Km by using Google maps and vehicle count $AADT$ on the said link is 27857 [150]. The average speed $\bar{s}$ of the vehicle on this link is 50 Km/h. Assume each vehicle has IEEE 802.11ay wireless interface and disk capacity is 256 GB. We assume that if only 20% ($\mu = 0.2$) drivers take part in this proposed framework, then by applying Equation (4.12) we can estimate the delay value of the proposed network. In the core network case, a cellular network, a wired network, or a wireless network can be used for data forwarding, and its delay can be estimated by using the Equation (4.16). We exchange the different information volume between these two locations and get the outcomes as appeared in Figure 4.6. For comparisons, we use the bandwidth 512 Mbps and 1 Gbps in traditional core network case. The outcome demonstrates that our proposed vehicular network outperforms the traditional core network. We enhance the data size and look at its impact on delay for data transmission. The results demonstrate that for huge information volume, the delay increases significantly for core networks as compared to our proposed vehicular network.

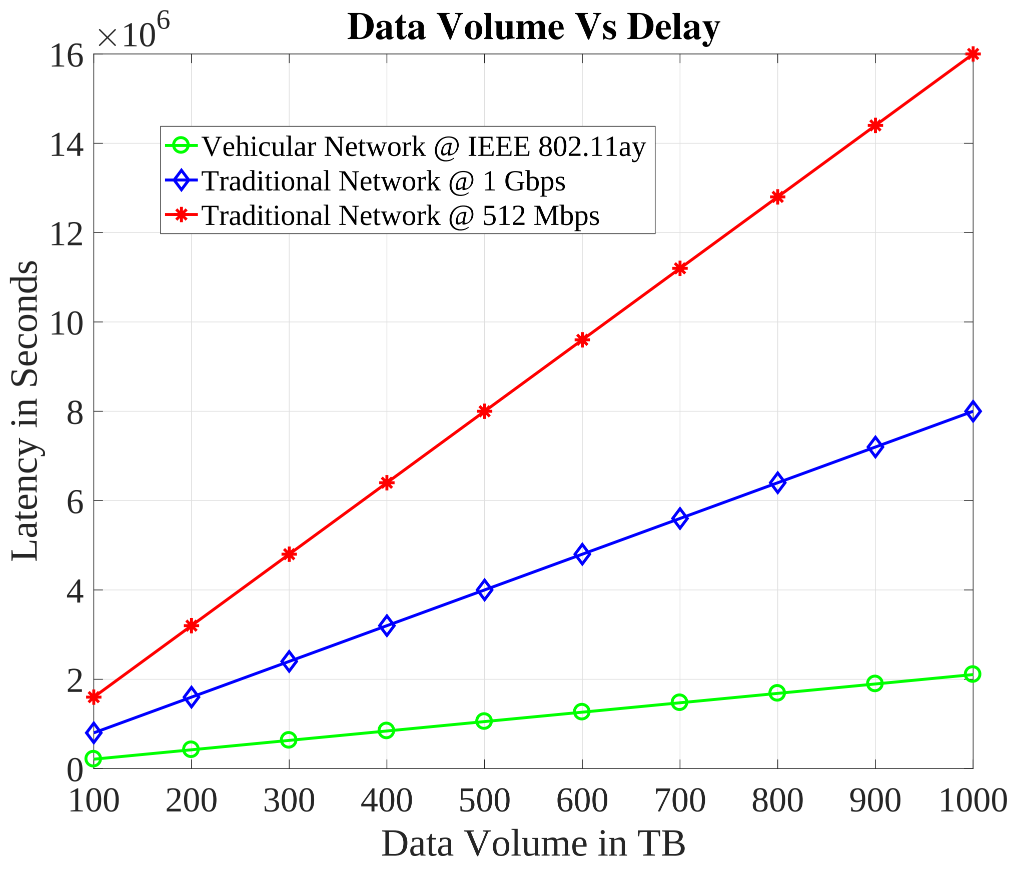


Figure 4.6: Data volume vs delay.

In the next case, as shown in Figure 4.7, we compare data transmission delay with distance for both cases. We transfer 100 TB data between two points by varying the distances. In this case, more routers and intermediate nodes will be used in the core network, and delay depends upon the distance between two points as well. This

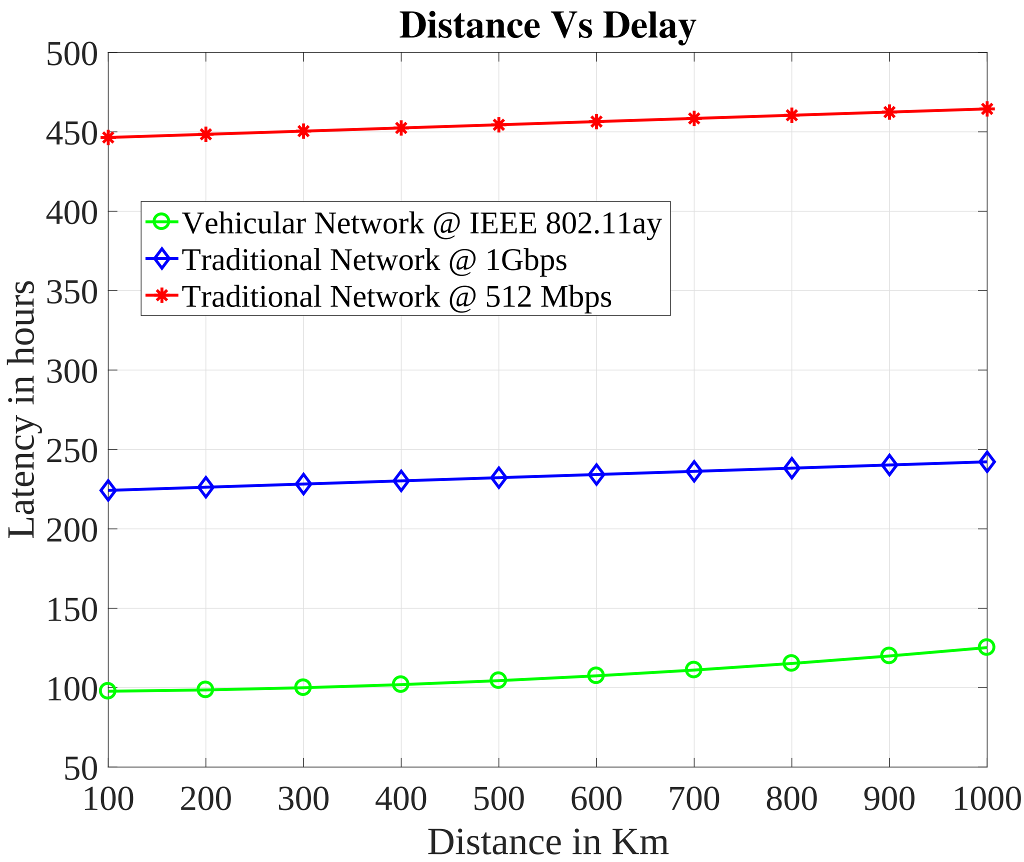


Figure 4.7: Distance vs Delay.

transmission delay depends upon queuing delay, propagation delay, transmission delay, and processing delay. We assume that there is no queuing delay on the core network when a packet arrives in at the router, there is no inbound and outbound queue. Also, assume that the core network is built on fiber. Hence the propagation delay can also be ignored. We consider only transmission and processing delay and assume that there is a router after every 100 Km [151]. Vehicular network delay also increases as the distance increases to the destination because each vehicle takes more time to reach the destination when we increase the distance. Similarly Figure 4.7, also shows that the delay in the case of core networks also increases as the number of routing devices keeps on increasing. The results show that our system outperforms the traditional core networks when we increase the distance.

4.5 Optimal code forwarding scheme (OCFS)

Smart sensors are smart devices and their functionality can be updated according to need. They will receive their software update code from MNs and update themselves automatically. A huge number of SS are densely deployed throughout the smart city closed to the road network. The working of these sensors can be adjusted according to application requirements. E.g, a temperature sensor can detect the temperature in a fire-prone area; a sensor can sense the filling level of garbage in the garbage can; for weather prediction wind sensors are required and a sensor can be used to determine building or road deformation structure. Suppose we have a set of sensors $S_{ss} = \{S_1, S_2, \ldots..S_k\}$. To update these sensors suppose we have a set of codes $S_{codes} = \{c_1, c_2, \ldots., c_m\}$. RSUs have a consistent power supply, are directly connected to CC and data center DC. They will collect the updated code from CC and forward it to SS by using MN. We have a set of RSU $S_{RSU} = \{r_1, r_2, \ldots.r_n\}$.

In this scheme, MNs are smart vehicles moving around the city, having storage, communication, GPS, and processing facility. These MNs get updated code for SSs from RSUs, work in a store carry forward manners and transfer the code to SS whenever they are in their wireless range. In general, these mobile nodes can be used without any change. They are not specifically responsible to forward the code to SSs. In other words, the trajectory of these MNs doesn't need to change by our proposed code forwarding scheme but it upon the destination of the rider. This non-scheduled and non-regular code forwarding communication between SSs and RSUs by using MNs is called vehicular opportunistic communication. The following are the main steps, how this scheme will work, as shown in Figure 4.8.

1. Whenever a user needs to update a set of sensors, he will send his request to CC, along with GPS locations of SS, updated code, and time interval D, the delay-tolerant indicator for task completion.

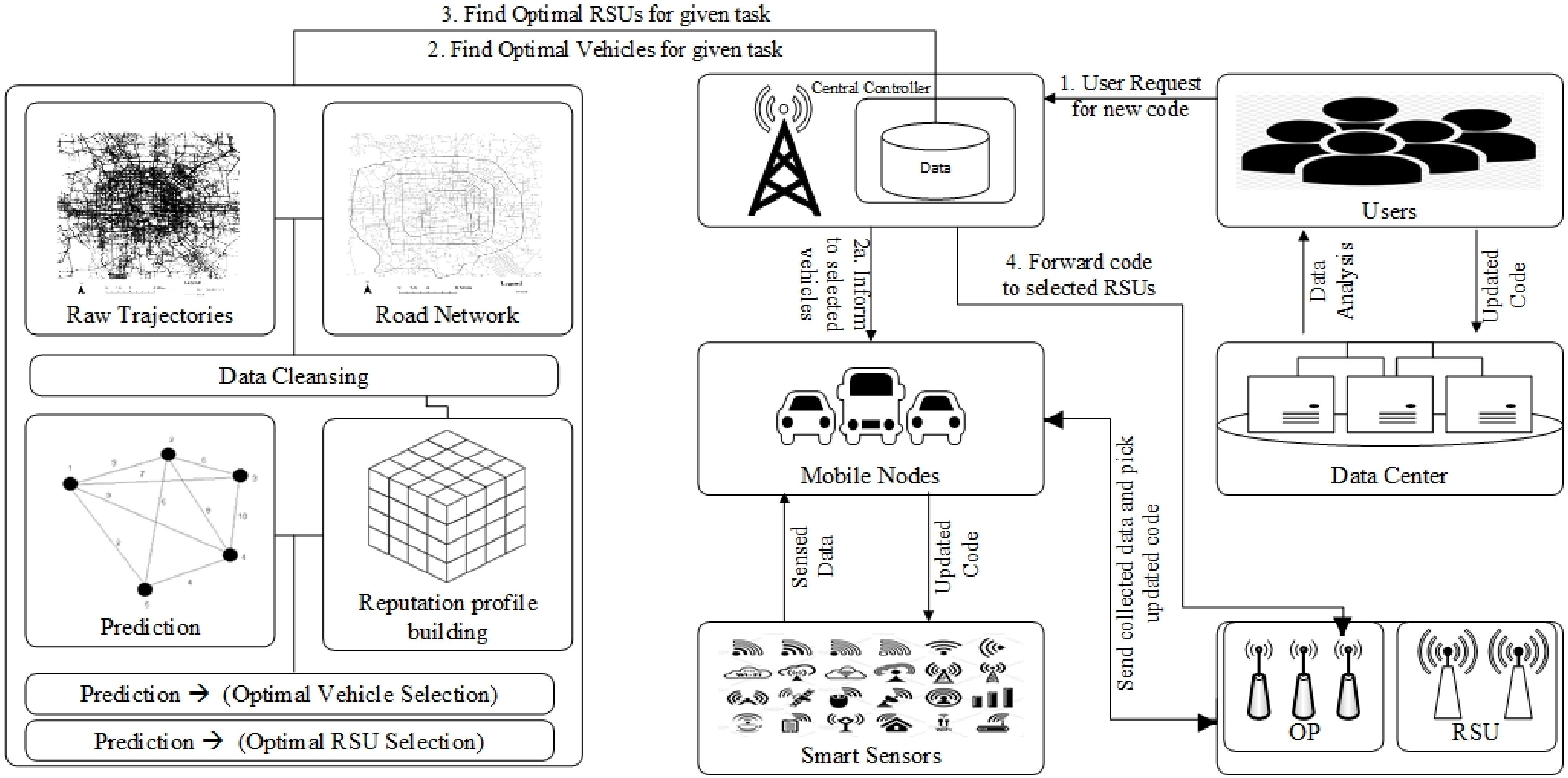


Figure 4.8: System overview.

2. CC finds the optimal set of vehicles for this task by using the vehicles trajectory history. Vehicle participating in this proposed system send their reputation profile to CC periodically for optimal decision making in vehicle selection. This reputation profile is based on the frequency of visits in different clusters.

3. Based on the optimal set of vehicles CC also calculates the optimal set of RSUs. CC finds which RSU or cluster in existing set of RSUs have maximum visits of these optimal vehicles.

4. CC forward the updated code to optimally selected RSUs.

5. All vehicles from the optimal vehicle selection set will start execution of the given task as follow

   (a) The updated code can be received from RSUs. MNs will get this code whenever they will be in the wireless range of any RSU.

   (b) After getting the code MNs transfer this code to required SSs wherever they are in their wireless range.

   (c) After each task completion, MN will inform to CC.

6. After delay-tolerant interval D, CC will inform to MN to stop executing the tasks.

4.5.1 Problem Statement

In our proposed scheme the trajectory of MN is not adjusted for code forwarding. Therefore, it is unpredictable that all sensors receive the updated code. Hence, we need to address some concerns: How to forward the code to maximum possible SSs. How to forward the code more quickly to SSs in a given time interval D. The problem definition is given in detail to solve these issues.

1. Code forwarding coverage rate

The quality of service of the proposed scheme depends upon the task completion rate. Code forwarding coverage rate (CFCR) is the ratio of the number of SSs who receive the updated code to the given number of SSs in a time interval T. Higher ratio will show the batter output of the proposed scheme. This ratio can be calculated by the following equation.

$$CFCR = \frac{SS_{Finish}}{SS_{Total}} \tag{4.17}$$

Where SS_{Total} is the given number of smart sensors need to update and SS_{finish} is the updated SS in time T.

2. Code forwarding time duration

It is a total time from start to end of the proposed scheme to forward the updated code. The completion time will reach when CFCR gives 100% coverage or it doesn't grow well after a certain time period. The system will outperform with a lower value of CFTD. It can be calculated by the following equation.

$$T_{total} = t_2 - t_1 \tag{4.18}$$

Where t_1 is the starting time and t_2 is the finishing time of the proposed scheme.

To get better performance and service quality of the proposed scheme we need to maximize the CFCR and minimize the CFTD as follow.

$$CFCR = \mathbf{max}(\frac{SS_{Finish}}{SS_{Total}}) \tag{4.19}$$

$$CFTD = \mathbf{min}(t_2 - t_1) \tag{4.20}$$

4.5.2 Design of proposed scheme

1. Motivation

To recognize SC, a huge number of SSs are need to install throughout the city. As stated earlier, the sensing requirements of different applications may change from time to time which leads to software code update in SSs. Many of the SSs in SC cannot be connected to DC clouds directly because of the high cost of infrastructure network and power resources. This is the reason, managing software update code transmission to SSs from DCs in traditional ways is challenging.

On the other hand, there exists a displacement of a huge number of mobile vehicles in SC. These vehicles travel on streets and roads almost everywhere and at any time. Many vehicles are equipped with storage, processing, and communication facilities. We can use this huge displacement of vehicles as mobile nodes to forward updated code to SSs.

The other important thing is, this displacement of vehicles have a certain similarity in trajectories in continuous periods of time. To examine the similarity, the Beijing taxi trajectory data set is used in this work. We use the following equation to measure the similarity of the vehicle displacement in the consecutive periods.

$$Similarity = \frac{\mid \lambda(x) \cap \lambda(y) \mid}{\mid \lambda(x) \cup \lambda(y) \mid} \tag{4.21}$$

This equation calculates the similarity between the periods x and y. Infect, this is a ratio of vehicles visited in common areas, to all areas visited by vehicles in time x and time y. We calculate the similarity of a given set of taxi trajectories, we divide the time into intervals of 24 hours each. Figure 4.9 shows that there exists a strong similarity in taxi displacement of taxi trajectories in an area of 25 $\times$25 m^2 .

2. Overview

We use mobile vehicles as mobile nodes MN in this proposed scheme of code

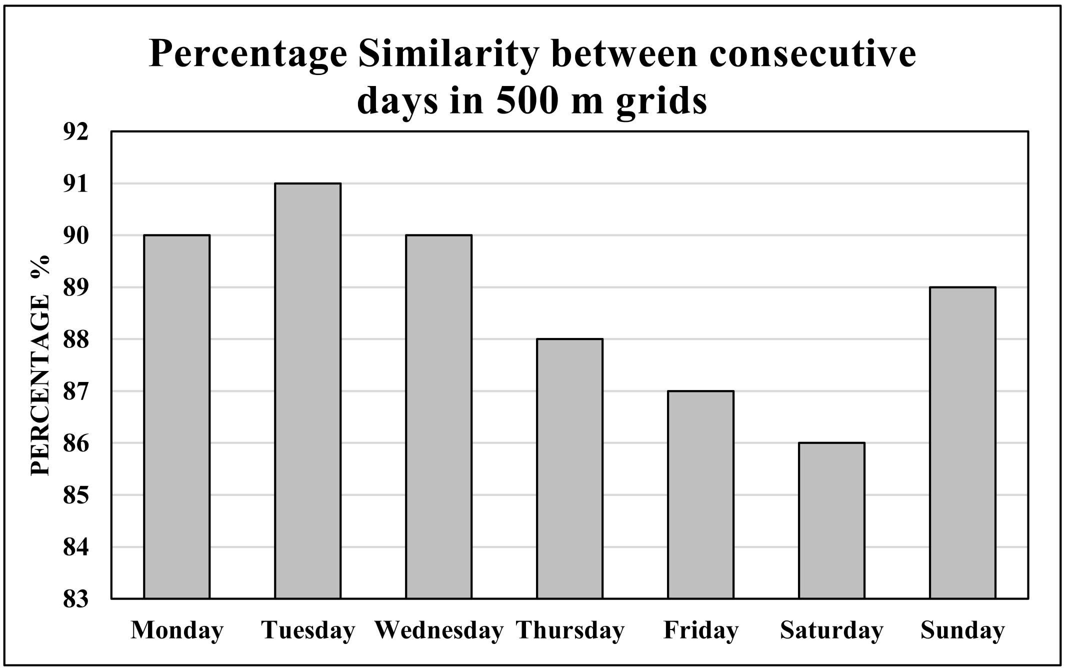


Figure 4.9: Similarity in consecutive days.

transmission to SSs in SC. This scheme manages a huge number of SSs and saves the infrastructure cost. In this scheme, SSs are scattered throughout the city at different locations to manage different sensing applications and they are installed near road networks. Although traditional cities have many sensors installed once they are installed, it is difficult to change there working status. But SSs can update themselves wirelessly after their installation and in many scenarios, it is necessary to update their working status. In many situations, environmental changes that are monitored by SSs, e.g high temperature in summer can become the reason to cause a fire. To monitor the fire cause more quickly in the summer season, the temperature of SSs can be adjusted for easy detection of fire by summer updated code. But the problem is how to transfer updated code to SSs without the internet. Our scheme uses smart vehicles as mobile nodes MN to forward this updated code to SSs. The following are the three main components used in this scheme.

a. Mobile Node

Algorithm 6: Algorithm at Mobile Node

```
while True do
    Enter Detecting Mode
    if Detect an RSU then
        R_RSU(aid) ← ++value
    else
        R_SS(aid) ← ++value
    end
    if MN in transmitting mode then
        Transmit a send request to RSU/SS
        receive a responce from RSU/SS
        if RSU/SS is ready then
            Transmit code/data and add log
            if finishing transmission then
                clear internal memory
            else
                attempt to transmit next time
            end
        else
            do nothing
        end
    else
        Switch to detecting mode
    end
    if MN in receiving mode then
        if RSU/SS is ready to send then
            MN receives the code/data packets
            if data is critical then
                upload to DC by using LTE connection
            else
                Wait for RSU/SS location
            end
        else
            do nothing
        end
    else
        Switch to detecting mode
    end
end
```

Mobile nodes in this work are the smart vehicles, which are extensively dispersed and frequently moving in the entire city. These vehicles are equipped with multiple wireless interfaces, storage, and processing units, so many of them can be used in this data communication between sensors and roadside units without any real modifications. We can say that the trajectory of the mobile node is not affected by our proposed model, but it depends upon the final destination of the rider is going to. This non-scheduled and non-regular communication between SSs and RSUs can be called as store carried and forward communication or opportunistic communication. Whenever a MN visits a cell area it updates its reputation parameter and profiles its reputation concerning encounters in different areas as shown in Figure 4.12. If MN is in transmitting mode then it sends a transmission request to SS/RSU, it transfers the data if the other device is ready to receive the data. If a device wants to send data to MN then it converts its mode to receiving mode and receive the data. It can transfer updated code from RSUs to SSs and sensed data from SSs to RSUs. To build the profile of a MN and forwarding code or data between RSUs and SSs, we use algorithm 6.

b. Smart Sensor

Smart devices and actuators installed in the infrastructure of the entire smart city are termed as the smart sensor in this work. These smart sensors are used for data collection and controlling of different applications. These sensors can be controlled and administrated by the central authority e.g by using software-defined networking. They can upgrade their IOS after receiving regular software updates from data centers and can become more intelligent. They are low powered devices, to save the cost; these devices are not directly attached to cellular networks. These smart sensors can communicate with MNs whenever they are in the wireless range of each other, by using different wireless technologies like, IEEE 802.11n, IEEE 802.11p, IEEE 802.16 and IEEE 802.11 ay, etc. After getting the data from control centers, these sensors can update their settings. For example, sensors that are previously working for data

Algorithm 7: Algorithm at Smart Sensor

```
while True do
    Enter Detecting Mode
    if SS detects MNs and have data to send then
        switch to transmission mode
        if m.ttl is not expired then
            1. Broadcast request for reputation index for each MN
            2. MNs send their reputation index
            3. Find the MN with highest reputation value
            if $R_{Value} \geq thresholdvalue$ then
                transfer data packets
            else
                wait for next comming MNs
            end
        else
            do nothing
        end
    else
        do nothing
    end
    if SS detects a MN having update code from $S_{MN}^{optimal}$ then
        1. Switch to receiving mode.
        2. Receive the code from MN.
        3. update its function by received code.
    else
        do nothing
    end
    Collect data from srounding environment.
end
```

collection of temperature can be reset to gather data for humidity in the vegetation field. The sensors controlling the working period of street lights can be reset with foggy weather conditions. These sensors can forward the collected data to RSUs by using MNs. SSs use algorithm 7 to perform its operations.

c. Road Side Units

They are controlled by the central authority, data center cloud, or control center. Data from all the sensors of the city is collected here with the help of mobile nodes.

Algorithm 8: Algorithm at RSU

Input: $S_{SS} \leftarrow$ set of smart sensors, Code $\leftarrow$ updated code from DC, $S_{MN}^{optimal} \leftarrow$ set of selected vehicle for code transmission

while *True* **do**
  Enter Detecting Mode
  if *RSU detect a MN having data from sensors.* **then**
    if *MN send a data forwarding request* **then**
      1. Switch to receiving mode.
      2. Receive the data from MN.
      3. Upload the data to DC.
    end
  else
    do nothing
  end
  if *RSU detects a MN from $S_{MN}^{optimal}$ for code transmission* **then**
    1. Switch to transmission mode.
    2. Forward the received code to MN.
    3. Update DC.
  else
    do nothing
  end
  update $RT_{MN}[tid][aid] \leftarrow ++value$ // update reputation value of vehicle at DC
  Switch to detecting mode
end

They are also responsible for downloading software updates from CC, for different sensors and forward to them by using MNs. Moreover, these RSUs have continuous power supply and connection with the core network. RSU uses algorithm 8 to perform its functionality

4.5.3 Code deployment

We use the RSU on the road network to transmit updated code to SSs by using MNs. All RSUs are attached to CC by using the internet. CC needs to select optimal RSUs among all, to transfer the updated codes and to maximize the coverage rate and minimize time and cost. We turned the RSU selection problem into two problems of minimum

Table 4.2: Symbols used in this scheme

Symbol	Meaning
S_{MN}	Set of Mobile Nodes
S_{SS}	Set of available Smart Sensors
S_{RSU}	Set of road side units
$Best_{MN}$	The MN covered the highest number of SSs
$S_{SS}^{bestcovered}$	Set of SSs that $Best_{MN}$ passed
$S_{MN}^{optimal}$	Set of optimally selected MNs
$S_{optimalMN}^{bestcovered}$	Set of optimal mobile nodes passed by $Best_{RSU}$
S_{Area}	Set of Areas where $S_{MN}^{optimal}$ passed
$Best_{RSU}$	The Area where the highest number of MNs passed in SArea
$RI(MN_i)$	Reputation index of mobile node MN_i,
$RI(RSU_i)$	Reputation index of cluster RSU_i,
$RFA(MN_i)$	Reputation function, can find the set of areas that the MN_i passed
$RFM(RSU_i)$	Reputation function, can find the set of MNs passed by RSU_i

set coverage. 1) Based on historical trajectories, find the optimal set of vehicles that can cover all the SSs of the given task. 2) based on optimal vehicle selection, find a minimum set of areas, where all selected vehicles travel through at least one selected area. This problem of set coverage is a distinctive NP-hard problem and our proposed algorithms can obtain a realistic solution set. The following are two algorithms are used for this selection.

a. Optimal vehicle selection

In the motivation section, the similarity between future and past vehicle trajectories have been observed. This similarity will build the reputation of vehicles w.r.t. area visit. Therefore, the reputation of vehicles has certain implications for future trajectories that can be used for code transmission. If we have an updated code for a given set of SSs, then we can use historical data to find the optimal set of vehicles that cover all areas of SSs by using algorithm 9 and $RFA(MN_i)$ function will calculate the reputation of mobile node i with respect to area visits, how many areas are covered by MN_i in the history and number of SSs covered in the given requirement, where they are located in a given set of clusters as follows.

$$RI(MN_i) = \sum_{i=1}^{n} \mid M(MN_i, aid) \mid \tag{4.22}$$

Where n is total number of clusters.

$$RFA(MN_i) = \{MN_i : RI(MN_i) > RI(MN_j) \therefore for\ all\ SSs\} \tag{4.23}$$

For all given set of SSs the CC calculate the set of optimal vehicles by 4.23 with respect to maximum coverage.

Algorithm 9: Optimal set of vehicles for code forwarding

Input: $S_{SS} \leftarrow$ set of smart sensors to update, $S_{MN} \leftarrow$ set of all mobile nodes.
Output: $S_{MN}^{optimal} \leftarrow$ set of optimal vehicles

while S_{SS} *is not* Φ **do**
  $Best_{MN}$ == NULL
  $S_{SS}^{bestcovered} = \Phi$
  for $MN_i \in S_{MN}$ **do**
    $S_{covered}^{MN_i}$= RFA(MN_i) $\bigcap S_{SS}$
    if $S_{covered}^{MN_i} > S_{SS}^{bestcovered}$ **then**
      $S_{SS}^{bestcovered}$ =$S_{covered}^{MN_i}$
      $Best_{MN} = MN_i$
    end
  end
  $S_{MN}^{optimal} \oplus Best_{MN}$
  $S_{SS} \ominus S_{SS}^{bestcovered}$
  $S_{MN} \ominus Best_{MN}$
end
return $S_{MN}^{optimal}$

b. RSU Selection.

To verify the performance of our proposed scheme, we following two more schemes for RSUs selection and compare the results with optimal RSU selection scheme. Here are the three schemes for RSU selection.

1. Optimal RSU Selection After getting the set of optimal vehicles, we need to find the areas where the selected vehicles travel through at least one time. Algorithm

10 is used to select the optimal set of RSUs for code transmission from CC to RSUs. $RFM(RSU_i)$ function will help to get the clusters where the mobile nodes in the set of optimal vehicles visited at least one time in the history. That code will be further forwarded to SSs by using the optimal set of MNs. $RI(RSU_i)$ will calculate the reputation index of RSU_i, how many vehicles passed in a given cluster. $RFM(RSU_i)$ calculate the optimal set of RSU_i with respect to given set of optimal vehicles.

$$RI(RSU_i) = \sum_{i=1}^{m} \mid M(aid, MN_i) \mid \tag{4.24}$$

$$RFM(RSU_i) = \{RSU_i : RI(RSU_i) > RI(RSU_j) \therefore for\ all\ MNs\} \tag{4.25}$$

Where m is total number of vehicles. For the given set of optimal vehicles CC calculates the optimal RSUs by 4.25 in algorithm 10.

Algorithm 10: b. Optimal RSU selection for code forwarding

Input: $S_{MN}^{optimal} \leftarrow Algorithm$ x, Set of Road Side Units = S_{RSU}
Output: $S_{RSU}^{optimal} \leftarrow$ set of optimal RSUs

while S_{RSU} *is not* Φ **do**
    $Best_{RSU}$ = NULL
    $S_{optimalMN}^{bestcovered}$ = NULL
    for RSU_i *in* S_{RSU} **do**
        $S_{covered}^{RSU_i} = RFM(RSU_i) \bigcap S_{MN}^{optimal}$
        if $S_{covered}^{RSU_i} > S_{optimalMN}^{bestcovered}$ **then**
            $S_{optimalMN}^{bestcovered} = S_{covered}^{RSU_i}$
            $Best_{RSU}$ = RSU_i
        end
    end
    $S_{RSU}^{optimal} \oplus Best_{RSU}$
    $S_{MN}^{optimal} \ominus S_{optimalMN}^{bestcovered}$
    $S_{RSU} \ominus Best_{RSU}$
end
return $S_{RSU}^{optimal}$

2. Adjacent Fixed RSU Selection.

By closely observing the heat table 3.13 in chapter 3, of vehicle trajectories, we have selected an area of adjacent cells, for RSU deployment, where the frequency of vehicle is very high.

3. Random Selection

In random selection, we have selected some areas randomly for RSU deployment that fulfil certain criteria.

4.5.4 Performance analysis of OCFS

In this analysis, we use 1000 m^2 cell size, and there are 625 cells in our selected area. The parameters used in this analysis are shown in table 4.3.

Table 4.3: Parameter settings for code transmission analysis.

Parameters	**Value**
Dataset	Beijing Taxi Traces (China)
Area	$25{\times}25Km^2$
Area division (Cell size)	1000 m^2
Number of cells	625
Wireless Technology	IEEE 802.11p
Number of GPS points reported	5,051,513
Longitude	116.24 ° E -116.5335° E
Latitude	39.8125° N-40.04° N
Number of Taxies	400
Average update time of Taxies	30 seconds
Number of smart sensors	1000,1200,1400,1600,1800,2000
Number of groups	6
Delay tolerant interval	12 hours, 24 hours
Number of repetitions for each group	5 times
Maximum execution time	1800 minutes

1. Area selection

To verify the performance of the proposed code transmission scheme we use the T-Drive taxi data set from the city of Beijing. For VADD contact estimation, we have selected an area of 25×25 km^2 having longitude between $116.24°$ E to $116.5335°$ E

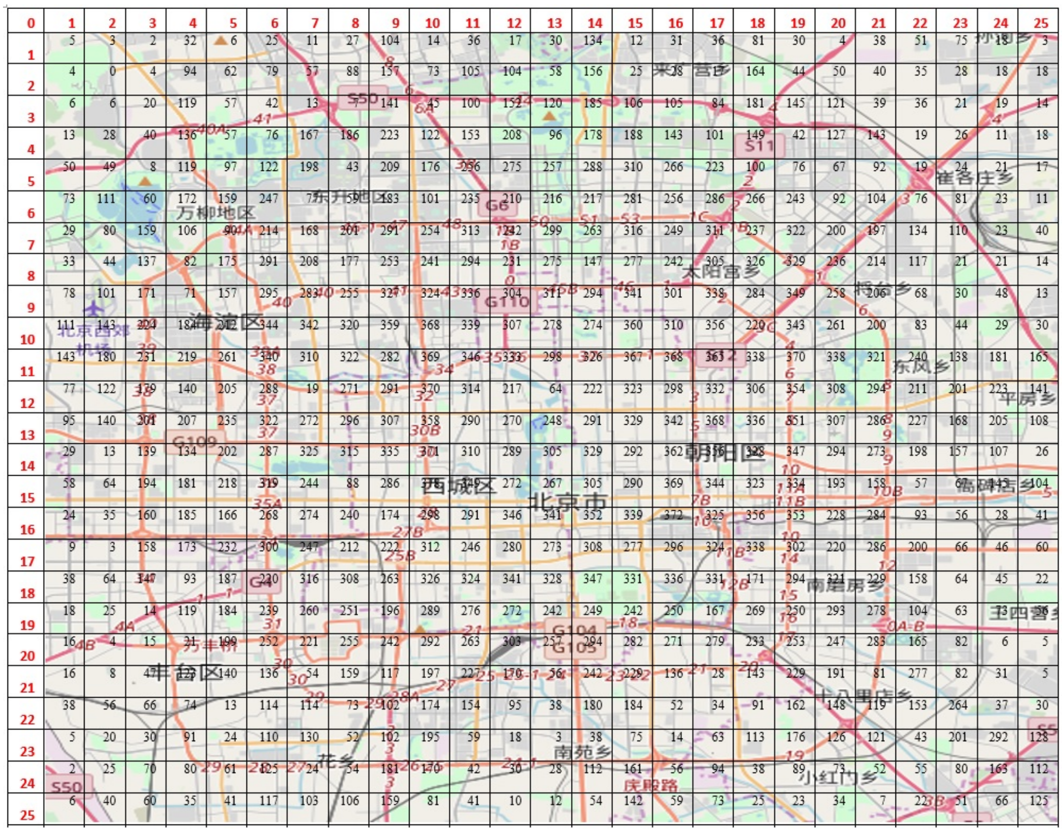


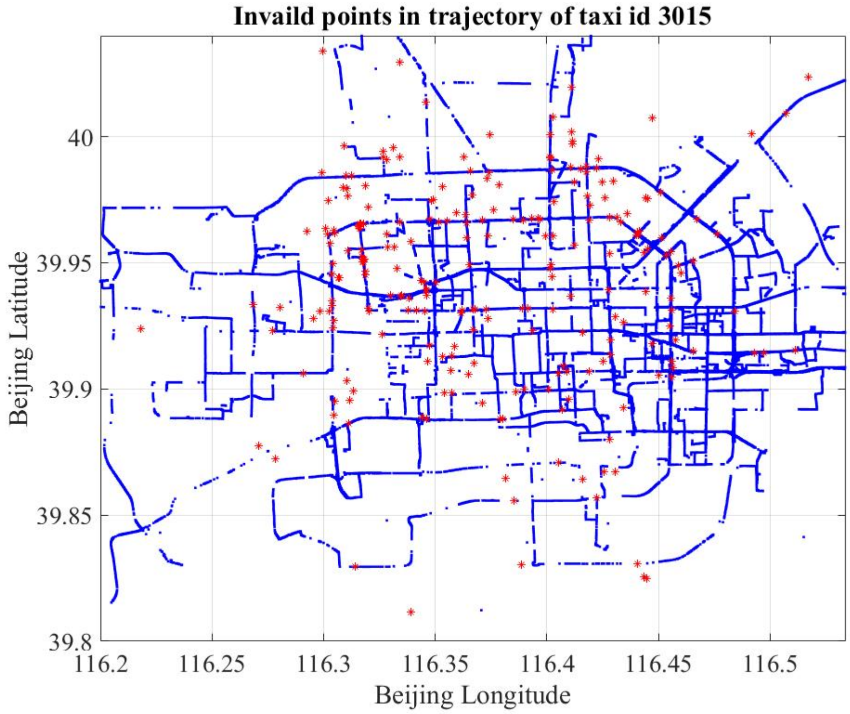
Figure 4.10: Number of vehicles visits in each cell of 1000 m^2.

and latitude between $39.8125°$ N to $40.04°$ N, as shown in figure 4.10 [139]. The exact measurements of all lengths between these longitudes and latitudes are calculated from a web page Moveable Type Scripts [152]. The number of vehicles visit from a set of 400 random selected vehicles is shown in figure 4.10, in each cell of 1000 m^2.

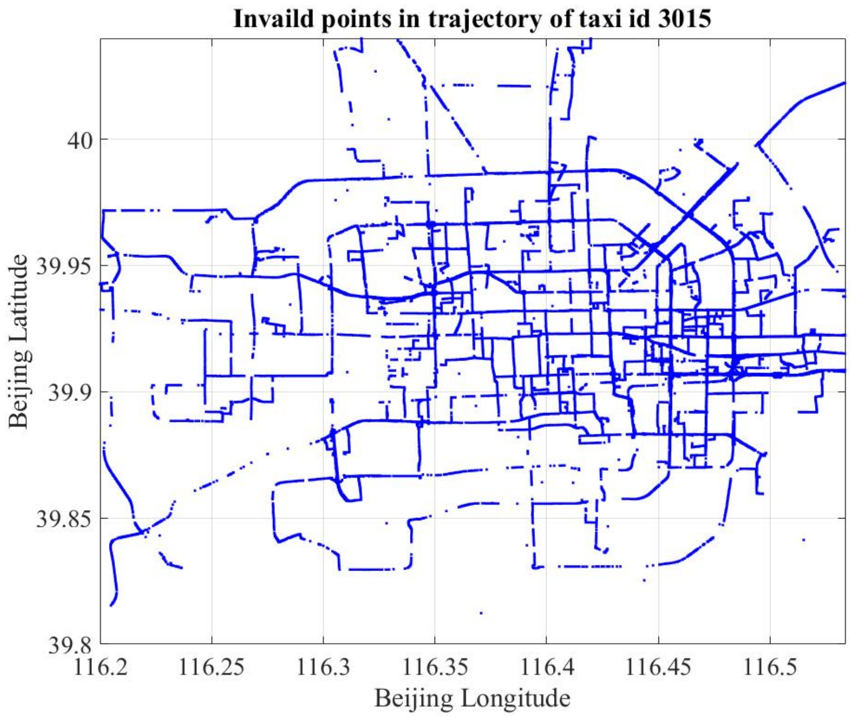
2. Data filtration

The remaining GPS locations are filtered out other than the selected area. To reduce the error, we filtered more points, deviated from the typical trajectory of each taxi. Figure 4.11(a) depicts the invalid points of the taxicab having ID 3015, with red colour.

The vehicles are broadcasting wrong GPS locations near the areas, vulnerable to GPS errors, like tunnels, areas having high buildings skyscrapers, etc. For this purpose, we calculate the average speed and remove those points where speed is much greater than the average speed. After this data filtration, we purify the data set that is shown in



(a) Trajectory with invalid points.



(b) Trajectory with valid points.

Figure 4.11: Trajectory of the taxi having ID 3015

figure 4.11(b)

3. Profile building

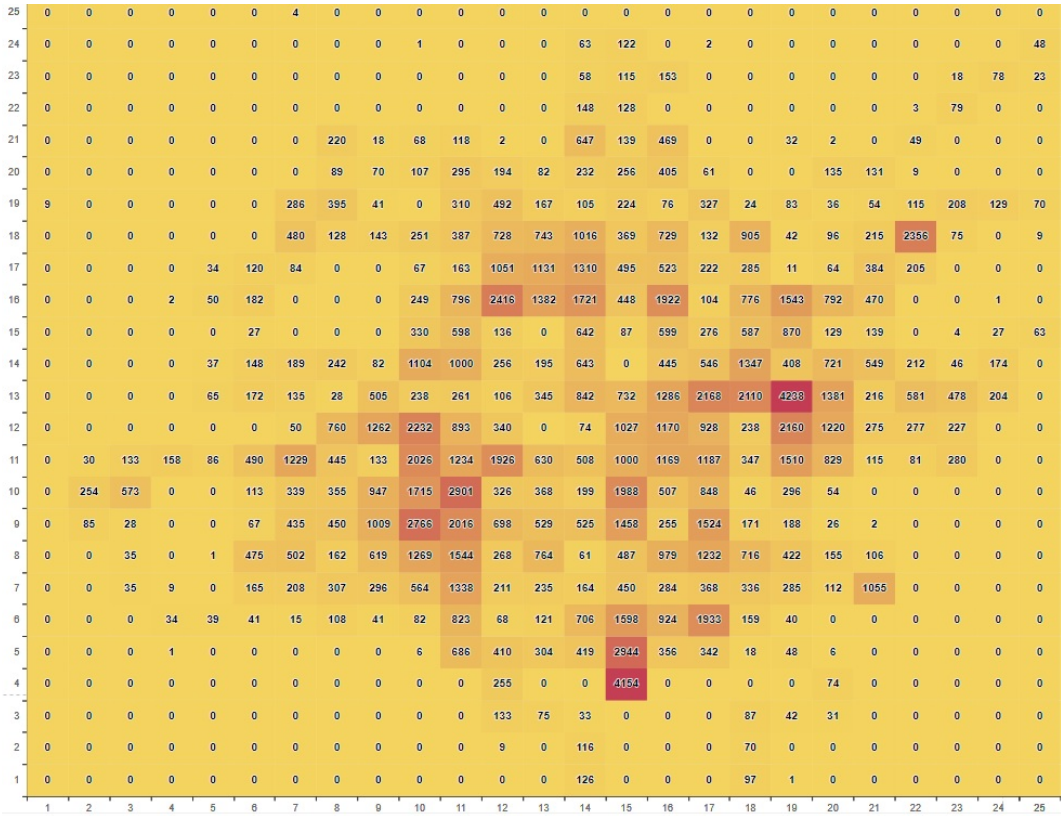


Figure 4.12: Reputation of taxi id 6275 in each cell of 1000 m^2.

After filtration of data for each taxi, we got 5,051,513 GPS trace points. We divided the data set into two parts, 70% of the data we have used for a historical time period T1 and the rest of the trajectory data has been used for code transmission in time period T2. In T1 each taxi will build his reputation profile as shown in figure 4.12. Each vehicle will maintain its trajectory history with respect to area visit and periodically update to CC with its heat table. Then for each given request CC can calculate reputation index with respect to number of SSs covered by each vehicle. A vehicle having highest index will be selected as an optimal vehicle. The heat table shows the movement of taxi ID:6275 in all cells of 1000 m^2 of selected area 25×25 km^2. Similarly, it can build its profile for 100 m^2, 250 m^2, and 500 m^2 cell sizes. Figure 4.12 shows that this taxi has more visits to the down-town area as compared to the peripheral areas. Similarly, all the taxies will build their profiles.

4. Smart sensors locations

To verify the proposed scheme, we need to find the locations of SS and RSUs. At first, we selected a set of 1000 SSs and find 1000 random locations for SSs near the road network of 100 m^2 cells of the selected area. To minimize the contingency of the experiment, each group of comparison experiments was performed 5 times for each set of SSs, for a total of 30 comparison experiments. The random locations of these 1000 SSs for 1^{st} experiment is shown in figure 4.13(a). In our proposed scheme, we use algorithm 9, to find the optimal MNs which can complete the largest number of tasks to transfer the updated code to these SSs.

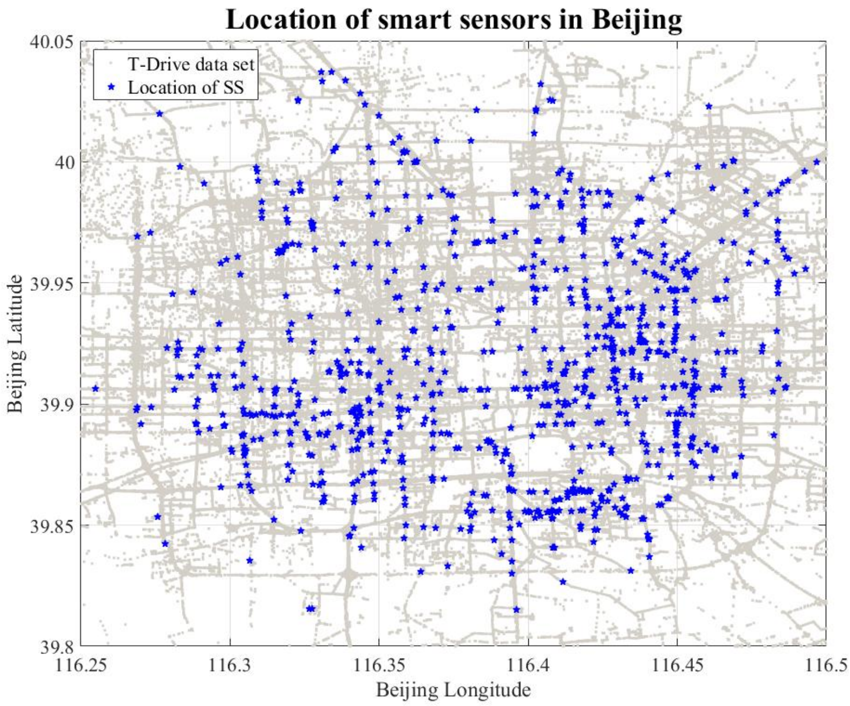
5. RSU deployment

Our next task is to find the locations for RSUs. In our proposed scheme, we use algorithm 8 to find the optimal locations of RSUs and we got 36 locations of RSU to implement this task. For comparison, we use the same number of RSU in fixed RSU selection, and Random RSU selection. For fixed RSU selection we selected 36 consecutive cells in the down-town area having the high number of taxi visits and for Random RSU selection we selected 36 random cells having taxi visits greater than 100. The RSU deployment by all three schemes is shown in figure 4.13(b).

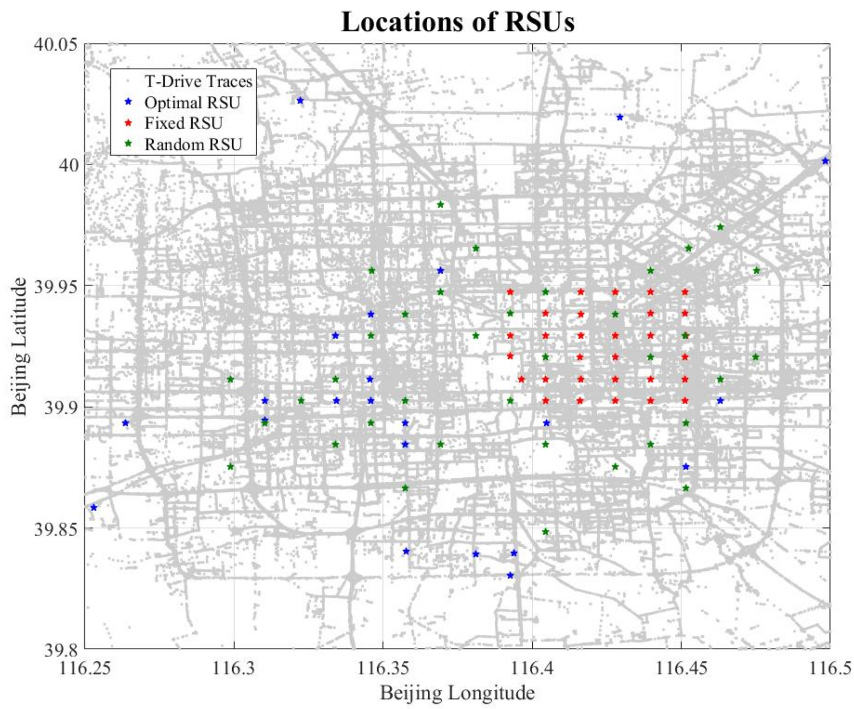
6. Results and discussion

For the 1st experiment, we execute the scheme for a delay-tolerant interval of 12 hours and compare the CFCR for all deployments. We got the results as shown in figure 4.14(a). Our proposed scheme outperforms in all 5 cases and obtains a CFCR rate of up to 85%. The fixed deployment scheme has also obtained batter CFCR it is because these locations have the highest frequency of MNs visits. On the other hand, random deployment obtains the highest value up to 70% in all 5 experiments.

In this experiment set, we increase the delay-tolerant interval equal to 24 hours and

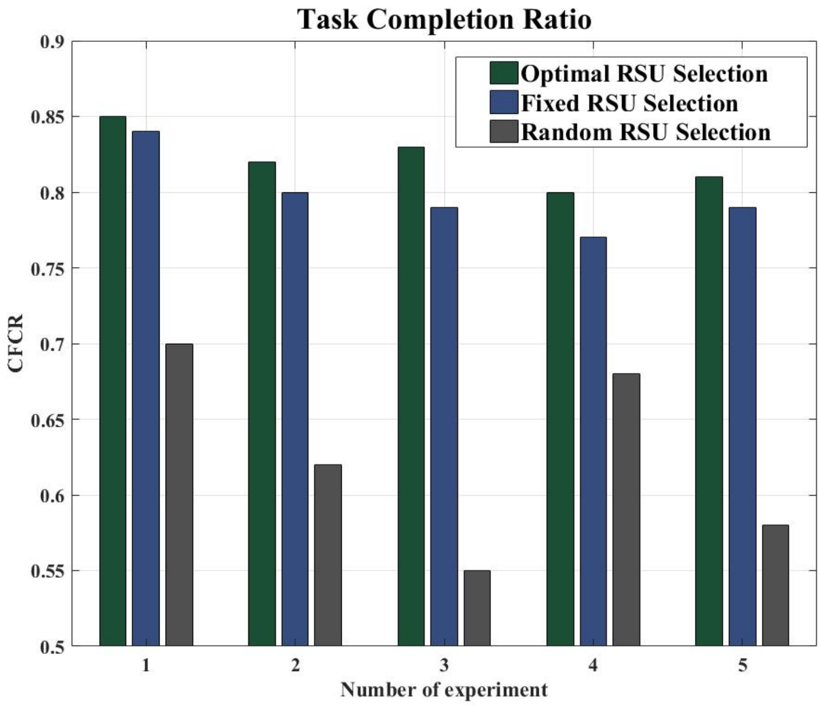


(a) Location of SS in Beijing City.

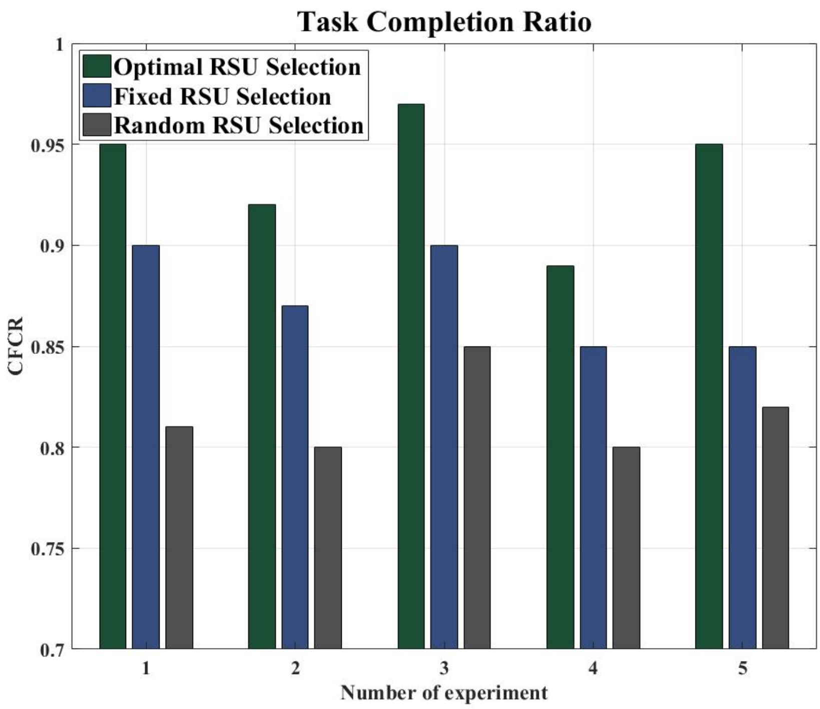


(b) Location of RSU in Beijing City.

Figure 4.13: Location of SS and RSU in selected area of $25 \times 25Km^2$.



(a) Delay tolerant interval = 12 hours and number of SSs = 1000.



(b) Delay tolerant interval = 24 hours and number of SSs = 1000.

Figure 4.14: Code forwarding coverage rate CFCR.

calculate the CFCR rate for all three schemes. When we increase the delay-tolerant interval then the CFCR rate also increases in all three cases. Our proposed scheme got the highest completion rate up to 97 %, whereas fixed and random deployment got the highest completion rate up to 90% and 85% as shown in figure 4.14(b).

The rest of the experiments are executed with 1200, 1400, 1600, 1800, and 2000 smart sensors. The results of the CFCR rate for all 30 experiments are shown in figure 4.15. our proposed scheme outperforms and got the highest completion rate as compared to fixed and random deployment.

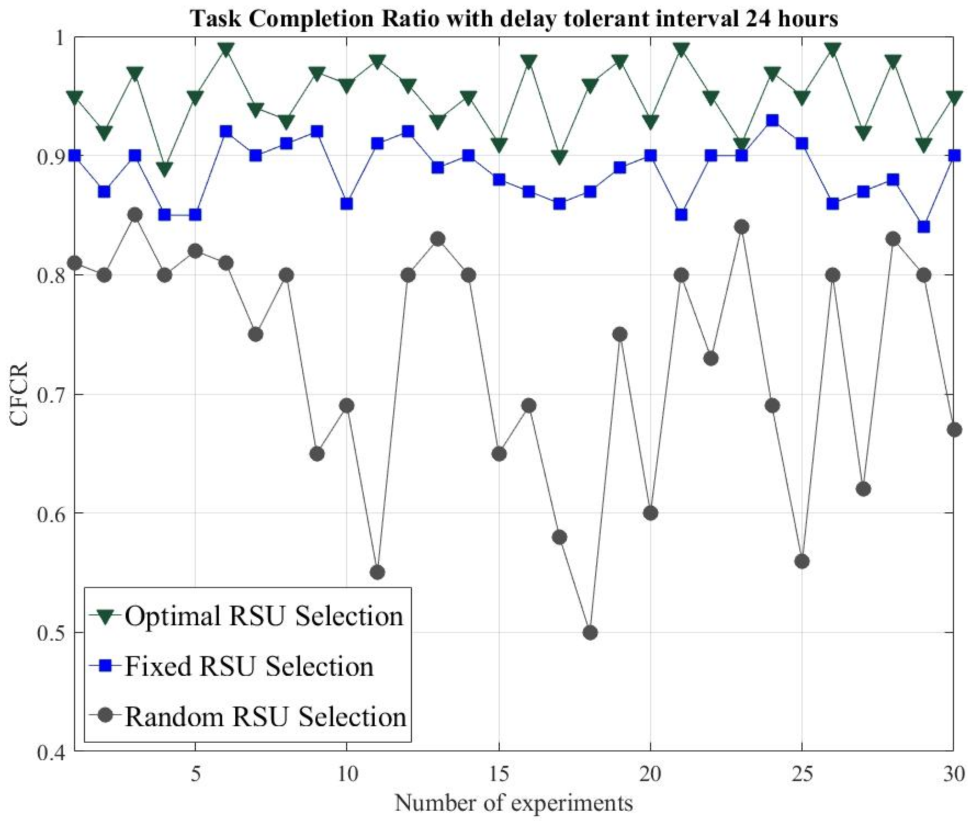


Figure 4.15: Task completion ratio for delay tolerant interval of 24 hours. (G1=1000 SSs→ experiment number1-5, G2=1200 SSs→ experiment number 6-10, G3=1400 SSs→ experiment number 11-15, G4=1600 SSs→ experiment number 16-20, G5=1800 SSs→ experiment number 21-25, G6=2000 SSs→ experiment number 26-30).

Now we assume that there is no limit on delay-tolerant interval and execute code

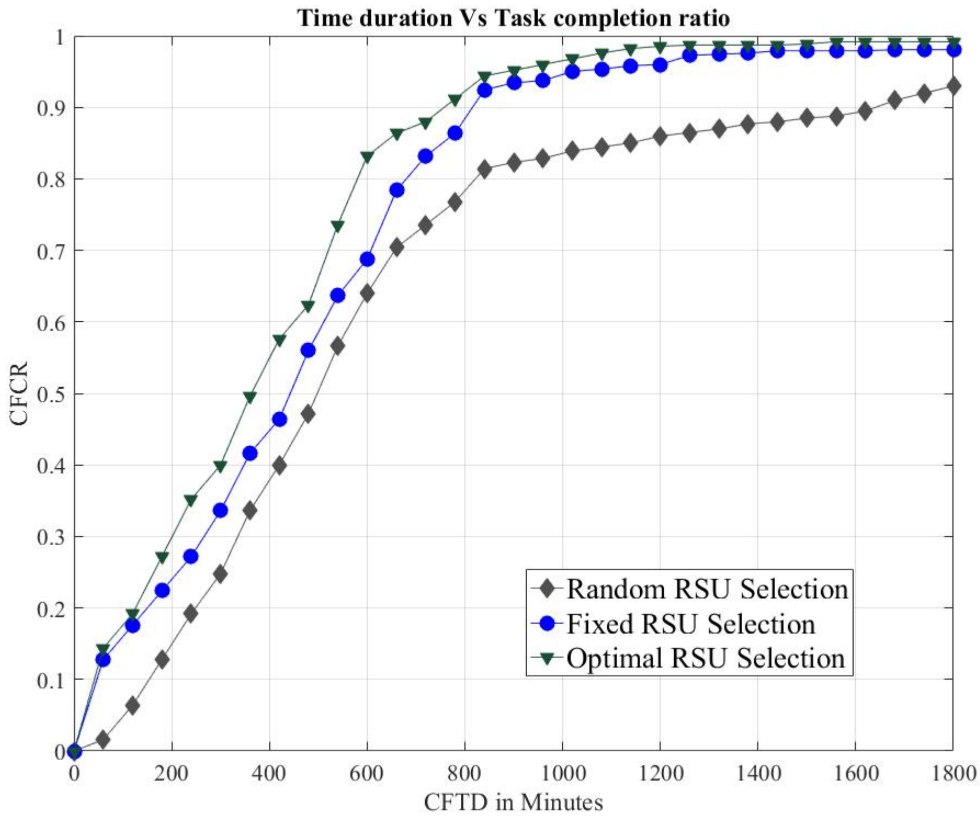


Figure 4.16: CFCR Vs CFTD in minutes for 400 vehicles in 1000 m^2 clusters.

dissemination for all there schemes for 1000 SSs. The results obtained are shown in figure 4.16, it compares the code dissemination rate CFCR over CFTD in minutes. The fixed deployment scheme got results closer to our proposed scheme of RSU deployment, after 1800 minutes.

Table 4.4 shows that our proposed code dissemination scheme obtains a 98.72% CFCR code completion rate in 1260 minutes. The figure shows that there is no change in the completion rate after this time. By observing traces it is found that a few MNs passed through these points but at that time they didn't visit any RSU location, i.e these MNs have no updated code when they visited these SSs locations. After 1800 minutes fixed and random deployment code dissemination archive 97.88% and 90% respectively. The number of sensors who didn't get the updated code is the smallest in our proposed

Table 4.4: Results of three RSUs deployments schemes.

RSU Selection	CFCR	CFTD (in minutes)	Un updated sensors
Optimal RSU	98.72%	1260	1.28%
Fixed RSU	97.88%	1800	2.12%
Random RSU	90%	1800	10%

scheme, which is 1.2% after 1260 minutes.

The results in the case of fixed RSU are closer and obtain in 1800 minutes, whereas in the case of random RSU selection after the same time 10% of the sensors didn't get the code. Our proposed OCFS achieves the highest code transmission coverage rate in less time.

4.6 Summary

In this chapter, we proposed data and code forwarding schemes for our proposed VADD framework. To evaluate the data dissemination scheme we develop a case scenario in Auckland city by using the real data of road traffic count and to verify the effectiveness of the code transmission scheme we use the real traces of Beijing city.

In this chapter, we developed delay models for sustainable data dissemination to be used to tackle typical SC daily data traffic. As a proof-of-concept of our work, we presented Auckland City case scenario where delay tolerant data is delivered to data centers. We found that our proposed system can effectively use the daily vehicle mobility of Auckland City for enormous data transmission from data sources to data centers. The results obtained show that the proposed system can offer up to four times better data transfer rate than the dedicated core network with a data rate of 1 Gbps.

In the 2nd part of this chapter, we proposed an optimal code forwarding scheme OCFS. In this strategy, we used thousands of vehicles as the mobile nodes to transfer code from the data center to SSs. RSUs collect the codes from CC and forward this code to optimally selected MNs. These MNs transfer the collected code to SSs whenever they are in their wireless range in store carry forward opportunistic communication. The huge displacement of vehicles in a city has a large number of vehicles that lead to high coverage of code. OCFS is low cost because we need to install the RSUs only, the other components MNs and SSs are already available in the proposed system. In OCFS we propose multi-objective optimization of an NP-hard problem, the minimum set coverage of the optimal selection of MNs, and RSUs. This optimal solution enhances the coverage ratio and reduces the time for this code transmission. The results of the experiments validated the proposed scheme of code forwarding from the users to SSs in SC.

In this chapter, the data and code forwarding schemes are proposed and in the next chapter, we discuss the usefulness of the proposed VADD framework in terms of energy consumptions and carbon emissions.

Chapter 5

Energy Consumption & Carbon Emissions by VADD

In chapter 3, we proposed a system overview of VADD and discussed data offloading in smart cities. In chapter number 4, we propose data and code forwarding schemes. In this chapter, we explore energy consumption models for data transmission. One of the main challenges of operating a smart city (SC) is collecting the massive data generated from multiple data sources (DSs) and transmitting them to the control units (CUs) for further data processing and analysis. These ever-increasing data demands require not only more and more capacity of the transmission channels but also results in resource over-provision to meet the resilience requirements, thus the unavoidable waste as a result of the data fluctuations throughout the day. Besides, the high energy consumption (EC) and carbon discharge from these data transmissions posing serious issues to the environment we live in. Therefore, to overcome the issues of intensive EC and carbon emission (CE) of massive data dissemination in SCs, we propose an energy-efficient and carbon reduction approach by using the daily mobility of the existing vehicles as an alternative communications channel to accommodate the data dissemination in SCs. To illustrate the effectiveness and efficiency of our approach, three numerical studies evaluated in this chapter confirm the noteworthy cost and energy saving can be obtained by utilizing existing vehicle volume on the road for big data dissemination in smart cities. The analysis of energy consumption and carbon emissions using traditional min-cost flow problem and multi-commodity flow problem is also included to optimize energy cost and carbon emissions.

A brief introduction is given in Section 5.1. Section 5.2, presents background and related work. Mathematical models are presented in Section 5.3. Three case studies are presented in Section 5.4, 5.5, and 5.6. A comparison of our proposed framework has been conducted with the existing techniques in section 5.7. Finally, the chapter is concluded in Section 5.8.

5.1 Introduction

In the near future, SCs are envisioned to provide services such as road lights conversing with the smart grids, urban parks associating with administrations, seasides conveying cautions on pollution levels, and flood alerts to disaster management. This results in generation of huge data, and they are stored in the data clouds or some data control centers for processing and analysis. A huge number of sensors, installed in each city will continually generate a tremendous amount of data [153]. For instance, Westminster City Council has introduced solar waste bins that can speak to city council workers and inform them how much full they are. The framework used infra-red and telemetry sensors and prompted a 60% decrease in cost for waste collection. Smart homes are now turning into reality, and it is expected that they will be available commercially by next year. Splunk predicts that one smart home today can create as much as 1 GB of data in a week. This means that all the UK smart homes may generate more than 26 million GBs of information [154]. Furthermore, video surveillance of the entire city, smart sensors, and smart girds also generate a big volume of data on each day for processing and analysis[155].

The transmission of a large amount of SCs data is going to swamp existing infrastructure networks, and it presents some unwieldy challenges. The operator's concern is to enhance the performance of their network by adding more capacity to their networks or by efficiently using the existing network resources. To address the issues of this explosion of data traffic we have different solutions. One of the expensive solutions is to elevate the existing networks to the next generation networks. Another expensive solution is to enhance the network capacity. However, the problem with these solutions is that it requires an enormous amount of cost for the operational expense (OPEX) and capital expenditure (CAPEX) [156].

To transmit huge information in SCs, one of the candidate networks is a power line communication (PLC), it may provide LAN connectivity, WAN access, and some command and control capabilities [157]. However, interoperability problems, low dispersion and short-range communication of PLC networks become weak points for its success in the market of data transmission system [158]. Other energy-efficient solutions may include the use of Zig-Bee, Bluetooth and WiFi along with PLC for smart metering and smart homes. Due to capacity and wireless range limitations, these solutions may be suitable for a short-range communication rather than long-distance communications [159]. Furthermore, fiber-optic in SC for huge data volume dissemination might be a good competitor. However, the city-wide deployment of optical-fiber system requires high budget [160].

Lately, Cellular Networks (CNs) appear to be an alternative solution for big data communication in SCs. However, huge mobile data demands have already created issues of devastating CNs [161]. Figure 5.1 shows that mobile data demands will grow up to 38 PB per month by 2021 in New Zealand, which is 380% more than 2010 [134]. Broadband Internet and traditional core networks could also be possible candidates for SC data transmissions, but these networks are also now congested networks, traffic on

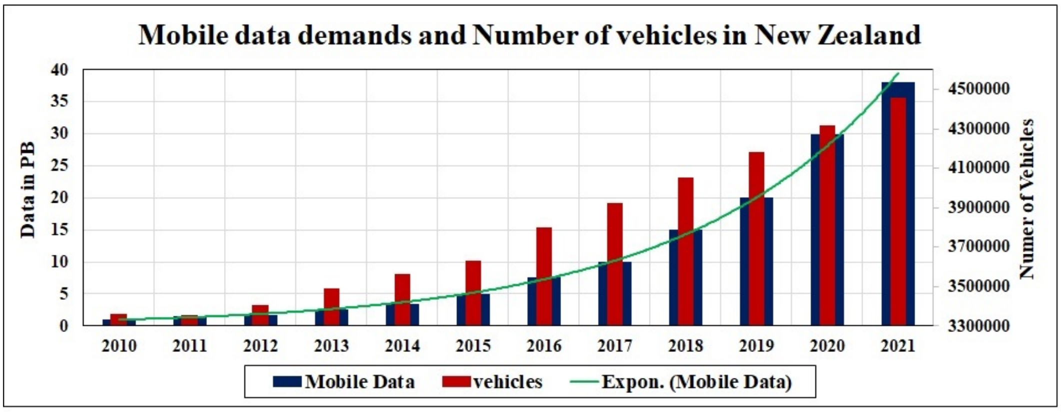


Figure 5.1: Mobile data demands and number of vehicles in New Zealand (NZ).

the Internet has increased more rapidly than its existing capacity [10, 162]. As a result, the problem of data dissemination between data sources and control units in SC ought to be solved by using some other types of hybrid networks instead of the using only Wi-Fi, 3G, LTE, Internet and so forth [163]. Under this topic, vehicular networks by using the existing routine rides in the city could be a possible solution to disseminate big data in SCs.

On the other hand, there is one more big challenge in energy consumption in information communication technology (ICT). Andrey gives his analysis of global EC, with the best average and worst-case scenarios. International Energy Agency (IEA) estimated that Global electricity consumption would grow by 2.8 to 3.4% per year and ICT EC will be a big ratio (51%) of global EC in 2030. The electricity demand of communication technology is increasing and will be 30,715 TWh of a worst-case scenario in 2030. Andrey predicts that depending on different scenarios of ICT, EC in 2010 is 8% to 14%, 2020 6% to 21% and 2030 8% to 51%, respectively. Hence, it is also a critical issue that needs to address [1, 164].

These factors of congested networks and EC in ICT create interest in alternate energy-efficient solutions to reduce the pressure of data traffic on traditional core networks. In this direction, Cho and Gupta [165] enhanced the data transfer performance by suggesting the concurrent use of conventional network and postal system. Some portion of the data is forwarded using the conventional network while some other delay-tolerant data is forwarded on hard disks through the postal system and it requires comprehensive data scheduling. In [11], Marincic and Foster enhance the same work for energy cost optimization and reduce the CE through various optimization scenarios. Munjal et al. proposed a centralized Software-Defined Vehicular Connectivity procedure that enables scalable and adaptive control of the scheduled vehicles to offload traffic [166].

Thus, it is useful to introduce an energy-efficient option by using the existing vehicular networks to address this problem of big data transmission and to reduce the load on existing infrastructure networks. In this chapter, we are proposing to use daily routine rides of SC's vehicles and annual average daily traffic (AADT) of Auckland city for delay-tolerant data transmission. The main contributions of this chapter are as follows:

- We develop an energy consumption model and an Auckland case is evaluated by using these models. Our proposed framework saves a huge amount of energy cost as compared to the traditional network.
- We use a min-cost flow problem to optimize energy consumption and carbon emissions by using vehicle count on the road.
- We design new algorithms using the multi-commodity flow problem to select an energy-efficient network for data transmissions to save more energy cost.

5.2 Background

Presently, different data clouds offer a data storage space on their web servers to their clients. Normally, numerous duplicates of data are stored on these servers [167]. The best example to transfer data from a client location to these servers is the data delivery services of Amazon Web Services (AWS). AWS offers data movement by using their data vehicles. For this purpose, AWS sends their vehicle to the client, the client loads required data onto the vehicle, and the vehicle is driven back to AWS for loading the client's data onto their web servers [168]. A digital map provider company Digital Globe used snowmobile services of AWS to transfer 70PB data onto the cloud of AWS [169]. This type of data transmission requires some fiscal cost because it needs a dedicated vehicle for this purpose. Moreover, it consumes energy and emits carbon in the environment.

A big data of delay-tolerant applications can be transferred by efficiently using the energy of daily routine trips of vehicles. Vehicular delay-tolerant network (VDTN) [79], is built on the theory of delay-tolerant network (DTN). It handles delay-tolerant applications at low cost and on unpredictable network conditions. Vehicles can be used as data carriers between terminal nodes either in rural areas or in emergency scenarios. This strategy can be helpful in both V2V and V2I communication. In this direction, Kashihara et al. [80], proposed a data offloading scheme to offload data, particularly to scheduled vehicles by using short-range and high-speed wireless communication. The motivation was to reduce traffic congestion because of the high data volume on traditional networks. Similarly, Hunjet et al. [170] and Usbeck et al. [171] uses VDTN and implement data forwarding schemes by using data farriers. A vehicular data dissemination project is implemented in France to reduce the load on conventional networks [172]. Dessler et al. [173] used a parked vehicle for data communication among other vehicles and proposed a protocol called vehicle cord where parked vehicles are used as RSUs.

In another work [174], Cho et al. use a hybrid network to upgrade the performance by proposing the concurrent use of traditional core network and postal framework. A portion of the time-critical data is exchanged by using the conventional core network while the postal system is used to transfer delay-tolerant data with the help of hard drives. It requires complex data scheduling for data forwarding. Marincic et al. upgrade a similar work to minimize the EC and reduce the CE [11].

Moving Vehicles on the road and roadside APs of WiFi Network can also be used as a practical solution for cellular data offloading in SCs, called vehicular Wi-Fi offloading. The research in this direction has an objective to enhance the cellular data offloading performance, particularly for delay-tolerant and non-interactive applications [88, 89]. At the same time, the advantages of opportunistic communication and D2D revolution are sound for some vehicular related use cases. It can empower location-based peer-to-peer applications and services. For example, thinking about the huge number of connected smart vehicles, a new software update of the smart vehicles can put a critical load on the cellular network, and cost huge money for vehicles owners. In this way, the new update can be downloaded by some designated vehicles by using D2D communication with RSUs, and then this update can be exchanged to other vehicles by D2D transmission. Along these lines, the majority of the cellular load can be shifted to V-D2D communication and accordingly cellular bandwidth, energy, and cost can be saved [175].

In this chapter, we are trying to describe how much data can be uploaded/downloaded to/from RSU while vehicle is on the move and how the proposed algorithm can be used to select an energy-efficient network mode to transfer big data between two points.

5.3 Energy Consumption Model

A detailed overview of our proposed framework VADD has already been given in 3.1. To transfer the data from source to destination, following are the two models of energy consumption for transport network and traditional core network.

5.3.1 Energy Model for Transport Network

EC for delay-tolerant information transporting between two points by using vehicles depends on two things. 1^{st} EC for data loading and 2^{nd} EC for transporting the weight of storage device from point A to B.

$$E_{Veh} = E_{DataLoad} + E_{Transport} \tag{5.1}$$

If k is the total number of offloading points between A and B then $E_{DataLoad}$ can be calculated as follows

$$E_{DataLoad} = 2 \times \sum_{i=1}^{k} E_{D2D_i} \tag{5.2}$$

$E_{transport}$ depends upon the distance between the sender and receiver, gross weight of the vehicle, and fuel economy of the vehicle. If n vehicles are used to transport total data between them and $W_{Storage}$ is the storage device weight then we can calculate the EC for whole data shipment as follows.

$$E_{Transport} = C_{fuel} \times W_{Storage} \times \sum_{i=1}^{n} E_{Shipment_i} \tag{5.3}$$

C_{fuel} is fuel constant, and it converts the volume of fuel into energy consumed i.e., from litters to joules. If FE is the fuel economy then $E_{Shipment}$ can be calculated as follows.

$$E_{Shipment} = \frac{d_{AB}}{FE \times V_{Load}} \tag{5.4}$$

If W_{Total} is the gross weight of the fully-loaded vehicle then V_{Load} can be calculated as follows.

$$V_{Load} = W_{Total} - W_{Empty} \tag{5.5}$$

5.3.2 Energy Model for Core Network

Data transmission begins from the 1st datagram at the source and ends when the last datagram is delivered at the destination. EC for conventional network depends upon the energy consumed at the sender, receiver, and all intermediate devices. These devices contain switches and routers. If D_{Vol} is the total data volume, E_{SR} is the EC at source and destination, and we consider the incremental EC at all k intermediate devices, then the EC for conventional network case can be calculated by using the following equation.

$$CnEc = E_{SR} + \sum_{i=1}^{k} E_{inc_i} \tag{5.6}$$

If B_{up} is the uploading and B_{down} is downloading bandwidth at sender and receiver respectively, $\triangle P$ is power change at sender/receiver while transmitting/receiving the data, then E_{SR} can be calculated as follows.

$$E_{SR} = max\left(\frac{D_{Vol}}{B_{up}}, \frac{D_{Vol}}{B_{down}}\right)(\triangle P_{Sender} + \triangle P_{Receiver}) \tag{5.7}$$

For intermediate devices, if E_{bit} is the energy cost per bit then energy consumption for total D_{Vol} data is calculated as follows.

$$E_{inc} = D_{Vol} \times E_{bit} \tag{5.8}$$

Energy consumed per bit is calculated as a fraction of the max power P_{max} and available bandwidth B.

$$E_{bit} = \frac{P_{max}}{B} \tag{5.9}$$

5.4 Case-I: EC & CE by VADD Framework

Energy consumption and carbon emission depend upon the data volume, disk weight, and the distance between source and destination in our proposed case. In the following case, we use our proposed EC model to calculate energy cost and carbon emission for data transmission.

5.4.1 Auckland Case Scenario

Proposed energy models show that Energy consumption depends upon the data volume and the distance between source and destination. For internet case, we consider 11 LAN switches, 2 core and edge routers with 0.1 Gbps uploading and downloading bandwidth at the source and sink respectively [11]. For vehicle case, we transfer 100TB data and vary the distance from 10 Km to 250 Km, by keeping the average speed of vehicles constant.

5.4.2 Results and analysis

Fig. 5.2, shows that energy consumption of the proposed case is better than the internet case when the distance is less than 150 Km, but when the distance is greater than 150 Km then the energy consumption of our proposed model is greater than the internet case. It shows that our proposed model is suitable for a smart city if it has a diameter of less than 150 Km. This assumption supports our proposed framework of “A Sustainable Vehicular Based Energy Efficient Data Dissemination Approach in Smart City”. Fig. 5.3, shows that if we increase the data volume the energy consumption increases in both cases. Our proposed model of energy consumption consumes less energy in case of the vehicular network. To transfer 100 TB data between the above-said source and destination (Telemetry Site 17 to CBD.) the energy consumption for the vehicular case is 89.60 MJ, which is much less than internet case that is 227.5 MJ.

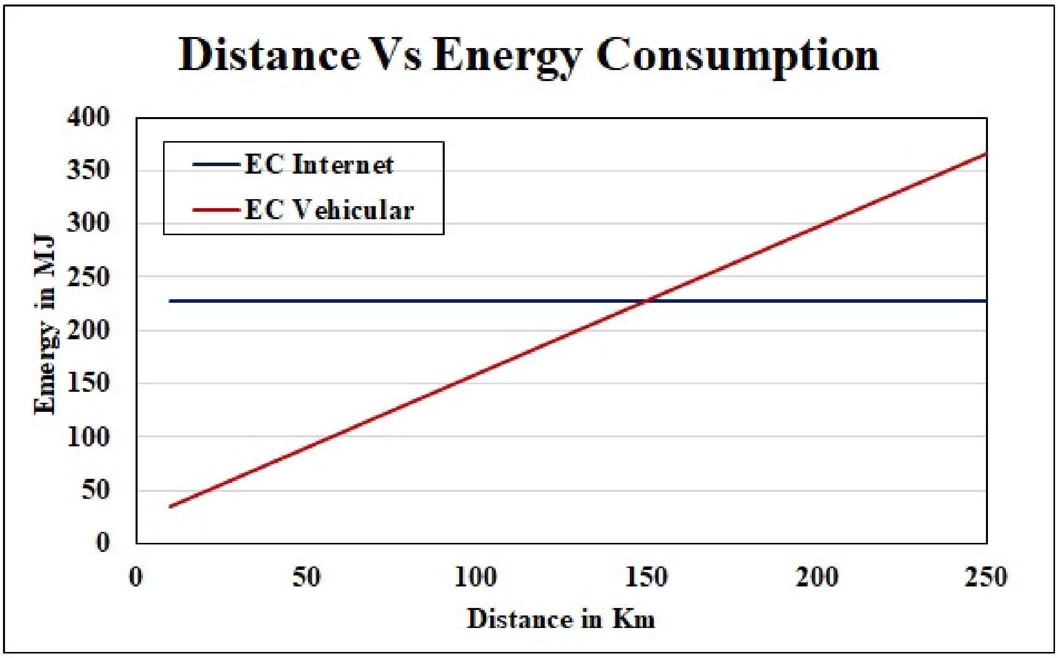


Figure 5.2: Distance vs Energy consumption.

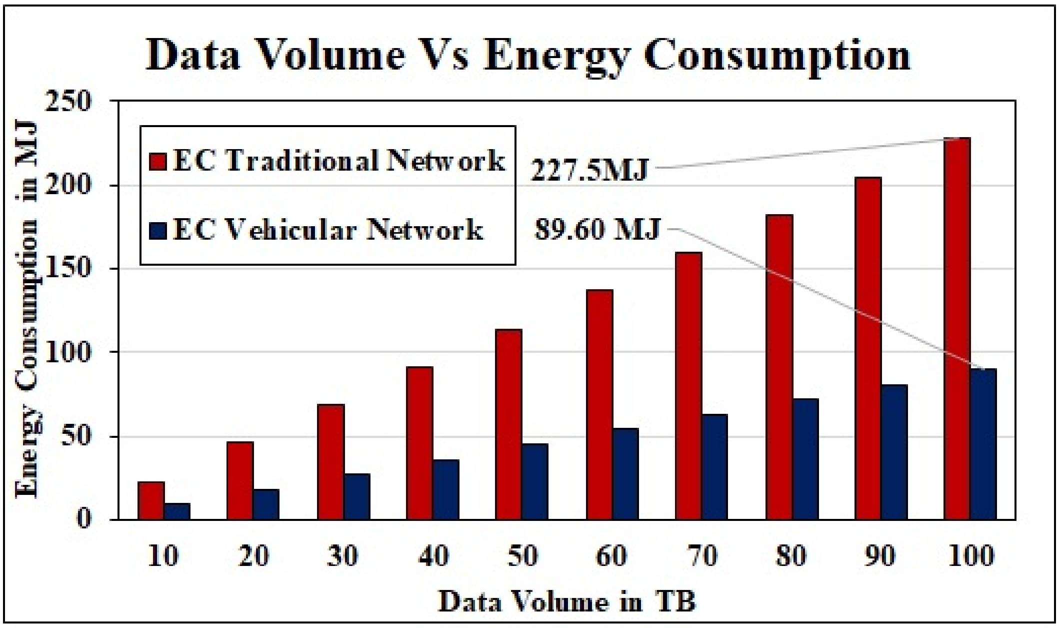


Figure 5.3: Data volume Vs Energy Consumption.

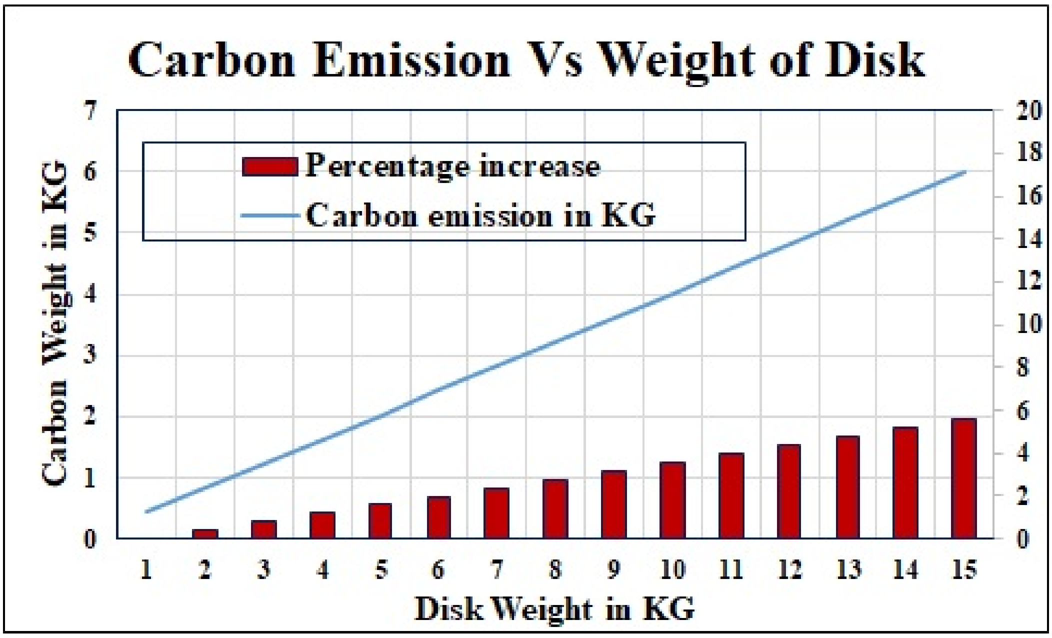


Figure 5.4: Carbon Emission vs Weight of disk.

Eq. 5.3 shows that energy consumption also depends upon the weight of the disk. If we increase the weight of the storage device then the energy consumption also increases. To transfer data between the same source and destination we increase the weight of the storage device and keep the remaining parameters constant. The results in figure 5.4 show the increase in carbon emissions while increasing the weight of the disk in the moving vehicles of our proposed framework VADD. The higher value of weight will consume more energy. This energy consumption emits more carbon into the atmosphere.

5.5 Case-II: Energy-Efficient Path Selection

In this case, we use the traditional min-cost flow problem to minimize the energy cost to transfer big data over the road network by using multi-hop communication in our proposed framework of VADD.

5.5.1 Problem formulation

We have a set of nodes N and set of edges E as shown in Fig. 5.5. Let $G = (N, E)$ be a directed graph a road network and a flow having sender node $s \in N$ and a receiver node $t \in N$. Each edge $(u, v) \in E$ having edge capacity $c(u, v) > 0$, energy cost $E_{Veh}(u, v)$ and vehicle flow as vehicle volume on the road $f(u, v) \geq 0$, with supporting negative cost edges as a receiver edge. The energy cost of a flow on a given edge (u, v) is $E_{Veh}(u, v).f(u, v)$. In this problem we need to send a data flow d, from source node s to sink node t.

The main objective of this problem over here is to minimize the total energy cost of a given flow over some given edges under some conditions of road capacity, which is represented here by the vehicle volume of the road.

Minimize :

$$\sum_{(u,v)\in E} E_{Veh}(s, t).f(s, t)$$

Subject to:

$$f(s, t) \leq c(s, t)$$

$$f_{(w,x)}(n) = -f_{(x,w)}(n)$$

$$\sum_{w\in N} f_{(w,x)}(n) = 0 \Longrightarrow \forall w \neq s, t$$

$$\sum_{(u,v)\in N} f_{(u,v)}(n) = d \text{ and,}$$

$$\sum_{(u,v)\in N} f_{(v,u)}(n) = d$$

5.5.2 Minimizing Energy cost

In this case study, we propose a problem having a road network as shown in Fig. 5.5. Each road has vehicle volume as its capacity and its energy cost to transmit 1TB data by using an embedded disk having weight 0.95 Kg in a car. In our calculations, we

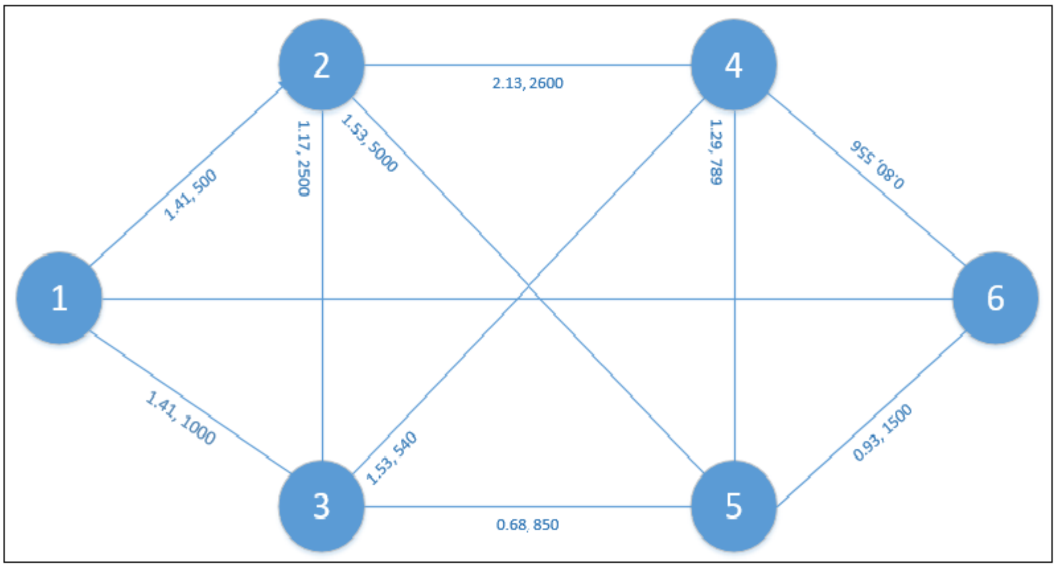


Figure 5.5: Graph of building network flow.

are using the Toyota Prius car and assume the fuel constant C_{fuel} = 37,624,722.29 J/L [176]. The cost calculations are shown in Table II. Here we solve our problem for energy consumption and carbon emissions by using traditional minimum cost flow problem and use IBM ILOG CPLEX Optimization Studio for an end to end energy cost calculations [14].

5.5.3 Energy Consumption

By using the above minimum cost flow model and the energy cost models from Eq. 5.3, we further minimize the energy consumption cost by transferring 20 TB data from source node 1 to sink node 6. Calculated results for both internet and vehicle case show that energy consumption is more efficient in the vehicle case as compared to the internet case. We consider 11 LAN switches, 2 core and edge routers with 0.1 Gbps uploading and downloading bandwidth at the source and sink respectively. Whereas in shipment case 60 vehicles are used to transfer the required data from source to destination and disk capacity of each vehicle is 1TB. The optimal route to minimize energy consumption is [1→3→5→6]. From source node 1, 20 vehicles load the data and drop at an offloading

Table 5.1: Summary of links and Energy Cost in MJ

Edge	Distance	speed	Veh. Volume	ECost(MJ)
1⟶2	50	50	500	1.41409939
1⟶3	50	50	1000	1.41409939
2⟶4	80	50	2500	1.172239512
2⟶3	40	50	2600	2.139679024
2⟶5	55	50	5000	1.535029329
3⟶2	40	50	850	1.172239512
3⟶4	55	50	540	1.535029329
3⟶5	20	50	850	0.688519756
4⟶5	45	50	789	1.293169451
4⟶6	25	50	556	0.809449695
5⟶4	45	50	2250	1.293169451
5⟶6	30	50	1500	0.930379634

point 3, then 20 vehicles transfer data from point 3 to 5. Finally, 20 more vehicles pick the data from offloading point 5 and drop at the final destination node 6. Fig. 5.6 shows that to transfer 20TB data from node 1 to node 6, the energy consumption for our proposed VADD framework is 60.66MJ, whereas in case of the traditional network it is 102.7 MJ. This case scenario shows that our proposed system is more energy-efficient as compared to the core network.

5.5.4 Carbon Emission

Data transfer for both internet and vehicular shipment consumes some energy from different resources like fuel and electricity. This energy consumption emits carbon into the atmosphere. Fig. 5.7, shows that carbon emission for VDTN case is less than internet case. The results show that the carbon emissions in this scenario is 0.592 Kg/TB for shipment case and 1.0025 Kg/TB for internet case. To get these results, we used the carbon emission 0.703 (CO_2)Kg / kWh [177] and 1 MJ = 0.2777778 kWh [178].

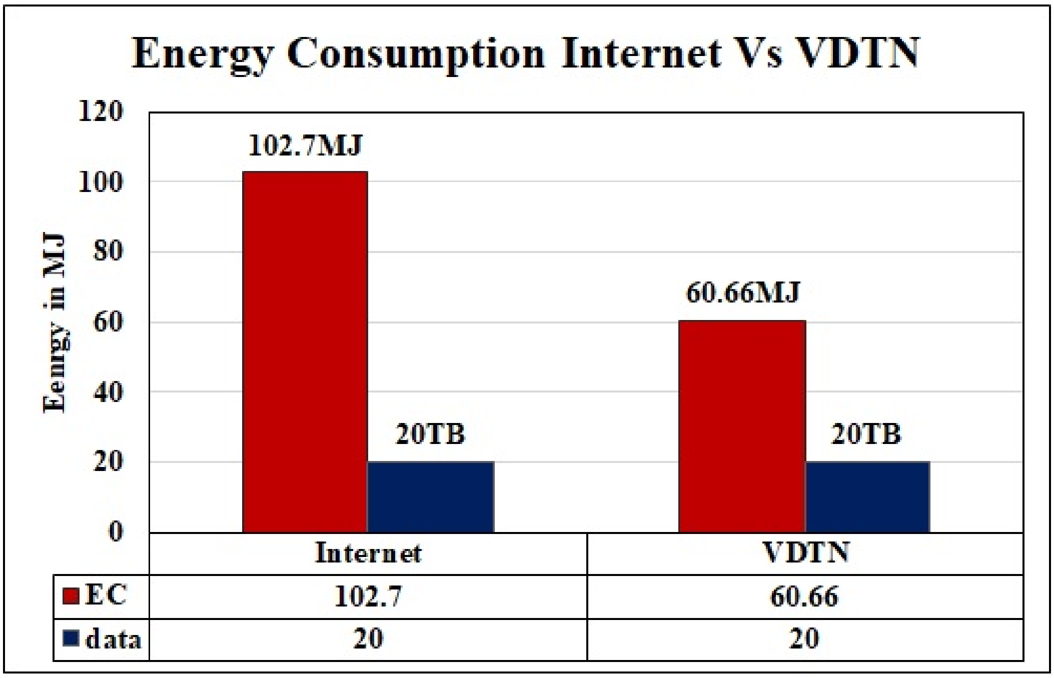


Figure 5.6: Energy consumption internet vs VDTN case.

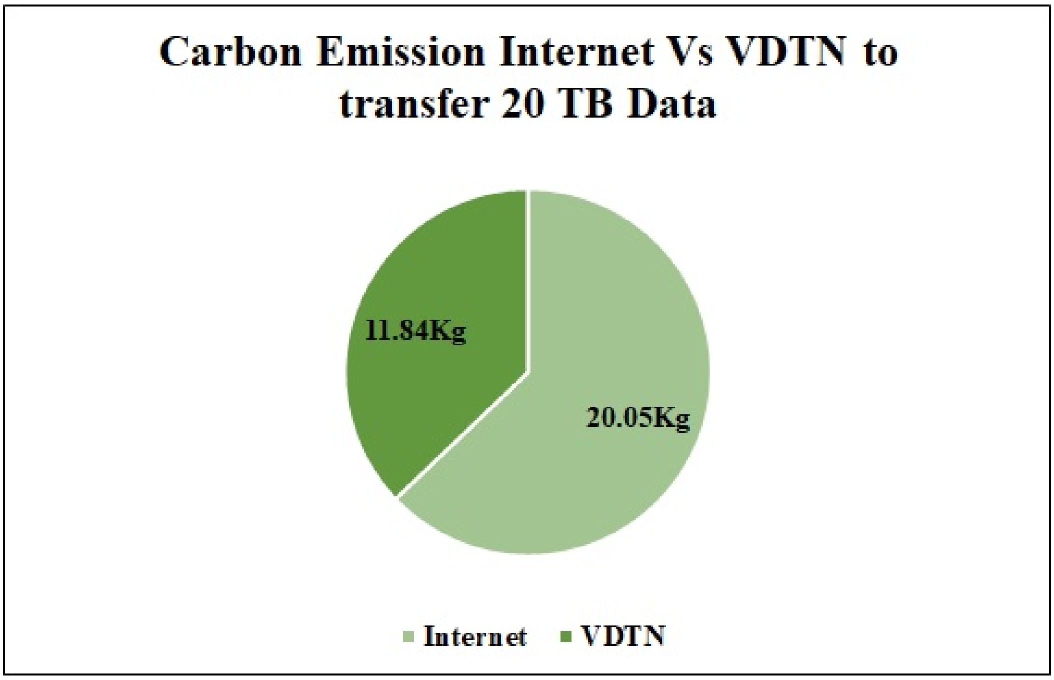


Figure 5.7: Carbon emission Internet vs VDTN case.

5.6 Case-III: Energy-Efficient Network Mode Selection

In this case, we use the optimal mode selection algorithm to select the energy-efficient network mode for a given task of data transmission.

5.6.1 Minimizing Energy Cost

The goal of this work is to minimize the total cost of sending the data requirements across the complex road network or traditional core network, while satisfying all demands/supplies, and respecting arc capacities. We solve the energy optimization problem for road network by using the multi-commodity flow problem and calculate the energy cost for optimal paths. Similarly, we calculate the energy cost for the core network and finally, we find that which network is suitable for data transmission concerning energy consumption.

5.6.2 Problem Formulation

Let $G = (V, E, C, A)$ be a capacitated undirected graph, where V is set of data sources or data centers locations, E is set of road links between data sources and destinations, C is the capacity of each road w.r.t vehicle count, and A is a set of cost per unit flow for a commodity b_i on each link $(i, j) \in E$. $R = \{(s_i, t_i, b_i)\}$ be a set of requirements, $s_i \in V$ is data source and $t_i \in V$ is destination data center for commodity b_i. For each edge $(i, j) \in E$ and each commodity r, associates a cost per unit of flow, designated by a^r_{ij} . The demand (or supply) at each node $i \in V$ for commodity r is designated as b^r_i, where $b^r_i > 0$ denotes a supply node and $b^r_i < 0$ denotes a demand node. We define decision variables x^r_{ij} that denote the amount of commodity r need to send from node i to node j. The amount of total flow, for all commodities, that can be sent across each link is bounded above by c_{ij}. We need to minimize the transport network energy consumption $TnEc$ by using the vehicle mobility of complex road network.

Minimize :

$$TnEc = \sum_{(i,j)\in A} \sum_{r\in R} a^r_{ij}.x^r_{ij}$$

Subject to :

$$\sum_{r\in R} x^r_{ij} \leq c_{ij} \qquad (i,j) \in E \qquad (Capacity)$$

$$\sum_{i,j\in E} x^r_{ij} - \sum_{i,j\in E} x^r_{ji} = b^r_i \quad i,j \in V, r \in R \qquad (Balance)$$

$$x^r_{i,j} \geq 0 \qquad (i,j) \in E, r \in R$$

$$\sum_{i,j\in E} x^r_{ij}(n) - \sum_{i,j\in E} x^r_{ji}(n) = \begin{cases} b^r_i & if\ n = s^r \\ -b^r_i & if\ n = t^r \\ 0 & otherwise \end{cases} \quad (Flow\ conservation)$$

$$\forall n \in V\ and\ r \in R$$

We use Algorithm 11, to solve the above energy optimization problem of multi-commodity flow, for multiple commodities of data transmission across the road network and calculate the energy-efficient paths.

5.6.3 Energy-Efficient Network Mode Selection

Let $G^a = (V^a, E^a, A^a)$ be a capacitated undirected graph, where V^a is a set of intermediate nodes (routers and switches), E^a is a set of the links between these intermediate nodes, and A^a is a set of unit energy cost to transfer a data commodity b_i on these links. We calculate the energy cost $CnEc$ for data transfer between the data source and data

Algorithm 11: Minimum cost multi-commodity flow

Input: A graph $G = (V, E, C, A)$, and set of requirements $R = \{(s_i, t_i, b_i)\}$
Output: Set of paths that meet the requirements

$UnMetR \leftarrow R$
$P = \emptyset$
$G_{Temp_1} \leftarrow G_{Temp_2} \leftarrow G$, and $P = \phi$
$G_{Temp_2} = G$
while $(|P| < |R|)$ **do**
  for $(r_i \in R)$ **do**
    $p_i = minimum\_cost\_flow(G_{Temp_1}, s_i, t_i, b_i)$
    if $(p_i == \phi)$ **then**
      $p_i^a = minimum\_cost\_flow(G_{Temp_1}, s_i, t_i, b_i)$
      for $(e, c_{ij} \in p_i^a)$ **do**
        if $(c_{ij} > Available\_Capacity\_in\_G_{Temp_1})$ **then**
          Increase cost of e in G_{Temp_1} and G_{Temp_2}
          for $(p \in P)$ **do**
            if $(e \in p)$ **then**
              $P \leftarrow P - p$
              Return the capacity used by p to G_{Temp_1}
            end
          end
        end
      end
    end
    else
      Reduce the capacity of edges in p_i from G_{Temp_1} $P \leftarrow P \cup p_i$
    end
  end
end
return P

center by using the traditional core network from Equation (5.6). For transport network energy cost $TnEc$ can be calculated from Algorithm 11. The central controller apply Algorithm 12 to take the decision which network is suitable for a given set of data demands. This decision is forwarded to all data sources and intermediate nodes. Finally, they select the appropriate energy-efficient network interface to forward the data.

Algorithm 12: Energy-efficient network mode selection

Input: Graphs $G = (V, E, C, A)$, $G^a = (V^a, E^a, A^a)$ and set of requirements $R = \{(s_i, t_i, b_i)\}$

Output: Energy-efficient network mode selection

$CnEc \leftarrow f(G^a, R)$

$TnEc \leftarrow f(P) \leftarrow multi\_commodity\_flow(G, R)$

for $r_i \in R$ **do**

  if $CnEc_i > TnEc_i$ **then**

    $mode_i$ = proposed network

  else

    $mode_i$ = traditional network

  end

end

return *mode*

5.6.4 Finding the Best Routes

In this case scenario, we proposed a solution, where a road network is investigated for data transfer assignments as shown in Figure 5.8. Here, nodes represent the data offloading locations and links represent the roads. Each road has the traffic count as the capacity of the road. We use an embedded storage device of weight 0.95 Kg in the vehicle and calculate the energy cost to transfer 1TB of data for each road. In these calculations, we use Toyota Prius as a relaying vehicle, and the value of fuel constant is assumed as C_{fuel} = 37,624,722.29 J/L [176] in Equation (5.3). These calculations are presented in Table 5.2. For each offloading request; we apply our Algorithm 11 to find the best route by solving the data transfer assignment as a traditional multi-commodity flow problem.

The data set in Table 5.2, provides the cost a^k_{ij} of sending a unit of commodity r along arc (i, j), with distance and speed of that particular road. To calculate the EC of the traditional core network we use Table 5.2

1. Energy Consumption

By using this multi-commodity flow model and our energy cost model, we evaluate

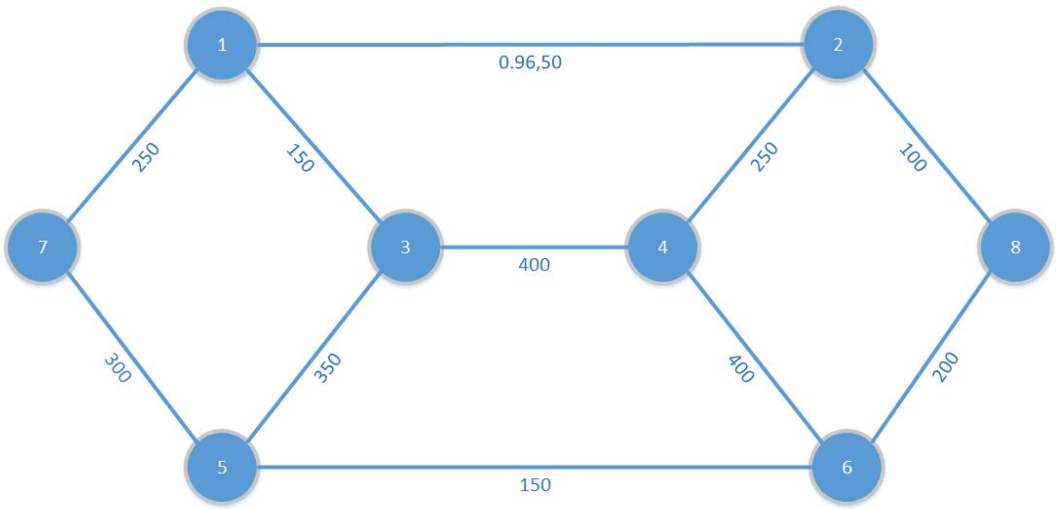


Figure 5.8: Graph of building network flow.

Table 5.2: Energy Cost per link in MJ/TB

From	To	Cost / TB	Distance (Km)	Speed (Km/h)
1	2	0.96	40	50
1	3	2.41	100	70
5	3	0.24	10	50
5	6	1.45	60	60
3	4	1.2	50	50
4	2	2.18	90	80
4	6	0.48	20	50
7	1	2.53	105	100
7	5	1.81	75	80
2	8	2.30	95	80
6	8	1.57	65	60

the EC cost for a given set of five commodities (Figure 5.9) $R = \{c1, c2, ..., c5\}$ of delay-tolerant data, with a total data flow of 580 TB. We execute Algorithm 11 on SaS optimization tool [15] to find a set P of energy-efficient paths and optimal solutions for a given set of commodities to transfer data by using the road network. To calculate the EC of the traditional core network, we use graph $G^a = (V^a, E^a, A^a)$ by applying Equation (5.6), and unit cost for EC to transfer a data demand on each edge in MJ/TB is calculated in Table 5.3. Calculated results for the core network and VDTN are shown in Figure 5.9. It shows that our proposed data transport model outperforms the traditional

Table 5.3: Core network cost to transfer 1 TB on each link.

From	To	Lan L3 Switches	Edge Routers	Core Routers	Up BW	Down BW	MJ/TB
1	2	9	2	15	0.1	0.1	5.095
1	3	6	2	3	0.1	10.0	4.355
5	3	8	2	14	1.0	1.0	2.03
5	6	11	2	3	1.0	1.0	2.275
3	4	6	2	5	0.1	1.0	4.4
4	2	9	2	6	1.0	10.0	2.02
4	6	8	2	6	0.1	0.1	4.74
7	1	11	2	7	1.0	1.0	2.36
7	5	6	2	7	0.1	0.1	4.44
2	8	11	2	5	1.0	1.0	2.32
6	8	13	2	14	0.1	0.1	5.71
1	6	8	2	9	1.0	10.0	1.925
2	6	14	2	6	1.0	1.0	2.635
3	6	9	2	15	1.0	1.0	2.215
4	8	11	2	2	0.1	0.1	5.135

core network, for all the commodities other than the commodity number 4. In this case, the distance between the data source and destination is high. That is why vehicles consume more energy to transfer data at the final destination. Moreover, for this demand, a huge amount of energy is also consumed at intermediate nodes for data offloading from vehicle to intermediate nodes and uploading from intermediate nodes to vehicles. To transfer 580TB data, the traditional core network consumes 1950MJ energy, and our proposed vehicular network consumes 1495MJ energy. In this way, our proposed vehicular network can save up to 24% of energy costs for these data assignments.

2. Energy-Efficient Network Mode Selection

Based on of the above calculated energy costs for traditional core network $CnEc$ and for transport network $TnEc$, the central controller apply Algorithm 12 to take the decision which network is suitable for each commodity in a given set of data demands. Central controller forwards this decision to concerning data sources and intermediate nodes. The data sources and intermediate nodes select appropriate energy-efficient

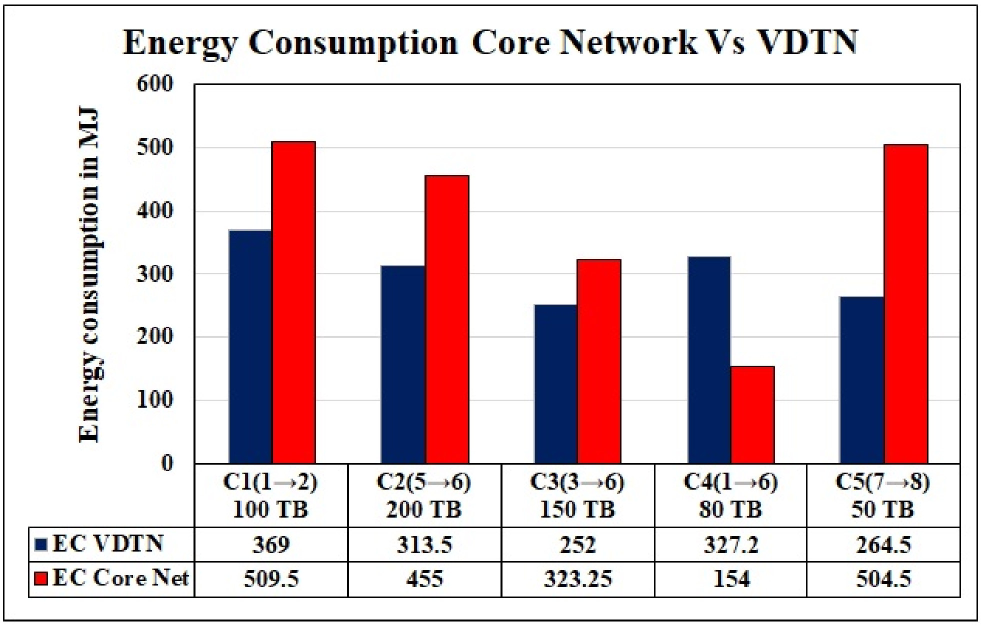


Figure 5.9: EC: A comparison of the core network and VDTN.

network interfaces to forward the data. With this optimal network mode selection, we can save the energy cost more than 32% for this given set of commodities.

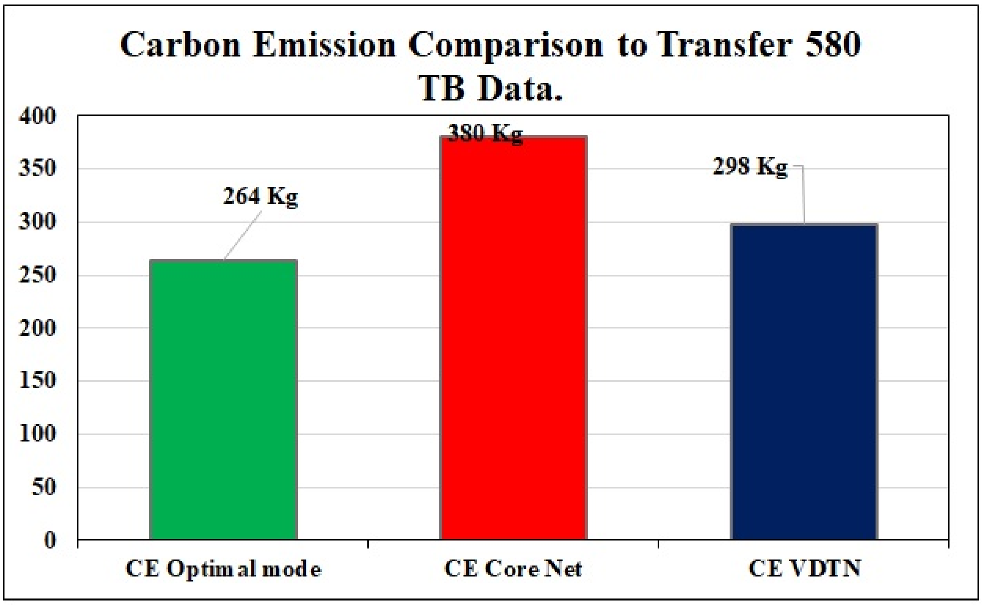


Figure 5.10: CE: A comparison of the core network, VDTN and optimal network mode.

3. Carbon Emission

By applying energy-efficient network mode selection, we can further reduce the energy cost and carbon consumption. Information exchange for both core network and VDTN consumes energy from various means like fuel and power. This usage of energy transmits carbon into the environment. Figure 5.10 demonstrates that the CE is less for our proposed optimal model as compared to traditional core networks and VDTN. To evaluate these outcomes, we used the carbon emission 0.703 CO_2 Kg/kWh [177] and the conversion unit 1 MJ = 0.2777778 kWh [178].

5.7 Comparison with existing techniques

Table 5.4 indicates our proposed framework, the vehicle assisted energy efficient data data dissemination framework, compared with other existing works. Apart from the Amazon, most of the existing works utilise radio communications for transferring data to or from the vehicles. Although Unmanned Aerial Vehicles (UAVs) is the comprehensive body of work for data transfer, this method is not sufficient for covering all areas due to inheriting battery restrictions. UAVs is not reasonably cost effective and they are inadequate in numbers. From the security point of view, questions have been raised as to whether it is secure to use UAVs for data mule [179]. Advanced scheduling is needed for Amazon snowball and snowmobile, while transferring data (of this method) is not free for the user. If data transferring via snowmobile is in its large scale, tens of days are taken to schedule this transfer for each appointment [180]. Researchers cannot benchmark their own route or algorithms of vehicle selection through other related contributions because these sources are not made freely available to the general public.

In order to transfer data from source to destination of a single hop or multiples', our framework has been proposed which is based in the vehicles and it uses the intermediate nodes RSUs. As smart vehicles and smart cities have become popular, vehicles can take the role of transferring data to spend less energy in network communication. If data is transferred through communication network, other applications are affected in addition to blocking the entire network. The average rates of uploading and downloading data in traditional data transferring network is also different from continent to continent. For instance, transferring 20TB of data would take around 38 months in Africa, whereas this amount of data is transferred by around 4 months in Australia. It is noted that only the time of uploading is considered at this case scenario and both time of transferring and downloading have been neglected [181].

Discover the optimal route between the source and the destination is the responsibility of the controller in the proposed framework. The controller also decides about suitability of transmitting data via traditional or vehicular networks. To decide, the time of transfer and the availability of volunteer vehicles close to the source for starting and close to the source or destination for continuing are considered. We have interest in adding more complex algorithms for the purpose of calculating the path of transferring from source to destination in the future.

Energy – to transfer data, Internet has been in use across the world traditionally. The energy which is consumed for communication through the Internet can be grouped as the end-user cost. While this energy is around 38% of the total energy consumption, the remaining of 62% energy supplies the network nodes of repeaters, routers, servers, switches and the data-centres [182].

Costenaro and Duer [182] had approximate calculation on the consumed energy by the Internet in kWh per GB as 5.12 kWh/GB according to the parameter values used for Internet data transfer and in relation to core network devices. Therefore, to transfer 20TB data, total of 102.4 MWh energy is estimated for consumption by the Internet data. From another standpoint, the average power consumption of a server equal to 145 watts per hour has been considered according to the research work of Pries et al. [183] in order to calculate the consumed energy using vehicles for transferring data. Whereas in our proposed case it consumed 60.66 MJ energy to transfer 20 TB of data by using vehicles. To determine the total cost in both scenarios, the average electricity cost as standardized by U.S. Energy Information Administration equal to i.e. \$0.112 per Kilowatt hour (kWh) has been taken into account [184]. As a result, the cost of energy of the system for 20 TB is estimated as \$11469 approximately according to the energy consumption for transferring data via Internet. Meanwhile, the cost for the proposed vehicular approach is \$1900 approx.

On the basis of the above discussion, Figure. 5.3 indicates the energy consumption by variation on the amount of transferred data, while utilising the Internet or the proposed vehicular approach. Thence, the performance of our vehicular based data dissemination approach is with a superior quality and excellence in comparison with the Internet data transfer in the context of both the energy and the delay.

One of the important implications of our proposed work is the business of sharing economy. The business of the sharing economy is a mode of purchasing services and goods that is different from the traditional model of business [185]. In the traditional model, we hire the employees to produce the products and then we sell to customers. On the other hand, in a shared economy, people are supposed to hire out their things like private time, cars, and home to other people in a peer-to-peer manner, e.g Uber and Airbnb. In our proposed work, people have their cars the can take part in this business of shared economy based on some incentives for this data and code dissemination in SCs.

Moreover we consider the bi-directional mode of communication in our proposed framework. In most of previous work data communication mode is unidirectional i.e from source to destination. Whereas in our work we consider big data migration, data transmission from SSs to service providers and code transmission from service provider to SSs.

Table 5.4: Comparison with existing techniques.

Previous Work	**Disk capacity**	**Transfer Frequency**	**Data Trans. Mode**	**Route requirement**	**Platform**	**Comm. between devices**	**Message Direction**	**Reliability of Trans. tasks**	**Energy Efficiency**	**Business of Sharing Economy**
Cheng et al. [124]	Undefined	Undefined	Unmanned Aerial Vehicles	Actual flight paths	-	Radio communication	Unidirectional $S \rightarrow D$	×	×	×
Palma et al. [179]	Undefined	Scheduled own resources	Unmanned Aerial Vehicles	Fixed route	Emulated Environment	Radio communication	Unidirectional $S \rightarrow D$	×	×	×
Usbeck et al. [171]	Undefined	Scheduled own resources	Unmanned Aerial Vehicles	Unspecified networks	Android Platform	Radio communication	Unidirectional $S \rightarrow D$	×	×	×
Hunjet et al. [170]	Undefined	Scheduled own resources	Unmanned Aerial Vehicles	Tactical network	MASON framework	Radio communication	Unidirectional $S \rightarrow D$	×	×	×
Amazon Snowmobile [169]	100PB	One time and Scheduled on demand	Large container truck	Limited specialised regions	Amazon Big Vehicle	Wired	Unidirectional $S \rightarrow D$	×	×	×
Amazon Snowball [169]	<100TB	One time and Scheduled on demand	Large container truck	Limited specialised regions	Amazon Big Vehicle	Wired	Unidirectional $S \rightarrow D$	×	×	×
Rashmi et al. [166]	2TB	Scheduled own resources	Public buses	Based on User Profiles	ONE Simulator	Wifi-direct	Unidirectional $S \rightarrow D$	×	✓	×
Coutinho et al. [180]	Undefined	Scheduled own resources	Boats	Unspecified network	ONE and NS2	Radio communication	Unidirectional $S \rightarrow D$	×	×	×
Proposed System	Independent of disk size	Scheduled by CC on each demand	Any vehicle can participate	Independent of vehicle trajectory	Modeling / Data driven approach	Virtus Chipsat/ WiFi-direct	Bidirectional $S \leftrightarrow D$	✓	✓	✓

5.8 Summary

In this chapter, the energy models for sustainable data dissemination suitable for daily data traffic in SC are proposed. A method of big data aggregation for smart cities using the existing vehicle volume on the road is investigated. For delay tolerant data needs, our proposed method efficiently utilizing the existing vehicles and their energy-efficient wireless links, for big data transmission to reduce the energy consumption cost and carbon emissions. Our essential and contextual investigations performed in the mathematical evaluation of these case studies give solid proof that substantial energy and cost savings can be accomplished while as yet ensuring the information delivery. Moreover, our minimum energy cost flow model shows that if we know vehicle volume on the road and transportation routes, we can identify the routes having the least cost for data transportation. Furthermore, a case is evaluated by using a traditional multi-commodity flow problem, energy-efficient network mode selection algorithm is used for data transmission scenarios, our proposed approach can offer about 32% better EC and CE than the traditional network.

In this chapter, we demonstrated that our proposed VADD framework is energy efficient than the traditional systems. Moreover, we presented the comparison of our proposed framework with the existing techniques. The conclusion and future research directions are discussed in in next chapter.

Chapter 6

Conclusions and Future Research Directions

The thesis is concluded in Section 6.1. The practical system implications for system planning and deployment are discussed in Section 6.2. Four future research directions are outlined in Section 6.3.

6.1 Conclusions

This thesis has focused on designing an energy-efficient sustainable vehicle assisted data dissemination framework for the collection of data from a massive number of smart devices in SCs to their designated control centres. The growing demands of big data dissemination, congested networks, a huge amount of energy consumption and carbon emissions by ICT networks were the main motivation for the development of VADD framework. To address the challenges of big data dissemination, this research proposes that an alternate vehicular delay-tolerant network exists, to connect such a huge number of smart devices with their data centres, by using daily mobility of smart vehicles.

In this thesis, a method of big data aggregation for smart cities using the existing vehicle volume on the road and vehicle displacement in the city is investigated by analytical modelling and data-driven approach. For delay tolerant data needs, our proposed method efficiently utilizing the existing vehicles and their energy-efficient wireless links, for big data transmission to reduce the energy consumption cost and carbon emissions. As a proof-of-concept of our algorithms, we presented Auckland City and Beijing City case scenarios where delay tolerant data is delivered to data centers. We found that our proposed system can effectively use the daily vehicle mobility of Auckland City for enormous information transmission to reduce the cost of EC and CE. Our essential and contextual investigations performed in the mathematical and analytical evaluation of different case studies give solid proof that substantial energy and cost savings can be accomplished while as yet ensuring the information delivery. This thesis has made several original contributions, which are reported in Chapters 3 to 5. A summary of the main contributions is outlined below.

In Chapter 3, we propose a VADD framework for data dissemination by using the existing vehicles in smart cities. It could best accommodate those massive and delay tolerant-able data transmission demands by utilizing efficiently the existing vehicles' mobility in the urban areas and could release the congestions in wireless and wired networks. We use the daily traffic count of Auckland city and taxi traces of Beijing city to explore the potential of our proposed VADD framework. In the Auckland case, we develop a model to calculate the delay for data transfer between two locations on a road by using the vehicle count on the road. We compare our results with a traditional network having a dedicated bandwidth of 1 Gbps and the disk capacity of each vehicle is 512 GB. Our proposed system reduces more than 50% latency as compared to the traditional network.

In the later part of Chapter 3, we use the displacement of Beijing taxies to collect the data from $25 \times 25km^2$ area of the city. We use a grid clustering algorithm and divide the whole city in different sizes of clusters or girds, to explore the coverage and capacity of the proposed system. In the case of 1000 m^2 area, the network coverage is more than 98$\%$ in 24 hours. Similarly, in the case of 500 m^2, 200m^2, and 100m^2, we got 90$\%$, 70$\%$, and 40$\%$ coverage respectively in 24 hours. All the grids near the roads got full network coverage in all cases. Similarly, Network coverage increases if we increase the number of taxies. We also explore the network capacity by varying the different wireless technologies and their wireless coverage sizes. The results show that the capacity depends upon the wireless technology used and the number of taxies. More than 50 TB data can be collected in an area of $25 \times 25km^2$ if we use IEEE 802.11ay and 400 taxi cabs.

Furthermore, we evaluate a case of data offloading and use a greedy algorithm for RSU placement. The results show that if we increase the delay-tolerant interval then we got good network coverage. The data from more than 98$\%$ of sensors can be collected in 24 hours if vehicles use their cellular links. To save the cost of cellular links we use the RSUs and got data from 96$\%$ sensors at the same time. In this way, our proposed VADD framework saves the infrastructure cost, reduce the congestion on the traditional networks, and provide good network coverage and capacity.

In Chapter 4, we propose data and code forwarding schemes. We updated the data forwarding model and use the effective range of wireless coverage of SSs or RSUs. We use the Poison arrival of vehicles to calculate the degree of data offloading under various vehicle densities and delivery probabilities. As a proof of concept for the effectiveness of the data forwarding model, we use the Auckland daily traffic count and compare the results with a traditional network having dedicated bandwidth of 512 Mbps and 1 Gbps. Our proposed network having IEEE 802.11ay wireless technology reduces the huge latency while transferring 100 TB data.

To transfer the software update code we propose the OCFS scheme by using the taxi displacement in Beijing city. To maximize the network coverage and minimize the time for code transmission, we propose an NP-hard minimum set coverage problem.

Our proposed algorithms select the optimal set of vehicles and RSUs for a given task of code transmission. To verify the effectiveness of our proposed scheme we compare our results with the fixed and random deployment of RSUs for code transmission. Our proposed scheme outperforms and reaches up to 98.72% of smart sensors in 1260 minutes. Whereas fixed and random RSU deployment got 97.88% and 90% transmission rate in 1800 minutes.

To overcome the problem of energy consumptions and carbon emissions, our proposed VADD framework saves the energy cost and reduce energy consumptions. In Chapter 5, we propose an energy consumption model, and to verify the effectiveness of our proposed model we evaluate three different cases to calculate energy consumption and carbon emissions. In 1st case, we use the daily vehicle count on Auckland road. Our proposed model outperforms and saves a huge amount of energy as compared to traditional data communication networks. Next, we formulate a problem by using a min-cost flow problem to transfer data by using multi-hop communication. Evaluate the results by using IBM ILOG CPLEX Optimization Studio. Our system outperforms in terms of energy consumption, save more than 40% energy cost, and reduce carbon emissions, to transfer 20 TB data. Finally, in the 3rd case, we formulate a problem by using a multi-commodity flow problem and use an algorithm of optimal network mode selection by the VADD framework for different data assignments. We evaluate our results by using the SaS optimization tool to find a set of energy-efficient paths and optimal solutions for a given set of commodities to transfer data by using the road network. Our system outperforms and save more than 32% energy cost and reduce carbon emissions.

6.2 Practical System Implications

The vehicle assisted data dissemination framework has been investigated in Chapter 3 to 5. The main objective of this thesis is the use of VADD for massive data dissemination in smart cities by using vehicle mobility. To achieve these objectives, system planners could apply the results evaluated in the main contribution chapters to design sustainable data dissemination in SC as an alternate channel by using existing transport. This segment delivers the directions based on the results and analysis derived in Chapter 3 to 5. The possible data dissemination scenarios and solid guidelines are debated next.

The target users of our proposed framework VADD are the departments that manage the various smart sensors in smart cities. Whenever a user wants to update the software of smart sensors that are managed by him, the user will initialize an update request to the central controller. The central controller receives the request and then find the locations of smart sensors. Based on these locations the CC will find the optimal vehicles based on the historical trajectories of vehicles. After finding the optimal vehicles for this task the CC will find the locations of RSUs based on the set of selected vehicles. CC will forward the updated codes to these RSUs and vehicles will pick the updated codes from these RSUs then transmit this code to the GPS locations where these smart sensors are installed. After updating each sensor vehicles will report to CC. In this way, CC will monitor task completion.

In the reverse direction, whenever an SS creates data to transfer to these departments, they send this data to the nearest vehicle with a high reputation value. The vehicle receives the data and transfers data to RSU whenever it is in its wireless range.

For big data migration in a smart city, a user can send a request to CC and send the location of its old and new data center. CC will find the set of vehicles for this data migration and start the task. For data migration, other solutions are the existing network or Amazon's service snowball. Existing networks are also very congested and consume more energy as compared to our proposed solution. On the other hand, Amazon's solution is very expensive and it is also not feasible to get service at any time. Our proposed solution can be available because we are using existing vehicle trips.

It is important, in the midst of rapid urbanization, to ensure the safety of people in future smart cities. Under such situations, video surveillance systems will not only improve the system's intelligence to pro-actively minimise the risk of multiple risks but will also play an essential role under detecting and prosecuting any crime. Studies [186] have shown that in different city centres, video surveillance systems minimise the crime rate. Also, some cameras have already been mounted by big cities to increase surveillance, e.g New York City surveillance cameras project [187] and 4000 surveillance cameras in Auckland City [188]. However, these cameras do not have good quality resolution and a lot of video processing is needed to work with any recognition device. Nonetheless, the network does not have enough cameras to provide complete video coverage of the various corners of the area. We envisage that future smart cities will be fitted with video surveillance cameras at all intersections of the cities to improve the coverage of video footage. Multiple high-definition cameras must be mounted at every corner of the street to protect 360 vision.

One of the important implications of our proposed work is the business of sharing economy. The business of the sharing economy is a mode of purchasing services and goods that is different from the traditional model of business. In the traditional model, we hire the employees to produce the products and then we sell to customers. On the other hand, in a shared economy, people are supposed to hire out their things like private time, cars, and home to other people in a peer-to-peer manner, e.g Uber and Airbnb. In our proposed work, people have their cars the can take part in this business of shared economy based on some incentives for this data and code dissemination in SCs.

6.3 Future Work

This thesis provides a VADD framework for sustainable data dissemination in smart cities. We use the daily vehicle count on the road and displacement of vehicles in a city to transfer data between different locations. The aim was to answer the research questions "What sustainable data dissemination framework can be developed for smart cities using the existing vehicular network?" and "What can be done to quantify the performance gains of the proposed VADD framework?". To address these questions we evaluated different scenarios in this thesis which open many future research issues. In this section, we highlight four key research problems that can be explored for the future extension of our proposed work.

- **Incentive-based communication:** We need to consider the incentive-based communication to motivate the drivers to take part in this proposed framework of VADD. Such motivation may contain low price rates, rewarded WiFi or cellular bandwidth, etc. Different pricing strategies have been considered in the literature, such as in [102]. Though, with flat rate appraising broadly engaged, such low-pricing strategies may not be viable. Consequently, a practical and smart incentive model should be considered for the potential marketing relationships between system users, cellular operators, and vehicle users. In the future, we are planning to design a framework that will consider a Bitcoin-based reliable and secure incentive scheme for VADD. Whenever the tasks will be assigned to a vehicle a smart contract will be initiated by CC and cost will be decided for each task completion. The CC will lock this transaction on the Bitcoin server, under the condition that a vehicle can consume the coins after the completion of given tasks. After the completion of the given tasks of data or code forwarding, the vehicle will send the proof of work to CC. The CC will forward this proof of work to the Bitcoin server and the incentive will be given to the vehicle base on the contract.

- **Advance prediction mechanisms:** The high mobility of vehicles in a city makes the precise prediction of vehicle availability and performance difficult, even with prior knowledge about the locations of SSs and RSUs or historical trajectories of vehicles are available. To optimize the data/code dissemination performance more accurate prediction mechanisms, more specifically the long prediction interval, are then essential for scheming more smart data dissemination schemes. Using deep learning neural network models such as CNN, DNN, RNN, or LSTM, etc, for more accurate predictions is our future work.

  In order to classify more suitable vehicles as data mules, the system will use a neural network, heading towards the data destination. This could potentially reduce the load from backend networks in terms of bandwidth usage and overall energy consumption.

- **Trust and security:** There are many issues involved in peer-to-peer communication of mobile vehicles related to privacy and security plan taking part in this propose VADD framework. Users hardly accept that anyone unknown or stranger to access the data available on their storage devices. More challenges take account of the development of a trust-based framework to guarantee distributed security and trust to terminals involved in the data dissemination process.

  With the more trustable and secure framework, there will be more chances that the vehicles will be available as data mules for this proposed network. This trust can enhance the potential of data offloading by motivating more users to take part in this process of data dissemination.

- **Augmented reality:** The data collection scheme of our proposed framework can be enhanced by adding with the modern expansion with Augmented Reality based trends and scenarios. In these applications, a video feed of different vehicles can be combined to manage a video view of the entire smart city. This type of data accumulation can be achieved either collaboratively between smart vehicles by using their displacement or in a centralized approach.

In the core of rapid urbanisation, it is necessary to ensure the protection of people in future smart cities. Video monitoring systems such as the augmented reality of SC will not only enhance the intelligence of the system in such circumstances to proactively reduce the likelihood of multiple threats but will also play an important role in the identification and prosecution of any crime.

References

[1] A. S. Andrae and T. Edler, "On global electricity usage of communication technology: trends to 2030," *Challenges*, vol. 6, no. 1, pp. 117–157, 2015.

[2] P. P. Engelen, *Wireless Radio, The Dutch WiFi Fella, WirelessIP, The Killer Application!?*, Nov. 2020. [Online]. Available: http://www.peterpaulengelen.com/2013/11/wireless-radio/

[3] A. Meola, *Automotive Industry Trends: IoT Connected Smart Cars & Vehicles*, Business Insider, Nov. 2020. [Online]. Available: http://www.businessinsider.com

[4] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper*, CISCO, Nov. 2020. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[5] *5G vs 4G: what is the difference?*, Raconteur, Nov. 2020. [Online]. Available: https://www.raconteur.net/technology/5g/4g-vs-5g-mobile-technology/

[6] G. Fliedner and G. DeHondt, "Lean productivity enhancements and waste elimination through emerging technology," *American Journal of Information Technology*, vol. 5, no. 1, 2015.

[7] G. Asada, M. Dong, T. Lin, F. Newberg, G. Pottie, W. Kaiser, and H. Marcy, "Wireless integrated network sensors: Low power systems on a chip," in *Proceedings of the 24th IEEE European solid-state circuits conference*, 1998, pp. 9–16.

[8] W. R. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks," in *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*, 1999, pp. 174–185.

[9] A. Boukerche and A. Darehshoorzadeh, "Opportunistic routing in wireless networks: Models, algorithms, and classifications," *ACM Computing Surveys (CSUR)*, vol. 47, no. 2, pp. 1–36, 2014.

[10] B. Baron, P. Spathis, H. Rivano, and M. D. de Amorim, "Vehicles as big data carriers: Road map space reduction and efficient data assignment," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, 2014, pp. 1–5.

[11] I. Marincic and I. Foster, "Energy-efficient data transfer: Bits vs. atoms," in *2016 24th IEEE International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, 2016, pp. 1–6.

[12] W. Wang, J. Yang, R. Muntz *et al.*, "Sting: A statistical information grid approach to spatial data mining," in *VLDB*, vol. 97, 1997, pp. 186–195.

[13] *Perform sensitivity analysis through random parameter variation*, Mathworks, Nov. 2020. [Online]. Available: https://www.mathworks.com/discovery/monte-carlo-simulation.html

[14] *IBM ILOG CPLEX Optimization Studio, IBM*, Nov. 2020. [Online]. Available: https://www.ibm.com/us-en/marketplace/ibm-ilog-cplex

[15] *The Power to Know: Statistical Analysis System (SAS) Customer Support.*, Nov. 2020. [Online]. Available: https://support.sas.com/en/support-home.html

[16] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Communications surveys & tutorials*, vol. 13, no. 4, pp. 524–540, 2011.

[17] G. Fettweis and E. Zimmermann, "Ict energy consumption-trends and challenges," in *Proceedings of the 11th international symposium on wireless personal multimedia communications*, vol. 2, no. 4. Citeseer, 2008, p. 6.

[18] M. Etoh, T. Ohya, and Y. Nakayama, "Energy consumption issues on mobile network systems," in *IEEE International Symposium on Applications and the Internet*, 2008, pp. 365–368.

[19] Y. Chen, "Great project overview,(slides) green touch open forum, seoul, korea," 2011.

[20] K. Sabhanatarajan and A. Gordon-Ross, "A resource efficient content inspection system for next generation smart nics," in *2008 IEEE International Conference on Computer Design*, 2008, pp. 156–163.

[21] G. Da Costa, J.-P. Gelas, Y. Georgiou, L. Lefevre, A.-C. Orgerie, J.-M. Pierson, O. Richard, and K. Sharma, "The green-net framework: Energy efficiency in large scale distributed systems," in *2009 IEEE International Symposium on Parallel & Distributed Processing*, 2009, pp. 1–8.

[22] J. Blackburn and K. Christensen, "A simulation study of a new green bittorrent," in *2009 IEEE International Conference on Communications Workshops*, 2009, pp. 1–6.

[23] A. Gladisch, C. Lange, and R. Leppla, "Power efficiency of optical versus electronic access networks," in *2008 34th European Conference on Optical Communication*, 2008, pp. 1–4.

[24] G. Griffa and T. A. WI, "Ngn2 and energy opex savings," *Giornata di studio su Efficienza Energetica*, 2008.

[25] J. Malmodin, "Carbon footprint of mobile communications and ict," in *Proc. of Joint International Congress and Exhibition Electronics Goes Green*, 2008.

[26] A. P. Jardosh, G. Iannaccone, K. Papagiannaki, and B. Vinnakota, "Towards an energy-star wlan infrastructure," in *Eighth IEEE Workshop on Mobile Computing Systems and Applications*, 2007, pp. 85–90.

[27] T. Edler and S. Lundberg, "Energy efficiency enhancements in radio access networks," *Ericsson review*, vol. 81, no. 1, pp. 42–51, 2004.

[28] M. A. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Optimal energy savings in cellular access networks," in *2009 IEEE International Conference on Communications Workshops*, 2009, pp. 1–5.

[29] J. T. Louhi, "Energy efficiency of modern cellular base stations," in *INTELEC 07-29th IEEE International Telecommunications Energy Conference*, 2007, pp. 475–476.

[30] L. Ceuppens, A. Sardella, and D. Kharitonov, "Power saving strategies and technologies in network equipment opportunities and challenges, risk and rewards," in *2008 IEEE International Symposium on Applications and the Internet*, 2008, pp. 381–384.

[31] J. Chabarek, J. Sommers, P. Barford, C. Estan, D. Tsiang, and S. Wright, "Power awareness in network design and routing," in *IEEE INFOCOM 2008-The 27th Conference on Computer Communications*, 2008, pp. 457–465.

[32] R. Bolla, R. Bruschi, and A. Ranieri, "Energy-aware equipment for next-generation networks," in *Proceedings of the 4th International Conference on Future Internet Technologies*, 2009, pp. 8–11.

[33] T. Chen, Y. Yang, H. Zhang, H. Kim, and K. Horneman, "Network energy saving technologies for green wireless access networks," *IEEE Wireless Communications*, vol. 18, no. 5, pp. 30–38, 2011.

[34] Y. Zhang, P. Chowdhury, M. Tornatore, and B. Mukherjee, "Energy efficiency in telecom optical networks," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 4, pp. 441–458, 2010.

[35] M. Gupta and S. Singh, "Dynamic ethernet link shutdown for energy conservation on ethernet links," in *2007 IEEE International Conference on Communications*, 2007, pp. 6156–6161.

[36] J. C. Restrepo, C. G. Gruber, and C. M. Machuca, "Energy profile aware routing," in *2009 IEEE International Conference on Communications Workshops*, 2009, pp. 1–5.

[37] L. Chiaraviglio, D. Ciullo, M. Meo, and M. A. Marsan, "Energy-efficient management of umts access networks," in *2009 21st IEEE International Teletraffic Congress*, 2009, pp. 1–8.

[38] H. Anand, C. Reardon, R. Subramaniyan, and A. D. George, "Ethernet adaptive link rate (alr): Analysis of a mac handshake protocol," in *Proceedings. 2006 31st IEEE Conference on Local Computer Networks*, 2006, pp. 533–534.

[39] C. Gunaratne, K. Christensen, B. Nordman, and S. Suen, "Reducing the energy consumption of ethernet with adaptive link rate (alr)," *IEEE Transactions on Computers*, vol. 57, no. 4, pp. 448–461, 2008.

[40] Y. Chen, T. X. Wang, and R. H. Katz, "Energy efficient ethernet encodings," in *2008 33rd IEEE Conference on Local Computer Networks (LCN)*, 2008, pp. 122–129.

[41] F. Blanquicet and K. Christensen, "Managing energy use in a network with a new snmp power state mib," in *2008 33rd IEEE Conference on Local Computer Networks (LCN)*, 2008, pp. 509–511.

[42] M. Jimeno and K. Christensen, "A prototype power management proxy for gnutella peer-to-peer file sharing," in *32nd IEEE Conference on Local Computer Networks (LCN 2007)*, 2007, pp. 210–212.

[43] M. Jimeno and K. Miguel Christensen, "P2p directory search: Signature array hash table," in *2008 33rd IEEE Conference on Local Computer Networks (LCN)*, 2008, pp. 506–508.

[44] W.-K. Park, C.-S. Choi, H. Lee, and K.-R. Park, "Energy efficient home gateway based on user service traffic in always-on home network environment," in *2008 IEEE International Conference on Advances in Electronics and Micro-electronics*, 2008, pp. 121–125.

[45] A. Wierman, L. L. Andrew, and A. Tang, "Power-aware speed scaling in processor sharing systems," in *IEEE INFOCOM 2009*, 2009, pp. 2007–2015.

[46] S. Zeadally, S. U. Khan, and N. Chilamkurti, "Energy-efficient networking: past, present, and future," *The Journal of Supercomputing*, vol. 62, no. 3, pp. 1093–1118, 2012.

[47] H. S. Hinton, "Photonic switching fabrics," *IEEE Communications Magazine*, vol. 28, no. 4, pp. 71–89, 1990.

[48] S. Nedevschi, L. Popa, G. Iannaccone, S. Ratnasamy, and D. Wetherall, "Reducing network energy consumption via sleeping and rate-adaptation." in *NsDI*, vol. 8, 2008, pp. 323–336.

[49] W.-T. Chen, J.-C. Liu, and H.-K. Huang, "An adaptive scheme for vertical handoff in wireless overlay networks," in *Proceedings. Tenth International Conference on Parallel and Distributed Systems, 2004. ICPADS 2004.*, 2004, pp. 541–548.

[50] H. Elaarag, "Improving tcp performance over mobile networks," *ACM Computing Surveys (CSUR)*, vol. 34, no. 3, pp. 357–374, 2002.

[51] G. Holland and N. Vaidya, "Analysis of tcp performance over mobile ad hoc networks," *Wireless Networks*, vol. 8, no. 2-3, pp. 275–288, 2002.

[52] S. Ganesh and R. Amutha, "Energy efficient transport protocol for wireless sensor networks," in *2009 2nd IEEE International Conference on Computer Science and Information Technology*, 2009, pp. 464–468.

[53] W. Lilakiatsakun and A. Seneviratne, "Enhancing tcp energy efficiency for mobile hosts," in *Proceedings of the 10th IEEE International Conference on Networks (ICON 2002). Towards Network Superiority (Cat. No. 02EX588)*, 2002, pp. 235–239.

[54] S. Ci, S. W. Turner, and H. Sharif, "An energy-efficient tcp quick timeout scheme for wireless lans," in *Conference Proceedings of the 2003 IEEE International Performance, Computing, and Communications Conference, 2003.*, 2003, pp. 193–197.

[55] H. Liu, C. Maciocco, V. Kesavan, and A. L. Low, "Energy efficient network selection and seamless handovers in mixed networks," in *2009 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks & Workshops*, 2009, pp. 1–9.

[56] M. Nam, N. Choi, Y. Seok, and Y. Choi, "Wise: energy-efficient interface selection on vertical handoff between 3g networks and wlans," in *2004 IEEE 15th International Symposium on Personal, Indoor and Mobile Radio Communications (IEEE Cat. No. 04TH8754)*, vol. 1, 2004, pp. 692–698.

[57] F. Siddiqui, S. Zeadally, H. El-Sayed, and N. Chilamkurti, "A dynamic network discovery and selection method for heterogeneous wireless networks," *International Journal of Internet Protocol Technology*, vol. 4, no. 2, pp. 99–114, 2009.

[58] E. Trevisani and A. Vitaletti, “Cell-id location technique, limits and benefits: an experimental study,” in *Sixth IEEE workshop on mobile computing systems and applications*, 2004, pp. 51–60.

[59] T. Origuchi, “Eco-efficiency evaluation of 3g service" foma",” in *Proceedings of the 7ˆ< th> International Conference on EcoBalance, 2006*, 2006, pp. 91–94.

[60] M. Jada, J. Hämäläinen, R. Jäntti, and M. Hossain, “Power efficiency model for mobile access network,” in *2010 IEEE 21st International Symposium on Personal, Indoor and Mobile Radio Communications Workshops*, 2010, pp. 317–322.

[61] S. Zhang, N. Zhang, S. Zhou, J. Gong, Z. Niu, and X. Shen, “Energy-aware traffic offloading for green heterogeneous networks,” *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1116–1129, 2016.

[62] H. Ma, G. Zheng, H. Wu, B. Ji, and J. Li, “Ebrp: An energy-efficient and buffer-aware routing protocol for mobile crowdsensing network,” *International Journal of Distributed Sensor Networks*, vol. 12, no. 3, p. 4972785, 2016.

[63] S. Singh, N. Saxena, A. Roy, and H. Kim, “Energy efficiency in wireless networks–a composite review,” *IETE Technical Review*, vol. 32, no. 2, pp. 84–93, 2015.

[64] L. Lilien, Z. H. Kamal, V. Bhuse, A. Gupta *et al.*, “Opportunistic networks: The concept and research,” in *Challenges in Privacy and Security,” Proc. NSF Intl. Workshop on Research Challenges in Security and Privacy for Mobile and Wireless Networks (WSPWN 2006*. Citeseer, 2006.

[65] C. Boldrini, M. Conti, and A. Passarella, “Context and resource awareness in opportunistic network data dissemination,” in *2008 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, 2008, pp. 1–6.

[66] K. Fall, “A delay-tolerant network architecture for challenged internets,” in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, 2003, pp. 27–34.

[67] L.-J. Chen, C.-H. Yu, C.-L. Tseng, H.-h. Chu, and C.-F. Chou, “A content-centric framework for effective data dissemination in opportunistic networks,” *IEEE Journal on selected Areas in Communications*, vol. 26, no. 5, pp. 761–772, 2008.

[68] S. K. Dhurandher, D. K. Sharma, I. Woungang, and H.-C. Chao, “Performance evaluation of various routing protocols in opportunistic networks,” in *2011 IEEE GLOBECOM Workshops (GC Wkshps)*, 2011, pp. 1067–1071.

[69] P. Papadimitratos and Z. J. Haas, “Secure message transmission in mobile ad hoc networks,” *Ad Hoc Networks*, vol. 1, no. 1, pp. 193–209, 2003.

[70] A. Vahdat, D. Becker *et al.*, "Epidemic routing for partially connected ad hoc networks," 2000.

[71] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," *Computer Networks*, vol. 51, no. 10, pp. 2867–2891, 2007.

[72] T. Matsuda and T. Takine, "(p, q)-epidemic routing for sparsely populated mobile ad hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 5, pp. 783–793, 2008.

[73] Y. Bejerano, "Efficient integration of multihop wireless and wired networks with qos constraints," *IEEE/ACM transactions on networking*, vol. 12, no. 6, pp. 1064–1078, 2004.

[74] L. Pelusi, A. Passarella, and M. Conti, "Opportunistic networking: data forwarding in disconnected mobile ad hoc networks," *IEEE communications Magazine*, vol. 44, no. 11, pp. 134–141, 2006.

[75] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Single-copy routing in intermittently connected mobile networks," in *2004 First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004.*, 2004, pp. 235–244.

[76] A. Lindgren, A. Doria, and O. Schelén, "Probabilistic routing in intermittently connected networks," *ACM SIGMOBILE mobile computing and communications review*, vol. 7, no. 3, pp. 19–20, 2003.

[77] F. Gil-Castineira, F. J. González-Castaño, and L. Franck, "Extending vehicular can fieldbuses with delay-tolerant networks," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 9, pp. 3307–3314, 2008.

[78] R. A. Gorcitz, Y. Jarma, P. Spathis, M. D. de Amorim, R. Wakikawa, J. Whitbeck, V. Conan, and S. Fdida, "Vehicular carriers for big data transfers (poster)," in *2012 IEEE Vehicular Networking Conference (VNC)*, 2012, pp. 109–114.

[79] V. N. Soares, F. Farahmand, and J. J. Rodrigues, "A layered architecture for vehicular delay-tolerant networks," in *2009 IEEE Symposium on Computers and Communications*, 2009, pp. 122–127.

[80] S. Kashihara, M. Sanadidi, and M. Gerla, "Mobile, personal data offloading to public transport vehicles," in *Proceedings of the 2012 IEEE International Conference on Mobile Computing and Ubiquitous Networking*, 2012.

[81] M. Sede, X. Li, D. Li, M.-Y. Wu, M. Li, and W. Shu, "Routing in large-scale buses ad hoc networks," in *2008 IEEE Wireless Communications and Networking Conference*, 2008, pp. 2711–2716.

[82] Y. Xian, C.-T. Huang, and J. Cobb, “Look-ahead routing and message scheduling in delay-tolerant networks,” *Computer Communications*, vol. 34, no. 18, pp. 2184–2194, 2011.

[83] *Data overload threatens mobile networks*, Nov. 2020. [Online]. Available: https://www.ft.com/content/caeb0766-9635-11e1-a6a0-00144feab49a

[84] F. Rebecchi, M. D. De Amorim, V. Conan, A. Passarella, R. Bruno, and M. Conti, “Data offloading techniques in cellular networks: A survey,” *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 580–603, 2014.

[85] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, “Mobile data offloading: How much can wifi deliver?” *IEEE/ACM Transactions on networking*, vol. 21, no. 2, pp. 536–550, 2012.

[86] N. Cheng, N. Lu, N. Zhang, X. S. Shen, and J. W. Mark, “Vehicular wifi offloading: Challenges and solutions,” *Vehicular Communications*, vol. 1, no. 1, pp. 13–21, 2014.

[87] *3GPP system to Wireless Local Area Network (WLAN) interworking; System description*, Nov. 2020. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=824

[88] A. Balasubramanian, R. Mahajan, and A. Venkataramani, “Augmenting mobile 3g using wifi,” in *Proceedings of the 8th international conference on Mobile systems, applications, and services*, 2010, pp. 209–222.

[89] X. Hou, P. Deshpande, and S. R. Das, “Moving bits from 3g to metro-scale wifi for vehicular network access: An integrated transport layer solution,” in *2011 19th IEEE International Conference on Network Protocols*, 2011, pp. 353–362.

[90] P. Deshpande, A. Kashyap, C. Sung, and S. R. Das, “Predictive methods for improved vehicular wifi access,” in *Proceedings of the 7th international conference on Mobile systems, applications, and services*, 2009, pp. 263–276.

[91] J. Eriksson, H. Balakrishnan, and S. Madden, “Cabernet: vehicular content delivery using wifi,” in *Proceedings of the 14th ACM international conference on Mobile computing and networking*, 2008, pp. 199–210.

[92] L. Hu, C. Coletti, N. Huan, P. Mogensen, and J. Elling, “How much can wi-fi offload? a large-scale dense-urban indoor deployment study,” in *2012 IEEE 75th vehicular technology conference (VTC Spring)*, 2012, pp. 1–6.

[93] S. Dimatteo, P. Hui, B. Han, and V. O. Li, “Cellular traffic offloading through wifi networks,” in *2011 IEEE eighth international conference on mobile ad-hoc and sensor systems*, 2011, pp. 192–201.

[94] N. Ristanovic, J.-Y. Le Boudec, A. Chaintreau, and V. Erramilli, “Energy efficient offloading of 3g networks,” in *2011 IEEE eighth international conference on mobile ad-hoc and sensor systems*, 2011, pp. 202–211.

[95] E. Bulut and B. K. Szymanski, “Wifi access point deployment for efficient mobile data offloading,” *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 17, no. 1, pp. 71–78, 2013.

[96] *Femtocells - natural solution for offload, Small Cell Fourm*, Nov. 2020. [Online]. Available: https://scf.io/en/documents/all_documents.php

[97] A. Aijaz, H. Aghvami, and M. Amani, “A survey on mobile data offloading: technical and business perspectives,” *IEEE Wireless Communications*, vol. 20, no. 2, pp. 104–112, 2013.

[98] *Femtocell: natural solution for offload,Femto Fourm*, Nov. 2020. [Online]. Available: https://www.scribd.com/document/212298645/Femtocell-naturalsolution-for-offload

[99] *Local IP Access and Selected IP Traffic Offload (LIPASIPTO),3GPP Portal, Huawei Technologies France*, Nov. 2020. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=909

[100] *UTRAN architecture for 3G Home Node B (HNB); Stage 2*, Nov. 2020. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=1220

[101] M. H. Cheung and J. Huang, “Dawn: Delay-aware wi-fi offloading and network selection,” *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 6, pp. 1214–1223, 2015.

[102] X. Zhuo, W. Gao, G. Cao, and S. Hua, “An incentive framework for cellular traffic offloading,” *IEEE transactions on mobile computing*, vol. 13, no. 3, pp. 541–555, 2013.

[103] J. Lee, Y. Yi, S. Chong, and Y. Jin, “Economics of wifi offloading: Trading delay for cellular capacity,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1540–1554, 2014.

[104] Y. Im, C. Joe-Wong, S. Ha, S. Sen, M. Chiang *et al.*, “Amuse: Empowering users for cost-aware offloading with throughput-delay tradeoffs,” *IEEE Transactions on Mobile Computing*, vol. 15, no. 5, pp. 1062–1076, 2015.

[105] C. K. Ho, D. Yuan, and S. Sun, “Data offloading in load coupled networks: A utility maximization framework,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 1921–1931, 2014.

[106] X. Chen, J. Wu, Y. Cai, H. Zhang, and T. Chen, "Energy-efficiency oriented traffic offloading in wireless networks: A brief survey and a learning approach for heterogeneous cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 4, pp. 627–640, 2015.

[107] X. Kang, Y.-K. Chia, S. Sun, and H. F. Chong, "Mobile data offloading through a third-party wifi access point: An operator's perspective," *IEEE Transactions on Wireless Communications*, vol. 13, no. 10, pp. 5340–5351, 2014.

[108] G. Iosifidis, L. Gao, J. Huang, and L. Tassiulas, "A double-auction mechanism for mobile data-offloading markets," *IEEE/ACM Transactions on Networking*, vol. 23, no. 5, pp. 1634–1647, 2014.

[109] Y. Sui, J. Vihriala, A. Papadogiannis, M. Sternad, W. Yang, and T. Svensson, "Moving cells: a promising solution to boost performance for vehicular users," *IEEE Communications Magazine*, vol. 51, no. 6, pp. 62–68, 2013.

[110] F. Haider, M. Dianati, and R. Tafazolli, "A simulation based study of mobile femtocell assisted lte networks," in *2011 7th IEEE International Wireless Communications and Mobile Computing Conference*, 2011, pp. 2198–2203.

[111] M. Z. Chowdhury, S. Q. Lee, B. H. Ru, N. Park, and Y. M. Jang, "Service quality improvement of mobile users in vehicular environment by mobile femtocell network deployment," in *IEEE ICTC 2011*, pp. 194–198.

[112] M. H. Qutqut, "Mobile small cells in cellular heterogeneous networks." Ph.D. dissertation, Queen's University at Kingston, Ontario, Canada, 2014.

[113] H. Soroush, P. Gilbert, N. Banerjee, B. N. Levine, M. Corner, and L. Cox, "Concurrent wi-fi for mobile users: analysis and measurements," in *Proceedings of the Seventh COnference on emerging Networking EXperiments and Technologies*, 2011, pp. 1–12.

[114] N. Golrezaei, P. Mansourifard, A. F. Molisch, and A. G. Dimakis, "Base-station assisted device-to-device communications for high-throughput wireless video networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 7, pp. 3665–3676, 2014.

[115] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 1801–1819, 2014.

[116] Y.-D. Lin and Y.-C. Hsu, "Multihop cellular: A new architecture for wireless communications," in *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No. 00CH37064)*, vol. 3, 2000, pp. 1273–1282.

[117] B. Kaufman and B. Aazhang, "Cellular networks with an overlaid device to device network," in *2008 42nd IEEE Asilomar conference on signals, systems and computers*, 2008, pp. 1537–1541.

[118] A. Asadi and V. Mancuso, "On the compound impact of opportunistic scheduling and d2d communications in cellular networks," in *Proceedings of the 16th ACM international conference on Modeling, analysis & simulation of wireless and mobile systems*, 2013, pp. 279–288.

[119] T. Peng, Q. Lu, H. Wang, S. Xu, and W. Wang, "Interference avoidance mechanisms in the hybrid cellular and device-to-device systems," in *2009 IEEE 20th international symposium on personal, indoor and mobile radio communications*. IEEE, 2009, pp. 617–621.

[120] Y. Pei and Y.-C. Liang, "Resource allocation for device-to-device communications overlaying two-way cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3611–3621, 2013.

[121] W. Peng, F. Li, X. Zou, and J. Wu, "The virtue of patience: Offloading topical cellular content through opportunistic links," in *2013 IEEE 10th International Conference on Mobile Ad-Hoc and Sensor Systems*, 2013, pp. 402–410.

[122] B. Han, P. Hui, V. A. Kumar, M. V. Marathe, J. Shao, and A. Srinivasan, "Mobile data offloading through opportunistic communications and social participation," *IEEE Transactions on mobile computing*, vol. 11, no. 5, pp. 821–834, 2011.

[123] T. H. Luan, X. Shen, F. Bai, and L. Sun, "Feel bored? join verse! engineering vehicular proximity social networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 3, pp. 1120–1131, 2014.

[124] X. Cheng, L. Yang, and X. Shen, "D2d for intelligent transportation systems: A feasibility study," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 1784–1793, 2015.

[125] W. Sun, E. G. Ström, F. Brännström, Y. Sui, and K. C. Sou, "D2d-based v2v communications with latency and reliability constraints," in *2014 IEEE Globecom Workshops (GC Wkshps)*, 2014, pp. 1414–1419.

[126] Y. Ren, F. Liu, Z. Liu, C. Wang, and Y. Ji, "Power control in d2d-based vehicular communication networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 12, pp. 5547–5562, 2015.

[127] S. T. Kouyoumdjieva and G. Karlsson, "Energy-aware opportunistic mobile data offloading under full and limited cooperation," *Computer Communications*, vol. 84, pp. 84–95, 2016.

[128] L. Pu, X. Chen, J. Xu, and X. Fu, "D2d fogging: An energy-efficient and incentive-aware task offloading framework via network-assisted d2d collaboration," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3887–3901, 2016.

[129] S. Raychaudhuri, "Introduction to monte carlo simulation," in *2008 Winter simulation conference*. IEEE, 2008, pp. 91–100.

[130] Nils, *Scatter Plot colored by Kernel Density Estimate*, Matlab, Nov. 2020. [Online]. Available: https://au.mathworks.com/matlabcentral/fileexchange/65728-scatter-plot-colored-by-kernel-density-estimate

[131] *Traffic counts*, Auckland Transport Agency, Nov. 2020. [Online]. Available: https://at.govt.nz/about-us/reports-publications/traffic-counts/

[132] *VIRTUS CHIPSET FACT SHEET*, May 2021. [Online]. Available: https://www3.ntu.edu.sg/CorpComms2/Documents/2012/May/FactSheet_120524_VirtusMicrochip.pdf

[133] *Internet service provider survey:*, Nov. 2020. [Online]. Available: https://www.stats.govt.nz/information-releases/internet-service-provider-survey-2018

[134] M. B. John Stephenson, Jean Pierre de Raad, *Mobile Industry in New Zealand Performance and Prospects*, Nov. 2020. [Online]. Available: https://www.sparknz.co.nz/content/dam/telecomcms/sparknz/content/news/NZIER-Mobile-Industry-in-NZ.pdf

[135] *A smart city of 1 million will generate 180 million gigabytes of data per day by 2019,*, Cisco global cloud index (2014-2019) ed., Cisco, Nov. 2020. [Online]. Available: http://www.dqindia.com/a-smart-city-of-1-million-will-generate-180-million-gigabytes-of-data-per-day-by-2019-predicts-cisco-study/

[136] *Transport volume : Fleet information*, Ministry of Transport, New Zealand, Nov. 2020. [Online]. Available: http://www.transport.govt.nz/ourwork/tmif/transport-volume/tv035/

[137] *Future Demand New Zealand transport and society: Trends and projections.*, Nov. 2020. [Online]. Available: https://www.transport.govt.nz/assets/Uploads/Our-Work/Documents/10760e4aa4/fd-trends-and-projections.pdf

[138] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM wireless communications with MATLAB*. John Wiley & Sons, 2010.

[139] Y. Zheng, *T-Drive trajectory data sample*, Nov. 2020. [Online]. Available: https://www.microsoft.com/en-us/research/publication/t-drive-trajectory-data-sample/

[140] G. Guilford, *(July 2, 2019). A big reason Beijing is polluted: The average car goes 7.5 miles per hour*, Quartz, Nov. 2020. [Online]. Available: https://qz.com/163178/a-big-reason-beijing-is-polluted-the-average-car-goes-7-5-miles-per-hour/

[141] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "Driving with knowledge from the physical world," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2011, pp. 316–324.

[142] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang, "T-drive: driving directions based on taxi trajectories," in *Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems*, 2010, pp. 99–108.

[143] *Atlas of urban expansion, Region: East Asia and the Pacific*, Beijing, China, Nov. 2020. [Online]. Available: http://www.atlasofurbanexpansion.org/cities/view/Beijing_Beijing

[144] A. M. Abdelgader and W. Lenan, "The physical layer of the ieee 802.11 p wave communication standard: the specifications and challenges," in *Proceedings of the world congress on engineering and computer science*, vol. 2, 2014, pp. 22–24.

[145] T. Vanhatupa, "Wi-fi capacity analysis for 802.11 ac and 802.11 n: Theory & practice," *Ekahau Inc*, 2013.

[146] Y. Ghasempour, C. R. da Silva, C. Cordeiro, and E. W. Knightly, "Ieee 802.11 ay: Next-generation 60 ghz communication for 100 gb/s wi-fi," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 186–192, 2017.

[147] S. Naseer, W. Liu, and N. I. Sarkar, "Energy-efficient massive data dissemination through vehicle mobility in smart cities," *Sensors*, vol. 19, no. 21, p. 4735, 2019.

[148] S. J. Elias, M. N. B. M. Warip, R. B. Ahmad, and A. H. A. Halim, "A comparative study of ieee 802.11 standards for non-safety applications on vehicular ad hoc networks: a congestion control perspective," in *Proceedings of the World Congress on Engineering and Computer Science*, vol. 2, 2014, p. 33.

[149] Y. Li, D. Jin, Z. Wang, L. Zeng, and S. Chen, "Coding or not: Optimal mobile data offloading in opportunistic vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 318–333, 2013.

[150] W. Kotahi, *State Highway Traffic Volumes*, Nov. 2020. [Online]. Available: https://www.nzta.govt.nz/resources/state-highway-traffic-volumes/

[151] *List of Cable Distance Limits: Ethernet, Fiber, HDMI, DVI and More*, May 2021. [Online]. Available: https://blog.router-switch.com/2020/11/list-of-cable-distance-limits-ethernet-fiber-hdmi-dvi-and-more/

[152] *Movable Type Scripts: Calculate distance, bearing and more between Latitude/Longitude points*, Movable type scripts ed., MTL, Nov. 2020. [Online]. Available: https://www.movable-type.co.uk/scripts/latlong-nomodule.html

[153] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things journal*, vol. 1, no. 1, pp. 22–32, 2014.

[154] J. Lima, *10 of the biggest IoT data generators*, Nov. 2020. [Online]. Available: https://www.cbronline.com/internet-of-things/10-of-the-biggest-iot-data-generators-4586937/

[155] P. Nyeng and J. Ostergaard, "Information and communications systems for control-by-price of distributed energy resources and flexible demand," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 334–341, 2011.

[156] Y. He, M. Chen, B. Ge, and M. Guizani, "On wifi offloading in heterogeneous networks: Various incentives and trade-off strategies," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2345–2385, 2016.

[157] S. Galli, A. Scaglione, and Z. Wang, "For the grid and through the grid: The role of power line communications in the smart grid," *Proceedings of the IEEE*, vol. 99, no. 6, pp. 998–1027, 2011.

[158] G. Deconinck, "An evaluation of two-way communication means for advanced metering in flanders (belgium)," in *2008 IEEE instrumentation and measurement technology conference*, 2008, pp. 900–905.

[159] P. Kinney *et al.*, "Zigbee technology: Wireless control that simply works," in *Communications design conference*, vol. 2, 2003, pp. 1–7.

[160] H. Liang, B. J. Choi, A. Abdrabou, W. Zhuang, and X. S. Shen, "Decentralized economic dispatch in microgrids via heterogeneous wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 6, pp. 1061–1074, 2012.

[161] A. Fehske, G. Fettweis, J. Malmodin, and G. Biczok, "The global footprint of mobile communications: The ecological and economic perspective," *IEEE communications magazine*, vol. 49, no. 8, pp. 55–62, 2011.

[162] S. Naseer, W. Liu, N. I. Sarkar, P. H. J. Chong, E. Lai, M. Ma, R. V. Prasad, T. C. Danh, L. Chiaraviglio, J. Qadir *et al.*, "A sustainable marriage of telcos and transp in the era of big data: Are we ready?" in *International Conference on Smart Grid Inspired Future Technologies*. Springer, 2018, pp. 210–219.

[163] C. P. Mayer and O. P. Waldhorst, "Offloading infrastructure using delay tolerant networks and assurance of delivery," in *2011 IFIP Wireless Days (WD)*. IEEE, 2011, pp. 1–7.

[164] A. Finlay and E. Adera, *Application of ICTS for Climate Change Adaptation in the Water Sector: Developing Country Experiences an Emerging Research Priorities*. Citeseer, 2012.

[165] B. Cho and I. Gupta, "Budget-constrained bulk data transfer via internet and shipping networks," in *Proceedings of the 8th ACM international conference on Autonomic computing*, 2011, pp. 71–80.

[166] R. Munjal, W. Liu, X. J. Li, J. Gutierrez, and M. Furdek, "Sustainable massive data dissemination by using software defined connectivity approach," in *2017 27th IEEE International Telecommunication Networks and Applications Conference (ITNAC)*, 2017, pp. 1–6.

[167] D. Sitaram and G. Manjunath, *Moving to the cloud: Developing apps in the new world of cloud computing*. Elsevier, 2011.

[168] A. W. Malik, I. Mahmood, N. Ahmed, Z. Anwar *et al.*, "Big data in motion: A vehicle-assisted urban computing framework for smart cities," *IEEE Access*, vol. 7, pp. 55 951–55 965, 2019.

[169] *Digital Globe Case Study: "DigitalGlobe Moves Petabytes of Data Quickly and Securely Using AWSSnowmobile*, Nov. 2020. [Online]. Available: https://aws.amazon.com/solutions/case-studies/digitalglobe/

[170] R. Hunjet, B. Fraser, T. Stevens, L. Hodges, K. Mayen, J. C. Barca, M. Cochrane, R. Cannizzaro, and J. L. Palmer, "Data ferrying with swarming uas in tactical defence networks," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 6381–6388.

[171] K. Usbeck, M. Gillen, J. Loyall, A. Gronosky, J. Sterling, R. Kohler, R. Newkirk, and D. Canestrare, "Data ferrying to the tactical edge: A field experiment in exchanging mission plans and intelligence in austere environments," in *2014 IEEE Military Communications Conference*. IEEE, 2014, pp. 1311–1317.

[172] B. Baron, P. Spathis, H. Rivano, M. D. de Amorim, Y. Viniotis, and M. H. Ammar, "Centrally controlled mass data offloading using vehicular traffic," *IEEE Transactions on Network and Service Management*, vol. 14, no. 2, pp. 401–415, 2017.

[173] F. Dressler, P. Handle, and C. Sommer, "Towards a vehicular cloud-using parked vehicles as a temporary network and storage infrastructure," in *Proceedings of the 2014 ACM international workshop on Wireless and mobile technologies for smart cities*, 2014, pp. 11–18.

[174] B. Cho and I. Gupta, "New algorithms for planning bulk transfer via internet and shipping networks," in *2010 IEEE 30th International Conference on Distributed Computing Systems*, 2010, pp. 305–314.

[175] N. Cheng, H. Zhou, L. Lei, N. Zhang, Y. Zhou, X. Shen, and F. Bai, "Performance analysis of vehicular device-to-device underlay communication," *IEEE Transactions on vehicular technology*, vol. 66, no. 6, pp. 5409–5421, 2016.

[176] P. McKendry, "Energy production from biomass (part 1): overview of biomass," *Bioresource technology*, vol. 83, no. 1, pp. 37–46, 2002.

[177] *Energy & the Environment: Emissions & Generation Resource Integrated Database (eGRID)*, Nov. 2020. [Online]. Available: https://www.epa.gov/energy/emissions-generation-resource-integrated-database-egrid

[178] *Convert MJ to kwh - Conversion of Measurement Units*, Nov. 2020. [Online]. Available: https://www.convertunits.com/from/MJ/to/kwh

[179] D. Palma, A. Zolich, Y. Jiang, and T. A. Johansen, "Unmanned aerial vehicles as data mules: An experimental assessment," *IEEE Access*, vol. 5, pp. 24 716–24 726, 2017.

[180] M. M. Coutinho, A. Efrat, T. Johnson, A. Richa, and M. Liu, "Healthcare supported by data mule networks in remote communities of the amazon region," *International scholarly research notices*, vol. 2014, 2014.

[181] *Speedtest Market Reports of Calculating Upload and Down-load Data Rates*, May 2021. [Online]. Available: http://www.speedtest.net/reports/

[182] D. Costenaro and A. Duer, "The megawatts behind your megabytes: going from data-center to desktop," *Proceedings of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings, ACEEE, Washington*, pp. 13–65, 2012.

[183] R. Pries, M. Jarschel, D. Schlosser, M. Klopf, and P. Tran-Gia, "Power consumption analysis of data center architectures," in *International Conference on Green Communications and Networking*. Springer, 2011, pp. 114–124.

[184] *Powerconsumption analysis of data center architectures*, May 2021. [Online]. Available: https://www.eia.gov/outlooks/aeo/pdf/aeo2018.pdf

[185] D. Robinson, *Sharing economy companies list: From Airbnb and Uber to Omni and Hubble, 15 companies disrupting their industries*, May 2021. [Online]. Available: https://www.ns-businesshub.com/technology/sharing-economy-companies-list/

[186] N. G. La Vigne, S. S. Lowry, J. A. Markman, and A. M. Dwyer, "Evaluating the use of public surveillance cameras for crime control and prevention," *Washington, DC: US Department of Justice, Office of Community Oriented Policing Services. Urban Institute, Justice Policy Center*, 2011.

[187] *NYC surveillance camera project*, Nov. 2020. [Online]. Available: http://www.mediaeater.com/cameras/

[188] C. Rapson, *2020 IEEE New Zealand Vehicle Technology (VT) 1st Workshop-Thursday, 29 of October 2020: Auckland Transport.*, Oct. 2020.