Scalable Urban Crowdsensing: Data Contributor and Consumer Dynamics

Elham Heydari-Gharaei

A DISSERTATION SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Graduate Program in Transporation Engineering,

YORK UNIVERSITY

TORONTO, ONTARIO

June 2025

Abstract

Dynamic transportation routing and parking management rely on real-time data on traffic conditions and curbside availability. Traditional data collection methods require significant infrastructure investments and processing capacity, making them costly and often impractical for large-scale implementation. In contrast, crowdsensing leverages data from users' smart devices, offering a cost-effective alternative for collecting real-time information. However, crowdsensed data is often noisy, inaccurate, unstructured, and heavily dependent on voluntary user contributions. This thesis is motivated by the central question: how can we effectively design and manage platforms, such as parking management and routing, that rely on such incomplete yet abundant data? To address this, several key studies are presented.

The first study, Leveraging Data Contributors to Enhance Social Welfare Through Crowdsensing, examines how user-generated data can improve decision-making at service facilities such as parking lots. By aggregating reported wait times, the platform provides real-time estimates that help incoming users decide whether to enter. This research determines the optimal fraction of data contributors needed to maximize the collective benefit, social welfare, of facility users.

The second study, Route Choice Using Crowd-Generated Travel Time Information, explores how crowdsensed travel times influence traffic assignment. Two critical factors, contribution ratio (the proportion of travelers sharing data) and observation window (the period over which data is collected), must be balanced to minimize network-wide travel times. This study analyzes their effects on perceived travel risks and proposes solutions to maintain their optimal values, ensuring efficiency in routing decisions.

The third study, Resource Allocation and Route Generation for Urban Mobile Sensing, focuses on optimizing the number and routes of sensing agents, users carrying data collection devices, to enhance efficiency. By determining the optimal number of agents and their navigation paths, this study minimizes total travel costs while ensuring that each parking spot is revisited within the specified headway constraints. A case study in

Toronto demonstrates the practical applicability of the proposed optimization framework for parking occupancy detection.

The fourth study, Implementing Parking Occupancy Detection Using Dashcam Footage, develops a sensing system that analyzes dashcam video to detect real-time parking occupancy. By applying advanced video processing techniques, this project provides dynamic and scalable insights into urban parking availability. It aims to overcome the limitations of traditional parking data collection tools, such as stationary cameras and sensors, which face challenges related to installation, maintenance, and regulation. Instead, mobile technologies like dashcams, LiDAR, and ultrasound sensors offer a more scalable solution for capturing on-street parking availability.

Together, these studies contribute to improving the efficiency of crowdsensing platforms by analyzing the impact of three types of crowdsensed data: wait times, travel times, and urban monitoring reports. These data types are examined across three distinct domains: service facilities (e.g., parking facilities), transportation networks (e.g., scenarios involving multiple modes of transportation or alternative routes with varying congestion levels), and urban monitoring (e.g., parking availability detection, pedestrian safety, and traffic flow monitoring). The research examines how access to such data influences individual decision-making, particularly in the presence of competing alternatives, as well as its impact on overall system performance. Furthermore, it investigates the factors affecting the accuracy and reliability of crowdsensed data in each domain and explores how optimizing these factors can enhance both user experience and system-wide efficiency.

Dedication

First and foremost, I thank Allah for granting me the strength, perseverance, and guidance to complete this work. Without His blessings and support, this achievement would not have been possible.

I dedicate this work to the love of my life, Fahad, whose unconditional support, understanding, and encouragement have carried me through every stage of this journey.

I also extend my heartfelt dedication to my parents, siblings, and friends, whose constant love and motivation have been an enduring source of strength and determination ♡.

Acknowledgments

I am deeply grateful to my supervisor, Dr. Mehdi Nourinejad, for his mentorship, wisdom, and encouragement throughout this journey. His invaluable guidance and wholehearted support have shaped my academic work and inspired me to explore new directions in transportation research.

I also wish to thank my co-authors, Dr. Hoessein Abouee Mehrizi and Dr. Sina Bahrami, for their support and for imparting the foundational principles of research.

Finally, I would like to express my gratitude to the faculty and staff in the Civil Engineering Department at York University for creating a supportive and enriching academic environment.

Thank you all for your support, encouragement, and belief in my work.

Table of Contents

List of Tables

List of Figures

Chapter 1

CHAPTER ONE:

INTRODUCTION

1.1 1.1 Background

This chapter introduces the concept of crowdsensing and explores its principles, advantages, and diverse applications, highlighting its transformative potential in modern data-driven environments. It also examines the importance and benefits of implementing efficient routing and parking management systems, which are crucial for addressing the dynamic challenges of urban mobility and enhancing overall transportation efficiency.

1.1.1 Crowdsensing

Crowdsensing is a population-level data collection strategy that leverages the widespread availability of sensing devices, such as smartphones, wearable gadgets, and onboard vehicle sensors. By observing a subset of users within a specified experiment or context, crowdsensing effectively gathers and aggregates real-time information on the system under investigation. One primary advantage of crowdsensing is its scalability and cost-effectiveness, as it eliminates the need for extensive infrastructure and significant capital investments. However, crowdsensed data is often noisy, inaccurate, unstructured, and heavily dependent on voluntary user contributions.

Crowdsensing has been adopted in various fields to improve data collection and analysis. For example, in clinical health trials, particularly cohort studies, research teams recruit patients and gather data over several years to examine the development of particular diseases or traits and identify common behaviors among participants. A 2019 Canadian study monitored home care clients over the age of 55 using wearable devices to assess frailty. These devices recorded daily step counts, sleep metrics, and heart rate. The study showed that analyzing crowdsensed data effectively assessed frailty in older home care clients and helped identify individuals requiring more intensive home care support. In contrast, traditional methods, such as interviews or surveys, have limited reach, accuracy, and efficiency.

Google Maps leverages crowdsensing by collecting real-time data from users, including location, travel speed, route information, and search queries, to generate network-level

insights for navigation, wait-time predictions, crowd assessments, and identifying peak hours for businesses. These insights directly support both individual and system-level decision-making. For instance, users can choose less congested routes, avoid delays, or adjust travel schedules based on predicted traffic conditions and crowd levels. Businesses can respond to search and visitation trends by modifying operating hours or resource allocation. Replicating such predictions without crowdsensing would involve considerable investments to install sensors that track user movement, traffic patterns, and business activities (Chadil et al. 2008).

Similarly, INRIX, a private location-based data and SaaS analytics provider, collaborates with BMW to gather real-time parking availability data. As a BMW vehicle travels, its onboard sensors capture information on open parking spaces. By aggregating the data from multiple BMWs, the platform creates a digital map of parking availability. This enables users to make informed decisions regarding where and when to search for parking, reducing search time and easing traffic congestion. These use cases illustrate how crowdsensing not only improves the efficiency of data collection but also plays a critical role in shaping real-time decision-making, leading to more adaptive, efficient, and user-centric transportation systems.

1.1.2 Efficient Routing and Parking Navigation

With the advent of the automobile in the early 20^{th} century, cities experienced a steady rise in vehicle numbers, leading to traffic congestion. Early transportation systems relied on rudimentary infrastructure, such as roads and railways, without real-time data integration. Traffic control was limited to fixed timetables, traffic lights, and manual oversight, making it difficult to adapt to sudden changes or unexpected incidents, which resulted in inefficient routes and longer travel times (Streets.mn 2019).

Beyond routing, parking management has become an equally critical challenge in urban mobility. Searching for on-street parking is time-consuming and frustrating, causing congestion and increased urban pollution. Conventional parking systems, which rely on static information, often fail to account for real-time availability, leading to inefficient

space utilization and prolonged search times.

Optimized traffic management systems distribute vehicles more evenly across city streets, preventing bottlenecks at intersections and reducing travel times, ultimately maximizing the effectiveness of existing transportation infrastructure. Additionally, optimized routing and parking strategies enhance the user experience by increasing convenience, saving time, and reducing travel-related stress. From an environmental perspective, these systems lower fuel consumption and emissions by directing drivers to optimal routes and available parking spaces, contributing to better air quality and a reduced transportation carbon footprint.

1.2 Research Motivation

In the following sections, we explain how crowdsensing overcomes traditional transportation system limitations. We first show that static, historical data-based models fail to adapt to real-time events and scale with urban growth, whereas crowdsensing offers continuous, dynamic data collection. We then discuss the balance between user privacy, participation, and system performance.

1.2.1 The Power of Crowdsensing

Traditional transportation systems that rely on historical data or pre-set models for decision-making often fail to account for real-time events—such as accidents, sudden traffic spikes, unplanned road closures, or fluctuations in curbside occupancy—that can quickly render previously efficient routes congested or blocked. Consequently, these systems lead to suboptimal routing decisions, inefficiencies, and increased travel times. In contrast, crowdsensing continuously collects up-to-date information on traffic conditions and parking availability, enabling routing systems to adjust dynamically to unexpected events, thereby reducing delays and mitigating congestion.

Moreover, traditional systems struggle to scale in the face of urban growth and increased vehicle volumes. Manual monitoring, semi-automated systems, and stationary sensors become quickly overwhelmed by the growing demand for both road capacity and parking spaces. These systems typically lack the capability to integrate diverse sources of real-time data, such as traffic sensors, GPS devices, or crowdsourced information, limiting their ability to offer comprehensive, dynamic solutions at scale. Crowdsensing overcomes these scalability challenges by leveraging the widespread use of smartphones, wearable devices, and connected vehicles. This decentralized approach not only captures granular, real-time data directly from individual users, drivers, bikers, and pedestrians, but also does so without the need for extensive infrastructure investments. As a result, crowdsensing naturally scales with urban growth and supports the development of adaptive, comprehensive transportation management solutions that are both responsive and cost-

effective.

Advancements in IoT, smartphone technology, and next-generation connectivity have further elevated the potential of crowdsensing in transportation. Modern smartphones, equipped with GPS, accelerometers, gyroscopes, and constant internet access, allow nearly every user to contribute detailed, real-time data on movement and road conditions. Moreover, IoT devices, including connected cars, smart parking sensors, and traffic infrastructure, integrate seamlessly with crowdsensing platforms, providing additional layers of data on traffic patterns and road quality. The advent of 5G networks has been particularly transformative, enabling faster data transmission and lower latency that support large-scale, instantaneous updates.

1.2.2 Balancing User Privacy, Participation, and Performance

The volume and consistency of collected data on crowdsensing platforms depend on users' willingness to contribute data. Growing concerns about user privacy have led companies like Apple and Google to implement strict controls over data sharing. For instance, iPhone users can select from multiple privacy options, including "Don't Allow" (which denies all access), "Allow Once" (granting temporary access), "Allow While Using the App" (permitting access only during app usage), and "Allow Always" (enabling continuous background access). While these measures empower users to protect their privacy, they also challenge planners to develop strategies to motivate users to agree on sharing data.

Beyond privacy, several other concerns complicate the widespread adoption of crowdsensing. A key issue is technology bias, as data contributors are more likely to use newer, more advanced devices. This skews the dataset and may lead to the under representation of users with older technologies, often including low-income populations or individuals in rural areas, thereby limiting the generalization, inclusiveness, and fairness of the insights derived from the data. Moreover, security risks extend beyond mere privacy violations. Crowdsensing systems are susceptible to malicious attacks, including data spoofing, manipulation, and even breaches of connected devices and vehicles.

Moreover, increasing the proportion data contributors aiming at closing the gap be-

tween crwodsensed data and real-time data does not always result in reduced social costs from a planner's perspective. Access to real-time travel data can cause many travelers to converge on the same routes, departure times, and available parking spots, a phenomenon known as the Informational Braess' Paradox (IBP) (Acemoglu et al. 2018), where additional information paradoxically worsens congestion and delays. Therefore, planners must understand the factors that determine the efficiency of a crowdsensing platform in order to manage the flow of user-generated data and assess how adjustments to these factors impact data accuracy and overall system performance.

1.3 Research Objectives

This thesis investigates innovative approaches to enhance parking management and routing applications by strategically deploying crowdsensing platforms. It is organized around the following objectives, each addressing key challenges in leveraging real-time, user-generated data to improve transportation system performance.

- Analyze the impact of disseminating user-generated wait time information (e.g., queue lengths in parking searches and travel times in routing) on user decision-making, overall system performance, and social welfare.
- Identify and optimize factors affecting the accuracy of crowd-generated wait time information. Establish the optimal level of user contribution in crowdsensing platforms to enhance social welfare, while safeguarding privacy and ensuring equitable access, and assess the effectiveness of contribution reward mechanisms in sustaining this optimal participation.
- Evaluate behavioral differences between users who access paid real-time wait time information and those who rely on free, crowd-generated data, with a focus on how risk sensitivity and wait time variability affect their trust and reliance on the information.
- Develop a practical framework for parking occupancy data collection using dashcam

footage and optimize resource allocation in crowd-generated data collection tasks.

- Provide strategic recommendations for service providers to effectively manage and enhance crowdsensing platforms.

The above objectives are examined through the following studies:

1.3.1 Leveraging Data Contributors to Enhance Social Welfare Through Crowdsensing

Many service facilities, such as parking lots, allow customers to report their waiting times via crowdsensing platforms. By aggregating these reports, the system produces real-time wait time estimates that help incoming users decide whether to enter the facility. This study focuses on understanding how timely information influences user decisions and on determining the ideal proportion of data contributors required to maximize overall social welfare.

In this study, the impact of customer access to user-generated queue length (QL) information on social welfare is analyzed when only a fraction of customers contribute their observed QL. A single-server queue is modeled in which strategic customers decide whether to join based on the information available through crowdsensing. In a two-stage model, customers first decide whether to check a public online platform (POP) containing shared QL data, similar to assessing a restaurant's reputation based on previous diners' feedback. Customers choose to observe the queue if the expected benefit of the crowdsensed information exceeds the cost of monitoring it. In the second stage, a decision is made to join or balk based on whether the service reward outweighs the expected waiting costs, with the possibility of sharing their QL after observing the queue. The analysis further examines how the data contributors ratio influences system dynamics and overall performance.

1.3.2 Route Choice Using Crowd-Generated Travel Time Information

Accurate travel time estimation in traffic management depends on two critical factors: the contribution ratio (the proportion of travelers sharing their data) and the observation

window (the duration over which data is collected). This objective aims to identify the optimal balance between these factors to minimize network-wide travel times. It also explores the impact of these parameters on travelers' perceived risk and develops strategies to sustain optimal data contribution levels, thereby enabling more efficient routing decisions.

This study focuses on transportation networks that rely on crowd-generated travel data (CTD) to minimize network-wide travel times while accounting for users' data-sharing preferences and the reluctance of some travelers to share their information. The network under consideration offers two options: one with a fixed and known travel time, termed the "safe road," and another with a stochastic travel time, referred to as the "random road" (Verhoef et al. 1996, Liu et al. 2018, Liu and Yang 2021). Although all travelers have free access to CTD, only a subset chooses to share their travel data. The analysis examines the impacts of free access to CTD on route choices by focusing on the contribution ratio and observation window. It also investigates whether the conclusions of Yang (1998) hold in a user-generated data-driven traffic network—that is, whether increasing the contribution ratio decreases overall travel time—and evaluates how traveler risk levels, particularly sensitivity to variability on the random road, affect decision-making. Furthermore, the study considers a scenario in which travelers can pay to access real-time travel information and analyzes how this option interacts with the contribution ratio and traveler behavior. Lastly, the effect of providing a contribution reward for those who share travel data is explored, with the goal of enforcing the optimal contribution ratio.

1.3.3 Resource Allocation and Route Generation for Urban Mobile Sensing

To monitor an urban area through crowdsensing, this study focuses on optimizing both the number of sensing agents equipped with sensing devices and their corresponding data collection routes. The objective is to minimize total travel costs while satisfying sensing frequency constraints, which are essential for efficient and reliable data collection. An urban area is modeled as a network, where nodes represent data collection locations and edges denote paths between them. These edges form closed routes that agents follow to

collect data. An optimization model is used to determine both the optimal number of agents and their routes, aiming to minimize total travel time while satisfying the required data collection frequency. Since obtaining an exact solution is computationally expensive and time-consuming, a heuristic algorithm is developed to efficiently find an approximate solution. The algorithm begins by applying K-Means clustering to partition the set of nodes into clusters. This clustering step groups nodes based on proximity, facilitating efficient route planning. Within each cluster, the Nearest Neighbor Algorithm is employed to determine the movement sequence by iteratively selecting the nearest unvisited node, thereby minimizing the total travel distance. Finally, a linear resource allocation model is solved to determine the optimal number of agents for each route, subject to headway constraints and the total number of available agents.

1.3.4 Implementing Parking Occupancy Detection Using Dashcam Footage

This study develops a sensing solution that analyzes dashcam footage to detect parking occupancy. By employing advanced video processing techniques, the system delivers dynamic, real-time parking insights, offering a scalable and effective tool for urban parking management. To perform this task, the YOLOv8 model by Ultralytics, trained on the COCO dataset for rapid and accurate object detection in video feeds is employed. This model detects various types of vehicles and people, while annotation tasks, such as defining detection zones, drawing bounding boxes, and blurring faces for privacy, are managed using the open-source Supervision library by Roboflow. Geofencing is implemented to create virtual boundaries around specific areas, such as streets between intersections, enabling more precise monitoring of parking availability. GPS metadata is extracted from videos using ExifTool and, when necessary, converted to the standard decimal degrees format used by Google Maps. Synchronizing the video and GPS data ensures that object detection is initiated and terminated at the correct moments.

Dashcam footage serves as input to the object detection model, with emphasis on the bottom half of each frame where parked vehicles are likely to appear. In designated geofenced areas, the detection model is activated to assign unique IDs to vehicles and count

them cumulatively, thereby providing an accurate measure of parked vehicles. When the camera exits these zones, such as at intersections, the detection model is deactivated and the accumulated count is reset. This process, combined with privacy measures such as blurring detected persons, enables continuous, accurate counting of occupied parking spaces while protecting individual privacy.

1.4 Review of Existing Literature

1.4.1 Crowdsensing in Transportation Systems

Crowdsensing, defined as data collection through individual contributions from a large user base, has increasingly been applied to monitor real-time transportation data. Studies have shown that crowdsensing provides a scalable and efficient alternative to traditional data collection methods, which are often hindered by high costs and limited by infrastructure constraints (Li et al. 2017). By leveraging users' mobile devices, GPS data, and vehicular sensors, crowdsensing applications can generate up-to-the-minute insights on traffic patterns, road incidents, and parking availability.

Li et al. (2017) highlight that while crowdsensing offers real-time traffic information, it also faces several notable challenges. Key among these are data accuracy, privacy concerns, and user participation levels. In particular, crowdsensed data may suffer from noise or inconsistencies due to variations in sensor quality or user engagement, making robust data processing algorithms essential (Ma et al. 2017b). Ensuring high-quality data often requires techniques such as outlier detection, sensor calibration, and data fusion algorithms capable of reconciling disparate sources (Ganti et al. 2011). Additionally, privacy concerns arise when users share location or movement data; methods like differential privacy, k-anonymity, and location obfuscation can help mitigate these risks (Shi et al. 2021, Zhang et al. 2020b).

Another critical factor in achieving effective crowdsensing is maintaining user participation at a sufficient level to produce representative insights. Existing research shows that gamification strategies, monetary rewards, or social incentives can encourage more users to share data, improving coverage and reliability (Du et al. 2017, Lee et al. 2018). For instance, Waze successfully employs a points-based system that rewards users for reporting traffic incidents, road closures, and other relevant information (Chen 2019). However, sustaining participation over time remains a challenge, particularly when users perceive minimal benefit from contributing (Lee et al. 2018).

Beyond traffic monitoring, crowdsensing has shown significant potential for improving

parking management. By aggregating data from drivers and their onboard sensors, real-time mapping of parking availability can be achieved, reducing the time spent searching for spaces and alleviating congestion (Liang et al. 2019, Shi et al. 2021). Crowdsensed parking solutions may integrate with existing navigation applications to direct users to vacant spots, thus enhancing both user satisfaction and overall traffic flow. Nonetheless, integrating such parking data often requires overcoming interoperability issues between various platforms and ensuring a consistent data format (Aguila et al. 2019).

Recent advancements in mobile computing, cloud infrastructure, and Internet of Things (IoT) devices have further enabled large-scale crowdsensing in transportation. For example, high-speed mobile networks (4G/5G) and edge computing architectures can accommodate the rapid exchange of large data volumes, reducing latency and improving the responsiveness of crowdsensed applications (Li et al. 2019). Connected vehicle technologies, such as Vehicle-to-Everything (V2X) communication, have also begun to enhance data sharing among vehicles, roadside units, and central servers, paving the way for more granular and adaptive traffic management systems (Abboud et al. 2016).

Taken together, these studies underscore the growing recognition of crowdsensing as a viable, cost-effective approach to modern transportation management. Despite its promises of scalability and real-time insight, widespread adoption hinges on addressing challenges related to data quality, privacy, interoperability, and sustained user participation. Consequently, research continues to evolve new incentive mechanisms, robust data-processing techniques, and privacy-preserving algorithms to fully harness the potential of crowdsensing in transportation networks.

1.4.2 Routing Optimization Through Crowdsensing

Routing optimization has been a central focus in transportation research, with notable advancements arising from the integration of crowdsensed data. According to Zhang et al. (2020a), routing applications that rely on real-time information contributed by road users enable more adaptive and efficient route planning. By continuously updating traffic conditions from multiple data sources, including smartphones and in-vehicle sensors, these

applications can rapidly detect congestion, accidents, or other disruptions, guiding travelers toward less congested routes. This dynamic approach is particularly advantageous in densely populated urban environments, where traffic patterns can change rapidly and unpredictably, often resulting in reduced travel times and more balanced traffic distribution across the network.

Chen et al. (2018) build on this concept by demonstrating how predictive models and machine learning algorithms can enhance adaptive routing systems. Their study shows that crowdsensing not only offers instant feedback on current traffic conditions, but also facilitates forecasting of emerging congestion patterns. By analyzing large volumes of user-generated data, the system can predict when and where bottlenecks are likely to occur, thereby suggesting alternate routes in advance of severe congestion. Such proactive routing strategies have been shown to outperform traditional, static routing algorithms. However, the authors emphasize that the effectiveness of predictive models is heavily dependent on sufficient data volume and quality, underscoring the need for robust user participation. Inconsistent or incomplete crowdsensed information can skew the predictions, ultimately limiting the practical benefits of adaptive routing solutions (Chen et al. 2018).

Recent research has further investigated the interplay between crowdsensing and advanced optimization techniques. For instance, some studies apply evolutionary algorithms or metaheuristics to identify near-optimal routing solutions in complex urban networks (Kim et al. 2019, Xu et al. 2020). Others highlight the importance of combining crowdsensing with edge computing infrastructures, enabling localized data processing that reduces latency and minimizes bandwidth requirements (Li et al. 2020). Despite the computational advantages, these approaches also introduce challenges related to data privacy and security, which can affect the willingness of users to participate and the accuracy of the information collected. Ensuring robust encryption and anonymization techniques can help overcome these obstacles, thereby fostering an environment in which crowdsensed data can be shared confidently (Wang et al. 2021).

Taken together, these studies highlight the considerable potential of crowdsensing for

dynamic and predictive routing optimization. By integrating up-to-date, user-generated data with advanced computational techniques, researchers and practitioners can develop systems capable of adapting to rapidly changing urban traffic conditions. However, the efficacy of these systems depends on consistent, high-quality data contributions from a sufficiently large user base. In turn, addressing privacy concerns, maintaining user engagement, and leveraging scalable architectures all remain critical factors for the long-term success of crowdsensing-based routing solutions.

1.4.3 Parking Management Using Crowdsensing

Parking availability and navigation remain critical challenges in urban mobility, often resulting in increased congestion and emissions as drivers circle in search of open spots. Crowdsensing has emerged as a promising solution by aggregating real-time data from individual drivers. Wang et al. (2019) demonstrate the effectiveness of a real-time parking guidance system that leverages crowdsensed data to predict parking availability in popular urban areas. Their findings indicate a significant reduction in both search time and emissions associated with parking-related congestion.

The collaboration between INRIX and BMW provides a practical illustration of crowdsensing in industry applications (INRIX). By installing sensors in BMW vehicles to detect open parking spaces, INRIX compiles a digital map of available spots in real time, thereby enhancing parking navigation. Other studies extend this approach by incorporating predictive analytics and machine learning to estimate future parking occupancy based on historical and crowdsensed data (Freedman et al. 2018, Liu et al. 2020). These methods can further optimize parking resource allocation, particularly when integrated into broader smart city systems.

Despite these advances, several challenges remain in implementing crowdsensed parking management. First, ensuring a sufficient number of participating users is vital, since the accuracy and coverage of crowdsensed data diminish when participation levels are low (Aguila et al. 2019). Second, privacy considerations arise when collecting and sharing real-time location data from drivers (Smith et al. 2020). Lastly, the interoperability of

different crowdsensing platforms can hinder the seamless exchange of parking information among diverse applications and services. Nevertheless, as more cities prioritize sustainable mobility and adopt innovative technologies, crowdsensing remains a powerful tool in addressing parking-related issues.

1.5 Structure of the Thesis

The remainder of this thesis is organized into four main chapters, each focusing on a key aspect of optimizing crowdsensing platforms for intelligent transportation systems. Each chapter begins with an overview, followed by an introduction outlining the research motivation, objectives, and a review of related literature. The main body presents the methodology, analysis, and key findings, and concludes with a discussion on practical implications and potential extensions. The thesis concludes with a final chapter summarizing the key insights, discussing broader implications, and outlining future research directions. Supplementary materials and additional results are provided in the appendix.

Chapter **??** examines the impact of user-generated data on system performance. It presents a theoretical framework that analyzes user behavior when accessing shared waiting time data. The chapter determines the optimal level of user participation and proposes mechanisms to balance social welfare and data privacy.

Chapter **??** focuses on optimizing traffic assignment by leveraging real-time, crowd-generated travel time data. It develops a route choice model that considers the trade-offs between data accuracy, user contribution levels, and system-wide efficiency. The chapter examines risk perception's impact on user decisions and evaluates strategies to optimize data contribution.

Chapter **??** presents an optimization framework for resource allocation in crowdsensed parking data collection. This framework determines the optimal number of mobile sensing agents and their navigation routes while minimizing travel costs and ensuring timely data updates.

Chapter **??** proposes a system that leverages dashcam footage from connected vehicles to detect parking occupancy in real time. The approach integrates video analytics, edge computing, and geofencing to automate data capture and enhance spatial coverage without requiring fixed infrastructure.

Chapter 2

CHAPTER TWO: LEVERAGING DATA CONTRIBUTORS TO ENHANCE SOCIAL WELFARE THROUGH CROWDSENSING

2.1 Chapter Overview

Problem definition: Crowdsensing collects data from individuals participating in an activity to obtain information about a system, eliminating the need for costly data collection infrastructure. Examples include Google Maps, which tracks users to infer network travel times for applications like navigation or clinical experiments where a population sample is monitored to track health trends, such as the spread of infectious diseases or lifestyle habits. This chapter seeks to answer whether increasing crowdsensing data providers is always socially beneficial in congestion-sensitive systems like queues. **Methodology/results:** We use a single-server queue with Poisson service demand and completion rates to examine how social welfare changes based on the ratio of crowdsensing data contributors. We assume all customers have free access to crowdsensing information, but only a subset reports their observed queue length. We analyze the impact of data contributors on customer decisions to join a queue in a two-stage decision-making process. **Managerial implications:** The effect of data contributors on social welfare is influenced by the system load factor. When the load factor is high, there exists an optimal ratio of data contributors. Increasing the data contribution beyond its optimal leads to frequent updates about the queue, which encourages customers to join at every opportunity, increasing wait times and reducing social welfare. Below the optimal value, the system cannot provide accurate enough information to customers, which leads to socially inefficient decisions. In contrast to high demands, social welfare increases monotonically with the ratio of data contributors in low-load factor systems as it avoids congested queues. Furthermore, the findings suggest that providing public access to crowdsensing information without mandatory data contributions maximizes social welfare. In contrast, restricting customer access to crowdsensing information diminishes social welfare and reduces system throughput.

2.2 Introduction

Crowdsensing is a data collection strategy that leverages the ubiquitous presence of individuals engaged in a particular activity. Crowdsensing aims to track a sample of agents in a defined experiment to collect and aggregate information about that experiment or system. The distinctive advantage of crowdsensing is its scalability and cost-effectiveness, eliminating the need for extensive infrastructure and substantial financial investments, often motivated by the collection of population-level information.

Google Maps leverages crowdsensing by tracking real-time data from users willing to share their information (including their location, activity types and duration) to obtain network-level information for applications such as navigation. Google Maps also predicts wait times, crowding levels, and peak hours for listed businesses, assisting users in planning their visits. Replicating Google's predictions without crowdsensing would require substantial investments in installing sensors across various locations, each tasked with gathering data on user movements, traffic patterns, and business activities (Chadil et al. 2008). Similarly, INRIX, a private location-based data and software-as-a-service analytics provider, uses crowdsensing in collaboration with BMW to detect and transmit real-time information about parking availability. As a BMW vehicle drives, its onboard sensors collect data on parking availability. By aggregating this information from cruising BMW vehicles, a digital map of parking availability is created, improving parking navigation for all BMWs.

Crowdsensing is also implemented in clinical health trials, specifically in cohort studies where a research team recruits patients and collects information related to specific diseases over several years to investigate the development of specific traits or diseases and discover common behaviors among participants. In a Canadian research study conducted in 2019, home care clients aged 55 years and older were recruited to be monitored by wearable devices to assess frailty. The wearable devices collected data, including daily step counts, sleep metrics, and heart rate. According to the study, the analysis of the crowdsensing data effectively assessed frailty in older home care clients, aiding in identifying vulnerable

individuals requiring additional home care services. In contrast, traditional methods like interviews or surveys exhibit limited reach, accuracy, and efficiency, potentially introducing biases (Kim et al. 2020).

At service facilities, a subset of customers who are willing to share their waiting times on a crowdsensing platform can assist others in making decisions about joining queues by helping them predict wait times. For instance, Abouee Mehrizi et al. (2021) consider a queue with an unknown service capacity, and customers contribute to reviews regarding service speed on a platform. The service reviews are accessible to subsequent customers and help them make informed joining decisions by inferring the service capacity. The sample size is the number of shared reviews incoming customers use to estimate service capacity. Their study investigates the impacts of sample size on throughput, social welfare, and revenue.

With a similar motivation, Wang and Hu (2020) study a service where customers gain access to shared queue length (QL) data from preceding customers before entering a system. In return, they are required to contribute QL data to assist future customers in their joining decisions. They show that the user-generated QL influences queue entry decisions and reduces waiting times by helping customers avoid congested queues. Furthermore, they demonstrate that sharing QL data enhances social welfare compared to scenarios without QL information. They show that in some instances, user-generated QL proves more advantageous to social welfare than providing real-time QL data to customers.

Data contributors, i.e., customers who share their observed QL after entering a system, serve as the primary resource in crowdsensing by ensuring reliable data flow. The efficacy of a crowdsensing platform and the credibility of its amassed data rely directly on data contributors. The active participation of customers in contributing data hinges on various factors, such as their access to mobile devices and their consent to share data. Moreover, the regulatory approach influences whether customers feel obligated to share their data in exchange for access to the inferences derived from the data provided by contributing customers. The existing literature does not thoroughly explore how the ratio of data contributors affects system performance, an aspect that could provide insights into

motivating more customers to contribute data. Additionally, with the widespread use of social media, it is irrelevant to make the data contribution mandatory to accessing shared data. Therefore, the focus should shift to recognizing the significance of data contributors over data consumers.

This study examines the impact of customer access to user-generated QL information on social welfare when only a fraction of customers visiting a system contribute data by sharing their observed QL. To achieve this, we consider a single server queue facing strategic customers who decide whether to join a system based on the information available through crowdsensing. We present a customer choice through a two-stage model, where customers strategically aim to maximize utility. In the first stage of "Enter-or-Leave," customers check a public online platform (POP) that contains shared QL data by previous contributors. This is similar to when customers assess the QL of a local restaurant based on information shared by others who have been there and use this feedback to decide whether to dine in or choose another food provider. Customers choose to observe the queue if the expected utility from the crowdsensing information outweighs the access cost, representing the effort required to monitor the queue. In the "Join-or-Balk" stage, customers observe the QL and join the queue if the service reward outweighs expected waiting costs; otherwise, they opt out. We assume customers may decide to share QL information on the POP after observing the queue.

We explore the impact of the data contributors ratio on the system dynamics and performance. Initially, we investigate how shared QL information influences a customer's decision to enter or bypass the system. Subsequently, we analyze the system's social welfare based on the data contributors ratio. We show that the impact of the data contributors ratio on social welfare depends on the load factor. In less popular services with lower load factors, data contributors enhance social welfare by providing frequent congestion updates on the crowdsensing platform, allowing customers to make informed choices and avoid high congestion. In contrast, services with high load factor benefit from maintaining a threshold on the ratio of the data contributors. Exceeding this threshold leads to longer queue wait times for customers. When the data contributors ratio surpasses

this threshold, customers are willing to pay the access costs, even during high congestion. This leads to a continuous influx of customers and prevents queue depletion. As a result, the findings suggest that less popular service providers should consistently encourage data contribution, while popular providers, facing a high load factor, should regulate the ratio of data contributors. This regulation eliminates the need for additional investments in universal data participation. Additionally, our research underscores the importance of acknowledging the reduction in social welfare that occurs when access to a crowdsensing platform is restricted.

The chapter is organized as follows: a review of related studies in Section 3.3, the mathematical model in Section 3.4, presentation of the results on social welfare in Section 2.5, a discussion on the heterogeneity among customers in terms of access to the POP in Section 3.5, and the conclusion in Section 3.6.

2.3 Literature Review

In queueing systems, the availability of QL information to customers before they decide to join the queue is explored in two streams of observable and unobservable queues. Naor (1969) pioneers the study of observable queues, where customers have real-time access to QL information and join only when QL falls below a specific threshold. Edelson and Hilderbrand (1975) introduce the concept of unobservable queues, where there is no access to QL information, and develop a Nash-equilibrium strategy in which each customer optimally decides whether to join based on expectations of congestion, assuming all others behave similarly.

Disseminating QL information offers several advantages to incoming customers. Whitt (1999) claims that QL-based predictors outperform delay-history-based predictors because they use real-time system data, including arrival rate, abandonment rate, and server count, whereas the latter relies just on recent customer delay history. Similarly, Pender et al. (2018) study how QL announcements provide customers with valuable information about their wait times, thus influencing their decision to either return for future service or continue waiting for their current service.

When real-time QL information provision is impractical due to the associated high costs, an alternative approach is to substitute real-time information with delayed information. For example, a theme park seeking to enhance visitor experiences by providing real-time QL information opts for cost savings by using delayed updates (e.g., every 15 minutes) instead of continuous monitoring. Delayed information refers to the time lag between obtaining real-time QL data and a customer's queue participation decision. Kaul et al. (2012) coin the delay in disseminating real-time QL as the age of information (AoI) and aim to determine its optimal value in an $M/M/1$ queue, a standard model where arrivals follow a Poisson process, service times are exponentially distributed, and a single server serves the queue. However, Li and Liu (2019) demonstrate that selfish user behavior can significantly increase AoI levels if the system is poorly designed. This underscores the importance of developing effective reward mechanisms to motivate cus-

tomers to share QL information promptly, with the goal of minimizing both AoI and queue congestion. Subsequently, Huang and Modiano (2015) extend AoI optimization to multiclass $M/G/1$ queue, allowing general service times, and in the capacity-constrained case, as an $M/G/1/1$ queue where arriving users are blocked if the server is busy.

To gain a deeper understanding of QL information distribution, it is beneficial to investigate its effects in the context of variability in customer information levels. Kremer and Debo (2016) examine a system where some customers (referred to as uninformed) lack knowledge about product quality, while others (referred to as informed) possess this knowledge and use QL as an indicator of service quality. They show that low-quality firms do not benefit from having more informed customers since low QL signals poor quality. Within the same context, Hu et al. (2018) investigate a system where not all customers have real-time access to QL information. They demonstrate that achieving optimal social welfare depends on maintaining a specific balance between the fraction of informed and uninformed customers.

As social media and information-sharing applications continue to expand, service providers increasingly rely on the voluntary participation of the user community to collect and share information regarding the service queue congestion levels (Li and Liu 2019). Using user-generated QL information aligns with incorporating delayed information into the decision-making process. Some studies assess the effectiveness of user-generated QL information compared to real-time QL dissemination, which involves providing users with continuously updated, current QL information. Wang and Hu (2020) compare the impact of three types of QL access on social welfare: real-time QL access, access to user-generated QL, and zero access to QL. They assume a sharing obligation exists for customers accessing user-generated QL, requiring them to contribute by sharing their observed QL. Their study shows that when more customers share QL information, it boosts social welfare by decreasing the number of customers entering the system during long queues. Additionally, they demonstrate that access to user-generated QL in specific settings is more beneficial than access to real-time QL.

Figure 2.1 summarizes the related literature on disseminating QL information to cus-

tomers and highlights the gap in the existing literature addressed by this study. The x-axis illustrates the proportion of customers contributing by sharing QL information, while the y-axis depicts the percentage of customers utilizing the provided QL information (whether by the service provider or prior customers) to decide on queue entry. The solid line, denoted as [4], outlines the focus of our study. Specifically, this study focuses on utilizing user-generated QL information in queue participation decisions. In real-world crowdsensing setups, not all users participate in sharing QL information, and there is no reciprocal requirement for those accessing the shared QL to contribute QL updates. Therefore, we assume all customers have homogeneous access to the shared QL, updated by a fraction of previous customers who have observed the queue. We investigate how variations in the ratio of QL information contributors impact social welfare. Furthermore, we simulate a system in which only a fraction of customers have access to the crowdsensing platform and use the available information to make their joining decisions, with some of them contributing QL data. The remaining customers decide whether to join the queue without prior access to this information. This scenario corresponds to the top triangle in Figure 2.1.

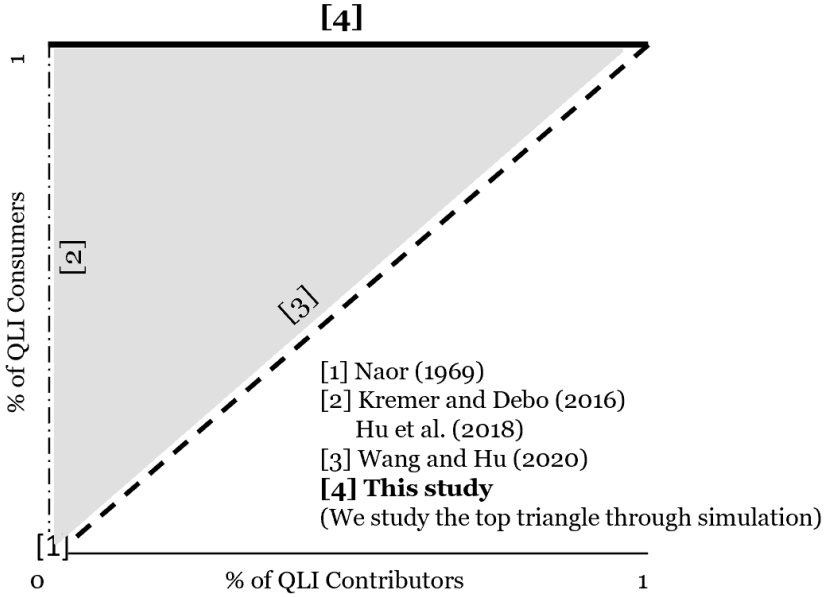


Figure 2.1: Study of customers access to and contribution in QL information in the literature

2.4 Queueing Model with Strategic Behavior and Information Sharing

We consider a single-server queue with service times distributed exponentially, having a mean of $1/\mu$, operating on a first-come-first-served basis. Service requests follow a homogeneous Poisson process with a rate of λ per unit of time. Customers are strategic and will balk if they see a crowded service area. Two classes of customers exist: contributors who share QL information and non-contributors who do not. Together, they make up a ratio of β and $(1-\beta)$ of the total service requests, respectively. Accordingly, the service request rate of contributors and non-contributors is $\lambda_e = \beta\lambda$ and $\lambda_u = (1-\beta)\lambda$, respectively.

Upon arrival at the queue, contributors post their observed QL, denoted as ϕ, including themselves, if they join the queue, along with the observation timestamp, represented as T, on the POP. From this point forward, the observed QL shared by a contributor, ϕ, is called the *reported congestion*. We assume there are no extra costs for contributors, and they are not charged an additional fee for posting ϕ and T on the POP. In contrast, non-contributors do not share their observed QL with others, regardless of their decision about joining the queue.

All customers have free access to the information contributors share on the POP, including the reported congestion and observation timestamp. If a customer decides to observe the QL, they will incur an access cost, denoted by h, regardless of their class. Access cost accounts for a customer's time and monetary expenses to observe the QL. After observing the QL, if a customer decides to join the queue based on the real-time QL, they will incur a marginal waiting cost per unit of time, denoted by c. This cost accounts for the time they spend waiting in the queue and receiving service and is applicable to both contributors and non-contributors. Upon service completion, customers receive a reward, denoted by R. The service provider charges no service fee. We assume the parameters are public knowledge and, without loss of generality, that $\mu = 1$.

2.4.1 Second Stage: Join or Balk

Given the reported congestion, customers will observe the real-time QL if they decide to enter the system. We assume that customers are strategic and aim to maximize their expected utility. The expected utility of joining a system with n customers, denoted as $U_{join}(n)$, is determined by subtracting the expected waiting cost of the sojourn time in the system from the service reward, thus

$$U_{join}(n) = R - \frac{c(n+1)}{\mu}. \tag{2.1}$$

If the number of customers in the queue, including the one currently receiving service, exceeds a predefined threshold, denoted as K, customers choose not to join the queue. This threshold on the QL ensures that the customer's waiting cost does not exceed the service reward. Thus, considering (2.1), we get

$$K = \left\lfloor \frac{R\mu}{c} \right\rfloor, K \geqslant 1, \tag{2.2}$$

where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

During the join-or-balk stage, contributors share the observed QL, including themselves, if they decide to join, along with their observation timestamp on the POP. Suppose a contributor enters the system and observes n other customers in the queue. If $n < K$, they will join the queue and report $\phi = n + 1$. If $n = K$, the contributor will post $\phi = K$ on the POP and balk. Therefore, $1 \leqslant \phi \leqslant K$.

2.4.2 First Stage: Enter or Leave

Upon a service request at time T', a customer decides whether to enter the system and observe the real-time QL or leave. They base this decision on the most recent QL information shared on the POP by a contributor at time T, where $T < T'$. Let δ denote the time difference between a service request and the most recent QL update on the POP. Note that the time difference $\delta = T' - T$ is an endogenous random variable specific to

each customer. As per the δ definition, only non-contributors enter the system during the time interval δ, while both contributors and non-contributors can receive service and depart. Consequently, if the reported congestion at time T is ϕ, the number of customers in the system at time T' ranges from zero to the maximum QL, K.

Provided with values of (ϕ, δ), customers perceive the probability of the queue containing fewer than K customers, allowing them to join the queue and receive the service. Additionally, they estimate their expected waiting time in the queue. These estimations enable them to assess their expected utility for entering the system after incurring the access cost, denoted as $U_{enter}(\phi, \delta)$. The expected entrance utility, $U_{enter}(\phi, \delta)$, captures the expected net value that a customer would gain from entering the system, assuming the most recent QL update ϕ occurred δ time units ago on the POP.

Let $\Gamma(t)$ be the random variable indicating the number of customers in the system at time t, where $0 \leqslant \Gamma(t) \leqslant K$. If $\Gamma(t) = n$ is less than K, the expected utility is the expected utility of receiving service from a system having n customers in the queue, minus the access cost. In contrast, when $\Gamma(t)$ equals K, a customer has to balk, wasting the access cost and earning a utility of -h. Therefore, given (ϕ, δ), the expected entrance utility is calculated as follows:

$$\begin{aligned} U_{enter}(\phi, \delta) = & \left[\sum_{n=0}^{K-1} p\Big\{ \Gamma(T') = n | \Gamma(T' - \delta) = \phi \Big\} \left(R - c\frac{n+1}{\mu} - h \right) \right] \\ & - p\Big\{ \Gamma(T') = K | \Gamma(T' - \delta) = \phi \Big\}\, h, \quad \delta \geqslant 0,\ 1 \leqslant \phi \leqslant K, \end{aligned} \tag{2.3}$$

where $p\Big\{ \Gamma(T') = n | \Gamma(T' - \delta) = \phi \Big\}$ denotes the probability that the number of customers in the system changes from ϕ to $n \in \{0, 1, 2, \ldots, K\}$ during the time interval δ. In the context of an $M/M/1/K$ queue, Morse et al. (1958) provided the first exact analysis for this transition probability. A proposed alternative solution by Takács (1962) can be found in Section A. 1 of the appendix.

When we set the timestamp of the most recent QL update, T, to zero, the service request time T' equals the time difference δ. After applying this simplification to (2.3),

we get:

$$U_{enter}(\phi,\delta) = R - \frac{c}{\mu}\Big(E[\Gamma(\phi,\delta)]+1\Big) - P_{\phi,K}(\delta)\left(R - \frac{c}{\mu}(K+1)\right) - h, \quad \delta \geqslant 0,\ 1 \leqslant \phi \leqslant K, \tag{2.4}$$

where $E[\Gamma(\phi,\delta)]$ denotes the *expected congestion*, i.e., the expected number of customers in the system δ time units after the most recent QL update ϕ. Additionally, the *transition blocking probability*, denoted as $P_{\phi,K}(\delta)$, signifies the probability of a customer encountering the maximum QL when requesting service δ time units after the most recent QL update ϕ on the POP.

A positive expected entrance utility means that the effort a customer invests in observing the QL by incurring the access cost and waiting in the queue will pay off, as the reward of receiving the service outweighs the hassle and waiting costs. Therefore, if a customer's expected entrance utility is greater than or equal to zero, they will progress to the second stage by paying the access cost h, which is incurred immediately after passing the first stage. On the other hand, if the expected costs outweigh the reward, a customer would have no motivation to enter the system. Hence, if a customer's expected entrance utility is negative, they will leave.

Only non-contributors might enter the system during the time interval δ. Therefore, we assume that the rate of customers who pass the first stage and enter the second stage during δ, denoted as $\lambda_u^{'}$, corresponds to the service request rate of non-contributors, $\lambda_u^{'} = (1-\beta)\lambda$. Note that customers' decisions are independent, and the decision of one individual to enter or leave the system at the first stage does not affect the arrival of future customers.

2.4.3 Transition Blocking Probability

Once a customer obtains the values of (ϕ,δ) from POP, they will choose to pay the access cost and enter the system if $U_{enter}(\phi,\delta)$ is positive. However, since non-contributors may have entered the system during the last δ units of time without updating the QL on the POP, there is a chance that a customer who has obtained a positive expected

entrance utility encounters a system with QL equal to K. This situation forces the individual to balk, resulting in a wasted access cost. Given (ϕ, δ), the transition blocking probability quantifies the risk of encountering the maximum QL and being forced to balk after incurring the access cost.

PROPOSITION 2.1. Given (ϕ, δ), the transition blocking probability, $P_{\phi,K}(\delta)$,

i. is a decreasing function of β.

ii. is zero when the contribution ratio β is one and the reported congestion ϕ is less than K.

Given the total service request rate λ, as the contribution ratio β increases, the QL updates on the POP occur more frequently, leading to a shorter expected δ time interval. Consequently, ϕ reflects the system states more accurately, enabling customers to make informed decisions. Furthermore, the ratio of non-contributors who might join the queue within δ units of time decreases with β. Therefore, the probability of reaching the maximum QL, K, during a shorter δ period decreases. The reduction in $P_{\phi,K}(\delta)$ positively affects customers by decreasing the probability of encountering a full system after passing the first stage and paying the access cost.

When β is one, all customers contribute to updating their observed QL on the POP upon arriving in the system. Wang and Hu (2020) consider this case in an $M/M/1/K$ setting and show that given the most recent QL update ϕ, the expected QL decreases with δ. Note that for the case with $\beta = 1$, δ represents the time interval between two consecutive service requests. Thus, no one enters the system during δ while some customers may receive service and leave the system if δ is sufficiently long. Therefore, if the reported congestion is less than K, δ time units later, the real-time QL is certainly smaller than K. Therefore, when all customers are contributors, and the reported congestion is less than K, no customer will be compelled to balk due to encountering a full system. Conversely, when $\phi = K$ is recorded on the POP, and a service request happens shortly after the QL update, customers might not have enough time to receive service and exit the system. Consequently, the system holds K customers, leading to a non-zero transition

blocking probability. However, if δ is large enough, even when the reported congestion is K, some customers will have enough time to receive service and depart from the system. Therefore, the number of customers in the system drops below K, reducing the transition blocking probability.

2.4.4 Expected Entrance Utility

In the first stage, the expected entrance utility is determined based on (ϕ, δ). However, if a service request occurs long after a QL update, the reported congestion ϕ may not accurately reflect the current state of the queue. Furthermore, a significant δ implies that the interarrival time of contributors is long, and the information on the POP is no longer trustworthy for customers, rendering the first stage ineffective.

Let P_K and L denote the blocking probability and the expected number of customers in an $M/M/1/K$ queue in the steady state, respectively. The following lemma states the condition where the expected congestion and the transition blocking probability converge to P_K and L.

LEMMA 2.2. When δ approaches infinity,

i. the transition blocking probability converges to P_K,

$$P_K = \lim_{\delta \to \infty} P_{\phi,K}(\delta), \quad \delta \geqslant 0,\ 1 \leqslant \phi \leqslant K. \tag{2.5}$$

ii. the expected congestion converges to L,

$$L = \lim_{\delta \to \infty} E[\Gamma(\phi, \delta)], \quad \delta \geqslant 0,\ 1 \leqslant \phi \leqslant K. \tag{2.6}$$

When the interval between a service request and the most recent QL update on the POP approaches infinity, tracking the arrival of non-contributors and the departure of customers who have received service becomes impractical. Consequently, the reported congestion is no longer helpful in estimating the real-time QL or the transition blocking probability. Therefore, the transition blocking probability and the expected num-

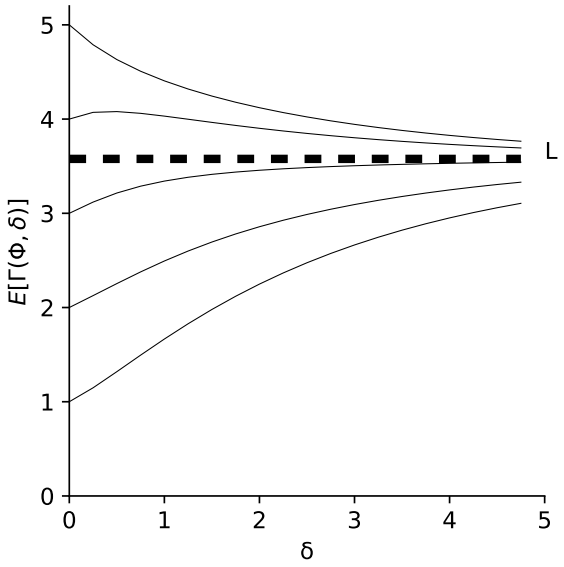


Figure 2.2: Expected congestion w.r.t δ, given the reported congestion is $\phi(T) = 1, 2, \ldots, 5$, $\phi_d = 4$.

ber of customers in the system converges to the steady-state blocking probability in an $M/M/1/K$ queue, independent of the reported congestion ϕ.

Abate and Whitt (1987) investigate the transient behavior of the expected number of customers in an $M/M/1$ queue as a function of δ. They introduce a threshold, referred to as "critical damping," related to the initial number of customers in the system, ϕ, at the beginning of the period δ. They observe that when δ is sufficiently large, the expected number of customers increases with δ for ϕ values smaller than the critical damping. Conversely, for ϕ larger or equal to the critical damping, the expected number of customers decreases monotonically over time.

Figure 2.2 illustrates the expected congestion as a function of δ for various ϕ values in an $M/M/1/K$ system. We observe that as δ exceeds a threshold δ_d, the expected congestion monotonically approaches its steady-state value, L. We define a threshold, ϕ_d, analogous to the concept of critical damping in an $M/M/1$ system. For $\delta > \delta_d$, the expected congestion increases with δ when ϕ is less than ϕ_d, and decreases with δ when $\phi \geq \phi_d$. In Figure 2.2, $\phi_d = 4$. In the rest of the chapter, to make the problem tractable and simplify the analysis, we assume δ_d is zero (numerical results to support this assumption are provided in Appendix A. 14). By applying (2.4) along with Lemma 2.2,

we derive Proposition 2.3, which explains the behavior of the expected entrance utility as δ approaches infinity.

PROPOSITION 2.3. If δ approaches infinity, the expected entrance utility, $U_{enter}(\phi, \delta)$, converges to a fixed value U^*, irrespective of the reported congestion ϕ,

$$U^* = R - h - c\frac{L+1}{\mu} - P_K(R - c\frac{K+1}{\mu}). \tag{2.7}$$

When δ approaches infinity, all customers obtain the same estimate for the expected entrance utility, denoted as U^*, regardless of ϕ recorded on the POP. As illustrated in (2.7), U^* is a function of the expected queue length, L, and the blocking probability, P_K, in the steady state of an $M/M/1/K$ queue.

According to (2.7), the highest value of U^* is achieved when both the steady-state expected QL, L, and the blocking probability, P_K, approach zero, which occurs in services with very low utilization. In contrast, customers have no interest in entering a system with a negative entrance utility. Therefore, the range of variations for the steady-state entrance utility is defined as

$$0 < U^* \leqslant R - \frac{c}{\mu} - h. \tag{2.8}$$

The above inequality highlights that the service reward must exceed the sum of the hassle and waiting costs of an empty system, i.e., $R \geqslant \frac{c}{\mu} + h$.

2.4.5 A Threshold on the Reported Congestion

Given δ, we analyze the behavior of the expected entrance utility for different ϕ values to explore the existence of a threshold on the reported congestion that determines whether a customer perceives a positive expected entrance utility and decides to enter the system. Lemma 2.4 explains the behavior of the expected entrance utility based on δ.

LEMMA 2.4. The expected entrance utility, $U_{enter}(\phi, \delta)$, is a monotonic function of δ for all ϕ values in the range of $[1, K]$.

The expected entrance utility either monotonically increases or decreases with δ, de-

pending on the reported congestion ϕ and the system parameters (the proof is provided in Appendix A. 6). Proposition 2.3 states that for significantly large values of δ, the expected entrance utility eventually converges to U^*, a positive value supported by (2.8). Thus, for a given ϕ, a positive expected entrance utility as δ goes to zero ensures a positive expected entrance utility for any values of δ.

Figure 2.3 illustrates the evolution of the expected entrance utility as a function of δ for different ϕ values, considering two distinct contribution ratios: $\beta = 0.01$ in Panel (a) and $\beta = 0.99$ in Panel (b). As depicted, for $\phi = 3, 4, 5$, the expected utility is positive at zero and remains so for all δ values. To calculate the expected entrance utility as δ

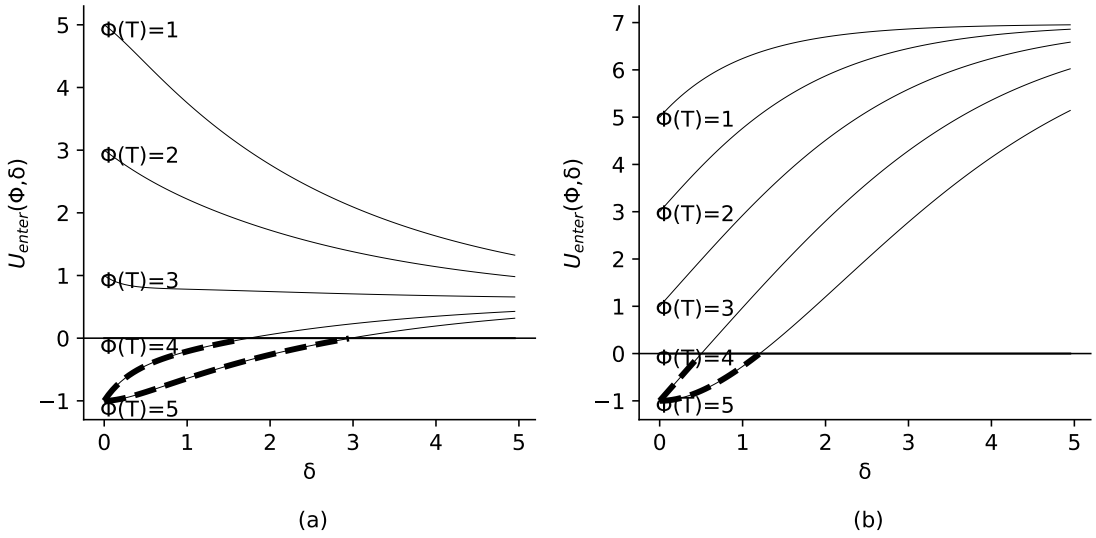


Figure 2.3: Expected entrance utility w.r.t. δ, given the reported congestion is $\phi(T) = 1, 2, \ldots, 5$. Parameters: $\lambda = 1.5$, $\mu = 1$, $R = 10$, $c = 2$, $h = 1$, (a) $\beta = 0.01$, and (b) $\beta = 0.99$

approaches zero, we first need to understand how the expected congestion behaves in this limit, as provided by Lemma 2.5.

Lemma 2.5. If a customer requests to enter the system immediately after a QL update on the POP, i.e., when the time interval δ goes to zero, their expected congestion goes to the reported congestion:

$$\lim_{\delta \to 0} E[\Gamma(\phi, \delta)] = \phi, \quad \delta \geqslant 0,\ 1 \leqslant \phi \leqslant K. \tag{2.9}$$

As δ approaches zero, non-contributors have insufficient time to enter the system, and customers in the queue do not exit. Consequently, the QL remains constant, matching the reported congestion. We employ Lemma 2.5 to determine the maximum reported congestion that generates a positive entrance utility when δ tends to zero, denoted by $\hat{\phi}$. After rewriting (2.4), we have

$$\lim_{\delta \to 0} U_{enter}(\phi, \delta) = R\Big(1 - P_{\phi,K}(0)\Big) + \frac{c}{\mu}\Big(K+1\Big)P_{\phi,K}(0) - \frac{c}{\mu}\Big(E[\Gamma(\phi,0)]+1\Big) - h. \quad (2.10)$$

Immediately after a QL below K is reported, the probability of the system reaching its maximum capacity, K, is zero. According to Lemma 2.5, as δ approaches zero, the expected congestion converges to the reported congestion, i.e., $E[\Gamma(\phi,0)] = \phi$. Therefore, the largest reported congestion that ensures a positive entrance utility when δ goes to zero is,

$$\hat{\phi} = \left\lfloor \frac{\mu}{c}(R-h) - 1 \right\rfloor, \ 1 \leqslant \hat{\phi} < K. \quad (2.11)$$

Given (2.8), Proposition 2.3, and Lemma 2.4, if the most recent reported congestion on the POP is less than or equal to $\hat{\phi}$, customers are guaranteed a positive entrance utility for all δ values, allowing them to pass the first stage and enter the second stage successfully. Proposition 3.7 discusses the case of $\hat{\phi} < \phi \leqslant K$.

Proposition 2.6. Given the most recent QL update, if $\hat{\phi} < \phi \leqslant K$, there exists a corresponding positive value of $\hat{\delta}(\phi)$ such that

$$U_{enter}(\phi, \delta) = \begin{cases} < 0, & \text{if } 0 \leqslant \delta < \hat{\delta}(\phi) \\ \geqslant 0, & \text{if } \delta \geqslant \hat{\delta}(\phi) \end{cases}. \quad (2.12)$$

If the reported congestion exceeds $\hat{\phi}$, the expected entrance utility becomes negative when δ approaches zero. For cases where $\hat{\phi} < \phi \leqslant K$, customers will perceive a negative entrance utility when δ goes to zero. However, as δ increases and reaches $\hat{\delta}(\phi)$, the expected entrance utility transitions from negative to zero. The expected entrance utility becomes positive for $\delta > \hat{\delta}(\phi)$, prompting customers to enter the system.

Using (2.4) and Proposition 3.7, we can determine $\hat{\delta}(\phi)$ as the unique solution to

$U_{enter}\left(\phi, \hat{\delta}(\phi)\right) = 0$ by solving:

$$R - h - \frac{c}{\mu}\left(E[\Gamma\left(\phi, \hat{\delta}(\phi)\right)] + 1\right) - P_{\phi,K}\left(\hat{\delta}(\phi)\right)\left(R - c\frac{K+1}{\mu}\right) = 0, \quad 1 \leqslant \phi \leqslant K. \tag{2.13}$$

In Figure 2.3, both panels show $\hat{\phi} = 3$. When $\phi = 1$, 2, or 3, which are smaller than or equal to $\hat{\phi}$, the expected entrance utility is positive for all values of δ. However, for $\phi = 4, 5$, values greater than $\hat{\phi}$, the expected utility $U_{enter}(\phi, \delta)$ is negative as δ goes to zero. The dashed lines indicate the time interval when the expected entrance utility is negative, resulting in no entries into the system. When δ surpasses $\hat{\delta}(\phi)$, the entrance utility turns positive, prompting customers to pass stage one and observe the QL.

2.5 Social Welfare

Social welfare pertains to the expected net benefit of all customers. A customer's decision during each stage determines one of three possible outcomes. If a customer leaves at the first stage, there is neither a loss nor a gain. Alternatively, if they enter the first stage but balk at the second stage, they have paid the access cost without receiving any gain. Finally, if a customer enters the first stage and proceeds to the second stage, they will receive the reward but must also incur the access and waiting costs.

Given $\mu = 1$, the load factor equals the service request rate λ. Using Proposition 3.7, we show that there exists a threshold $\hat{\lambda}$ for the service request rate, which determines the behavior of social welfare as a function of β.

Proposition 2.7. For a given set of cost parameters, the social welfare is:

i. a monotonically increasing function of the contribution ratio β, if $\lambda \leqslant \hat{\lambda}$.

ii. a unimodal function of the contribution ratio β, if $\lambda > \hat{\lambda}$.

In a low-demand system where $\lambda \leqslant \hat{\lambda}$, social welfare increases with the contribution ratio β due to the decrease in waiting time. As β increases, customers make more informed decisions using frequently updated ϕ on the POP. Therefore, they avoid entering the system whenever there is a long queue. This strategic and state-dependent entrance reduces customers waiting times as they enter the system only when it is not excessively congested. Since the system's request rate is below $\hat{\lambda}$, indicating low demand, even though the throughput reduces due to some customers leaving to avoid congestion, a significant fraction of customers will find time slots with low congestion and will be served. Consequently, the reduced waiting time dominates the decrease in throughput, increasing social welfare with β.

In a high-demand system where $\lambda > \hat{\lambda}$, there exists a local maximum for the contribution ratio, denoted as $\beta^*(\lambda)$. When $\beta \leqslant \beta^*(\lambda)$, customers lack sufficient information to make precise decisions due to infrequent ϕ updates on the POP. Additionally, they are aware of the large service request rate, so at their service request time, they may conser-

vatively choose to leave at the first stage to avoid wasting the access cost, even if there is available space in the queue. As a result, although the system's throughput decreases, the customers who enter the system experience significantly shorter waiting times, leading to an increase in social welfare for the contribution ratio up to $\beta^*(\lambda)$. Conversely, when $\beta > \beta^*(\lambda)$, customers have greater confidence in their decisions based on the frequent updates of ϕ on the POP. Therefore, they are not conservative when it comes to paying the access cost. Furthermore, since the service is either popular or necessary, indicated by a significant service request rate, customers seize every opportunity to join the system if there is an empty space in the queue. As a result, the queue begins to fill up shortly after each service completion. Consequently, more customers are served, increasing the system's throughput. However, customers spend more time waiting in a crowded queue, considerably increasing the waiting time. Therefore, social welfare decreases for the contribution ratio larger than $\beta^*(\lambda)$. Figure 2.4 illustrates the social welfare for different values of the service request rate. In Panel (a), where the service request rate is less than $\hat{\lambda}$, the social welfare increases with β. However, in Panel (b), where the service request rates exceed $\hat{\lambda}$, the social welfare is a unimodal function of β. In both panels, social welfare decreases with the service request rate.

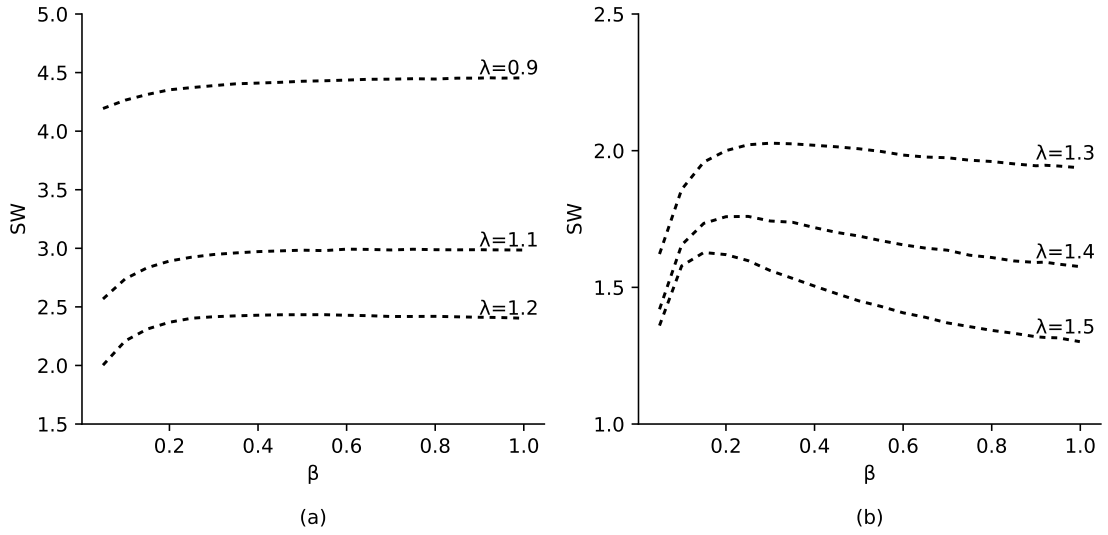


Figure 2.4: Social welfare w.r.t the contribution ratio β, given the demand rate, (a) $\lambda \leqslant \hat{\lambda}$, and (b) $\lambda > \hat{\lambda}$.

2.5.1 Access Cost

Access cost is a key factor that influences a customer's decision regarding entering or leaving the system in the first stage. When there is no access cost, $h = 0$, our model is equivalent to an observable queue. The social welfare in an observable queue, with the maximum QL of K, denoted as S_O, is given by Hassin and Haviv (2003) as

$$S_O = \lambda R \frac{1-\rho^K}{1-\rho^{K+1}} - c\Big[\frac{\rho}{1-\rho} - \frac{(K+1)\rho^{K+1}}{1-\rho^{K+1}}\Big], \tag{2.14}$$

where $\rho = \lambda/\mu$. Figure 2.5 illustrates the changes in social welfare with respect to the contribution ratio β for various access cost values. In Panel (a), as an example of a low-demand system, the maximum social welfare is achieved without the access cost, where $h = 0$. We also observe that social welfare declines with access costs. This decline is attributed to customers' reluctance to make extra effort for a service that is either less popular or not deemed essential. As a result, with the increase in the access cost, more customers leave during the first stage, reducing the system's throughput and the expected waiting time for those who join the queue. Since the service is not in high demand, the expected waiting time reduction outweighs the throughput reduction, and the social welfare increases with β, for all values of the hassle cost. In Panel (b), depicting a high-demand system, since the service offered is either highly popular or deemed essential, for small access cost values relative to the service reward, such as $h = 0.3$ and 0.6, a substantial proportion of customers pass the first stage and enter the system. However, a considerable fraction of them are compelled to balk during the second stage due to the system holding maximum QL, resulting from the high demand. In this dynamic, the social welfare increases with β, attributed to a reduction in the expected waiting time experienced by those customers who join the system during the second stage. In cases where the access cost is higher, exemplified by $h = 0.9$ and 1.2, the social welfare demonstrates unimodal behavior concerning the contribution ratio for the same reason discussed in Section 2.5.

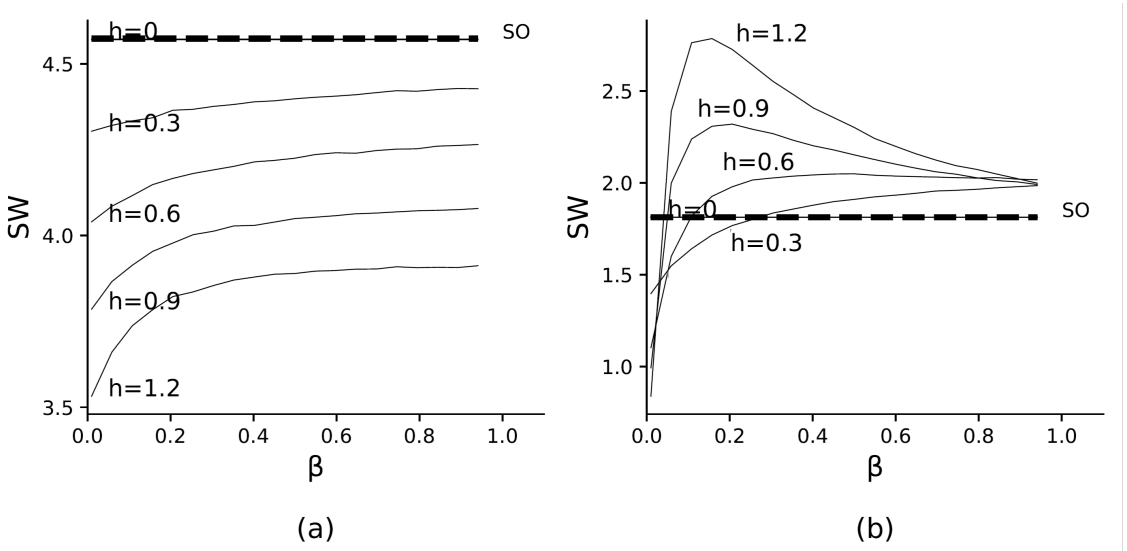


Figure 2.5: Social welfare w.r.t the contribution ratio, β, given the access cost Parameters: $\mu = 1$, $R = 10$, $c = 1$, (a) $\lambda = 0.9$, and (b) $\lambda = 1.5$

2.6 Restricted Access to a Crowdsensing Platform

This section examines a scenario where access to a crowdsensing platform is restricted. This restriction can be due to subscription-based models, quality control, exclusive access for trusted entities, resource management considerations, and compliance with regulations or service optimization in specific areas. Thus, only a fraction of customers, denoted by α, are granted access, while the remaining $(1-\alpha)$ have no prior access to the shared QL on the crowdsensing platform. In a setup with restricted access, only α percentage of customers face two stages, the enter-or-leave and the join-or-balk stages, before receiving the service. Among the customers with access to the crowdsensing platform, β ratio contributes to updating the QL, i.e., $0 \leqslant \beta \leqslant \alpha \leqslant 1$. The remaining $(1-\alpha)$ percentage deals with only one stage, which is the join-or-balk stage.

When access to the crowdsensing platform is restricted, the arrival rate of customers who directly enter the system is $(1-\alpha)\lambda$. Among the remaining customers with access to the crowdsensing platform, only those expecting a positive entrance utility pass the first stage and enter the system. Consequently, the effective entrance rate of customers into the system, denoted as $\lambda_{s\alpha}$, is

$$\lambda_{s\alpha} = (1-\alpha)\lambda + \alpha\lambda \left(p\big(U_{enter}(\phi,\delta)\big) \geqslant 0\right). \tag{2.15}$$

Proposition 2.8. The effective entrance rate of customers to the system, $\lambda_{s\alpha}$, decreases with the access level to the crowdsensing platform, α.

Among customers with access to the shared QL information, those who expect a negative entrance utility leave the system. Meanwhile, the fraction of $(1-\alpha)$ of customers without prior shared QL information enter the system, where their proportion decreases with α. Hence, the total effective entrance rate into the system decreases with α. Furthermore, the effective entrance rate determines how many customers visit the system, referred to as foot traffic. According to Proposition 2.8, service providers can achieve foot traffic goals by managing access to the crowdsensing platform.

We next explore the social welfare implications of a crowdsensing platform with restricted access. Social welfare for $(1-\alpha)$ percentage of customers without prior access to the shared QL information, denoted as $SW_{(1-\alpha)}$, equals

$$SW_{(1-\alpha)} = (1 - P_B)(R - h - cN) - P_B h, \tag{2.16}$$

where P_B denotes the probability of having K customers in the queue, and N represents the expected number of customers in the queue when the crowdsensing platform is not publicly available. The total social welfare of customers with and without access to the crowdsensing platform is defined as follows

$$SW = SW_\alpha + SW_{(1-\alpha)}, \tag{2.17}$$

where SW_α represents the social welfare of customers with access to the crowdsensing platform as a function of α, and SW_α refers to the social welfare of customers without access to the platform.

Section 2.5 elaborated on SW_α for $\alpha = 1$, where the crowdsensing platform is publicly available and everyone has access to it, i.e., $SW = SW_\alpha$. On the other hand, when $\alpha = 0$, customers have no prior information on QL before entering the system since $\beta = 0$, and no one shares the QL information. Thus, our proposed model in section 3.4 is equivalent to an observable $M/M/1/K$ queue with a service reward of $R - h$.

Proposition 2.9. Social welfare increases with the access level to the crowdsensing platform, α.

Increasing the access level to a crowdsensing platform, α, yields a dual positive effect. When more customers check the shared QL information before joining the queue, it reduces the chance of encountering an overcrowded queue and minimizes the waste of the hassle cost. Secondly, with fewer customers directly entering the queue, the information on the crowdsensing platform gains increased reliability. This reliability results from fewer customers entering the queue between the arrivals of two data contributors, reducing the difference between the reported congestion and the real-time QL.

The described dynamics above are visually represented in Figure 2.6, where the plots depict social welfare based on the contribution ratio for various values of α. This figure illustrates how controlling access levels to the crowdsensing platform impacts social welfare and provides insights into the system's efficiency under different contribution ratios β. Both panels show that the system with the highest access rate, $\alpha = 1$, provides the highest social welfare. We also observe that the optimal contribution ratio, which maximizes social welfare, decreases with α. This suggests that as more customers access the crowdsensing platform, a smaller percentage of them is needed to share QL information to achieve optimal social welfare.

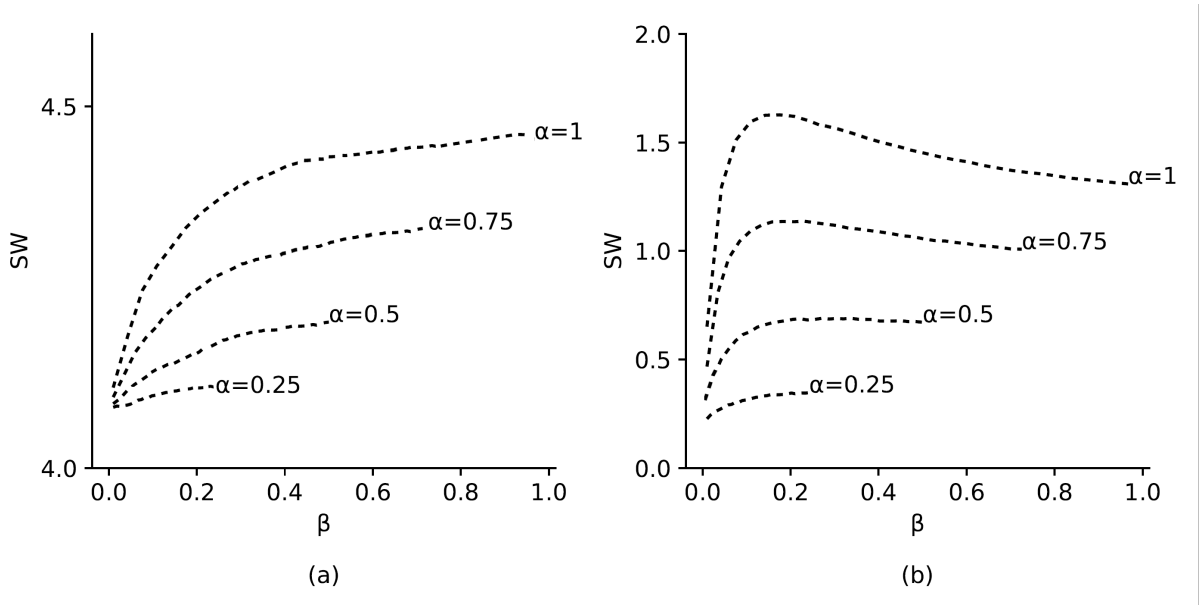


Figure 2.6: Social welfare at the contribution ratio β, given the access rate $\alpha = \{0.25, 0.5, 0.75, 1\}$
. Parameters: $\mu = 1$, $R = 10$, $c = 1$, $h = 1$, (a) $\lambda = 0.9$, and (b) $\lambda = 1.5$

2.7 Chapter Summary

Crowdsensing platforms are popular tools for collecting data on a large scale across diverse domains such as urban planning, environmental monitoring, and disaster response. Data contributors directly influence the quality of the collected data and its impact on the system. In this chapter, we study the application of a crowdsensing platform in operational management using a single-server queue. Initially, we assumed the platform was available to the public for free, but only a fraction of customers contributed data by sharing their observed QL and the timestamps of their observations. This information helps future customers determine the expected utility of receiving the service, allowing them to avoid wasting effort by entering a system with the maximum QL.

We examined the impact of QL contributions by a subset of customers on social welfare. The findings show that changes in social welfare, relative to the ratio of data contributors, depend on the load factor. Social welfare improves with the ratio of data contributors in less popular services characterized by a low load factor. This is because more frequent QL updates on the crowdsensing platform help customers avoid entering the system when the queue is congested and join when it is less crowded. In popular and high-demand services, there is an optimal fraction of data contributors that maximizes social welfare. Beyond this optimal threshold, extended wait times within the longer queue negatively affect social welfare. When the ratio of the data contributors exceeds the optimal level, customers are more willing to trust the QL recorded on the POP and pay the access cost to enter the system, even if a congested queue has been reported. This continuous influx of customers prevents the queue from depleting, resulting in an immediate refill after each service completion.

This study demonstrated that less popular service providers benefit from continuously encouraging customers to contribute data. In contrast, for popular systems, service providers should focus solely on maintaining the optimal fraction of data contributors rather than exerting additional effort to motivate universal data contribution. Finally, the study showed that if service providers limit access to a crowdsensing platform, they

must anticipate a subsequent reduction in foot traffic and social welfare. To mitigate the reduction in social welfare, they should ensure that a higher proportion of the customers with access to the platform contribute data.

Chapter 3

CHAPTER THREE: ROUTE CHOICE USING CROWD-GENERATED TRAVEL TIME INFORMATION

3.1 Chapter Overview

Smartphones and onboard vehicle sensors are increasingly used to gather crowd-generated travel data (CTD), providing a cost-effective alternative to traditional data collection methods such as stationary sensors and cameras. However, the reliability of CTD depends on travelers' willingness to share their data. This chapter examines how CTD can be strategically utilized to optimize network-wide travel times while promoting equitable information access and preserving user privacy. Specifically, we highlight the role of two factors in decision-making and the accuracy of the travel time estimation: the *contribution ratio*, which represents the proportion of travelers sharing their data, and the *observation window*, which defines the period over which CTD is collected and analyzed. We consider a transportation network with two routes: one with fixed travel times, referred to as the *safe road*, and another with variable, stochastic travel times, called the *random road*, which is modeled as an $M/M/1$ queue. The expected travel time on the random road is estimated using the central limit theorem, which is applied to a sample of CTD collected within the observation window. A choice model is proposed to calculate the probability of travelers selecting each road. The study examines how variations in contribution ratios and observation windows influence route choices and overall travel times. An optimal contribution ratio is identified that minimizes overall travel time in a fixed-demand setting. Furthermore, it is demonstrated that the optimal proportion of travelers selecting the random road increases with the square root of the travel time on the safe road. Additionally, the impact of risk on route choice is explored by incorporating a risk aversion function linked to the contribution ratio. To tackle this risk, a scenario is studied where travelers can pay to access real-time travel information on the random road. We establish a threshold on the contribution ratio that determines when travelers shift from paying the access cost to using free CTD, which is a function of the access cost and the variability of travel time on the random road. Finally, it is demonstrated that introducing a reward for travelers who share their data can lead to a further reduction in overall travel time compared to the base model, where the reward is zero.

3.2 Introduction

Access to travel-time information enables travelers to make informed routing decisions, avoid delays, and reduce the stress of congestion. Traditional data collection methods rely on fixed roadside sensors such as cameras and loop detectors, which are costly to maintain and limited in coverage throughout a network. The emerging alternative is crowd-generated data collection, known as crowd-sensing, which takes advantage of the widespread adoption of smartphones and vehicle sensors to obtain information on the entire network using shared travel time experiences. Platforms like Google Maps and Waze collect smartphone data, such as speed, route trajectory, and travel time to estimate congestion and travel times. Transportation corporations like GeoTab and HERE leverage crowd-generated travel data (CTD), advanced analytics, and real-time inputs from connected devices, such as GPS-enabled vehicles and mobile apps, to improve the accuracy and reliability of travel information (Geotab 2017, Here 2020, 2022). Original Equipment Manufacturers in the auto industry also equip vehicles with more onboard sensors to provide real-time updates and enable enriching transportation networks with accurate and timely data (Raza et al. 2020, Bridgelall and Tolliver 2020). The success of crowd-sensing travel time data depends on the active participation of travelers in data sharing, which determines the accuracy and reliability of the information.

Growing concerns about user privacy have led companies like Apple and Google to implement strict controls over data sharing. For instance, iPhone users can select from multiple privacy options, including "Don't Allow," which denies all access; "Allow Once," granting temporary access; "Allow While Using the App," permitting access only during app usage; and "Allow Always," enabling continuous background access. While these measures empower users to protect their privacy, such privacy protections reduce the volume and consistency of shared data. We investigate two factors that influence the effectiveness of CTD: the "contribution ratio", the proportion of travelers who agree to share data, and the "observation window", the duration over which data is collected. A higher contribution ratio reduces the uncertainty in travel time predictions by increasing

the collected data volume. At the same time, the observation window also affects the accuracy of forecasts; a window that is too short may fail to capture temporary delays, and an excessively long window introduces outdated data, leading to unreliable predictions. This study addresses two key questions for effectively managing transportation systems by optimizing the collection of CTD. First, the goal is to determine the existence of an optimal interior contribution ratio below one that minimizes network-wide travel time. Second, if such an optimal threshold exists, what practical strategies, policies, or incentives can be implemented to encourage users to contribute data at the optimum level while ensuring fairness and compliance? By answering these questions, our work seeks to provide managerial insights to balance the collection of CTD and network efficiency.

A seminal study by Yang (1998) introduces a sensor-based framework that provides real-time travel information for various links within a transportation network, enabling travelers to select user-optimal routes. A key finding from the study is that increasing the proportion of travelers with access to travel-time information does not always lead to reduced social costs from a planner's perspective. This is because access to real-time travel data can cause many travelers to converge on the same routes and departure times, a phenomenon known as the Informational Braess' Paradox (IBP) (Acemoglu et al. 2018), where additional information paradoxically worsens congestion and delays. To mitigate the effects of IBP, information design has been used to strategically control the disclosure of information in transportation networks. Information design operates on two levels: (1) controlling the type and amount of information to be distributed and (2) classifying travelers to offer personalized travel information, ensuring that each traveler receives the most relevant and precise data tailored to their conditions (Tavafoghi and Teneketzis 2019, Acemoglu et al. 2016). However, information design still relies on real-time travel data as input, which is often impractical to collect and raises equity concerns by limiting user access to travel information. To address these issues, CTD provides a feasible alternative, with research highlighting its reliability and potential for effective application (Tavafoghi and Teneketzis 2019, Acemoglu et al. 2016). However, previous studies overlook the role of data contributors and fail to examine how adjusting key elements of a CTD platform, the

contribution ratio and observation window, can optimize transportation networks without manipulating data or imposing restrictive access controls. These two elements directly influence the quantity and consistency of data collected from users, which determines the reliability of congestion estimates. Investigating this gap explores whether IBP occurs in a crowd-generated, data-driven setup and examines the role of adjusting the contribution ratio and observation windows in addressing it. Additionally, it assesses the effective mitigation of congestion in such a transportation network, particularly when travelers selfishly choose the shortest route based on the available information.

This study focuses on transportation networks that rely on CTD to minimize network-wide travel time while considering users' data-sharing preferences and the reluctance of some travelers to share their data. We consider a network with two options: one with a fixed and known travel time, termed the "safe road," and the other with a stochastic travel time, referred to as the "random road" (Verhoef et al. 1996, Liu et al. 2018, Liu and Yang 2021). All travelers have free access to CTD, but only a subset chooses to share their travel data. We examine the impacts of free access to CTD on route choices, focusing on the contribution ratio and observation window. We investigate whether the conclusion of Yang (1998) holds in a crowd-generated data-driven traffic network. Specifically, we analyze whether increasing the contribution ratio decreases overall travel time and evaluate how traveler risk levels, particularly their sensitivity to variability in travel times on the random road, affect their decisions. We also examine the scenario where travelers can pay to access real-time travel information and analyze how this option interacts with the contribution ratio and traveler behavior. Additionally, we investigate the consideration of a contribution reward for those who contribute travel data, aiming to enforce the optimal contribution ratio.

This study's results demonstrate that while higher contribution ratios enhance traveler confidence in route selection, excessively high contribution ratios at the network level can lead to congestion, ultimately increasing overall travel times. We identify the optimal contribution ratio that minimizes the overall travel time in a fixed-demand setting. This finding emphasizes the importance of maintaining the contribution ratio below one to

ensure system-wide efficiency. Additionally, it underscores the balance between optimizing network performance and respecting travelers' privacy, showing that the network can still operate efficiently even when not all travelers are comfortable sharing their travel data. The optimal proportion of travelers selecting the random road decreases in proportion to the square root of the ratio between the free-flow travel time on the random road and the fixed travel time on the safe road. To minimize the network travel time, we introduce a threshold for the minimum observation window, ensuring that travelers select the random road only when its expected travel time is, on average, shorter than that of the safe road. Incorporating risk into the model reveals that the probability of selecting the random road consistently decreases compared to the base model, where travelers are assumed to be risk-neutral. We also analyze a scenario where real-time travel information on the random road is available after paying an access fee. In this case, we find that an increase in the access fee leads travelers to require a lower contribution ratio to prefer free CTD over paid information. Conversely, lower access fees make paying for precise travel information more attractive. Additionally, a greater variance in travel time increases the required contribution ratio for travelers to opt for free CTD. Finally, we demonstrate the existence of an optimal contribution reward tied to network travel demand. Implementing this reward reduces both the optimal contribution ratio and overall travel times. In addition, introducing a contribution reward further decreases overall travel time compared to the base model, where no reward is provided, offering a practical strategy to improve network performance.

The remainder of the chapter is organized as follows. Section 3.3 provides an overview of the related literature. Section 3.4 presents the choice model, a detailed explanation of its underlying assumptions and components. Section 3.5 explores extensions to the base model, including the incorporation of traveler risk levels, the option to pay for real-time travel information, and incorporating incentives for travelers to contribute their data. Finally, Section 3.6 summarizes the findings.

3.3 Literature Review

This literature review explores the impact of access to travel information and strategies for optimizing traffic flow using real-time data and CTD. It then discusses the challenges and opportunities posed by CTD, emphasizing concerns related to data accuracy, reliability, and privacy, and how this research addresses those gaps.

3.3.1 Optimizing Traffic Flow with Real-Time Information

Real-time travel time information helps travelers plan their trips more efficiently, but its impact on overall network performance can be complex. Yang (1998) explores a traffic assignment problem with two traveler classes: those with access to real-time travel information and those making stochastic decisions without such information. They find that while real-time congestion information benefits individual travelers, it can worsen total network travel time if the proportion of informed travelers exceeds an optimal threshold. This occurs because access to real-time information leads travelers to converge on the same routes and departure times, increasing overall congestion. This effect is also observed in dynamic routing, where travelers adjust their routes mid-journey based on real-time updates. Mahmassani and Jayakrishnan (1991) assess the effects of real-time in-vehicle navigation data and find that equipping 25% to 50% of travelers with navigation systems yields the best performance.

Pricing strategies and traveler preference analysis offer practical solutions to mitigate congestion and optimize network performance. Assigning costs, such as tolls or parking fees, makes travelers cautious of the impact of their travel choices on the network. Verhoef et al. (1996) find that flat tolls, combined with real-time travel time information, enhance system efficiency. Yu et al. (2023) demonstrate that responsive pricing, adapting to real-time traffic conditions, consistently outperforms fixed pricing in reducing congestion, particularly as the accuracy of travel information increases. Travelers' diverse risk tolerances and preferences further influence congestion. Liu et al. (2018) show that accurate travel information benefits risk-averse individuals but may exacerbate delays

for risk-seekers. Yu et al. (2023) compare responsive pricing, where tolls vary based on real-time or predicted traffic conditions, with fixed pricing schemes. They show that responsive pricing can raise welfare and lower travel costs, especially as the quality of travel information improves. However, the performance gap between the two pricing approaches remains small even under high uncertainty, highlighting the role of information quality and system variability in pricing strategy effectiveness.

Alternative solutions focus on optimizing traffic flow and network efficiency by controlling access to real-time information through adjustments in timing, accuracy, and level of detail in distributed travel information. Zhang and Nie (2018) show that guiding just 10% of vehicles toward system-optimal routes can significantly improve network performance. Similarly, Acemoglu et al. (2018) find that restricting users' knowledge to specific road segments prevents congestion resulting from access to real-time information. Spana and Du (2022) propose an optimal information perturbation strategy to mitigate network congestion caused by selfish routing decisions. They demonstrate that strategically altering traffic information can improve overall system performance. Their analysis focuses on a homogeneous perturbation scheme applied uniformly across all network links, balancing the trade-off between enhancing system efficiency and minimizing the loss of user optimality due to perturbation.

We have so far presented traffic demand management strategies, such as controlling real-time information, imposing tolls, and accounting for behavioral differences, that optimize network performance using real-time travel data. However, CTD takes a decentralized approach, enabling dynamic traffic adjustments without imposing external controls or requiring micro-level analysis of traveler behavior. The next section explores how CTD enhances congestion mitigation and routing efficiency.

3.3.2 Traffic Optimization With Crowd-Generated Travel Data

Crowd-generated travel data leverages mobile devices, such as smartphones and on-board sensors, to provide a cost-effective solution for traffic management. The widespread deployment of sensors, wireless communication modules, and data storage devices in vehicles

enables the development of robust platforms that are not constrained by device energy consumption or infrastructure requirements (Ganti et al. 2011, Al-Sakran et al. 2015). Unlike the strategies discussed in Section 3.3.1, which focus on controlling the flow of real-time information, CTD allows for dynamic real-time adjustments that optimize traffic conditions. These systems regulate traffic flow by providing real-time information, including travel updates, road incidents, and unexpected weather conditions. They also estimate travel times and alleviate congestion by guiding commuters to more efficient routes. Additionally, CTD can predict user behavior and public transit passenger flow (Zhang et al. 2016), detect traffic anomalies, and suggest alternative routes to avoid congestion (Pan et al. 2013). Furthermore, CTD supports dynamic rerouting, enabling travelers to select the shortest path, thereby improving traffic flow and reducing congestion (Yu et al. 2021). This data can also enable customized routing options by considering road conditions, driver expertise, and the current state of the network (Abdelrahman et al. 2020).

Crowd-generated travel information is transmitted through two primary communication modes: vehicle-to-vehicle and vehicle-to-server. These network topologies have proven highly effective in managing traffic congestion through rerouting solutions and dynamic traffic assignment strategies (Yu et al. 2021). For instance, Pan et al. (2016) explore vehicle-to-vehicle rerouting strategies to mitigate congestion, demonstrating that real-time traffic predictions based on CTD can proactively address congestion. Patni and Jain (2015) propose a vehicle-to-vehicle-based system that analyzes vehicle speeds and transmitted data to enable real-time traffic adjustments. Their results show improved traffic flow by alerting drivers to potential breakdowns, which facilitates better decision-making and alleviates congestion. In smart city initiatives, Gomides et al. (2020) use a vehicle-to-server network to analyze traffic data, enabling vehicles to evaluate alternative routes based on distributed information. Their findings demonstrate a reduction in traffic congestion and improved flow by providing vehicles with real-time traffic information, enhancing decision-making and optimizing route selection. Ali et al. (2021a) propose a framework that employs CTD within a vehicle-to-server network to direct vehicles to the

fastest routes. This system ensures secure data handling without sharing information between vehicles. It evaluates vehicle distribution across routes, considers route capacities, and assigns the most efficient path based on estimated arrival times. Simulation results show that this dynamic routing approach significantly enhances traffic flow and reduces congestion, outperforming vehicle-to-vehicle systems by an average of 92%. Moreover, it reduces computational load and network traffic compared to traditional centralized systems (Pan et al. 2016).

3.3.3 Optimizing Crowd-Generated Travel Data Collection

While CTD improves traffic management by enabling real-time dynamic adjustments, its effectiveness depends on the accuracy, reliability, and spatial-temporal coverage of the collected data (Wang et al. 2017). A well-designed data collection system requires a sufficient and diverse pool of contributors to enhance representativeness and minimize biases to provide comprehensive and accurate insights for decision-making (Girolami et al. 2021). To ensure data sufficiency, Huang et al. (2018) propose a flexible sweep coverage model for mobile sensing devices carried by participants. This model guarantees comprehensive coverage of the designated area while optimizing resource allocation and minimizing the total costs paid to contributors. To ensure the reliability of the collected data, Wang et al. (2017) develop a reputation system to validate shared data, complemented by a payment system to incentivize contributors. Performance evaluation demonstrate the effectiveness of these systems. In practice, leveraging for-hire vehicles, such as taxis, rideshare vehicles, and delivery fleets, has emerged as a scalable solution for urban sensing. These vehicles can efficiently collect data across large areas, overcoming the hardware and software limitations of private vehicles. Ding et al. (2021) suggest that equipping for-hire vehicles with sensing devices provides the dual benefits of mobility and extensive coverage for travel data collection. However, integrating these data collection tasks with vehicles' primary job of serving orders is challenging. To address this challenge, they propose a framework that enables vehicles to make coordinated routing decisions, balancing the need to serve orders with the data collection task.

CTD collection has primarily focused on increasing data contributors to ensure a consistent flow of data and geographical coverage. However, existing studies have not examined how the ratio of data contributors impacts traveler decisions and congestion levels. We analyze the contribution ratio as a critical factor in balancing data accuracy and system efficiency while highlighting travelers' privacy concerns in data sharing. Our findings reveal that an optimal contribution level exists, beyond which additional data sharing may not further reduce congestion and, in some cases, may even increase total travel time. Furthermore, this study identifies an unexplored strategy: optimizing the proportion of travelers who share travel time information to enhance system performance while ensuring equitable access to crowd-generated travel information. We enforce this optimal contribution ratio through a contribution reward. Additionally, we introduce the observation window to assess how the temporal coverage of the collected data influences its reliability and usability for real-time decision-making. We recognize that relying solely on historical data can be misleading due to unforeseen conditions and that real-time data collection is often infeasible. We employ queuing theory to model the interaction between data sufficiency, traveler decision-making, and system-wide efficiency, ultimately determining the ideal conditions under which CTD can minimize congestion and maximize social welfare.

3.4 Model

3.4.1 Preliminaries

We examine a network of a single source-sink pair connected by two routes, as illustrated in Figure 3.1.

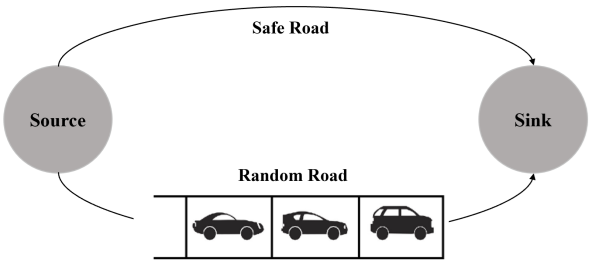


Figure 3.1: Schematic of the network.

A homogeneous Poisson process generates travel demand from the source to the sink at a rate of Λ travelers per time unit. Travelers choose between a "safe road" with a fixed travel time τ and a "random road" with congestion-sensitive travel time. The random variable T represents the travel time on the random road, with mean $M_T(\lambda) = \mu$ and variance $V_T(\lambda) = \sigma^2$. The rate and proportion of travelers choosing the random road are denoted by λ and α, respectively, where $\lambda = \alpha\Lambda$.

To capture the effect of congestion, the random road is modeled as a single-server queue, where the queue dynamics reflect the delay caused by traffic buildup. In this analogy, the server represents the random road, and travelers are the customers waiting for the service of traveling through it. Assuming a homogeneous Poisson arrival rate of $\lambda = \alpha\Lambda$, travelers arrive randomly but at a constant average rate. Travelers access the random road based on arrival orders following a first-come-first-served service priority rule.

In steady state, the expected waiting time due to congestion, as well as the mean and variance of the travel time on the random road, are the same for all travelers, based on the properties of an $M/M/1$ queue (Larson and Odoni 1981). This uniformity occurs

because the system reaches equilibrium, meaning that every traveler experiences the same stochastic conditions regardless of when they enter the system. As a result, variations in travel times are uniformly distributed among all travelers. The expected travel time on the random road, μ, consists of the expected free-flow travel time, denoted by $1/\theta$, and the additional expected delay due to congestion, denoted by W_q, as shown below

$$\mu = \frac{1}{\theta} + W_q. \tag{3.1}$$

Given $M/M/1$ properties described by Larson and Odoni (1981), we have

$$W_q = \frac{\lambda}{\theta(\theta - \lambda)}, \tag{3.2}$$

$$M_T(\lambda) = \mu = \frac{1}{\theta - \lambda}, \tag{3.3}$$

and

$$V_T(\lambda) = \sigma^2 = \frac{1}{(\theta - \lambda)^2}. \tag{3.4}$$

The congestion level in queuing theory, equivalent to the volume-to-capacity ratio on the random road, is defined as $\rho = \lambda/\theta$. Here, ρ must be less than one, i.e., $\lambda < \theta$. Otherwise, the number of travelers on the random road will tend to infinity, and the queue on the random road will not reach a steady-state condition (Larson and Odoni 1981). A weak stability condition for the queue on the random road is $\Lambda < \theta$.

3.4.2 Crowd-Generated Travel Data

We assume that a fraction of travelers, denoted by $\beta > 0$, agree to contribute their travel information when using the random road. This fraction, referred to as the contribution ratio, reflects the portion of travelers participating in data sharing. The expected travel time on the random road is derived from the sample of travel times contributed by previous travelers. The sample size, represented by the random variable S, indicates the number of travelers within a given observation window who have chosen the random road and

agreed to share their travel data. The observation window, denoted by δ, defines the time period during which a moving average analysis of collected travel times is performed and is treated as an exogenous parameter.

Given that travel demand follows a Poisson distribution, the sample size S collected within the observation window also follows a Poisson distribution with a mean of $\bar{S}$. The expected sample size is defined as

$$\bar{S} = \alpha\Lambda\beta\delta, \tag{3.5}$$

because out of $\lambda = \alpha\Lambda$ travelers who choose the random road within the observation window δ, a ratio β contributes the travel data.

For a given realization s of the sample size random variable S, if $\{t_1, t_2, ..., t_s\}$ are the random samples of the CTD drawn from a population with a mean of μ and a variance of σ^2, the average sample travel time is

$$\bar{T}_s = \frac{\sum_{i=1}^{s} t_i}{s}. \tag{3.6}$$

According to statistical theory, when the sample size s becomes sufficiently large, the distribution of the sample mean $\bar{T}_s$ converges to a normal distribution, regardless of the underlying distribution of the shared travel times. This result is formalized by the Central Limit Theorem, as presented in the following lemma.

LEMMA 3.1. The Central Limit Theorem states that if $\{t_1, t_2, ..., t_s\}$ are independent and identically distributed random variables drawn from a population with a mean of μ and a finite variance of σ^2, and the sample mean is denoted by $\bar{T}_s$, then if s is large enough, the standardized sample mean, $\lim_{s\to\infty} \frac{\bar{T}_s - \mu}{\sigma/\sqrt{s}}$, follows a standard normal distribution (Montgomery and Runger 2010).

According to Lemma 3.1, when s is large enough, the distribution of $\bar{T}_s$ approximately follows a normal distribution with a mean of μ and a variance of σ^2/s, demonstrated as

$$\bar{T}_s \sim N(\mu, \frac{\sigma^2}{s}). \tag{3.7}$$

Given that the average sample travel time is non-negative, we model it using a truncated normal distribution confined to the interval $[0, \infty)$. Specifically, we define the random variable $\widehat{T}_s$ as the expected travel time on the random road derived from a sample of size s as

$$\widehat{T}_s \sim \text{Trunc}\left(\bar{T}_s, [0, \infty)\right). \tag{3.8}$$

3.4.3 Route Choice Model

We assume that travelers make a risk-neutral choice between the random road and the safe road by selecting the option with the shorter expected travel time. Later, in Section 3.5.1, risk is incorporated into the model. As time passes and travelers enter the random road and contribute their travel times, a steady-state condition emerges where the ratio of travelers choosing the random road stabilizes and equals the probability of choosing the random road by each traveler. Consequently, the ratio of travelers who choose the random road, α, is given by

$$\alpha = \text{Probability}\left(\widehat{T}_s \leqslant \tau\right) = \Psi(\tau), \tag{3.9}$$

where $\Psi(\cdot)$ denotes the cumulative density function of $\widehat{T}_s$.

Lemma 3.2. The cumulative density function of the variable $\widehat{T}_s$ at the point τ, which follows a normal distribution with mean μ and variance σ^2/s, truncated within the range $[a, b]$, is defined as

$$\Psi(\tau) = \frac{\Phi(\frac{\tau-\mu}{\sigma/\sqrt{s}}) - \Phi(\frac{a-\mu}{\sigma/\sqrt{s}})}{\Phi(\frac{b-\mu}{\sigma/\sqrt{s}}) - \Phi(\frac{a-\mu}{\sigma/\sqrt{s}})}, \tag{3.10}$$

where $\Phi(\cdot)$ shows the cumulative density function of the standard normal distribution with a mean of zero and a variance of one.

By considering Equation (3.8) and applying Lemma 3.2, we can express Equation (3.9) as

$$\alpha = \frac{\Phi(\frac{\tau-\mu}{\sigma/\sqrt{s}}) - \Phi(\frac{-\mu}{\sigma/\sqrt{s}})}{1 - \Phi(\frac{-\mu}{\sigma/\sqrt{s}})}. \tag{3.11}$$

To ensure analytical tractability, the sample size s is substituted with its mean value

$\bar{S}$ in Equation (3.11). Additionally, applying the cumulative density function property $\Phi(-x) = 1 - \Phi(x)$ yields

$$\alpha = 1 - \frac{\Phi\left(\sqrt{\bar{S}}\frac{\mu-\tau}{\sigma}\right)}{\Phi\left(\sqrt{\bar{S}}\frac{\mu}{\sigma}\right)}. \tag{3.12}$$

In the presented queue-based model, we observe that $\mu = \sigma$, given in Equations (3.3) and (3.4). Consequently, the choice model in Equation (3.12) simplifies to

$$\alpha = 1 - \frac{\Phi\big(\sqrt{\bar{S}}\,(1 - \frac{\tau}{\mu})\big)}{\Phi(\sqrt{\bar{S}})}. \tag{3.13}$$

The choice model proposed in Equation (3.13) exhibits the following characteristics:

I. For non-negative values of $\{\tau, \mu\}$, the variable α changes in the range of $[0, 1]$.

II. When the travel time on the safe road is significantly larger than the expected travel time on the random road, $\tau \gg \mu$, almost all travelers choose the random road, i.e., α approaches one.

III. When the expected travel time on the random road is significantly larger than the travel time on the safe road, $\mu \gg \tau$, almost all travelers choose the safe road, i.e., α approaches zero.

IV. When the travel time on the safe road equals the expected travel time on the random road, $\tau = \mu$, the choice model in Equation (3.13) equals

$$\alpha = 1 - \frac{0.5}{\Phi(\sqrt{\bar{S}})}. \tag{3.14}$$

As the sample size increases, the uncertainty in the travel time on the random road decreases. If the sample size is sufficiently large, the condition $\tau = \mu$ leads to the probability of selecting the random road approaching 0.5. At this point, travelers become indifferent between the two roads.

Substituting Equations (3.3), (3.4), and (3.5) into Equation (3.13) yields the proba-

bility of selecting the random road as

$$\alpha = 1 - \frac{\Phi\Big(\sqrt{\alpha\Lambda\beta\delta}\,\big(1-\tau(\theta-\alpha\Lambda)\big)\Big)}{\Phi\Big(\sqrt{\alpha\Lambda\beta\delta}\Big)}. \tag{3.15}$$

PROPOSITION 3.3. For a set of given parameters $\{\Lambda, \theta, \beta, \delta\}$, Equation (3.15) has exactly one positive solution, given $\Lambda < \theta$.

Based on Proposition 3.3, established in Appendix B. 1, for a given contribution ratio β and observation window δ, a unique positive value of α determines the probability of selecting the random road. Solving Equation (3.15) analytically is cumbersome. Figure 3.2 presents numerical examples to illustrate the α value across different parameters with an error tolerance 0.001. The solid lines depict the right-hand side of Equation (3.15) as a function of α. The dashed line represents the identity function. The intersection between each solid and dashed line denotes the value of α, where (3.15) holds.

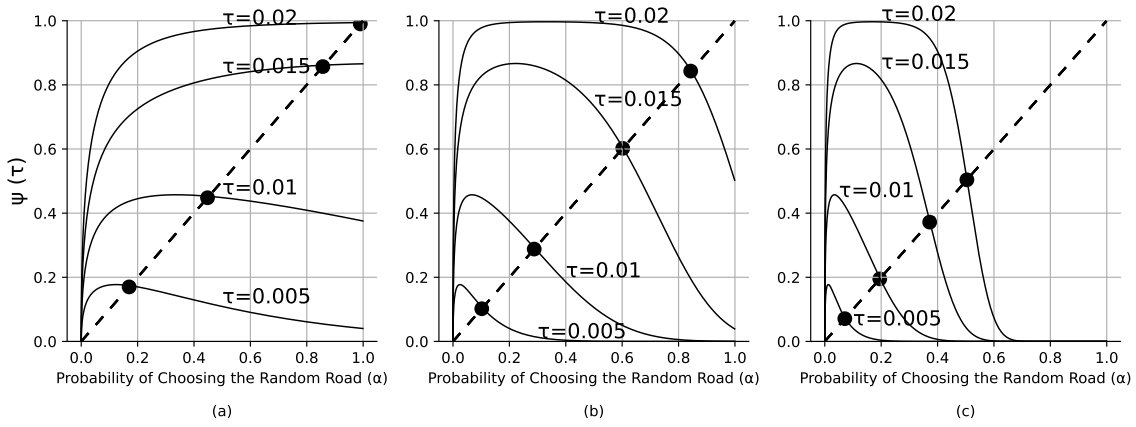


Figure 3.2: The probability of selecting the random road, α, under traveler equilibrium. Parameters: β=1, δ=1, θ=100, with Λ taking values of (a) 10, (b) 50, and (c) 99.

In each panel, the average free-flow travel time on the random road, $1/\theta$, is fixed, while the travel time on the safe road, τ, varies. Each panel corresponds to a specific travel demand, Λ, increasing from left to right. In all panels, α increases with τ. As we move across the panels from left to right at a given τ value, the probability of selecting the random road decreases with increasing travel demand Λ. When τ is less than or equal to the expected free-flow travel time on the random road, i.e., $\tau \leqslant 1/\theta$, it follows that

$\tau \leqslant \mu$. This means that more travelers choose the safe road, which is the shorter route, resulting in $\alpha \leqslant 0.5$. In cases where τ exceeds $1/\theta$, we must also consider the waiting time in traffic before estimating the expected travel time on the random road. Depending on the parameters $\{\Lambda, \theta, \tau, \beta\}$, the value of α may be either less than or greater than 0.5.

Proposition 3.4 explains how α changes in response to variations in the contribution ratio β, with the proof provided in Appendix B. 2. Proposition 3.4 also explores the existence of an optimal contribution ratio, denoted as β_I^*, to determine the maximum probability of selecting the random road.

PROPOSITION 3.4. When $\tau > \mu$, the probability of selecting the random road increases with β. Conversely, when $\mu \geqslant \tau$, it increases below a threshold β_I^* and decreases for β above β_I^*.

The travel time on the safe road, τ, is known to all travelers. When τ is greater than μ, an increase in β leads to a greater number of travelers contributing their travel times within the observation window δ. As a result, the probability of incoming travelers selecting the random road as the shorter route consistently increases. In contrast, when the expected travel time on the random road exceeds τ, a threshold β_I^* arises. For β values below β_I^*, the travel time contributed by a small fraction of travelers does not accurately represent the actual travel time on the random road, as data contribution occurs infrequently. The limited sample size leads incoming travelers to perceive the random road is faster than the safe one, thereby increasing the probability of choosing the random road. However, once β surpasses the β_I^* threshold, a sufficient amount of travel time data is contributed within the observation window δ, ensuring travelers recognize that the random road takes longer than the safe road, prompting them to take the safe road.

As shown in Figure 3.3, when the expected travel time on the random road is less than τ (e.g., $\tau = 0.015, 0.02$), an increase in the contribution ratio β enables more travelers to recognize that the random road is faster than the safe one. However, when the expected travel time on the random road exceeds τ (e.g., $\tau = 0.005, 0.01$), a local maximum occurs for the probablity of choosing the random road.

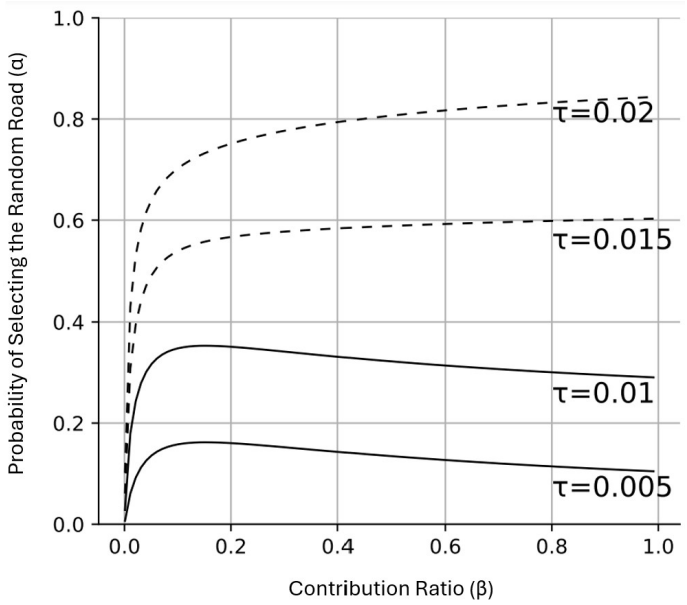


Figure 3.3: The probability of selecting the random road, α, w.r.t the contribution ratio, β. Parameters: θ=100, δ=1, and $\Lambda = 50$. The dashed lines represent scenarios where $\tau > \mu$, while the solid lines correspond to cases where $\mu \geq \tau$.

3.4.4 System Optimal

In the previous section, Equation (3.15) established that the distribution of the travel demand between the safe and random road is a function of the contribution ratio, β. This section investigates the role of a social planner aiming to minimize the overall travel time by selecting an optimal contribution ratio, denoted as β_N^*. We explore whether adjusting β allows the social planner to influence α, referred to as $\alpha(\beta)$ hereafter, and regulate the traffic distribution to achieve the condition of minimum overall travel time, as obtained from the following mathematical model.

$$\min_{\beta} \quad E(\beta) = \Lambda \left[\alpha(\beta) M(\lambda) + \big(1 - \alpha(\beta)\big)\tau \right], \tag{3.16a}$$

$$\text{s.t}: \quad \alpha(\beta) = 1 - \frac{\Phi\Big(\sqrt{\alpha\Lambda\beta\delta}\,\big(1 - \tau(\theta - \alpha\Lambda)\big)\Big)}{\Phi\Big(\sqrt{\alpha\Lambda\beta\delta}\Big)}, \tag{3.16b}$$

$$M(\lambda) = \frac{1}{\theta - \lambda}, \tag{3.16c}$$

$$\lambda = \alpha(\beta)\Lambda, \tag{3.16d}$$

$$0 < \Lambda < \theta, \tag{3.16e}$$

$$0 \leqslant \alpha(\beta) \leqslant 1, \tag{3.16f}$$

$$0 < \beta \leqslant 1. \tag{3.16g}$$

The objective function $E(\beta)$ represents the overall travel time, considering travelers who take the random road and those who choose the safe road. Constraint (3.16b) defines the probability of selecting the random road, as given by Equation (3.15). Constraint (3.16c) specifies the average travel time on the random road, while constraint (3.16d) describes the ratio of travelers choosing this route. Lastly, constraint (3.16e) imposes the weak stability condition for the queue on the random road.

Upon solving model (4.1), the value of β_N^* is determined. Proposition 3.5 explains the behavior of the overall travel time based on the contribution ratio, β.

PROPOSITION 3.5. When $\mu \geqslant \tau$, the overall travel time decreases with β. Conversely, when $\tau > \mu$, it decreases with β below a threshold β_N^* and increases for β above β_N^*.

According to Proposition 3.4, when the travel time on the safe road is shorter than the expected travel time on the random road, the maximum ratio of travelers taking the random road is limited to $\alpha(\beta_I^*)$. As a result, the random road would not be overcrowded, and the overall travel time would decrease with β. However, when the safe road takes longer than the expected travel time on the random road, an increase in β leads to an increase in the ratio of travelers choosing the random road. Under this condition, when the contribution ratio is less than β_N^*, overall travel time decreases due to travelers saving

time by taking the random road. However, when the random road becomes overcrowded, indicated by the road selection probability larger than $\alpha(\beta_N^*)$, selfish routing increases the overall travel time.

To minimize overall travel time, the proportion of travelers taking the random road, $\alpha(\beta_N^*)$, is established with β_N^* through model (4.1), in alignment with Proposition 3.3. Proposition 3.6 illustrates the value of $\alpha(\beta_N^*)$ under network equilibrium condition.

PROPOSITION 3.6. To minimize the overall travel time, the optimum ratio of travelers who take the random road is

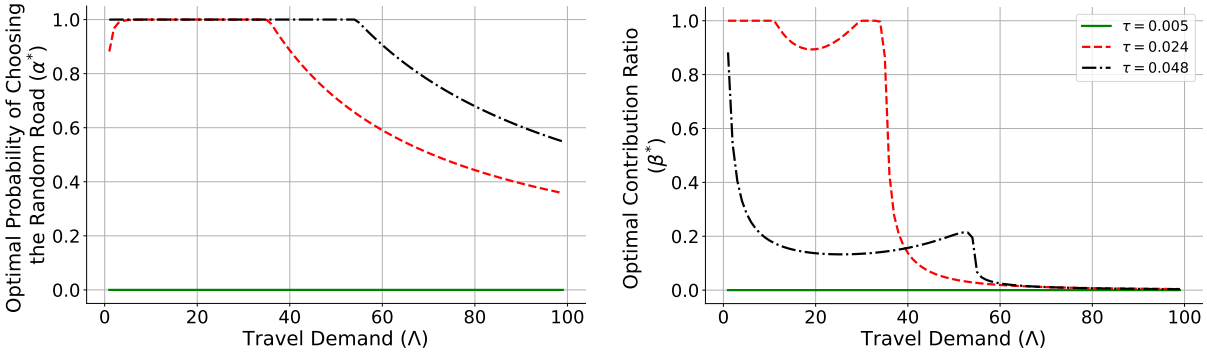
$$\alpha(\beta_N^*) = \begin{cases} 0 & 0 \le \tau \le \frac{1}{\theta}, \\ \frac{\theta - \sqrt{\frac{\theta}{\tau}}}{\Lambda} & \frac{1}{\theta} < \tau < \frac{\theta}{(\theta-\Lambda)^2}, \\ \alpha(\beta') & \tau \ge \frac{\theta}{(\theta-\Lambda)^2}, \end{cases} \tag{3.17}$$

where $\beta' =_\beta (\alpha)$.

When the random road is congestion-free with no queue, $W_q = 0$, the expected travel time becomes free-flow, $1/\theta$. If the free-flow travel time on the random road exceeds that of the safe road, all travelers should take the safe road to minimize travel time, resulting in $\alpha = 0$, the first case. When the free-flow travel time on the random road is less than τ, and $\tau < \theta/(\theta - \Lambda)^2$, the optimal α is given by the second term in Equation (3.17), derived from solving the optimization model (4.1). If τ exceeds the threshold $\theta/(\theta - \Lambda)^2$, the optimal value of α is its maximum achievable value. Therefore, in the third case, $\beta' = \beta_N^*$; and the argmax function is employed for enhancing clarity.

Figure 3.4 illustrates the optimal α and the corresponding β values as a function of travel demand Λ for different τ values. When τ is less than the free-flow travel time on the random road, no traveler chooses the random road, and $\alpha^* = 0$. In this case, the optimal contribution ratio, shown in panel (b), is also zero. For $\tau > \frac{1}{\theta}$, the proportion of travelers using the random road is determined by network demand, as stated in Proposition 3.6. When demand is low, travelers prefer the faster random road, and the optimal α is 1. However, as demand increases, congestion on the random road raises travel time,

prompting more travelers to choose the safe road, which reduces the proportion of travelers on the random road.



(a) (b)

Figure 3.4: (a) Optimal α values and (b) corresponding β values for minimizing overall travel time across different travel demand levels and τ values. Parameters: $\theta = 100$ and $\delta = 1$.

Special Case: $\alpha(\beta_N^*) = 1$. Looking at the proposition 3.6, we examine whether a scenario exists in which no traveler chooses the safe road, $\alpha(\beta_N^*) = 1$. This implies the condition that all travelers should choose the random road to achieve the minimum overall travel time. To achieve this, we shift our focus from the contribution ratio to the observation window, which plays an equally important role in configuring a crowd-sensing platform. Proposition 3.7 describes this scenario and defines an optimal threshold for the observation window, denoted as δ_N^*, for a given β.

PROPOSITION 3.7. When $\mu < \tau$, for a given β, a minimum observation window δ_N^* exists, such that for $\delta > \delta_N^*$, all travelers choose the random road to minimize overall travel time. The threshold δ_N^* is given by

$$\delta_N^* = \frac{1}{\beta\Lambda}\left(\frac{\nu}{\tau(\theta - \Lambda) - 1}\right)^2, \tag{3.18}$$

where ν represents $\Phi^{-1}(1)$.

The value of ν can be set based on the acceptable error level. For instance, with an acceptable error of 2×10^{-5}, ν equals 4.09. Equation (3.18) shows that higher data contributions (β) or increased travel demand (Λ) allow for a shorter optimal δ while still

providing sufficient CTD to minimize overall travel time. Figure 3.5 shows the simulated probabilities of selecting the random road based on the number of travelers traveling from the sink to the source node for different values of δ. The expected travel time on the random road varies between 0.01 and 0.02, which is less than the travel time on the safe road, $\tau = 0.03$. The optimal observation window δ_N^* is calculated from Equation (3.18). For $\delta \geq \delta_N^*$, the simulated α converges to one, whereas for $\delta < \delta_N^*$, α converges to a value below one.

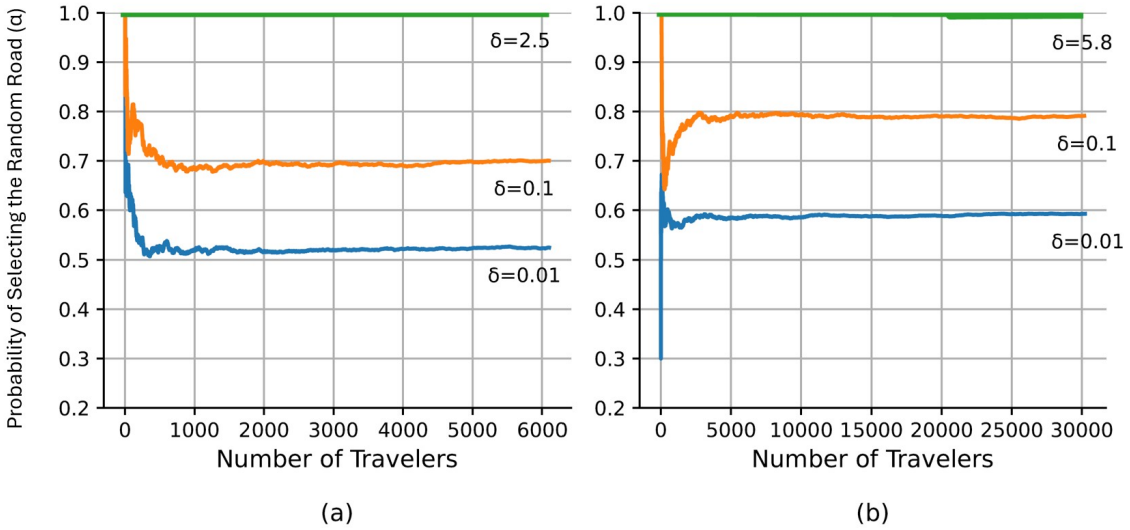


Figure 3.5: Simulated probability of selecting the random road, α, w.r.t the number of travelers. Parameters: $\theta = 100$, $\tau = 0.03$, $\beta = 1$, and $\nu = 8.5$. For (Λ, δ_N^*): (a) $(10, 2.5)$, (b) $(50, 5.78)$.

3.5 Extensions to Analysis User Behavior

This section extends the base model to incorporate additional complexities that influence traveler route choice behavior. A risk-sensitive framework is introduced to capture traveler aversion to risk on a random road with uncertain travel times and to assess how the contribution ratio mitigates the perceived risk. Next, the impact of access to real-time travel information is analyzed to compare decision making based on precise travel time data with that based on CTD. Finally, a contribution reward is introduced to incentivize data sharing, and its effect on the overall travel time is investigated.

3.5.1 Access to Accurate Travel Data in a Risk-Incorporated Model

Traveler route choices are influenced by the uncertainty of travel times. When traffic conditions are unpredictable, a common strategy to mitigate the risk of delays is to allocate extra time beyond the expected travel duration (Nikolova and Stier-Moses 2014). This section examines risk-averse travelers and explores strategies to tackle uncertainty in stochastic travel times on the random road. A risk-aversion function is introduced to model the impact of the contribution ratio on perceived risk.

The disutility of selecting the safe and random roads, denoted as U_{Safe} and U_{Random}, is expressed as

$$U_{\text{Safe}} = \tau, \tag{3.19}$$

and

$$U_{\text{Random}} = \widehat{T}_s + \gamma(\beta)V_T(\lambda), \tag{3.20}$$

where $\gamma(\beta)$ is the risk aversion function, where $\gamma(\beta) = 0$ corresponds to the base choice model defined in Equation (3.9). A higher $\gamma(\beta)$ indicates greater risk aversion, meaning that travelers prefer the more predictable route to avoid uncertainty. The following lemma establishes the key properties of the risk aversion function, highlighting its relationship

with β.

LEMMA 3.8. The risk aversion function $\gamma(\beta)$ is a non-negative, concave, and decreasing function of β.

In a CTD setup, we assume that the level of risk aversion, $\gamma(\beta)$, depends on the contribution ratio β. Since β is publicly available from historical data, it directly influences traveler trust in the CTD and, consequently, their level of risk aversion. The frequency of CTD updates increases with the contribution ratio β. As updates become more frequent, the accuracy of travel-time estimates improves. This reduction in perceived uncertainty makes the random road a more attractive option.

Considering the risk associated with the uncertainty of the travel time on the random road, the probability of selecting the random road, denoted as α_{Risk}, is

$$\alpha_{Risk} = \Pr\Big(U_{Random} \leqslant U_{Safe}\Big), \tag{3.21}$$

which simplifies to

$$\alpha_{Risk} = \text{Probability}\Big(\widehat{T}_s \leqslant \tau - \gamma(\beta)V_T(\lambda)\Big) = \Psi\Big(\tau - \gamma(\beta)V_T(\lambda)\Big). \tag{3.22}$$

Since $\gamma(\beta)$ decreases with β, the choice model in Equation (3.22) suggests that the maximum probability of choosing the random road occurs when $\beta = 1$.

We propose a specific functional form for $\gamma(\beta)$ that is compatible with Lemma 3.8. A power-law function is used to capture the diminishing effect of β on the level of risk aversion, as given by

$$\gamma(\beta) = \gamma_0(1-\beta)^k, \tag{3.23}$$

where γ_0 represents the maximum risk aversion level when the contribution ratio β is zero. The parameter $k > 1$ determines the rate at which $\gamma(\beta)$ decreases with β.

Figure 3.6 illustrates the proportion of travelers who select the random road under two scenarios: risk-averse, dashed lines, and risk-neutral, solid lines.

The intersection of the dashed curve with the identity function represents α_{Risk}, while

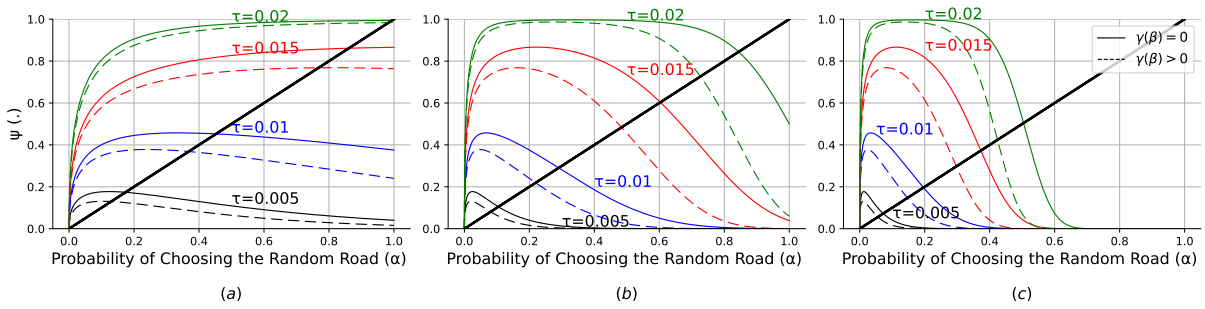


Figure 3.6: Comparison of the probability of choosing the random road with (dashed lines) and without (solid lines) incorporating risk. Parameters: $\theta = 100$, $\beta = 1$, $k = 1$, $\gamma_0 = 30$. For Λ: (a) 10, (b) 50, (c) 99.

the intersection of the solid line indicates α. Given that $\Psi(\cdot)$ is an increasing function and both $V_T(\lambda)$ and $\gamma(\beta)$ are positive, a comparison of Equations (3.9) and (3.22) suggests that the probability of selecting the random road decreases in the presence of risk, i.e., $\alpha_{Risk} \leqslant \alpha$. This observation is confirmed in Figure 3.6 across varying levels of travel demand and values of τ, consistently demonstrating that $\alpha_{Risk} < \alpha$.

In this scenario, the minimum overall travel time is obtained from the following mathematical model

$$\min_{\beta} \quad E_{Risk}(\beta) = \Lambda \left[\alpha_{Risk}(\beta) M(\lambda) + \big(1 - \alpha_{Risk}(\beta)\big)\tau \right], \tag{3.24a}$$

$$\text{s.t}: \quad \alpha_{Risk}(\beta) = 1 - \frac{\Phi\Big(\sqrt{\lambda\beta\delta}\,\big(1 - \big(\tau - \gamma(\beta)V_T(\lambda)\big)(\theta - \lambda)\big)\Big)}{\Phi\Big(\sqrt{\lambda\beta\delta}\Big)}, \tag{3.24b}$$

$$M(\lambda) = \frac{1}{\theta - \lambda}, \tag{3.24c}$$

$$\lambda = \alpha_{Risk}(\beta)\Lambda, \tag{3.24d}$$

$$0 \leqslant \alpha_{Risk}(\beta) \leqslant 1, \tag{3.24e}$$

$$0 < \beta \leqslant 1, \tag{3.24f}$$

where $\gamma(\beta)$ is the risk aversion function, and $E_{Risk}(\beta)$ is the overall travel time with incorporating risk.

Using $\gamma(\beta)$ given in Equation (3.23), model (3.24) is solved numerically. Figure 3.7 shows how the optimal contribution ratio, β^*_{Risk}, varies with travel demand, Λ, under different risk levels.

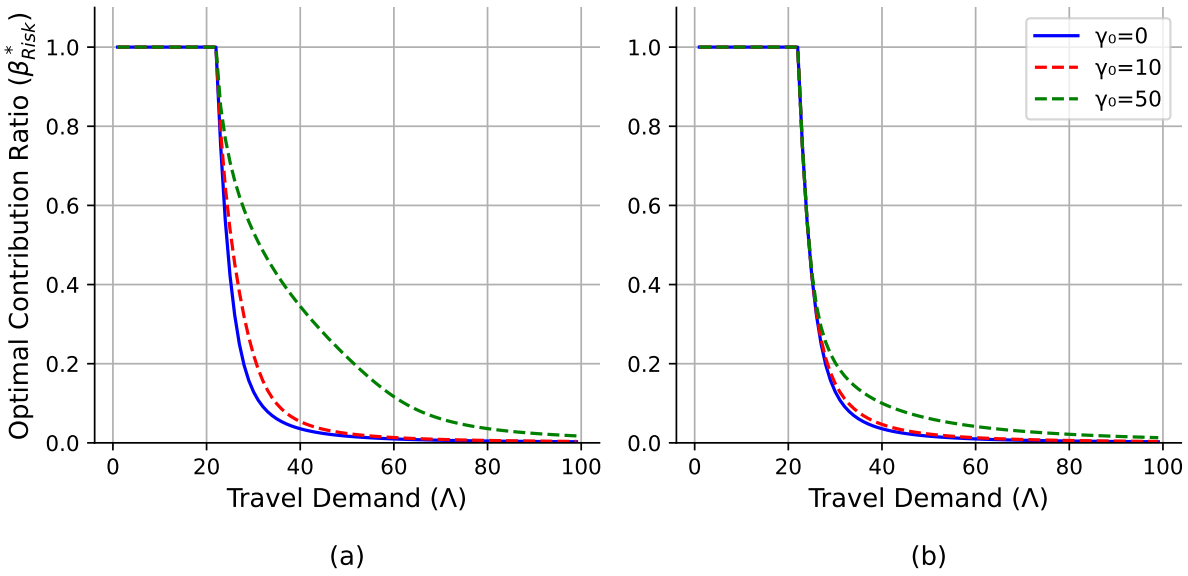


Figure 3.7: Optimal contribution ratio β^*_{Risk} as a function of Λ for different risk levels (a) $\gamma(\beta) = \gamma_0(1-\beta)^2$ and (b) $\gamma(\beta) = \gamma_0(1-\beta)^{10}$. Parameters: θ=100, $\tau = 0.015$, $\gamma_0 = 0, 10, 50$.

The solid blue curve represents the base model where $\gamma(\beta) = 0$, while the dashed curves correspond to risk-averse scenarios where $\gamma(\beta) > 0$. The two panels display the results for different values of k. The results indicate that when Λ is sufficiently large, introducing risk increases the optimal β, particularly for lower values of k. This observation suggests that a higher maximum risk level, γ_0, leads to more cautious route selection, therefore a higher level of optimum user contribution β is required to mitigate potential losses due to uncertain travel times on the random road. Furthermore, the panels demonstrate how increasing k accelerates the decline of the optimal β, leading to lower β values across all Λ for the greater k value in panel (b) compared to panel (a). These observations demonstrate that less sensitivity to travel time variability, higher k, and lower γ_0 result in a lower optimum contribution ratio.

Previously, we demonstrated that to mitigate the uncertainty associated with fluctuating travel times on the random road, the optimization model increases the optimal contribution ratio for risk-averse travelers. Now, we explore how access to real-time travel information can offer an alternative solution to tackle risk. While CTD provides estimates of travel times, the variability inherent in these estimates may lead risk-averse travelers to search for other sources of information. Accurate travel time data alleviate this uncer-

tainty, allowing travelers to make more confident and better informed route choices.

To meet the demand for an accurate source of travel time information, significant investments have been made in deploying stationary sensors and cameras along roads to provide accurate travel time data. In this analysis, we assume that travelers must pay an access fee to obtain accurate travel data. The impact of accurate travel information on route choice behavior is examined to weigh the trade-off between paying for the certainty offered by accurate data and accepting the uncertainty inherent in free-of-charge CTD.

The disutility of taking the random road for travelers with access to accurate travel information, denoted by U_{Exact}, is defined as

$$U_{Exact} = T_E + h, \tag{3.25}$$

where T_E is the accurate travel time on the random road, and h is the access cost to receive the real-time travel information.

To obtain the travel time on the random road, travelers face two options: they can either rely on the free CTD with inherent uncertainty or pay for accurate travel time information. By comparing the disutility functions U_{Random}, given in Equation (3.20), and U_{Exact}, the conditions under which travelers prefer one option over the other can be determined. Specifically, travelers choose free CTD if

$$U_{Random} < U_{Exact}. \tag{3.26}$$

After deciding between the free CTD and paid accurate information, travelers then compare the obtained travel time on the random road with that on the safe road, following the same decision process outlined in the base model.

By substituting Equations (3.20) and (3.25) into Equation (3.26) and solving for β, we obtain the following inequality:

$$\beta > 1 - \left(\frac{h - (\widehat{T}_s - T_E)}{\gamma_0 V_T(\lambda)} \right)^{\frac{1}{k}}. \tag{3.27}$$

We denote the right-hand side of Equation (3.27) by $\hat{\beta}$:

$$\hat{\beta} = 1 - \left(\frac{h - (\widehat{T}_s - T_E)}{\gamma_0 V_T(\lambda)} \right)^{\frac{1}{k}}. \tag{3.28}$$

Based on Equation (3.27), for a given sample of collected travel times, as the access fee h increases, travelers require a smaller $\hat{\beta}$ to choose free CTD over paid information. Conversely, lower access fees make paying for accurate travel information more appealing. Furthermore, a higher uncertainty of the travel time on the random road, a higher variance $V_T(\lambda)$, increases the $\hat{\beta}$ threshold for travelers to choose free CTD.

Since α varies between 0 and 1, $V_T(\lambda)$ lies in the range $[1/\theta, 1/(\theta - \Lambda)]$ (see Equation (3.4)). To obtain a lower bound for $\hat{\beta}$, and thereby define a threshold beyond which travelers prefer free CTD over accurate travel information, we substitute the maximum value of $V_T(\lambda)$, which is $1/(\theta - \Lambda)$. This yields:

$$\hat{\beta} = 1 - \left(\frac{(h - (\widehat{T}_s - T_E))(\theta - \Lambda)}{\gamma_0} \right)^{\frac{1}{k}}. \tag{3.29}$$

Given that $\widehat{T}_s$ is a normally distributed random variable, the difference $\widehat{T}_s - T_E$ follows a normal distribution with mean $\mu - T_E$ and variance σ^2/s. For a given access fee h, the threshold $\hat{\beta}$ varies depending on the realization of $\widehat{T}_s$. To visualize the distribution of $\hat{\beta}$ based on the access fee, we simulate 1000 random realizations of $\widehat{T}_s$, where each realization is drawn from a normal distribution with mean μ, and standard deviation $\sigma/\sqrt{s}$, where $\mu = \sigma$, given Equations (3.3) and (3.4). For each realization of $\widehat{T}_s$ and the corresponding value of h, the associated $\hat{\beta}$ is calculated. The results are shown in Figure 3.8, which illustrates the mean $\hat{\beta}$ values along with the 10th and 90th percentiles, highlighting how uncertainty in $\widehat{T}_s$ influences $\hat{\beta}$. Since $\hat{\beta}$ is constrained between zero and one, any values below zero are adjusted to $\hat{\beta} = 0$. As we move from left to right, travel-time variance increases. As a result, the mean value of $\hat{\beta}$ increases, indicating that travelers perceive the random road as more uncertain and therefore require a higher contribution ratio to choose CTD. Additionally, the maximum access fee at which travelers choose CTD for

any contribution ratio also increases. This is represented by the intersection of the solid line, the mean $\hat{\beta}$, with the line $\hat{\beta} = 0$.

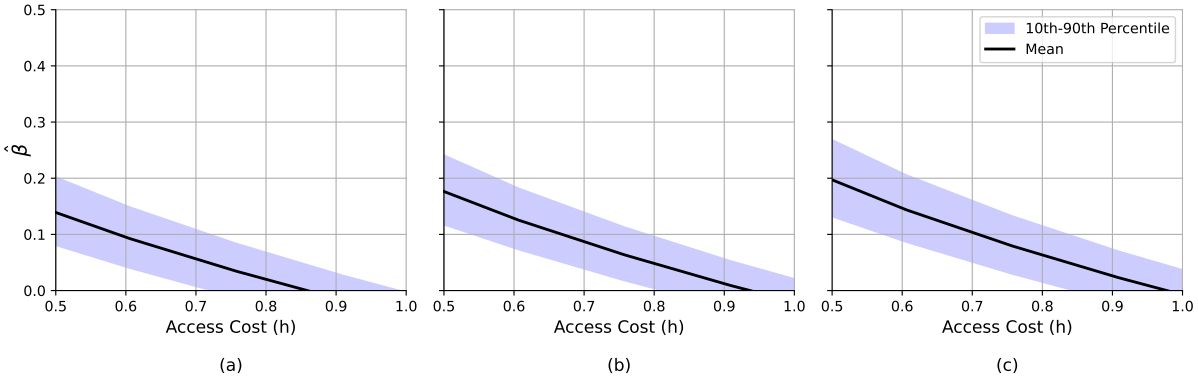


Figure 3.8: The distribution of $\hat{\beta}$. Parameters: (a) σ_1, (b) σ_2, (c) σ_3, where $\sigma_1 < \sigma_2 < \sigma_3$.

Given (3.27), travelers pay the access cost h only if the contribution ratio β falls below the threshold $\hat{\beta}$. To determine the optimal contribution ratio, denoted as β_h^*, the following optimization model is formulated:

$$\min_{\beta} \quad E(\beta, h) = \Lambda \left[\Pr(\beta < \hat{\beta}) \Big(h + \alpha_h(\beta) T_E + (1 - \alpha_h(\beta))\tau \Big) + \Pr(\beta \geqslant \hat{\beta}) \Big(\alpha_h(\beta) M(\lambda) + \big(1 - \alpha_h(\beta)\big)\tau \Big) \right] \tag{3.30a}$$

$$\text{s.t.} \quad \alpha_h(\beta) = \Pr(\beta < \hat{\beta}) \Pr(U_{Exact} \leqslant U_{Safe}) + \Pr(\beta \geqslant \hat{\beta}) \Pr(U_{Random} \leqslant U_{Safe}) \tag{3.30b}$$

$$M(\lambda) = \frac{1}{\theta - \lambda}, \tag{3.30c}$$

$$\lambda = \alpha_h(\beta)\Lambda, \tag{3.30d}$$

$$0 \leq \alpha_h(\beta) \leq 1, \tag{3.30e}$$

$$0 < \beta \leq 1, \tag{3.30f}$$

where T_E represents the average accurate travel time on the random road. The objective function minimizes the expected network-wide travel cost. When the contribution ratio is below $\hat{\beta}$, travelers purchase the accurate travel information by incurring the cost h.

Figure 3.9 illustrates how the optimal contribution ratio varies with travel demand under three scenarios. In the Baseline Model (4.1) travelers are risk-neutral, and the optimal contribution ratio is β^*. In the Risk-Aversion Model (3.24) risk-averse travelers

require a higher contribution ratio to compensate for travel time uncertainty, i.e., $\beta^*_{Risk} \geqslant \beta^*$. Finally, the Access Cost with Risk Aversion Model (3.30) considers the case in which risk-averse travelers must pay an access fee for accurate travel-time information; here, the optimum contribution ratio is β^*_h, and the expected lower bound on the contribution ratio, $\hat{\beta}$, is derived from simulation.

The figure is presented in two panels. Panel (a) corresponds to an access cost of h_1, while panel (b) corresponds to an access cost of h_2, where $h_1 < h_2$. The results emphasize that as the access cost increases from panel (a) to panel (b), the expected lower bound threshold $\hat{\beta}$ decreases. Moreover, the behavior of $\hat{\beta}$ as a function of travel demand Λ shows an interesting dynamic: as travel demand increases, risk-averse travelers require a higher contribution ratio threshold to rely on CTD. Although a high-demand network could theoretically achieve the same level of information reliability with a lower percentage of contributors than a low-demand network, risk-averse travelers only trust CTD in high-demand networks if the contribution ratio is sufficiently high; in low-demand networks, they are more tolerant of travel time uncertainty and prefer free CTD. A comparison

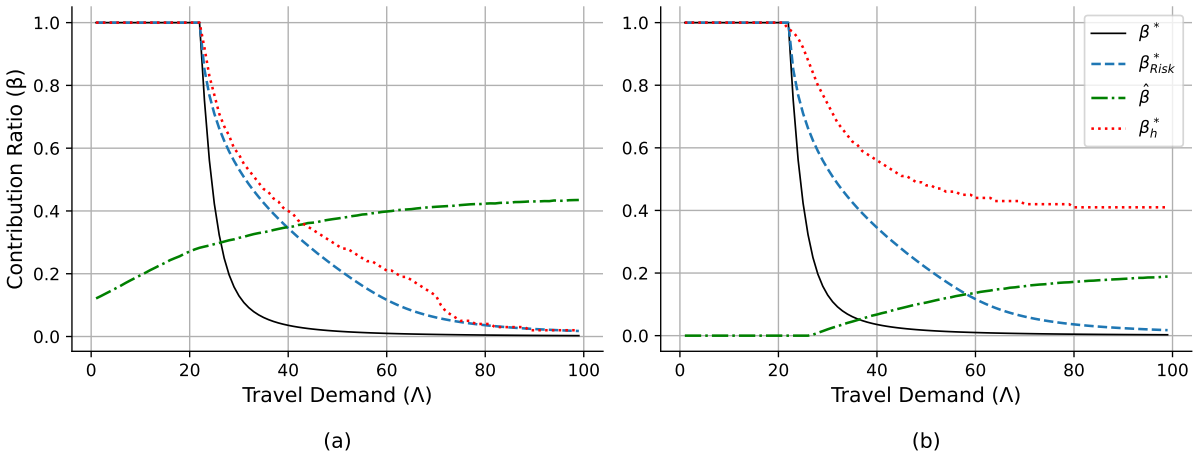


Figure 3.9: Comparison of the optimal contribution ratio, β^*, and the mean value of $\hat{\beta}$ under different scenarios for a given access cost h. The figure illustrates three cases: (1) β^*, the baseline model, in which travelers are risk-neutral and accurate travel information is not available ($\gamma(\beta) = 0$ and h is not applicable); (2) β^*_{Risk}, the risk-aversion model, in which travelers exhibit risk aversion ($\gamma(\beta) > 0$ and $h = 0$) but accurate travel information is still not available; and (3) β^*_h, the access cost model, where risk-averse travelers can access accurate travel-time information by paying a fee ($\gamma(\beta) > 0$ and $h > 0$). Access cost levels are given as (a) h_1 and (b) h_2, where $h_1 < h_2$.

between $\hat{\beta}$ and the optimum contribution ratio under the risk-aversion scenario, β^*_{Risk},

further clarifies this behavior. In low-demand networks where the contribution ratio exceeds the mean $\hat{\beta}$, travelers opt for free CTD because the optimal contribution ratio meets their minimum threshold. In such cases, introducing an access fee discourages travelers from purchasing accurate travel information due to their cost sensitivity. However, in high-demand networks where $\hat{\beta}$ exceeds the optimal contribution ratio, travelers face a choice: either pay for accurate travel-time information or require a higher contribution ratio to maintain trust in CTD. Increasing the contribution ratio beyond the optimal level in these conditions, however, results in longer overall travel times. Thus, offering risk-averse travelers the option to purchase accurate information in high-demand networks is the optimal alternative. The threshold for classifying a network as low- or high-demand depends on the access cost; as the access cost increases from panel (a) to (b), this threshold also rises in both the β^* and β^*_{Risk} scenarios.

Introducing a positive access fee increases the optimal contribution ratio relative to the baseline and risk-aversion scenarios, i.e., $\beta^*_h \geqslant \beta^*$ and β^*_{Risk}. With a higher access cost in panel (b), the optimum contribution ratio exceeds that of the baseline and risk-aversion models and lies above $\hat{\beta}$, the minimum contribution ratio for using CTD. This means that when the access cost is high, optimizing network performance requires the contribution ratio to be above $\hat{\beta}$ so that travelers trust free CTD to choose their travel path rather than paying for expensive, accurate travel information. In contrast, when the access cost is lower in panel (a), the optimum contribution ratio β^*_h depends on travel demand. When network demand is low, the optimum contribution ratio remains above $\hat{\beta}$, allowing travelers to effectively use free CTD and maintain network performance. However, when demand increases under a low access fee regime, the benefits of purchasing accurate travel-time information outweigh the need to raise the contribution ratio, thereby avoiding additional congestion, so β^*_h falls below $\hat{\beta}$.

If policymakers or service providers aim to encourage the use of accurate travel information, they must consider factors such as the access fee, contribution ratio, and network demand. Analyzing these patterns helps determine whether investing in infrastructure to obtain accurate real-time information will significantly influence traveler behavior and

network performance and, if so, guide the formulation of pricing policies.

3.5.2 Contribution Reward

It has been demonstrated that the contribution ratio is crucial in determining optimal policies for traffic management and improving overall system efficiency. However, many conservative travelers may hesitate to share their data due to concerns about privacy, security, and the perceived lack of personal benefit. Introducing a contribution reward system can address these concerns by offering tangible incentives in exchange for valuable traveler data. These rewards enhance the traveler experience and can take various forms, such as discounts on future travel, priority access to preferred routes, or participation in exclusive loyalty programs. Ultimately, the contribution reward system aims to create a mutually beneficial ecosystem: travelers receive rewards while contributing to the improvement of the transportation system (Wei et al. 2022).

In this section, to motivate travelers to choose the random road and agree to contribute travel data, the social planner offers a contribution reward, denoted as R. We examine how different levels of contribution reward influence the contribution ratio and traveler behavior. Given there are no budgetary constraints, the following mathematical model is proposed to determine the optimal contribution ratio, and the optimal proportion of travelers who choose the random road to minimize the overall travel time

$$\min_{\beta,R} \quad E_{Reward}(\beta) = \Lambda \left[\alpha(\beta) M(\lambda) + \big(1 - \alpha(\beta)\big)\tau \right], \tag{3.31a}$$

$$\text{s.t}: \quad \alpha(\beta) = (1-\beta)\ \Pr\Big(\widehat{T}_s \leqslant \tau\Big) + \beta\ \Pr\Big(\widehat{T}_s - R \leqslant \tau\Big), \tag{3.31b}$$

$$M(\lambda) = \frac{1}{\theta - \lambda}, \tag{3.31c}$$

$$\lambda = \alpha(\beta)\Lambda, \tag{3.31d}$$

$$0 \leqslant R \leqslant R_{max}, \tag{3.31e}$$

$$0 \leqslant \alpha(\beta) \leqslant 1, \tag{3.31f}$$

$$0 < \beta \leqslant 1, \tag{3.31g}$$

where $E_{Reward}(\beta)$ is the overall travel time with incorporating the contribution reward, and R_{max} represents the maximum reward that the planner can assign to each traveler. Equation (3.31b) captures the overall probability that travelers choose the random road over the safe road. The reward R is an incentive provided to travelers when they share their travel time data, and β reflects the proportion of travelers who consider this reward in their decision-making process. The first term, $(1-\beta)\Pr\Big(\widehat{T}_s \leqslant \tau\Big)$, represents the probability that a traveler, not influenced by the reward, chooses the random road if the estimated travel time is lower than the travel time on the safe road. The second term, $\beta\Pr\Big(\widehat{T}_s - R \leqslant \tau\Big)$, accounts for those travelers who consider the reward, effectively reducing their perceived travel time on the random road by R. In other words, a traveler chooses the random road if either the estimated travel time is already favorable without the reward or if the inclusion of the reward makes the travel time on the random road effectively lower than that of the safe road. This equation thus provides a measure of traveler behavior by weighting both decision scenarios according to the proportion β of travelers considering the reward.

To determine the probability of taking the random road, as defined in constraint (3.31b), the following equation is used

$$\alpha(\beta) = (1-\beta)\Psi(\tau) + \beta\Psi(\tau + R), \tag{3.32}$$

where the value of $\Psi(\cdot)$ is determined by Equation (3.10), with $a = 0$ and $b = +\infty$.

Figure 3.10 compares the base model (4.1) with model (3.31), which includes the contribution reward, where $\mu < \tau$.

Panel (a) shows the probability of selecting the random road based on travel demand. At low demand levels, the reward motivates more travelers to choose the random road compared to the base model, which is also, on average, the shorter path. However, as travel demand increases and the network becomes crowded, the contribution reward no longer motivates travelers to take the random road and does not increase α, as the optimal contribution reward does not compensate for the extra time spent waiting in congestion on the random road. Panel (b) illustrates the optimal contribution ratio to

minimize overall travel time. When more travelers opt for the random road, a smaller percentage needs to contribute data to minimize the overall network cost. Panel (c) shows the reduction in overall travel cost achieved by incorporating the contribution reward as a function of travel demand. As shown, the effectiveness of the contribution reward varies with network demand levels. Providing the contribution reward is most effective in low-demand networks, but it becomes less impactful in high-demand networks, where it may not substantially reduce overall travel time. Panel (d) shows the optimal contribution reward based on Λ. When $\mu < \tau$, providing the contribution reward in a congested, high-demand network is not a practical option to minimize overall travel time.

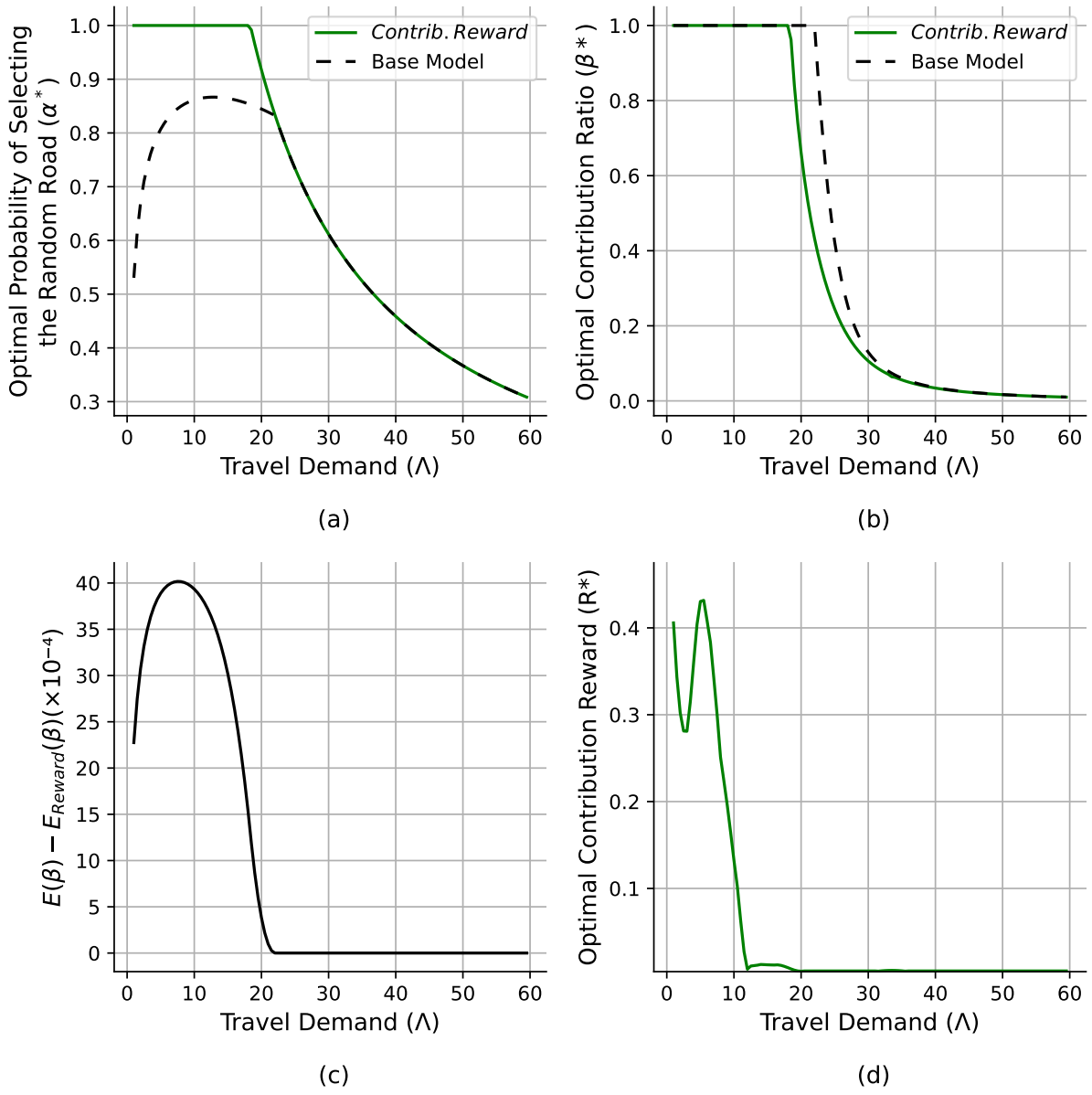


Figure 3.10: Comparison of the base model (4.1) and the model with the contribution reward (3.31): (a) Value of β, (b) Ratio of travelers choosing the random road, (c) Reduction in overall travel time, and (d) the optimum reward. Parameters: $\delta = 1$, $\theta = 100$, $\tau = 0.015$, $R_{max} = 1$.

3.6 Chapter Summary

This study explored the role of CTD in optimizing route choices and minimizing network-wide travel times while ensuring equitable access to information and preserving user privacy. We introduced two key factors, contribution ratio and observation window, that influence the reliability of travel time estimations. By modeling a transportation network with a safe road, with fixed travel time, and a random road , with stochastic travel time, we analyzed how these factors impacted traveler route choices and overall system efficiency.

Our findings revealed that an optimal contribution ratio exists, below full participation, which minimizes network-wide travel time in a fixed-demand setting. While higher contribution ratios improved travel time predictions, they led to congestion due to a disproportionate shift of travelers onto the random road, increasing the overall travel time. Our results showed that when the travel time of the safe road is greater than the free-flow travel time on the random road, travel demand determines the optimal ratio of travelers taking the random road. Additionally, we determined a threshold for the minimum observation window, ensuring that travelers opted for the random road only when its expected travel time was shorter than that of the safe road.

Incorporating risk aversion into traveler decision-making demonstrated that increased uncertainty in travel times discourages travelers from selecting the random road, thereby decreasing the ratio of travelers who take it. Moreover, we explored the scenario in which real-time travel information was available for a fee, showing that travelers' willingness to pay depended on the contribution ratio and network demand. Lastly, we demonstrated that incentivizing data sharing through rewards could further reduce overall travel time, particularly in low-demand networks, offering a practical policy tool to encourage optimal participation levels.

These insights contribute to transportation planning and policy by highlighting how strategic management of CTD can enhance network efficiency without compromising traveler privacy. Future research could extend this framework by incorporating dynamic de-

mand conditions. Our study underscores the importance of balancing participation incentives, privacy considerations, and network demand in designing effective crowd-generated data-driven transportation systems.

Chapter 4

CHAPTER FOUR: RESOURCE ALLOCATION AND ROUTE GENERATION FOR URBAN MOBILE SENSING

4.1 Chapter Overview

Mobile sensing platforms, such as dashcams and vehicle-mounted ultrasonic sensors, offer a scalable and cost-effective alternative to traditional systems, including in-ground sensors and fixed cameras for urban data collection. However, optimizing the deployment of mobile sensors is critical to balancing data quality, spatial coverage, and operational cost. This study introduces an optimization framework for urban data collection that jointly determines the optimal number of agents, i.e., users equipped with sensing devices, and their routing paths. The framework integrates sensor allocation and route planning while enforcing headway constraints to maintain data recency and avoid redundancy. The city is modeled as a graph, where nodes represent facilities of interest for data collection and edges denote travel paths, with the objective of minimizing the total travel time of agents. We develop an exact mixed-integer programming model and a scalable heuristic algorithm, the latter incorporating Lagrangian-inspired refinements to enhance solution quality via node reassignments. Numerical experiments validate the heuristic performance against the exact model, showing that it yields near-optimal solutions with substantially lower computational cost. Results indicate that travel time decreases with more routes and looser headway constraints, while the required number of agents declines with headway but varies non-monotonically with the number of routes. We apply the proposed heuristic to parking facilities in the City of Toronto to demonstrate its scalability and practical applicability in real-world settings.

4.2 Introduction

Monitoring urban environments is essential for ensuring efficient city management and improving quality of life. However, traditional stationary infrastructure used for urban data collection is expensive to scale, challenging to install, and difficult to maintain, given issues like limited coverage, strict city regulations, and power supply constraints. In contrast, mobile sensing devices, such as dashcams, vehicle-mounted ultrasound sensors, LiDAR-equipped autonomous vehicles, and wearable smart gadgets, provide a scalable, flexible, and cost-effective alternative for gathering real-time data across large urban areas. As individuals equipped with these sensing devices, referred to as agents, move through city streets, the devices collect valuable data on environmental and traffic conditions, including noise levels, air pollution, vacant parking spaces, and congestion patterns. This information is transmitted to a central database for processing and analysis. Sharing the processed data with the broader community increases public awareness, reduces traffic congestion, and health risks. Nevertheless, achieving comprehensive and efficient urban data collection requires the strategic deployment of agents to maximize coverage, minimize travel time, and avoid redundant data gathering.

A key challenge in deploying agents is balancing data accuracy with operational efficiency. While multiple agents collecting data at the same location can enhance accuracy, excessively frequent visits increase hardware costs and impose a greater computational burden without significantly improving data quality or precision. For instance, parking turnover is typically higher between 6–8 AM and 5–7 PM, remains relatively stable from 8 AM to 5 PM, and is less frequent during other periods (Pang et al. 2023). Similarly, noise and air pollution levels fluctuate more significantly during rush hours or near high-traffic intersections, while remaining relatively stable during off-peak times. Therefore, determining a headway range, i.e., revisitation thresholds which are defined as the minimum and maximum time intervals between successive visits to a location, is essential for maintaining data reliability and avoiding redundant sensing, thereby improving overall resource efficiency. is essential to ensure data reliability and resource efficiency. In addi-

tion to establishing visit frequencies, an effective deployment strategy must optimize the routing of agents to maximize coverage while minimizing data redundancy. This study addresses two key questions: (1) how to determine the optimal number of agents required to effectively cover a designated urban area, and (2) how to identify agent routes that balance travel efficiency with data accuracy. In response, we develop an optimization model that minimizes total travel time by determining the optimal number of agents and their routes, while ensuring the required data collection frequency.

Previous research has explored the optimization of mobile sensor allocation in urban settings. For example, Hu et al. (2023) proposed a Partially Observable Markov Decision Process framework to optimize the path planning of an autonomous vehicle for collecting parking occupancy data. Their model considers a two-dimensional parking lot with a fixed number of spaces, assuming that the vehicle has access to a detailed parking lot map and is equipped with sensors capable of detecting occupancy within its field of view. The objective is to generate optimal paths for continuous and accurate occupancy estimation. However, this approach is limited to a controlled parking lot environment and does not address the deployment of multiple agents or the need for coordinated real-time data collection across larger urban areas. Another study by Pang et al. (2023) formulated an integer programming model to minimize the number of sensing units installed on city buses while ensuring that consecutive detections occur within a specified time threshold. Their approach assumes the availability of bus routes, schedules, and on-street parking geolocations through open data platforms. While the model effectively optimizes sensor allocation, it does not incorporate agent routing, thus limiting its applicability for dynamic real-time mobile data collection in complex urban environments.

Beyond parking, mobile sensor optimization techniques have also been applied to environmental monitoring tasks. Chen et al. (2022) applied reinforcement learning techniques to guide mobile sensors through pollution hotspots, and optimizing routes to maximize the collection of informative air quality data. Although effective for hotspot navigation, their method assumes real-time pollution gradient information, which may not always be readily available in practice. Additionally, optimization techniques such as mixed-integer

programming (Betalo et al. 2025) and genetic algorithms (Marques et al. 2020) have been employed to enhance the spatial deployment and routing of mobile sensors. However, these approaches often prioritize spatial coverage without explicitly considering the trade-offs between data freshness, agent mobility constraints, and real-time operational efficiency.

This research optimizes both the number of agents and their routes for mobile urban sensing (MUS). Our approach offers a scalable and cost-effective solution by minimizing agent count and travel time while maximizing spatial coverage. Unlike prior work, we introduce minimum and maximum revisitation thresholds to balance data accuracy with operational efficiency, ensuring that locations are monitored frequently enough without redundant visits that increase costs. These thresholds offer planners flexibility to tailor monitoring frequency based on specific urban needs. The proposed framework incorporates real-world mobility constraints, including network connectivity, travel time, and limits on the number of available agents, enabling practical deployment across diverse city environments. Our approach supports a range of sensing platforms, such as autonomous vehicles, shared mobility services, and dedicated monitoring fleets. By integrating sensor allocation with dynamic agent routing, the model provides a structured, real-time solution for mobile sensing in urban environments.

To improve computational performance for practical deployment, we propose a heuristic algorithm that reliably produces near-optimal solutions while significantly reducing runtime. The algorithm begins with k-means clustering to partition the city into distinct routes, ensuring each cluster contains at least two points. Within each cluster, the Nearest Neighbor Algorithm establishes an initial visitation sequence to minimize total travel distance. A linear resource allocation model then assigns the optimal number of agents to each route, subject to headway constraints and the total available agents. To further enhance efficiency, Lagrangian-inspired intra-route and inter-route swapping heuristics iteratively refine both route structures and agent distribution. Numerical experiments on the Solomon dataset demonstrate that the heuristic closely approximates exact solutions while achieving substantial reductions in computational time. Results show that increas-

ing the number of routes leads to a more balanced workload among agents, reducing total travel time while maintaining full spatial coverage. Additionally, the analysis quantifies the effects of headway constraints, revealing that relaxing maximum allowable headways decreases the required number of agents and total travel time by extending intervals between successive visits. Finally, we apply the proposed heuristic to parking facilities in the City of Toronto, demonstrating its scalability and practical applicability: the algorithm successfully covers all permit parking zones using fewer than 25% of the city's bike-share fleet.

The remainder of the chapter is organized as follows: Section 4.3 reviews the related literature. Section 4.4 presents the optimization model, and Section 4.5 introduces the heuristic algorithm. Section 4.6 reports numerical results based on the Solomon dataset, while Section 4.7 demonstrates the model's application to on-street permit parking in Toronto. Finally, Section 4.8 concludes the study and discusses directions for future research.

4.3 Literature Review

Mobile Urban Sensing has emerged as a decentralized and scalable approach for real-time data acquisition. Leveraging the proliferation of smartphones, sensor-equipped vehicles, and IoT devices, MUS enables continuous monitoring of diverse urban phenomena, including traffic dynamics, environmental pollution, infrastructure health, and public safety. Compared to traditional fixed-sensor networks, MUS provides enhanced spatial-temporal flexibility, faster deployment, and greater responsiveness to dynamic conditions. This section reviews advancements in MUS research across three dimensions: (1) foundational frameworks and architectural innovations for efficient data collection, (2) domain-specific applications in environmental sensing, mobility tracking, and parking detection, and (3) optimization strategies aimed at improving performance and scalability.

4.3.1 Frameworks and Strategies for Mobile Urban Sensing

Recent advancements in MUS focus on building scalable, decentralized frameworks that enhance real-time data acquisition, processing efficiency, and user privacy. To highlight early developments, Liu et al. (2019) provided a foundational review identifying core architectural models, participation strategies, and data quality control. Since MUS systems produce substantial volumes of data, ensuring efficient processing is essential for maintaining system performance. To address latency and alleviate the burden on central servers, Montori et al. (2017) proposed a collaborative framework for mobile sensors that incorporates edge computing to enable localized decision-making. To further enhance scalability and privacy, Ali et al. (2021b) discussed emerging trends such as blockchain and edge-based learning. Similarly, Zhang et al. (2023) explored distributed edge computing solutions for real-time environmental sensing, and Jones et al. (2024) investigated privacy-preserving techniques like differential privacy and secure multi-party computation in decentralized MUS platforms. DiChiappari et al. (2016) developed a real-time alert system combining cloud analytics, edge computing, and mobile agents to enable faster responses to pollution events.

An important area of research in MUS focuses on adaptive and geofencing-enabled sensing to dynamically adjust data collection strategies based on environmental and social contexts, thereby improving the relevance and quality of the collected data. Geofencing, creating virtual geographic boundaries that trigger actions when a device enters, exits, or dwells within a specific zone, plays a critical role in automating interactions and enhancing spatiotemporal coverage. Cheerkoot-Jalim and Purahoo (2020) introduced SenseAPP, an IoT-enabled platform that integrates geofencing and interactive user feedback to adapt sampling based on urban density and activity patterns. Similarly, Fiandrino et al. (2017) proposed an adaptive sensing framework that adjusts data collection frequency based on social density and environmental triggers, leveraging fog computing. Miranda et al. (2024) demonstrated that incorporating real-time geofenced triggers into sensing workflows significantly enhances data resolution, particularly for mobility and safety applications. Likewise, Cardone et al. (2014) combined activity recognition with geofencing to manage environmental monitoring during mass urban events like festivals and parades, where rapid shifts in crowd density demand adaptive sensing. Expanding further, Tasgaonkar et al. (2024) employed geofenced and vehicular telemetry-based systems to detect traffic incidents and urban anomalies, while Belwafi et al. (2022) surveyed UAV-based mobile sensing solutions that autonomously navigate within geofenced boundaries to collect real-time environmental data.

To address spatial and temporal gaps in user participation, hybrid MUS models integrate multiple data sources, such as fusing macro-level satellite imagery with micro-level mobile or IoT data. Santhosh et al. (2024) proposed a hybrid platform that combines satellite and smartphone data to monitor urban heat, air quality, and congestion. Gkoumas et al. (2023) deployed micro-mobility units augmented with geospatial satellite data for infrastructure scanning, while Liu et al. (2024) fused satellite imagery with crowd-sourced traffic data to enhance urban mobility analysis. Similarly, Mohamed et al. (2024) integrated multimodal sensor data from drones, vehicles, and wearable devices to improve environmental and mobility monitoring. Lu et al. (2023) integrated telemetry data from bike-mounted and vehicle-mounted sensors. Simulation results indicate that multimodal

strategies significantly outperform single-mode systems in terms of coverage efficiency and detection times.

4.3.2 Applications of Mobile Urban Sensing

MUS is an effective tool for tracking mobility, air quality, noise pollution, and thermal hotspots in real time. Longo et al. (2017) and Cardone et al. (2014) pioneered MUS systems where users contribute air and noise pollution data through mobile applications. Nemati et al. (2017) provided a critical review of smartphone-based air sensing platforms, highlighting challenges related to battery consumption, data fidelity, and privacy. Liu et al. (2021) demonstrated how public buses equipped with sensors can collect pollution data across entire urban networks, offering scalable solutions for city-wide air quality mapping. Zappatore et al. (2017) developed an MUS platform that generates real-time urban noise maps, supporting zoning policy enforcement through distributed smartphone-based sensing. In the context of infrastructure monitoring, Matarazzo et al. (2022) leveraged smartphone-mounted sensors to detect structural stress via vibration data, demonstrating the potential of MUS for monitoring urban infrastructure.

MUS has also been applied to real-time transportation analysis. Darsena et al. (2023) reviewed sensing frameworks designed to estimate public transport occupancy, integrating geofencing and user behavior recognition for adaptive scheduling and load balancing. Rodrigues (2020) developed a smartphone-based MUS architecture to infer urban mobility patterns with minimal privacy intrusion. Meanwhile, Hasan (2023) presented an MUS platform that deploys real-time alerts for pedestrians in risk-prone zones using geofenced location verification. Parking detection represents one of the earliest and most practical applications of MUS in transportation. Liao et al. (2016) and Roman et al. (2018) proposed vehicle-mounted sensor systems to identify roadside parking availability. Mathur et al. (2010) introduced ParkNet, a system leveraging ultrasonic and GPS-equipped taxicabs to map parking occupancy city-wide. Shah et al. (2021) integrated geofencing with automated event detection by leveraging smartphone localization services, including GPS, Wi-Fi, and cellular data, to monitor and record vehicle entry and exit events across vir-

tual parking zones. Bock et al. (2019) demonstrated that GPS traces from ride-hailing vehicles could provide reliable parking occupancy estimates across spatially diverse urban regions, a finding further validated by Cherian et al. (2016) and Liniger and Stiller (2015) in structured parking facilities.

4.3.3 Optimization Approaches for Mobile Urban Sensing

Optimization techniques support the intelligent deployment of MUS by enabling efficient path planning, determining the optimal number of agents, scheduling data collection, and coordinating system operations to maximize performance and resource utilization. Wang et al. (2024) provided a comprehensive survey of algorithmic strategies for MUS, focusing on optimization under constraints such as spatial heterogeneity and limited resources. Their review classifies methods by sensing objectives and agent coordination styles, covering reinforcement learning, task allocation strategies, and federated learning approaches.

Within the context of smart mobility and environmental sensing, Lin et al. (2017) examined optimization challenges in smart parking systems, focusing on dynamic sensor allocation to adapt to fluctuating demand patterns. The study highlighted that optimizing sensor placement and resource allocation significantly improves the efficiency and scalability of smart mobility solutions. Complementarily, Radmanesh and Bani Younes (2018) explored Unmanned Aerial Vehicle (UAV)-based path planning for urban sensing in cluttered, obstacle-dense environments, reviewing motion planning techniques such as potential fields, genetic algorithms, and graph-based methods. Rather than presenting experimental results, their review highlights the trade-offs among different approaches and emphasizes that hybrid strategies may offer the most robust solutions for real-world aerial navigation.

Hu et al. (2023) optimized path planning for autonomous vehicles conducting parking occupancy sensing. They applied a Partially Observable Markov Decision Process (POMDP) framework to balance exploration and exploitation under uncertainty in a controlled parking lot. Their policy prioritized high-interest areas while minimizing re-

dundant paths, improving detection accuracy and operational efficiency compared to fixed-route models. Nevertheless, their evaluation was limited to a controlled environment, which may not capture the full complexity and unpredictability of real-world urban settings, such as dynamic obstacles, variable parking behaviors, and inconsistent sensing conditions. Pang et al. (2023) minimized the number of sensors needed for MUS to collect on-street parking data. They applied integer programming to optimize sensor placement and ensure full coverage within revisit constraints, using public buses as mobile sensing agents. However, their method assumed static bus schedules and fixed routes, limiting adaptability in dynamic urban settings where route changes, delays, or traffic variability could affect sensing performance.

While optimization techniques for mobile urban sensing have been extensively applied to parking occupancy detection, parallel efforts have emerged in other domains such as air quality and noise pollution monitoring. In mobile air sensing, optimization strategies focus on routing agents to maximize spatial-temporal coverage of pollution hotspots, often using adaptive sampling or reinforcement learning methods to prioritize areas of high variability (Lameski et al. 2023, Fadhel et al. 2024). Similarly, in noise pollution monitoring, dynamic path planning and clustering techniques are employed to optimize coverage in high-risk zones such as construction areas and traffic intersections (Thakuriah et al. 2020). Broader environmental sensing initiatives involving multiple data streams (e.g., air quality, noise, temperature) further introduce multi-objective optimization challenges, including balancing energy consumption, sampling frequency, and spatial coverage (Rowley and Karakuş 2023).

In contrast to prior work, this study presents a comprehensive model that simultaneously optimizes the number of agents and their dynamic routing paths, given the data collection schedule. By modeling the city as a graph, it enables geofencing, allowing data to be collected within defined spatial boundaries, which improves operational efficiency, ensures coverage compliance, and reduces unnecessary resource usage. The model applies an exact optimization approach and then develops a heuristic method to enhance computational scalability, supporting deployment across large urban networks under real-

istic operational constraints. Table 4.1 compares recent MUS studies by sensing domain and technical features. Although all use mobile sensing platforms, they differ in geofencing, optimization methods, and city-scale deployments. This comparison highlights the distinct contributions of the present study.

Table 4.1: Comparative analysis of MUS studies across application domains and key technical dimensions. **Geo.** indicates the use of geospatial triggers or zone-based sensing, geofencing. **Opt. Agents** denotes studies that optimize the number of deployed sensing agents. **Opt. Routing** refers to optimization of agent trajectories or patrol paths. **City Dep.** indicates deployment at city-wide scale or inclusion of large-scale simulation.

Study	**Application Domain**	**Geo.**	**Opt. Agents**	**Opt. Routing**	**City Dep.**
Liu et al. (2019)	General MCS Framework	✗	✗	✗	✓
Montori et al. (2017)	Pollution and IoT Monitoring	✗	✗	✗	✓
Santhosh (2024)	Hybrid Heat / Air Monitoring	✓	✗	✗	✓
Zappatore et al. (2017)	Urban Acoustic Monitoring	✓	✗	✗	✓
Matarazzo et al. (2022)	Infrastructure Vibration	✗	✗	✗	✓
Fiandrino et al. (2017)	Fog Computing + Pollution	✗	✓	✗	✗
Rodrigues (2020)	Mobility Flow Tracking	✓	✗	✗	✓
Darsena et al. (2023)	Public Transit Occupancy	✓	✗	✓	✓
Lu et al. (2023)	Multimodal Parking Detection	✗	✗	✓	✓
Bastos et al. (2023)	Adaptive Parking Allocation	✗	✗	✓	✓
Wang et al. (2024)	Cross-domain Urban Sensing	✓	✓	✓	✓
Hasan (2023)	Pedestrian Safety Monitoring	✓	✗	✗	✗
This Study	General Urban Data Collection	✓	✓	✓	✓

4.4 Optimization Model for Tour-Based Sensing

Consider an urban environment comprising a set of discrete locations that need to be visited at a defined frequency. Every visit allows the agent to collect data from that location, e.g., parking occupancy from visited parking facilities. The operational cost of agents increases, and the data quality (in terms of predictive confidence level) decreases with the visit frequency. Therefore, in principle, there exists an optimal visit frequency within allowable limits. In addition to the visit frequency, the visits are conducted by agents in the form of tours, such that each location is visited once and in one tour, and the frequency of that tour is inversely related to the duration of completing a lap. From hereon, we use the term "facility" instead of location for convenience.

We formulate the sensing operation as an optimization problem aimed at minimizing the total travel time of agents while satisfying visit frequency constraints. The urban environment is represented as a directed graph $\mathcal{G} = (V, E)$, where V denotes the set of facilities and E the set of road segments connecting them. Travel between any two facilities $i, j \in V$ incurs a symmetric travel time, denoted by D_{ij}, such that $D_{ij} = D_{ji}$. The sensing operation is conducted in K closed routes, each covering a subset of facilities. Each facility must be visited on a given frequency, with inter-visit intervals constrained between a minimum and maximum headway, denoted by $H_{\min}$ and $H_{\max}$, respectively. A fleet of F agents are present to perform the data collection.

Model (4.1) captures the sensing agent routing and scheduling problem, incorporating the operational constraints and objectives described above. Indices $i, j \in \{1, \ldots, N\}$ refer to nodes (parking facilities) in the graph $\mathcal{G}$, and $k \in \{1, \ldots, K\}$ indexes the sensing routes. The objective Z denotes the total weighted travel time of agents across all routes. The binary variable x_{ij}^k indicates whether an agent on route k travels directly from node i to node j, while y_i^k indicates whether node i is included in route k. The variable f^k represents the number of agents assigned to route k, and M is a large constant used to enforce logical constraints.

$$\text{Minimize} \quad Z = \sum_{k=1}^{K} \left(\sum_{i=1}^{N} \sum_{j=1}^{N} D_{ij} \cdot x_{ij}^{k} \cdot f^{k} \right) \tag{4.1a}$$

Subject to:

$$\sum_{k=1}^{K} y_i^k = 1, \quad \forall\, i \tag{4.1b}$$

$$\sum_{i=1}^{n} y_i^k \geq 2, \quad \forall\, k \tag{4.1c}$$

$$\sum_{j \neq i} x_{ij}^k = \sum_{j \neq i} x_{ji}^k = 1, \quad \forall\, i, k \tag{4.1d}$$

$$\frac{\sum_{i,j} D_{ij} x_{ij}^k}{H_{\max}} \leq f^k \leq \frac{\sum_{i,j} D_{ij} x_{ij}^k}{H_{\min}}, \quad \forall\, k \tag{4.1e}$$

$$y_i^k \geq \frac{1}{2} \sum_{j} (x_{ij}^k + x_{ji}^k), \quad \forall\, i, k \tag{4.1f}$$

$$\sum_{k=1}^{K} f^k \leq F \tag{4.1g}$$

$$\sum_{i=1}^{N} r_i^k = 1, \quad \forall k \tag{4.1h}$$

$$t_i^k = 1 + M(1 - r_i^k), \quad \forall i, k \tag{4.1i}$$

$$t_i^k + 1 \leq t_j^k + M(1 - x_{ij}^k), \quad \forall i, j, k \tag{4.1j}$$

$$1 \leq t_i^k \leq N, \quad \forall i, k \tag{4.1k}$$

$$t_i^k \in \mathbb{Z}_{\geq 1}, \quad \forall i, k \tag{4.1l}$$

$$x_{ij}^k, y_i^k, r_i^k \in \{0, 1\}, \quad \forall i, j, k \tag{4.1m}$$

$$f^k \in \mathbb{Z}_{\geq 1}, \quad \forall\, k. \tag{4.1n}$$

Constraint (4.1b) ensures that each facility is included in exactly one route. Constraint (4.1c) ensures that every route covers at least two facilities to form a valid path. Constraint (4.1d) enforces flow continuity, requiring each visited node to have one incoming and one outgoing arc. Constraint (4.1e) links route length and frequency, enforcing that the allocated number of agents per route maintains the required headway between facility observations. Constraint (4.1f) guarantees node inclusion consistency, if a node is traversed, it must be marked as part of the route. Finally, Constraint (4.1g) limits the total number of deployed agents across all routes to not exceed the available fleet F.

To eliminate sub-tours and enforce sequential visits, we introduce the variable t_i^k, which represents the position of node i in the visiting sequence of route k. To establish a clear starting point for each route, we define a root node, a node from which the traversal of route k begins. This root is selected using the binary variable r_i^k, where $r_i^k = 1$ if node i is chosen as the root of route k, and $r_i^k = 0$ otherwise. Constraint (4.1h) ensures that

exactly one root node is selected for each route. Constraint (4.1i) ensures that if node i is selected as the root of route k (i.e., $r_i^k = 1$), then its visit order variable t_i^k is set to 1, indicating that node i is the starting point of route k. Sequential consistency is enforced by Constraint (4.1j), which ensures that if route k includes an arc from node i to node j, then node j must be visited after node i. Constraint (4.1k) bounds the visit order variables within the valid range of node indices, and Constraint (4.1l) restricts these variables to be positive integers. Constraint (4.1n) ensures that the number of agents assigned to each route k is a positive integer.

The proposed mathematical model is nonlinear due to the term $x_{ij}^k \cdot f^k$ in the objective function, which involves the product of a binary and an integer decision variable. This nonlinearity prevents the use of standard mixed-integer linear programming (MILP) solvers, and the model must be linearized to ensure computational tractability. To address this, an auxiliary variable $w_{ij}^k \in \mathbb{Z}_{\geq 0}$ is introduced to represent the product $x_{ij}^k \cdot f^k$. The variable w_{ij}^k captures the effective weighted flow on arc (i, j) in route k, accounting for the number of agents assigned to that route. This linearization approach is a widely used technique in MILP to handle products of binary and integer variables (see, e.g., Wolsey (1998)). It ensures that $w_{ij}^k = f^k$ when $x_{ij}^k = 1$, and $w_{ij}^k = 0$ when $x_{ij}^k = 0$. To ensure the correct representation of w_{ij}^k, the following constraints are introduced in the model:

$$w_{ij}^k \leq x_{ij}^k \cdot F, \quad \forall\, i, j, k, \tag{4.2a}$$

$$w_{ij}^k \leq f^k, \quad \forall\, i, j, k, \tag{4.2b}$$

$$w_{ij}^k \geq f^k - (1 - x_{ij}^k) \cdot F, \quad \forall\, i, j, k, \tag{4.2c}$$

$$w_{ij}^k \in \mathbb{Z}_{\geq 0}, \quad \forall\, i, j, k. \tag{4.2d}$$

Constraints (4.2a)–(4.2d) ensure the intended product behavior: when $x_{ij}^k = 0$, constraints (4.2a) and (4.2c) force $w_{ij}^k = 0$; when $x_{ij}^k = 1$, constraints (4.2b) and (4.2c) imply $w_{ij}^k = f^k$. The original bilinear term $x_{ij}^k \cdot f^k$ in the objective is then replaced by w_{ij}^k, and the linear constraints (4.2a)–(4.2d) are added to the model to obtain the linearized

formulation:

$$\min \quad Z = \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{j=1}^{N} D_{ij} \cdot w_{ij}^{k} \tag{4.3a}$$

$$\text{s.t.} \quad \text{Constraints (4.1b)–(4.1n)}, \tag{4.3b}$$

$$w_{ij}^{k} \leq x_{ij}^{k} \cdot F, \quad \forall\, i, j, k, \tag{4.3c}$$

$$w_{ij}^{k} \leq f^{k}, \quad \forall\, i, j, k, \tag{4.3d}$$

$$w_{ij}^{k} \geq f^{k} - (1 - x_{ij}^{k}) \cdot F, \quad \forall\, i, j, k, \tag{4.3e}$$

$$w_{ij}^{k} \geq 0, \quad \forall\, i, j, k, \tag{4.3f}$$

The revised agent routing and scheduling model (4.3) includes three binary and three integer variables. Specifically, the model uses binary variables to represent pairwise transitions between nodes, route-facility assignments, and root node selections. Integer variables are used to capture visiting sequences, number of agents assigned to each route, and the auxiliary variable to linearize the model.

The computational complexity of the model scales with the number of facilities N and the number of routes K, and is driven largely by the binary and integer components of the formulation. Table 4.2 summarizes the asymptotic growth of the variable sets and constraint groups. The most computationally intensive elements are the binary transition variables and the sequencing-related integer variables, which together lead to a complexity of order $\mathcal{O}(K \cdot N^2)$. These components contribute to a combinatorial explosion in the solution space, especially as N increases.

Integer programming is NP-hard in general, and the presence of routing-related binary variables (e.g., for sequencing and sub-tour elimination) significantly increases the solution complexity. Even for moderately sized instances, solving such models can be computationally intensive due to the need to explore a vast number of discrete configurations. While the proposed model remains tractable for small- to medium-sized networks, real-world applications involving large urban areas require a faster approach to achieve reasonable runtime.

Table 4.2: Asymptotic Complexity of Model (4.1)

Component	**Description**	**Asymptotic Count**
Decision Variables		
x_{ij}^k	Travel arc from node i to j in route k	$\mathcal{O}(K \cdot N^2)$
y_i^k	Node i included in route k	$\mathcal{O}(K \cdot N)$
r_i^k	Node i is root of route k	$\mathcal{O}(K \cdot N)$
f^k	Agents assigned to route k	$\mathcal{O}(K)$
t_i^k	Visit order of node i on route k	$\mathcal{O}(K \cdot N)$
w_{ij}^k	Linearization of $x_{ij}^k \cdot f^k$	$\mathcal{O}(K \cdot N^2)$
Constraints		
Routing logic	Node assignment, flow continuity	$\mathcal{O}(K \cdot N)$
Subtour elimination	Order preservation for tour feasibility	$\mathcal{O}(K \cdot N^2)$
Linearization rules	Ensure correct agent-travel interaction	$\mathcal{O}(K \cdot N^2)$
Frequency bounds	Visit interval control per route	$\mathcal{O}(K)$
Total Model Complexity	Dominated by arc and routing interactions	$\mathcal{O}(K \cdot N^2)$

4.5 Heuristic Algorithm

Exact solutions are computationally expensive and time-consuming, particularly for real-world instances involving a large number of facilities N and routes K. To address this scalability challenge, a heuristic algorithm is proposed to efficiently generate high-quality approximate solutions.

The algorithm begins by applying k-means clustering to partition the set of facilities into K spatially coherent clusters. This step reduces problem complexity by enabling localized route planning within each cluster. To ensure feasibility, each cluster is required to contain at least two facilities, thereby avoiding configurations that are too small to support valid route construction. If any cluster contains only a single facility, the clustering process is re-initialized until a valid partition is achieved. Once clustering is complete, an initial route is constructed within each cluster using a Nearest Neighbor (NN) algorithm. To reduce sensitivity to the starting point, a known limitation of NN, routes are generated from each node in the cluster. For each starting node, a complete tour is built, and among these route candidates, the one with the shortest total travel distance is selected.

4.5.1 Node Swapping Procedure

After constructing initial routes, the algorithm enters a two-step node swapping phase to globally refine the solution. This phase involves both intra-cluster route adjustments and inter-cluster node reassignments. Guided by a Lagrangian relaxation framework, the algorithm evaluates the effect of moving nodes within or between clusters with the goal of minimizing overall travel time while satisfying headway and feasibility constraints. Dual variables associated with connectivity and visit order constraints quantify the marginal cost of constraint violations and help identify high-impact node movements. Nodes are then iteratively reassigned to further refine the routing plan, ensuring compliance with all model constraints.

4.5.1.1 Intra-cluster Route Optimization

The intra-cluster optimization phase improves the initial NN-based routes by reordering nodes within each cluster to minimize travel distance and Lagrangian penalties, while maintaining connectivity and feasibility. To achieve this, the classic 2-Opt local search algorithm is applied, a widely used technique in route optimization problems such as the Traveling Salesman Problem (TSP) (Croes 1958). In 2-Opt, the algorithm iteratively examines all pairs of non-adjacent edges in a given route. For each pair, it removes the two edges and reconnects the segments by reversing the order of the intermediate nodes. This process can eliminate edge crossings and shorten the route. After each potential swap, the total route distance is recalculated. If the swap results in a shorter route, the update is accepted. The process continues until no further improvement is observed or a predefined limit of 1000 iterations is reached. This refinement step ensures that intra-cluster routing is as efficient as possible before any inter-cluster changes are made. Without this, global improvements could be based on suboptimal local routes, reducing the overall solution quality.

To illustrate the intra-cluster optimization process, a simplified example with thirty dummy nodes is presented. These nodes are first clustered into three groups using k-means. Figure 4.1 shows the clustering result, where each cluster is indicated by a distinct marker shape.

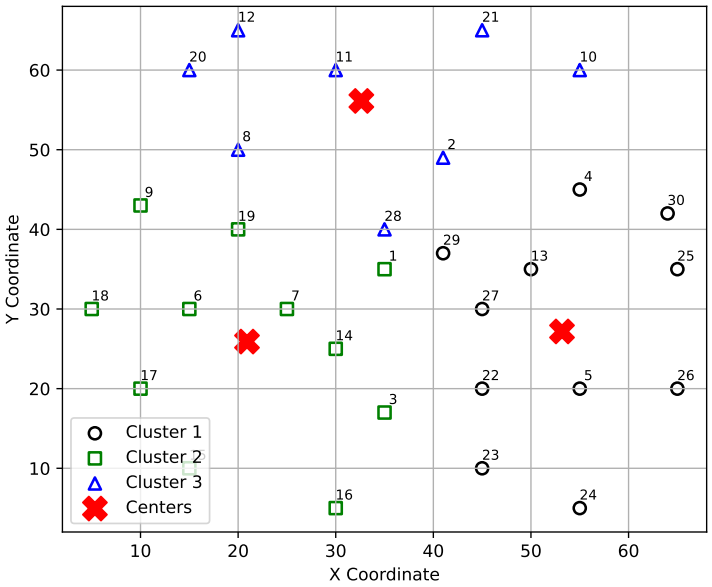


Figure 4.1: Clustering of 30 spatial points into three groups using k-means clustering. Each cluster is represented with a distinct marker shape and cluster centroids are marked with X symbols.

Node labels correspond to their dataset indices, and cluster centroids are marked with **X** symbols. The k-means algorithm forms spatially compact groups by minimizing the distance between nodes and their respective centroids. Figure 4.2 demonstrates how intra-cluster optimization through reordering nodes within clusters can enhance route efficiency while preserving feasibility. Arrows indicate the direction of travel. In Panel (a), initial routes are constructed using the NN algorithm within each cluster, which results in paths with unnecessary edge crossings and suboptimal ordering. Panel (b) shows the result of applying the 2-Opt algorithm within each cluster, which systematically removes edge crossings and reconnects segments in a more efficient order. As a result, the total travel distance is reduced from 381.87 units to 376.56 units. This example highlights how intra-cluster optimization improves the initial routing solution produced by NN and prepares the clusters for potential improvements through inter-cluster node swapping.

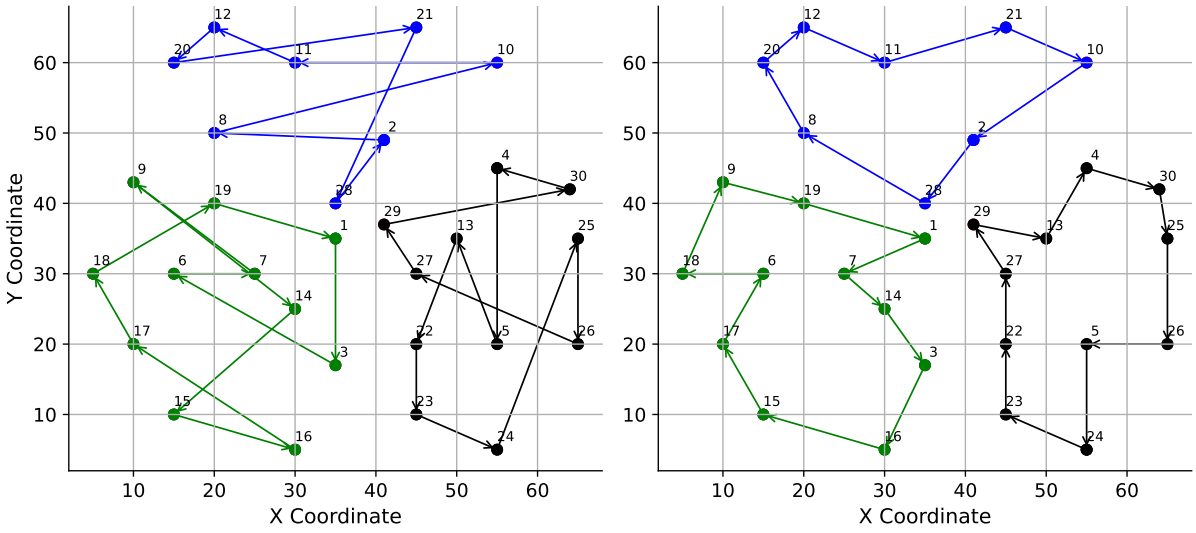


Figure 4.2: An example of intra-cluster node swapping. (a) The initial route generated by the NN algorithm. (b) The optimized route after applying the 2-Opt algorithm.

4.5.1.2 Inter-cluster Node Reassignment

The inter-cluster optimization phase aims to improve overall routing efficiency by reassigning nodes between clusters to reduce the total travel distance of agents. While the initial k-means clustering groups nodes based on proximity to centroids, it does not consider route continuity or visit order. Consequently, nodes near cluster boundaries or those that disrupt internal route flow may introduce inefficiencies, even if they are spatially close to a centroid. Inter-cluster swapping addresses this limitation by re-evaluating node assignments based on their contribution to route-level efficiency. Nodes that contribute less efficiently to intra-cluster routing or lie near the edge of a cluster may be reassigned to neighboring clusters where their inclusion results in shorter overall paths. This adjustment better aligns spatial grouping with actual routing performance, an alignment that centroid-based clustering alone cannot guarantee.

To implement this refinement, the algorithm performs a greedy search over all pairs of clusters, systematically evaluating possible node swaps. For each unique cluster pair, it examines all node pairs, one from each cluster, and temporarily exchanges them. After each trial swap, the updated routes are recalculated within the affected clusters, and the total travel distance is compared to the original. If the swap yields a strictly lower total distance, it is accepted; otherwise, the nodes revert to their original clusters.

The swap procedure is governed by two stopping criteria designed to balance optimization quality with computational efficiency. The first criterion limits the total number of improvement iterations, where each iteration involves evaluating possible node reassignments between clusters. This prevents the algorithm from running indefinitely and promotes timely convergence. The second criterion restricts the number of accepted swaps allowed within each iteration. By limiting the extent of changes per cycle, the algorithm avoids excessive reassignments that could compromise previously optimized intra-cluster routes. The process terminates early if no beneficial swaps are identified in an iteration or if both stopping conditions are met, ensuring that the refinement step remains both effective and computationally manageable. Figure 4.3 illustrates an example of inter-cluster optimization process, building upon the intra-cluster optimization results shown in Figure 4.2.

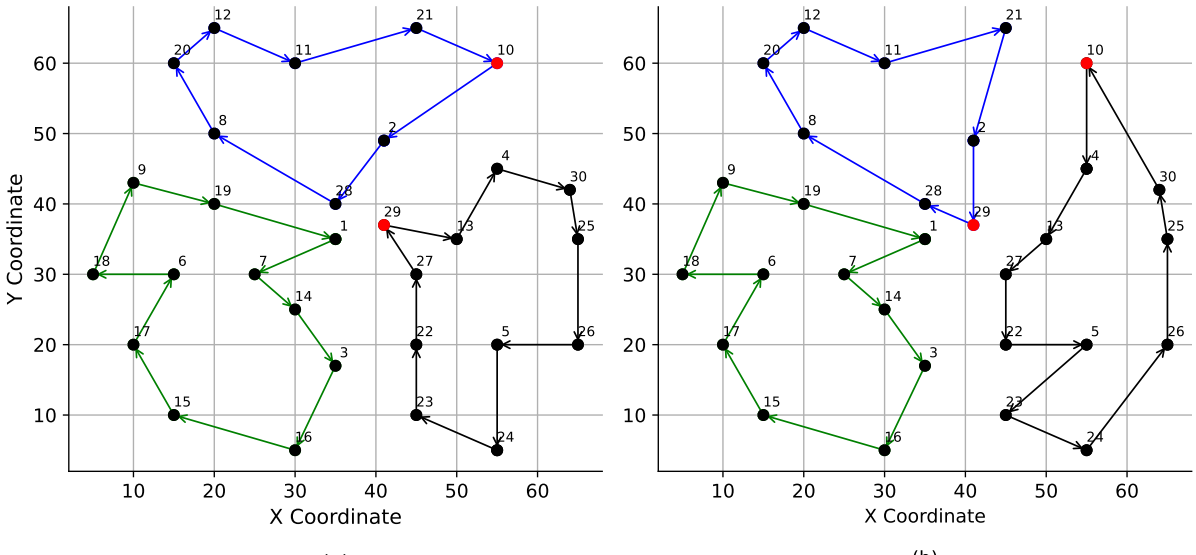


Figure 4.3: An example of inter-cluster swapping. Panel (a) shows the route configuration after the 2-Opt step. Panel (b) illustrates the result of swapping Node 10 from Cluster 3 with Node 29 from Cluster 1.

In Panel (a), Node 10 is assigned to Cluster 3 and Node 29 to Cluster 1. While locally efficient, this configuration remains globally suboptimal. Panel (b) shows the outcome after swapping the two nodes across clusters, resulting in a 30.52-unit reduction in total travel time. The complete swapping process that generates the route configurations is detailed in Algorithm 1.

Algorithm 1 Node Swapping Procedure

Input: Initial route assignments from Nearest Neighbor algorithm
Output: Refined route assignments
(a) Intra-route Swapping
for each route k **do**
    repeat
        Apply 2-Opt to reorder nodes within route k
        Update travel time for the route
    until convergence is achieved **or** maximum number of iterations is reached
end for
(b) Inter-route Swapping
repeat
    for each pair of routes (k, k') **do**
        for each pair of nodes $(i \in k, j \in k')$ **do**
            Temporarily swap nodes i and j
            Recalculate total travel time of affected routes
            if swap reduces total travel time **then**
                Accept the swap
            else
                Revert to original assignment
            end if
        end for
    end for
until no further improvement is observed, the maximum number of iterations is reached, or the allowed number of swaps in an iteration has been exceeded
Return: Updated route assignments

After the node swapping procedure is completed and the routes are finalized, the routing configuration yields fixed values for y_i^k and x_{ij}^k, indicating whether node i is covered by route k, and whether edge (i, j) is included in route k, respectively. These binary variables, now treated as known parameters, define the finalized structure of each route. Based on this configuration, the following linear resource allocation model is solved to determine the optimal number of sensing agents (agents) assigned to each route. Model (4.4) ensures that the headway constraints at each node are satisfied and that the total number of deployed agents does not exceed the available fleet capacity.

$$\min \quad Z = \sum_{k} \left(\sum_{i} \sum_{j} D_{ij} \cdot x_{ij}^{k} \cdot f^{k} \right) \tag{4.4a}$$

$$\text{s.t.} \quad \frac{1}{H_{\max}} \sum_{i=1}^{n} \sum_{j=1}^{n} D_{ij}\, x_{ij}^{k} \leq f^{k} \leq \frac{1}{H_{\min}} \sum_{i=1}^{n} \sum_{j=1}^{n} D_{ij}\, x_{ij}^{k}, \quad \forall\, k, \tag{4.4b}$$

$$\sum_{k} f^{k} \leq F, \tag{4.4c}$$

$$f^{k} \in \mathbb{Z}_{\geq 1}, \quad f^{k} \leq F, \quad \forall\, k. \tag{4.4d}$$

Since the routing structure x_{ij}^{k} is fixed, the model reduces to an integer linear program defined over the decision variables f^{k}, allowing for efficient computation. Establishing the route configuration prior to resource allocation simplifies the problem and enables rapid determination of the optimal agent distribution. The complete heuristic procedure, including clustering, route construction, node swapping, and resource optimization, is summarized in Algorithm 2.

The proposed heuristic reduces computational complexity by dividing the problem into sequential stages: clustering, route generation, and iterative refinement. Within this framework, the node swapping phase plays a critical role in enhancing solution quality through targeted local adjustments. By incorporating both intra- and inter-route swapping heuristics, the algorithm incrementally improves route structure and agent allocation while maintaining tractable computational requirements.

Algorithm 2 Agent Allocation

Input: Total agents F, number of routes K, headway bounds $H_{\min}$, $H_{\max}$, and dataset df with node coordinates (X, Y)
Output: Agent allocation per route f^k, final route assignments x_{ij}^k, y_i^k
Step 1: Clustering
repeat
    Partition nodes into K clusters using k-means
    if any cluster has fewer than 2 nodes **then**
        Merge with nearest neighbor or reinitialize
    end if
until all clusters have at least 2 nodes
Step 2: Initial Route Construction
for each route k **do**
    Compute travel time matrix D_{ij}
    Construct route using Nearest Neighbor (NN)
end for
Step 3: Swapping Phase
Refine routes using Swapping Procedure (Alg. 1)
Step 4: Agent Allocation
Solve model (4.4) to find optimal f^k for each route
Return: Optimal f^k, and route assignments x_{ij}^k, y_i^k

4.6 Analysis

This section presents a comprehensive evaluation of the proposed heuristic algorithm and its comparison with the exact optimization model. We begin with a baseline assessment on small-scale instances to validate the exact model. We then compare performance across scenarios, analyze heuristic scalability, and conduct sensitivity analyses on key design parameters such as revisit constraints and route count. The section concludes with visual illustrations of heuristic-based routing patterns.

4.6.1 Experimental Setup and Dataset Description

The $R1$ dataset, introduced by Solomon (1987), is a standard benchmark for evaluating and comparing routing algorithms. It consists of 100 randomly distributed points within a square area, with distances computed using the Euclidean metric. This dataset is utilized in our numerical experiments to assess the performance of both the exact optimization model and the proposed heuristic approach.

4.6.2 Baseline Evaluation: Exact Optimization Results

To establish a performance baseline and better understand how the optimization model responds to different revisit and clustering constraints, we first examine exact solutions on a small, tractable subset of the problem. Specifically, the model is applied to the first ten nodes of the Solomon dataset and evaluated under varying values of the number of clusters K and maximum headway $H_{\max}$.

Figure 4.4 illustrates the optimal routing configurations and the corresponding number of sensing agents (agents) assigned to each route, as determined by the exact optimization model.

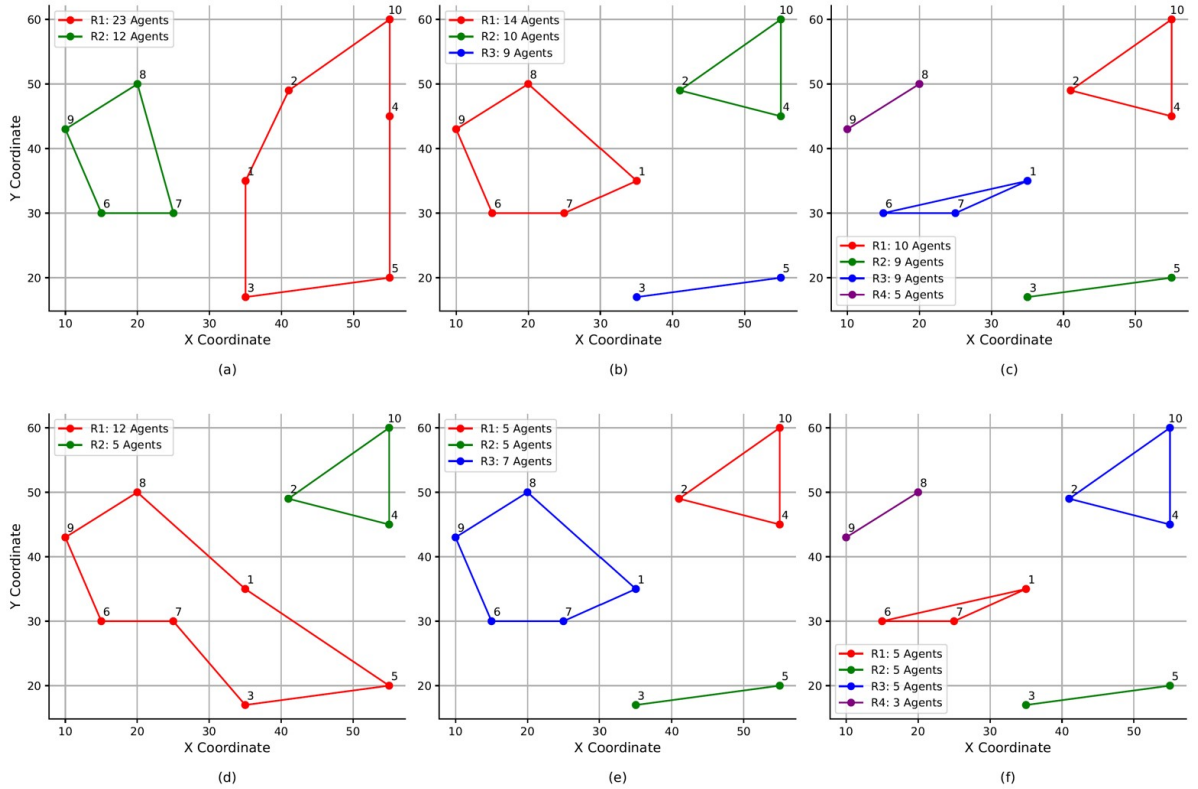


Figure 4.4: Visualization of optimization results using the first 10 points of the dataset. Subfigures (a)–(c): $K = 2, 3, 4$, $H_{\max} = 5$; (d)–(f): $K = 2, 3, 4$, $H_{\max} = 10$.

In each row of the figure, the number of clusters K increases from left to right, while $H_{\max}$ remains fixed. From top to bottom, the value of $H_{\max}$ increases, enabling comparisons across different revisit frequency constraints.

Panels (a)–(c) correspond to $H_{\max} = 5$, while panels (d)–(f) correspond to $H_{\max} = 10$. As the number of clusters increases from 2 to 4 in each row, the routes become more spatially localized, and the number of agents per route decreases. This reflects improved spatial partitioning and reduced travel burden per agent. Meanwhile, increasing $H_{\max}$ from the top to the bottom row allows greater flexibility in scheduling, reducing the total number of required agents and, in some cases, such as panels (a) versus (d), prompting a reconfiguration of routes that further lowers total travel time.

These results illustrate how the optimization model balances agent deployment and route design under competing operational constraints, and they serve as a reference point for evaluating the performance of the proposed heuristic approach in larger instances.

4.6.3 Heuristic Validation: Comparison Against Exact Model

To evaluate the effectiveness of the proposed heuristic algorithm, we compare its performance against the exact optimization model across a range of experimental settings. Table 4.3 summarizes the results, including key performance metrics such as the total number of agents assigned, total travel time, computational runtime, and the relative speedup achieved by the heuristic.

The comparison highlights the influence of problem parameters, particularly the revisit constraints $H_{\min}$ and $H_{\max}$, the number of routes K, and the number of nodes N, on optimization outcomes. As $H_{\max}$ increases, allowing longer intervals between visits, the number of assigned agents and total travel time decreases, since fewer agents are needed to satisfy less frequent monitoring requirements. Increasing the number of routes K improves spatial coverage and reduces travel distance by enabling more localized assignments. However, as the number of nodes N grows, the computational time required by the exact model increases rapidly, often becoming impractical for real-time applications. In contrast, the heuristic algorithm remains computationally efficient across all scenarios. The heuristic achieves substantial runtime reductions, often by several orders of magnitude, compared to the exact model. This speedup increases with problem size and complexity. While the exact model scales exponentially, the heuristic algorithm maintains polynomial-time performance and produces near-optimal solutions that closely approximate the exact results. This makes it particularly well-suited for large-scale or time-sensitive deployments in urban environments. Figure 4.5 further illustrates the comparison by plotting the total number of assigned agents $\sum_k f^k$ and the total travel time Z for both approaches. The plots confirm that the heuristic algorithm delivers solutions with minimal deviation from the exact model while drastically reducing computation time. This trade-off between solution quality and efficiency makes the heuristic a practical tool for real-time decision-making, large-scale simulations, and sensitivity analyses where extensive parameter tuning may be required.

Table 4.3: Comparison of optimization results between the exact model and heuristic algorithm, evaluating agent allocation, travel time, computational efficiency, and speedup across various scenarios.

Scenario	N	K	$H_{\min}$	$H_{\max}$	**Exact Model**			**Heuristic Algorithm**			
					$\sum_k f^k$	Z	Run Time (sec)	$\sum_k f^k$	Z	Run Time (sec)	Speedup (%)
1	10	1	1	3	55	9040.2	21	55	9040.2	0.34	209838
2	10	1	1	5	33	5425.2	33	36	5424.1	0.11	329967
3	10	1	1	10	17	2794.8	32	17	2794.2	0.11	319966
4	12	1	1	3	62	11477.4	64	62	11475.0	0.14	639978
5	12	1	1	5	38	7034.6	247	38	7033.1	0.13	2469995
6	12	1	1	10	19	3617.3	460	19	3516.5	0.13	4599997
7	14	1	1	3	64	12144.6	253	68	13861.1	0.19	2529992
8	14	1	1	5	38	7210.9	482	41	8357.4	0.23	4819995
9	20	1	1	3	85	21471.0	1231	91	24814.3	0.22	12309998
10	10	2	1	3	57	5307.1	752	62	5634.5	0.21	7519997
11	10	3	1	5	33	1797.6	1082.73	36	2110.3	0.22	10827298

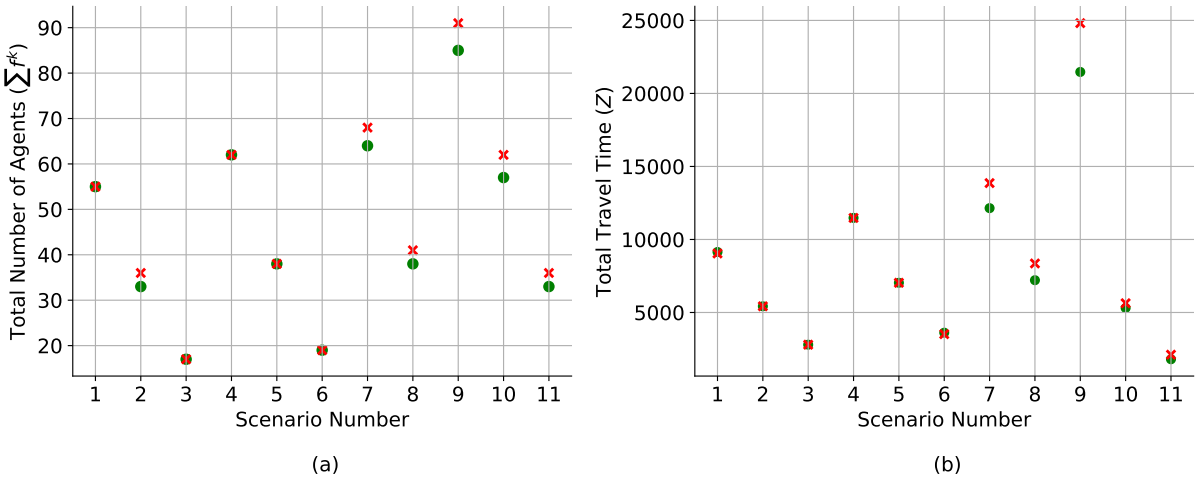


Figure 4.5: Comparison of agent count $\sum_k f^k$ and total travel time Z for exact and heuristic methods. The green circles represent the results obtained from the exact model, while the red crosses indicate the results produced by the heuristic algorithm.

4.6.4 Scalability and Efficiency of the Heuristic Algorithm

To evaluate the scalability of the proposed heuristic algorithm, we apply it to a larger problem instance consisting of the first 50 nodes from the Solomon dataset. Figure 4.6 illustrates the optimized routes produced under varying numbers of clusters K, while Table 4.4 reports the corresponding number of sensing agents (agents) assigned to each route.

The results demonstrate that the heuristic consistently generates spatially efficient and well-balanced routing configurations across a wide range of cluster counts. Moreover,

the algorithm maintains high computational efficiency as problem size increases, enabling practical deployment in large-scale urban environments.

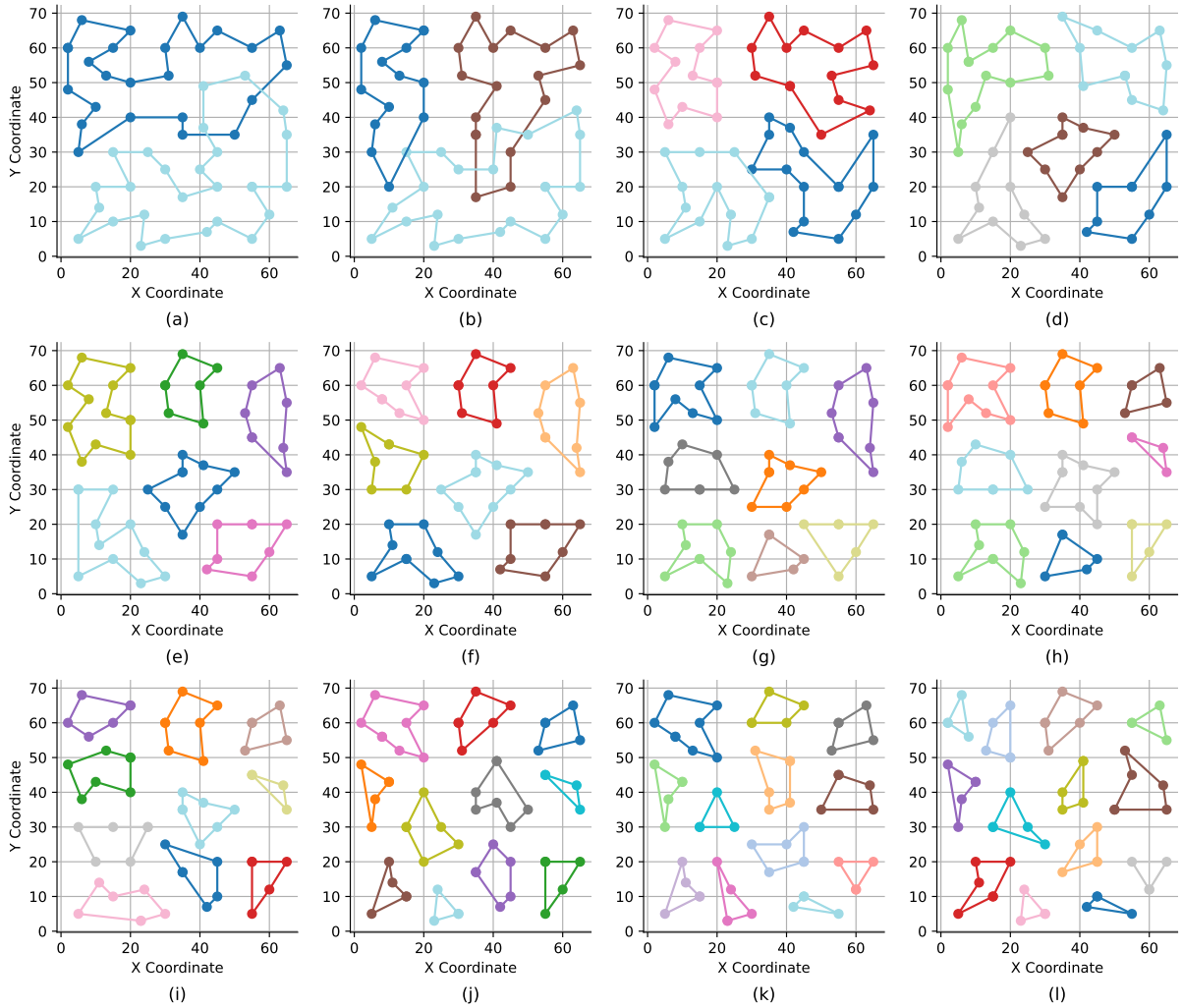


Figure 4.6: Heuristic-optimized routes for a 50-node instance with varying values of K.

Table 4.4: Number of agents assigned to each route, f^k, for the example shown in Figure 4.6. Dashes indicate infeasible cases where k exceeds K.

K	f^1	f^2	f^3	f^4	f^5	f^6	f^7	f^8	f^9	f^{10}	f^{11}	f^{12}	f^{13}
1	103	—	—	—	—	—	—	—	—	—	—	—	—
2	51	52	—	—	—	—	—	—	—	—	—	—	—
3	29	34	43	—	—	—	—	—	—	—	—	—	—
4	24	26	27	30	—	—	—	—	—	—	—	—	—
5	15	18	23	24	27	—	—	—	—	—	—	—	—
6	14	14	14	15	21	23	—	—	—	—	—	—	—
7	12	13	14	14	14	15	15	—	—	—	—	—	—
8	9	9	10	13	14	14	15	17	—	—	—	—	—
9	7	9	9	9	12	14	14	14	17	—	—	—	—
10	7	9	9	10	10	11	11	12	13	14	—	—	—
11	6	7	8	9	9	9	9	10	11	12	13	—	—
12	6	6	7	8	8	9	9	9	9	11	11	13	—
13	6	6	6	6	7	8	8	9	9	10	10	10	12

4.6.5 Sensitivity Analysis and Design Insights

This section investigates how key model parameters influence agent allocation, routing efficiency, and overall system performance. In practice, operational constraints such as revisit frequency, route count, and fleet size must be carefully tuned to balance resource limitations with monitoring coverage. To support informed decision-making, we conduct a series of sensitivity analyses focusing on three core design parameters: the headway bounds $H_{\min}$ and $H_{\max}$, and the number of routes K for different number of parking facilities. The results provide insights into the trade-offs between service frequency, travel time, and resource utilization, offering practical guidance for configuring sensing operations in dynamic urban settings.

Impact of Headway Parameters: We examine how variations in the headway constraints, $H_{\min}$ and $H_{\max}$, affect agent allocation and total travel time. Figure 4.7 shows the impact of $H_{\max}$ on these outcomes across different node sizes. Panel (a) reports the total number of agents, and panel (b) shows the corresponding travel time, both as functions of $H_{\max}$. In both panels, it is observed that for any fixed value of $H_{\max}$, increasing the number of nodes leads to higher resource requirements and longer total travel time. Conversely, as the maximum allowable headway increases, i.e., when the time interval between successive visits to a node can be relaxed, the total number of agents and the overall travel time decrease. This trend reflects the system's ability to economize on agents when the frequency of observations is reduced. The rate of reduction is not linear. The decline in both metrics follows an exponential pattern, exhibiting diminishing returns. Specifically, there exists a threshold beyond which increasing $H_{\max}$ yields marginal changes in either the number of required agents or total travel time.

This insight is particularly critical for decision-makers tasked with deploying agents in heterogeneous urban environments. In practice, the required frequency of data collection is not uniform across all regions of a city. High-turnover areas, such as commercial districts, downtown cores, or zones near transit hubs, experience rapid changes in parking occupancy and therefore demand more frequent monitoring to ensure accurate real-time availability data. In these locations, stricter headway requirements (i.e., lower $H_{\max}$) are

necessary, which in turn necessitates a larger number of agents to maintain coverage.

Conversely, in suburban areas or low-turnover zones where parking durations are typically longer and changes in occupancy occur less frequently, a larger headway value may be acceptable. This flexibility allows planners to allocate fewer agents without significantly compromising data accuracy or service quality in those regions. Thus, the trade-off between $H_{\max}$ and resource allocation should not be treated uniformly but rather adapted based on localized parking dynamics.

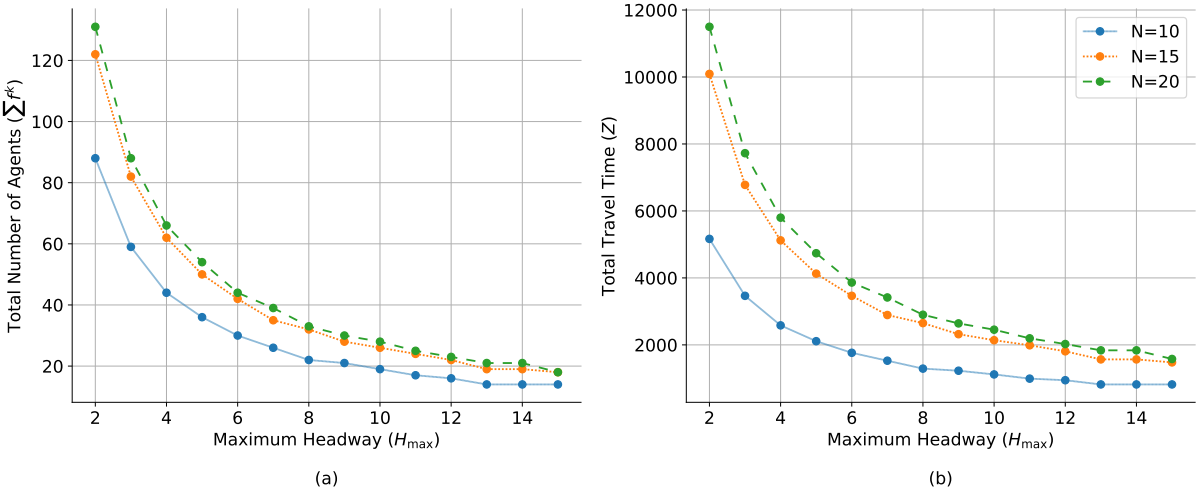


Figure 4.7: Effect of $H_{\max}$ on agent count and travel time across different node sizes.

To complete the headway analysis, Figure 4.8 shows the optimization results under varying minimum headway values, $H_{\min}$. Since the model minimizes total travel time while satisfying the upper bound $H_{\max}$, the minimum headway constraint remains non-binding across a wide range of lower $H_{\min}$ values. During this phase, both the total travel time and the number of agents remain unchanged, indicating that small increases in $H_{\min}$ have no effect on the optimum solution. However, when $H_{\min}$ nears $H_{\max}$, the revisit interval becomes more flexible, allowing greater freedom in scheduling. This shift triggers a sharp drop in both total travel time and agent allocation, signaling a sudden change in the optimal configuration. This behavior underscores the importance of carefully calibrating the headway window. A well-chosen gap between $H_{\min}$ and $H_{\max}$ ensures reliable coverage without incurring unnecessary sensing costs, particularly in areas where high-frequency monitoring is not required.

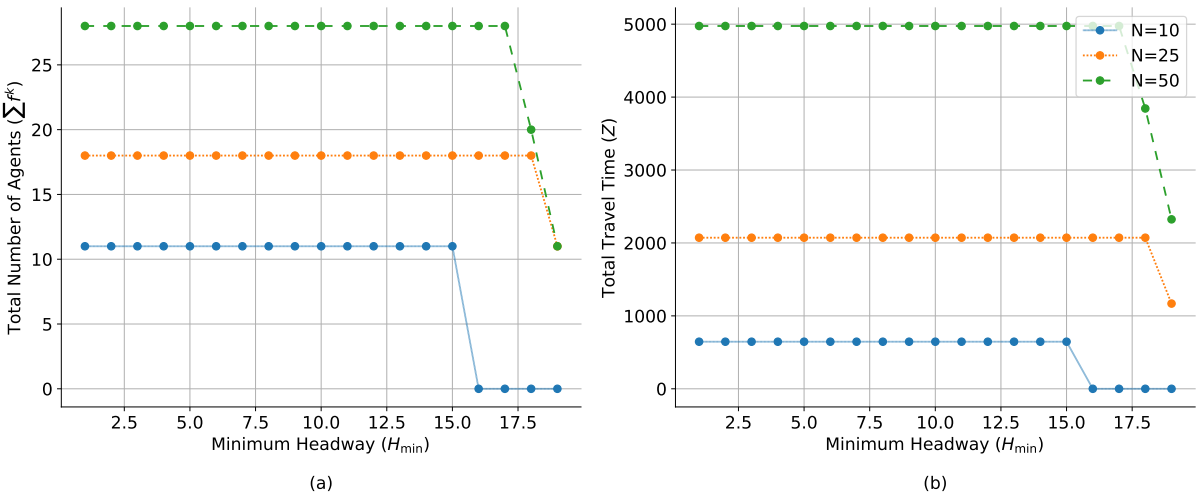


Figure 4.8: Effect of $H_{\min}$ on optimization outcomes.

Impact of Route Count on Sensing Efficiency: The number of routes, K, is a key design parameter determined by the planner based on operational constraints and desired coverage granularity. Figure 4.9 presents the optimization results as a function of the number of routes. Panel (a) shows the total number of agents required, while panel (b) illustrates the corresponding total travel time. Increasing the number of routes generally leads to a reduction in total travel time, as routes become shorter and more localized. However, the behavior of the optimal number of agents is less predictable. For small values of K, the number of agents increases with the number of routes. Beyond a certain point, however, the pattern becomes irregular and highly dependent on specific problem parameters. These findings indicate that the number of routes should not be arbitrarily increased. Instead, planners should tune K to balance travel efficiency with agent availability, aligning route design with resource constraints and overall system performance goals.

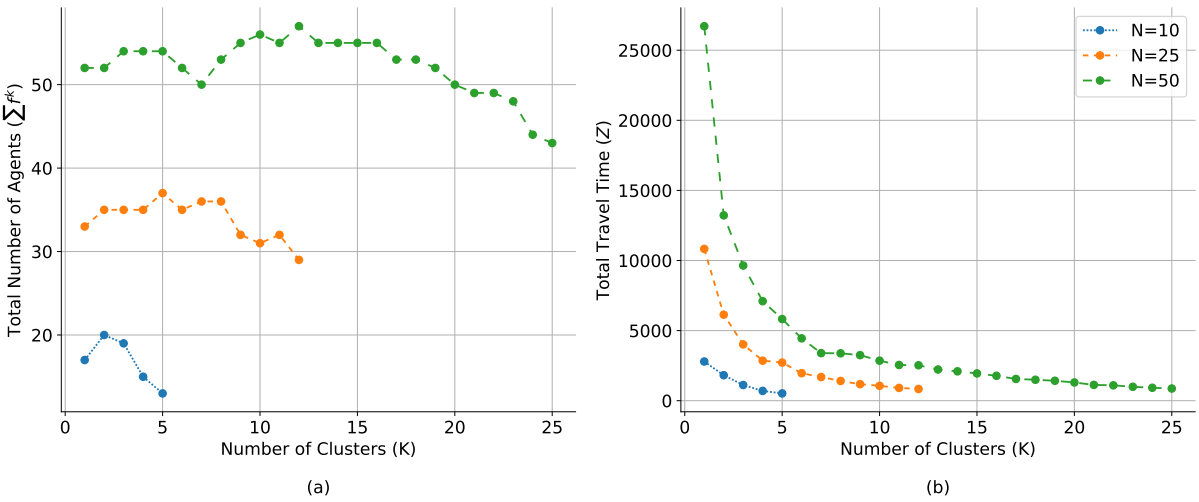


Figure 4.9: Comparison of optimization results as a function of K for different number of nodes. (a) Total number of agents required across routes, (b) The corresponding travel time. Parameters: $K = 3$, $H_{min} = 1$, $F = 500$.

4.7 Case Study: On-Street Permit Parking in Toronto

To assess the practical utility of the proposed optimization framework, we conduct a case study based on the *On-Street Permit Parking Area* dataset published by City of Toronto (2024b). This dataset includes the boundaries of 97 officially designated on-street permit parking zones distributed throughout the City of Toronto, with each zone represented as a distinct polygonal geometry (see Figure 4.11-(a)).

All spatial features are encoded using the WGS 84 coordinate reference system (EPSG:4326) and represented as nested polygons suitable for geospatial analysis. Each polygon is associated with a unique alphanumeric identifier and accompanied by structured metadata, including short labels, descriptive area names, and textual annotations aligned with the classification schema adopted by the Toronto Parking Authority.

4.7.1 Fleet Capacity

To estimate the upper bound on deployable agents, we draw on data from Bike Share Toronto, a program operated by the Toronto Parking Authority. The system comprises a fleet of approximately 6,850 shared bicycles distributed across more than 700 docking stations[1].

Bike share networks span large urban areas and generate continuous streams of mobility data, making them well-suited for city-scale sensing applications. Beyond promoting sustainable transportation, these systems have also been investigated as mobile sensing platforms for urban infrastructure monitoring. For example, He et al. (2021) introduced a crowdsensing approach that identifies illegally parked vehicles by detecting deviations in GPS traces of shared bikes moving through designated cycling lanes.

Assuming each bicycle is equipped with sensing hardware capable of capturing parking occupancy data, we define the upper bound of deployable agents as $F = 6{,}850$. For resource-constrained scenarios, smaller subsets of the fleet (e.g., 10%) may be selected to evaluate the framework's performance under limited deployment conditions. To comple-

[1]Fleet data from: `https://open.toronto.ca/dataset/bike-share-toronto/`

ment this, we incorporate spatial data from the City of Toronto's official cycling network, which delineates the city's bikeway infrastructure,[2] as shown in Figure 4.10.

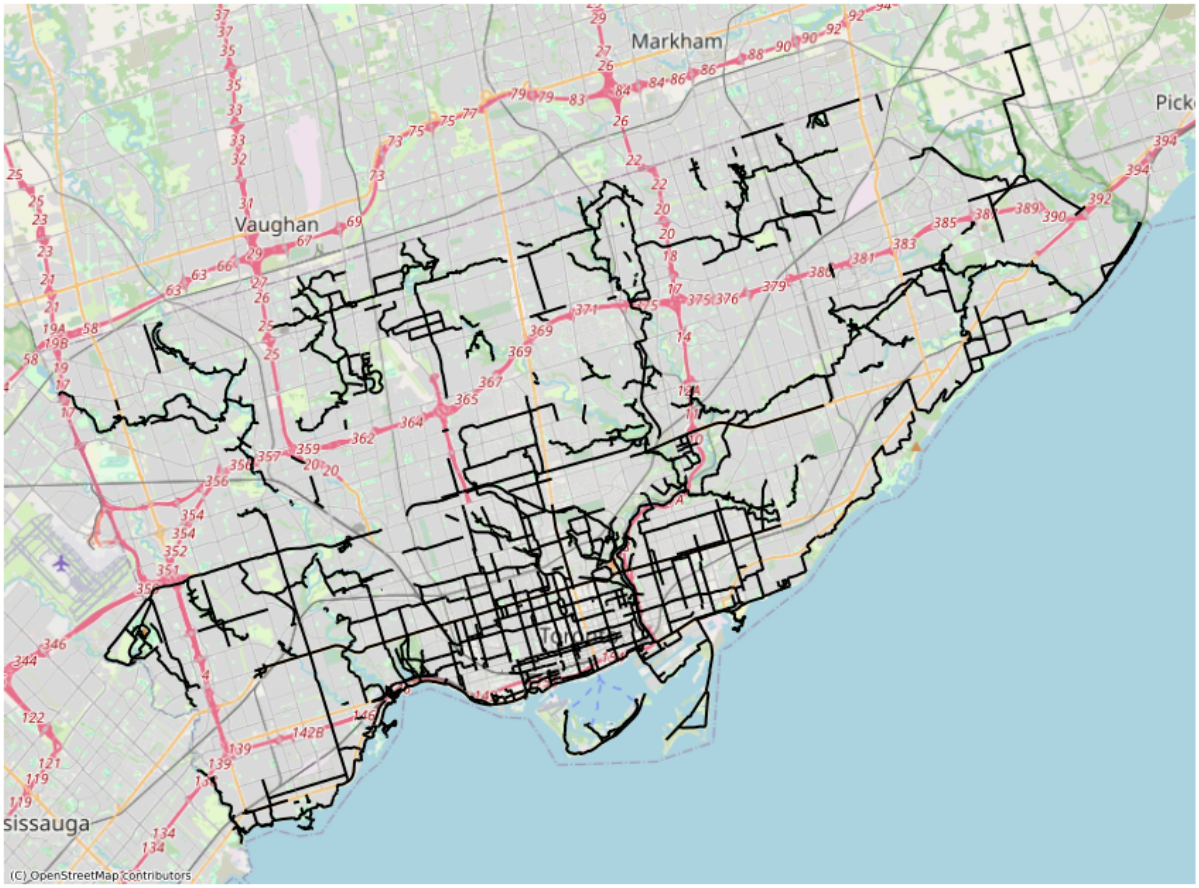


Figure 4.10: Toronto cycling network overlaid on an OpenStreetMap basemap. The network includes cycle tracks, bike lanes, sharrows, multi-use trails, and signed routes.

The dataset includes various types of bikeways: cycle tracks, bicycle lanes (including buffered and contra-flow lanes), neighborhood routes with sharrows (including wayfinding sharrows), multi-use trails (off-road and in-boulevard), and signed cycling routes. These classifications provide insight into likely bike travel paths and enable integration of realistic mobility constraints into the sensor deployment model.

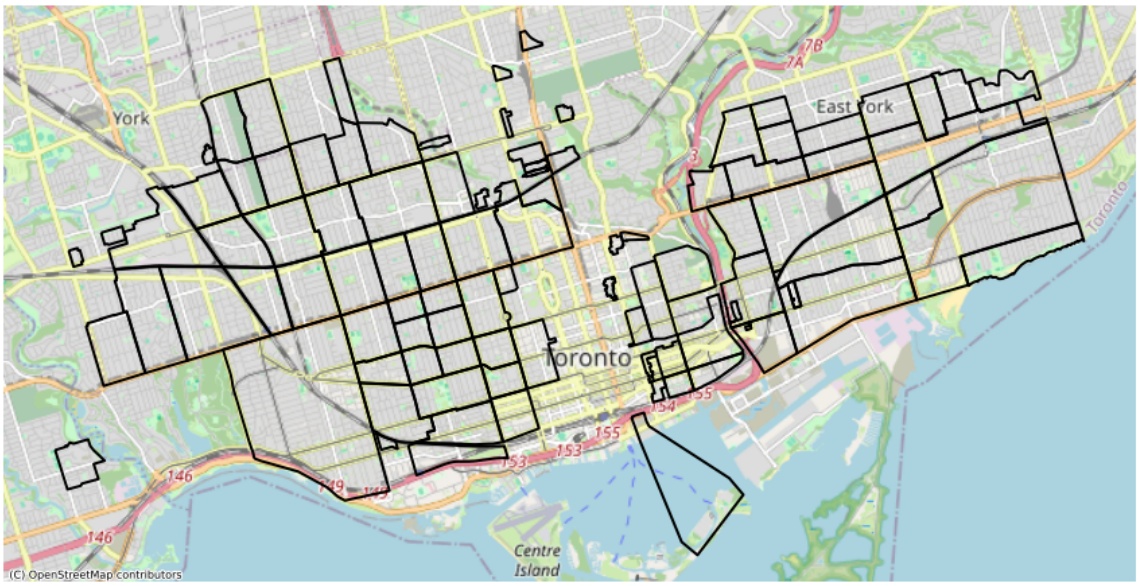
4.7.2 Headway Parameter Calibration

We begin our spatial analysis by examining the geographic distribution of the 97 permit parking zones, as illustrated in Figure 4.11-(a). To quantify spatial relationships between

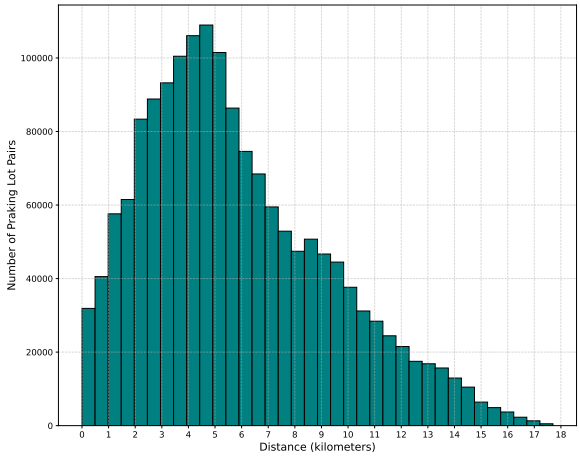
[2]Cycling network data: `https://open.toronto.ca/dataset/cycling-network/`

zones, we compute all pairwise haversine distances, accounting for Earth's curvature, based on the latitude and longitude coordinates provided in the dataset. The resulting histogram in Figure 4.11-(b) reveals a high density of facility pairs within a 2–6 km range, peaking around 4.5 km. This distance distribution forms the basis for setting service coverage and visit frequency parameters in the optimization model.

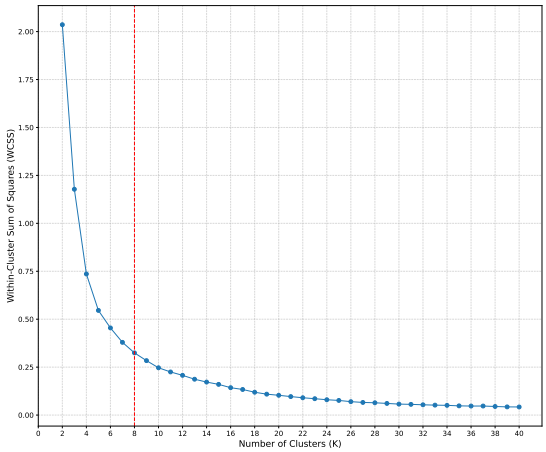
Assuming an average urban cycling speed of 15 km/h, each kilometer corresponds to roughly 4 minutes of travel time. Accordingly, the modal pairwise distance of 4.5 km implies a travel time of approximately 18 minutes. Based on this estimate, we recommend setting the minimum headway parameter $H_{\min}$ between 15 and 25 minutes to avoid excessive revisits to nearby locations. Conversely, the maximum headway $H_{\max}$ can range from 50 to 60 minutes to ensure sufficient coverage of more spatially dispersed areas. Although granular turnover data specific to Toronto are unavailable, studies in urban mobility suggest average turnover periods of 1–2 hours in high-demand urban districts and 3–6 hours in suburban or lower-demand zones (Shoup 2011). These findings support the use of determined headway values. In cases where real-time or historical usage data become available, headway calibration can be further refined based on context-specific factors such as time of day, demand intensity, parking zone capacity, and operational priorities.



(a) Geographic distribution of on-street permit parking zones in Toronto.



(b) Histogram of pairwise haversine distances between permit zones.



(c) Elbow plot of within-cluster sum of squares (WCSS) for K-means clustering with $K \in [1, 40]$.

Figure 4.11: Spatial analysis and clustering results of on-street parking zones in Toronto.

4.7.3 Clustering for Route Generation

To apply the optimization model, we first compute the geographic centroid of each parking zone based on its polygon geometry. We then perform k-means clustering on these centroids to divide the service area into spatially coherent sub-regions, facilitating efficient route generation.

To determine the optimal number of clusters K, we apply the elbow method by plotting the within-cluster sum of squares (WCSS) across a range of K values, as illustrated

in Figure 4.11-(c). To automate the identification of the elbow point, we employ the KneeLocator algorithm (Satopaa et al. 2011), which detects $K = 8$ as the optimal number of clusters, corresponding to the point of diminishing returns in WCSS. Setting $K = 8$ balances route compactness and spatial coverage, thereby enhancing operational efficiency. Figure 4.12 presents the optimized service routes for the eight clusters, each covering a distinct subset of Toronto's on-street permit parking zones.

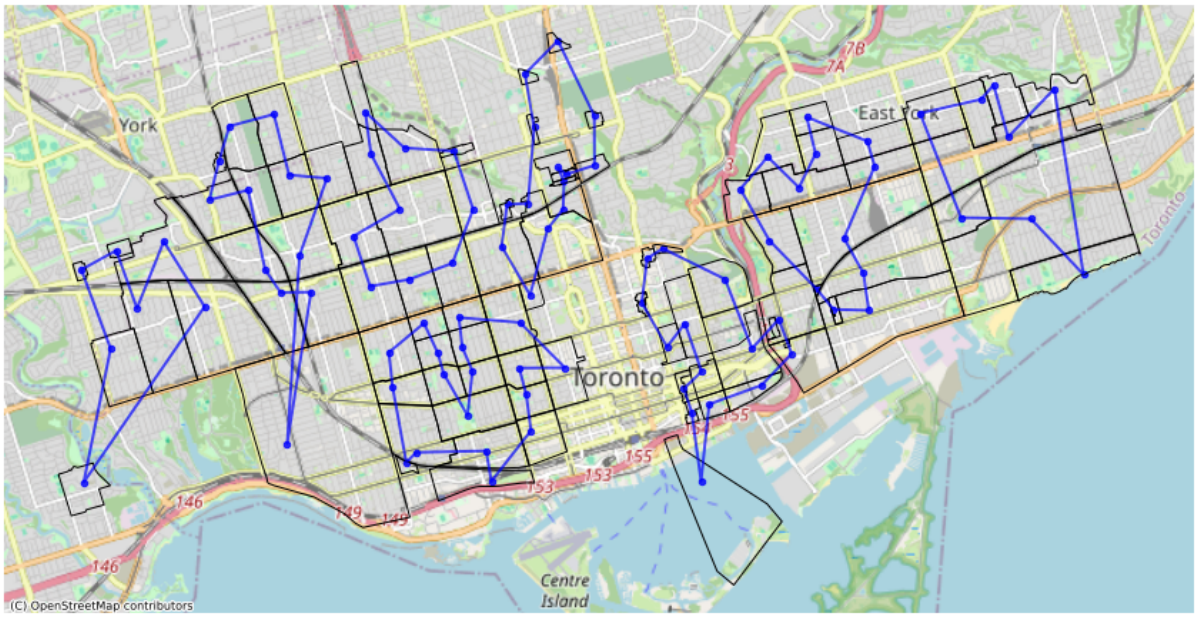


Figure 4.12: Optimized data collection routes for on-street permit parking zones in Toronto. Parameters: $K = 8$, $F = 6,850$, $H_{\min} = 20$ minutes, and $H_{\max} = 55$ minutes.

4.7.4 Performance Evaluation

To assess the effectiveness of the proposed framework, we evaluate the impact of key parameters on agent allocation and routing efficiency. Figure 4.13 summarizes the model's output in terms of total agent count and route travel time across varying numbers of clusters K. Both the number of required agents and the total travel time initially decrease with increasing K, as clustering enables more localized routing. However, beyond a certain threshold, identified as $K = 8$ using the elbow method, further increases yield diminishing returns. Specifically, the number of required agents remains relatively stable beyond this point, indicating that additional clustering does not significantly enhance efficiency.

While our framework determines optimal routes based on spatial clustering and headway constraints, future applications could benefit from integrating empirical mobility data. For instance, bike share trip patterns, such as the frequency of trips between origin–destination pairs, could be used by transportation planners to align sensing routes with the city's most active corridors. This alignment would help prioritize areas with high usage and support data-driven decision-making in real-world deployments.

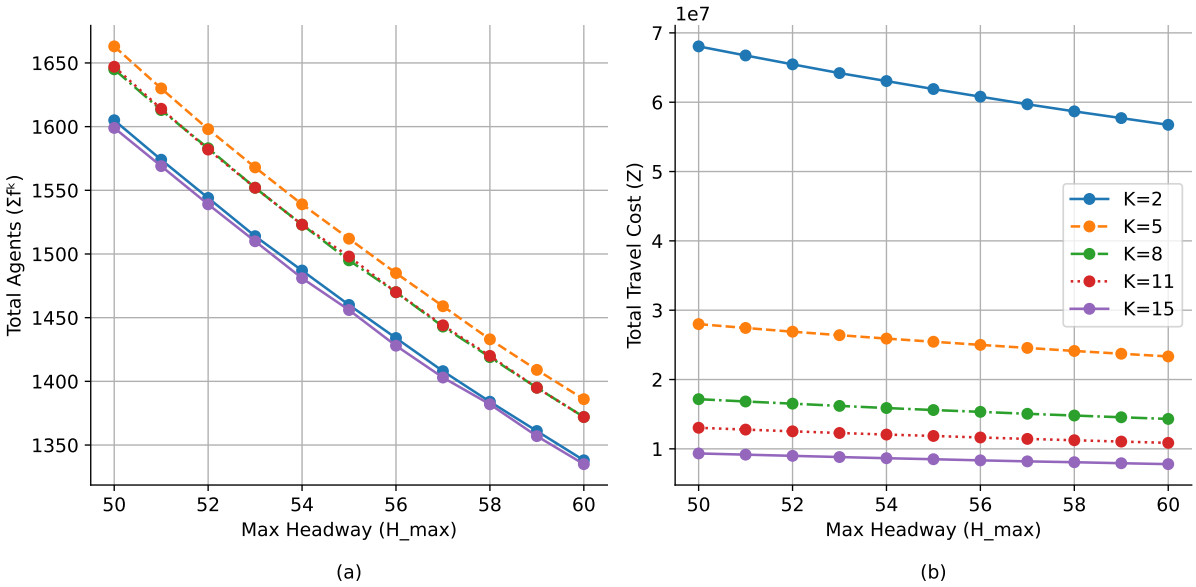


Figure 4.13: Impact of varying parameters on agent allocation and travel time. Here, $H_{min} = 20$ minutes. (a) Optimized number of agents and (b) travel time.

Our results show that performing crowdsensing-based parking occupancy monitoring in Toronto is feasible using approximately 25% of the current Bike Share Toronto fleet. Equipping this subset of shared bicycles with sensing hardware, such as cameras, LiDAR, or ultrasonic sensors, can enable a continuous and scalable flow of real-time parking data across the city.

4.8 Chapter Summary

This study presents a comprehensive optimization framework for mobile crowdsensing in urban parking environments, addressing the dual challenges of agent routing and allocation under practical operational constraints. We formulate a mixed-integer programming model with explicit headway bounds and propose a scalable heuristic algorithm that delivers flexible, efficient, and high-performing solutions suitable for real-world deployment.

Results from both the exact and heuristic models show that increasing the maximum allowable revisit interval, $H_{\max}$, significantly reduces the number of required agents and total travel time. Optimizing the number of service routes, K, improves spatial coverage and agent distribution, although excessive segmentation yields diminishing returns. The proposed heuristic approach consistently achieves near-optimal performance, within 1–2% of the exact model, while reducing computation times by several orders of magnitude, making it well-suited for large-scale urban networks.

The case study conducted in Toronto demonstrates the practical applicability of the proposed optimization framework in a real-world urban environment. By leveraging open data on on-street permit parking zones and the city's cycling network, we simulate a realistic deployment scenario for sensing agents using shared bicycles. Through spatial clustering, headway calibration, and route optimization, the model effectively minimizes the number of required agents and total travel time while maintaining sufficient revisit frequency. The results show that the entire city can be monitored using only half of the existing Bike Share Toronto fleet, suggesting a scalable and sustainable solution for parking occupancy detection. While the model operates independently of usage data, future integration with bike share trip patterns may further enhance route prioritization and policy alignment, providing a valuable tool for city planners and transportation agencies.

This framework opens several promising avenues for future research. Integrating dynamic traffic and parking demand data could enable real-time, adaptive routing. Incorporating predictive models to identify high-demand or frequently occupied zones may improve resource targeting. Additionally, extending the model to account for heteroge-

neous agent capabilities, battery or energy constraints, and multimodal mobility behaviors would increase realism and operational fidelity.

Overall, this research contributes a robust and adaptable solution to the problem of city-scale parking occupancy sensing. By jointly optimizing agent allocation, routing, and visit frequency under real-world constraints, the proposed framework supports the design of scalable, cost-effective, and sustainable smart city monitoring systems.

Chapter 5

CHAPTER FIVE: IMPLEMENTING PARKING OCCUPANCY DETECTION USING DASHCAM FOOTAGE

5.1 Chapter Overview

The absence of real-time on-street parking availability data can result in extended parking search times and increased traffic congestion. The primary solution to this issue is to provide a real-time parking availability map that assists drivers in locating vacant parking spaces. Various devices can be used to collect parking availability data, each with its own set of advantages and drawbacks. Traditional parking sensors like magnetic pucks face high maintenance costs, low accuracy, and weather sensitivity. Steady cameras and sensors have limited sight ranges, which means that deploying these devices in specific locations will only cover a restricted number of parking spots. An alternative solution is crowdsensing, in which drivers utilize smartphones or onboard devices to share on-street parking availability. Crowdsensing offers flexibility and scalability, reducing the need for expensive infrastructure and resource allocation. Recently, the use of dashcams in vehicles for recording road events has seen significant growth. Our goal is to harness the data captured by these dashcams to create a digital parking availability map. In the initial step, we employ a pre-trained model to perform object detection and classification tasks on the raw videos recorded by dashcams. Subsequently, we must filter out unnecessary detected objects while safeguarding sensitive data and respecting individuals' privacy. The final step involves extracting metadata from the dashcam recordings to retrieve GPS data and accurately pinpoint the detected cars, allowing us to generate a parking occupancy map. From a technical perspective, this approach involves object detection and real-time parked car counting using the YOLOv8 model from the Ultralytics library. Video management and annotation tools from the supervision library provided by Roboflow and the OpenCV library are utilized to create this program. Using Exiftool, we make use of embedded GPS data in dashcam footage to identify legal parking spots. Finally, we develop a prediction model to estimate the probability of finding parking spots in a given area, further enhancing the efficiency of urban parking management.

5.2 Introduction

Limited on-street parking facilities and the increasing reliance on private cars have strained the balance between supply and demand for parking spots (Box 2004). Searching for parking can account for anywhere from 8% to 74% of traffic, incurring substantial congestion costs (Shoup 2006, Zhu et al. 2020). In the Greater Toronto and Hamilton Area, parking-related congestion is estimated to cost $3.3 billion annually, while insufficient curbside parking results in $2.7 billion in economic losses as businesses lose customers and jobs (Metrolinx 2008). Rather than expanding costly and often impractical parking facilities, a more effective approach is to optimize existing parking resources by disseminating real-time parking availability information (Lin et al. 2017). This information helps drivers navigate directly to vacant spots, plan their trips more efficiently, and in some cases, reconsider their mode of transport. It also aids law enforcement by pinpointing density violations or illegal parking.

Crowdsensing provides a cost-effective approach to collect real-time parking data by leveraging mobile devices, such as smartphones or vehicle-based sensors, instead of costly stationary infrastructure (Ganti et al. 2011). Crowdsensing supports intelligent transportation systems by supplying traffic flow information, travel patterns, and carbon footprint estimates (Pilloni 2018). Motivated by these benefits, we propose employing crowdsensing in parking availability data collection by equipping shared bikes with a *smart box*, including a dashcam, power supply, and an edge processor, to detect on-street parking vacancies. Bikeshare systems traverse extensive urban areas, producing continuous data feeds while promoting sustainable transport. We therefore investigate two main questions: how dashcam footage from shared bikes can produce a real-time parking occupancy map?, and how such data can provide reliable, forward-looking predictions?

Image-based methods for parking occupancy detection have been extensively explored to provide automated solutions for monitoring parking availability. Grbić and Koch (2023) applied deep learning to images from fixed cameras for parking lot monitoring, which works well for structured facilities but requires predefined locations and fixed infrastruc-

ture. To expand beyond static setups, Wu and Yeh (2019) developed a neural-network-based method that uses dashboard camera footage to identify available parking spaces in real time as a vehicle moves. Their model processes these video frames in real time, extracting spatial and temporal cues to identify vacant spots before the vehicle arrives at them. Beyond image-based methods, bikeshare systems have also been used for parking occupancy detection. For example, He et al. (2021) demonstrated a crowdsensing-based approach that identifies illegally parked vehicles using shared bikes' GPS traces, detecting obstructions by analyzing path deviations in designated cycling lanes.

We use footage from moving dashcams mounted on shared bikes to generate real-time parking availability maps and forecasts. The mobility of bikeshare users provides continuous coverage in areas lacking fixed surveillance. For example, the Toronto Parking Authority's Bike Share program operates 6,850 bikes across 625 docking stations, running 24/7 year-round. Since its launch in 2011, the program has recorded over 15.9 million rides, with ridership increasing by 21% from 2019 to 2020 and by 17% from 2020 to 2021 (City of Toronto 2024a). Furthermore, using bikes aligns with sustainability goals and reduces privacy concerns compared to relying on dashcam footage from private vehicles. By performing on-bike edge processing, we reduce the computational burden associated with handling the large volumes of data collected by riders.

The remainder of this research is organized as follows. Section 5.3 provides an overview of the related literature. Section 5.4 presents a detailed explanation of the methodology, including video analysis, geo-fencing, and edge processing.

5.3 Literature Review

This section reviews existing literature on parking occupancy detection, with a particular focus on crowdsensing-based approaches. First, we discuss traditional parking occupancy detection technologies, highlighting their advantages and challenges. Next, we examine crowdsensing methods that utilize sensor-equipped vehicles, smartphones, and video-based detection systems to improve real-time parking monitoring. Finally, we explore predictive modeling techniques, including queuing theory-based approaches and machine learning algorithms, used to forecast parking availability based on historical and real-time data.

5.3.1 Parking Occupancy Detection Technologies

Accurate and efficient parking availability detection is essential for improving urban mobility, reducing congestion, and optimizing space utilization. However, selecting a practical, scalable, and cost-effective method remains a challenge, as different technologies come with trade-offs in accuracy, implementation costs, and real-time adaptability (Revathi and Dhulipala 2012, Sarker et al. 2020). This section explores various parking occupancy detection technologies, including in-ground sensors, camera-based systems, GPS and Bluetooth tracking, and transaction data analysis, highlighting their advantages, limitations, and suitability for large-scale deployment.

In-ground Sensors

One approach involves installing in-ground, laser, or magnetic sensors in parking spots, which provide binary outputs indicating whether a space is occupied or available (Zhang et al. 2015). Several projects have implemented these sensors. For instance, the "SFpark" project in San Francisco, launched in 2010, successfully reduced drivers' parking search time by 50% through in-ground sensor technology (SFpark 2014). Similarly, in China, a wireless network of in-ground magnetic sensors achieved 99% accuracy by neutralizing the impact of temperature fluctuations on sensor performance (Gu et al. 2012). How-

ever, despite their effectiveness, in-ground sensors present scalability challenges. Their high installation and maintenance costs make widespread deployment difficult, and their accuracy depends on drivers parking in designated areas.

Camera-Based Parking Detection

Stationary cameras offer an alternative approach to collecting parking occupancy data. In this method, cameras mounted on poles record vehicle movements to determine parking availability. Compared to in-ground sensors, cameras provide a more cost-effective and scalable solution, as they can monitor multiple parking spaces simultaneously and collect data for various urban applications (Farley et al. 2021).

Chen et al. (2020) leveraged existing roadside surveillance equipment and real-time processing to capture parking occupancy. Their proposed model employs one-shot object detection to identify vehicles and uses image processing techniques to determine parking space occupation from a sequence of images. During nighttime operations, streetlights play a crucial role in object detection, providing on-demand adaptive lighting by adjusting brightness based on the presence of pedestrians, cyclists, or vehicles.

Despite its advantages, video-based parking detection has several limitations. While machine learning algorithms can extract parking availability information from video footage, video analysis is computationally expensive and requires advanced models for real-time processing. Additionally, camera placement is constrained by legal and logistical challenges, such as obtaining installation permits from urban authorities and ensuring a reliable power supply.

GPS and Bluetooth-Based Detection

GPS and Bluetooth sensors have also been explored for parking occupancy detection. In this method, Bluetooth sensors embedded in mobile phones detect whether a vehicle is in driving or parking mode, while GPS data tracks the parking location. However, a significant drawback of GPS and Bluetooth-based approaches is the potential invasion of drivers' privacy, as continuous location tracking raises concerns about data security and

user anonymity (Xu et al. 2013).

Transaction Data for Parking Monitoring

Analyzing parking payment transaction data is another method used to estimate parking occupancy, as it reflects legitimate parking activity. However, this approach has significant limitations. First, transaction data does not account for illegal parking, such as when drivers fail to pay or park in unauthorized zones (e.g., loading areas, fire lanes, or no-parking zones). These activities remain unrecorded, even though they significantly impact parking availability and enforcement. Second, transaction data fails to capture vehicles that overstay their paid parking time. Once a driver makes an initial payment, the system assumes compliance, but vehicles that exceed their paid period remain invisible in the dataset. This can lead to reduced turnover and lost revenue for parking operators. Similarly, early departures distort occupancy rates. When a driver leaves before their paid time expires, the system incorrectly records the spot as occupied for longer than it actually is. This misrepresentation of real occupancy levels creates inefficiencies in parking management. While transaction data offers insights into paid parking behavior, it provides an incomplete view of overall parking dynamics, particularly regarding illegal parking and real-time availability.

5.3.2 Crowdsensing-Based Parking Occupancy Detection

Crowdsensing enables real-time parking occupancy detection by leveraging data collected from various sensing devices carried by users as they move through city streets. As a user equipped with a sensing device passes by parking spaces, the device detects availability status and transmits the data to a central database, as illustrated in Figure 5.1. The processed data is then used to generate an up-to-date parking availability map, providing valuable information for drivers and urban planners.

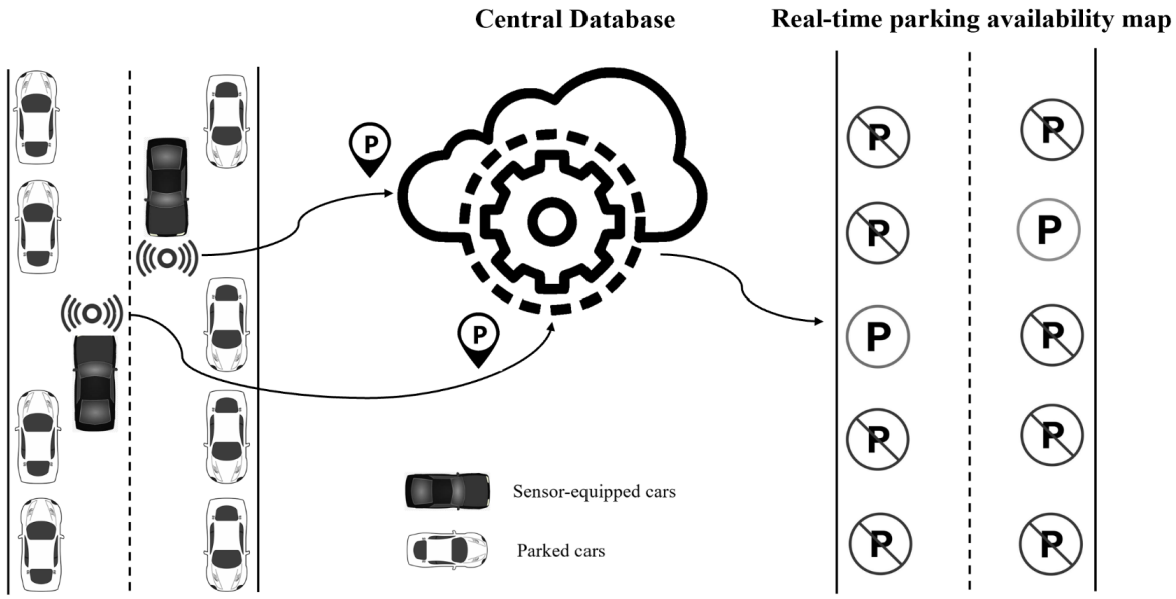


Figure 5.1: Crowdsensing-based Parking Availability Detection by Onboard Sensors

Sensor-based Solutions

The "ParkNet" project in 2010 equipped vehicles with GPS receivers and ultrasonic range finders. These vehicles transmitted data to a central server, providing real-time updates on vacant and occupied spaces. ParkNet employed mobile sensing and statistical methods to optimize sensor deployment, achieving 95% accuracy due to the effectiveness of ultrasonic sensors day and night, and their affordability compared to alternatives like laser range finders or radars (Mathur et al. 2010). Similarly, Roman et al. (2018) combined an ultrasound sensor mounted on the passenger side of a car with GPS data. The sensor measures the distance from the vehicle to the nearest roadside obstacle, and when combined with accurate GPS readings through a map-matching technique, parking availability data is collected. They developed a supervised learning algorithm to estimate roadside parking availability and a model to determine the number of sensing units needed for urban coverage.

The Bosch-Mercedes Benz and INRIX-BMW teams initiated a collaborative project to deliver a premium crowdsensing-based parking service using onboard ultrasonic sensors. Installing these sensors on cars transforms each vehicle into a parking spot locator. The sensor-equipped cars transmit data regarding the availability of on-street parking

spots to a central database. This aggregated data is then shared with drivers, allowing them to identify streets with the highest probability of available parking spots. As of 2018, the BMW On-Street Parking Information service is available in 68 cities across the US, Germany, the United Kingdom, France, Spain, the Netherlands, Russia, and Norway (Bosch 2016, INRIX 2018). Compared to a fixed sensing solution, their model shows that the crowdsensing approach performs equally well when accurate location information is available but requires significantly fewer sensors. However, using ultrasound sensors can result in noisy data due to vehicle lane changes.

Smart Phone-based Solutions

Building on the success of onboard ultrasonic sensors, another crowdsensing-based approach for parking availability detection leverages smartphones (Nawaz et al. 2013, Nandugudi et al. 2014). In this method, smartphone sensors and WiFi signatures detect drivers' parking and driving modes, providing real-time parking availability data.

One notable implementation is the "ParkMaster" project, introduced in 2017, which utilized drivers' smartphones mounted on windshields to capture video footage for parking occupancy detection. The project achieved 90% accuracy, with smartphones transmitting real-time data to a central server, where specialized algorithms processed the footage to detect parked vehicles (Grassi et al. 2017). Expanding on this concept, Rahman et al. (2020) developed an Android application that integrates sensor data from a smartphone's accelerometer, gyroscope, and GPS. Their approach constructs a probability model to estimate the occupancy probability of each parking spot based on a driver's search trajectory and destination. This method improves the accuracy of real-time parking detection by utilizing multi-sensor fusion techniques. To further enhance smartphone-based detection accuracy, Xu et al. (2013) introduced a historical availability profile that analyzes past parking patterns to refine real-time predictions. By incorporating historical occupancy trends, their model compensates for occasional detection errors in smartphone-based systems, making the predictions more reliable and robust.

Video-based Solutions

The most relevant work supporting the feasibility of our approach is done by Higuchi and Oguchi (2020). They propose vision-based vehicular crowdsensing to determine parking availability. They use a monocular RGB camera installed in a vehicle and a computer vision-based object tracking mechanism to detect surrounding parked vehicles. By analyzing the time series of object tracking results and correlating them with a digital map of a parking facility, they identify the availability of each parking spot. Simulation results demonstrate that this sensing system can detect parking availability at the spot level with an average accuracy of 83%.

5.3.3 Parking Occupancy Prediction

There are two primary approaches to forecasting parking availability.

Model-Based Approach. The first approach establishes a predictive model for parking occupancy, generally involving estimating model parameters and predicting occupancy. Given the time-dependent nature of parking occupancy, these models often benefit from queuing theory. For example, Du and Gong (2016) introduces a stochastic Poisson game based on M/M/c queue theory, assuming that each vehicle's parking time follows an exponential distribution and that the arrival of vehicles at parking facilities follows a Poisson process. They employ a multinomial logit model to capture travelers' parking choices. They propose a decentralized coordinated parking mechanism (DCPM) to reduce parking congestion by organizing vehicles searching for parking into coordination groups. The DCPM models competition for parking spaces within a central business district (CBD) by categorizing travelers based on destination proximity, treating them as groups sharing the same parking priority, and assigning them a probabilistic utility function based on the likelihood of being forced to cruise for parking. Using queuing models, arrival and departure rates are estimated from historical data. One approach fits curves to the occupancy probability distribution, which is practical when parking is under-saturated. At the same time, the other uses maximum likelihood or least squares estimation for over-saturated conditions. Similarly, Xiao et al. (2018) applies an M/M/C/C queuing model to forecast

occupancy, validating the approach with simulated and real data to accurately estimate arrival and departure rates. Other studies, such as Ma et al. (2017a) and Tavafoghi et al. (2019), apply variations of queuing models with transient and non-homogeneous rates to predict dynamic parking occupancy in real time.

Machine Learning Approach. The second approach uses statistical and machine learning (ML) methods to predict occupancy directly from observed data. The ML-based framework typically involves training an algorithm on historical parking data and then applying survival analysis to forecast the likelihood of a space being free in future intervals. For example, in Vlahogianni et al. (2016), a neural network is trained on data from magnetic in-ground sensors in Santander, Spain, to predict short- and long-term parking availability. Although effective, sensor-based solutions can be costly, face battery life limitations, and may be disrupted by environmental magnetic fields. Video-based methods address some of these issues. For instance, Bulan et al. (2013) and Provoost et al. (2020) use traffic camera footage to train support vector machines (SVM) and neural networks, respectively, to predict occupancy, with Provoost et al. (2020) using random forests alongside neural networks for enhanced prediction accuracy. Further, Amato et al. (2016) and Amato et al. (2017) apply convolutional neural networks (CNNs) on smart camera footage to monitor real-time parking occupancy, achieving over 91% detection accuracy, which surpasses in-ground sensor accuracy. Their video-based approach includes offline training to identify regions of interest in parking scenes. Subsequently, it applies an SVM classifier to detect occupied spaces in real-time, validated through installations of multiple cameras over a five-day trial.

5.4 Dashcam-Enabled Smart Box for Shared Bikes

Bikeshare systems, widely used for daily commutes, provide continuous, city-wide data collection, keeping the sensing system responsive to urban changes. Their mobility allows access to areas less frequented by cars, ensuring comprehensive coverage. To leverage shared bikes for parking availability detection, we equip them with a smart box containing a dashcam, power supply, and edge processor.

A crucial factor in selecting a dashcam for this experiment is ensuring consistent footage resolution under various lighting conditions, both day and night. Therefore, we focused on dashcams equipped with night vision or High Dynamic Range to enhance visibility in low-light environments. While advancements in artificial intelligence have introduced features, such as voice warnings and license plate recognition, parking occupancy detection task only requires a dashcam capable of recording GPS data embedded in the video. This feature ensures that the dataset includes precise location details, providing accurate information on when and where each video is recorded.

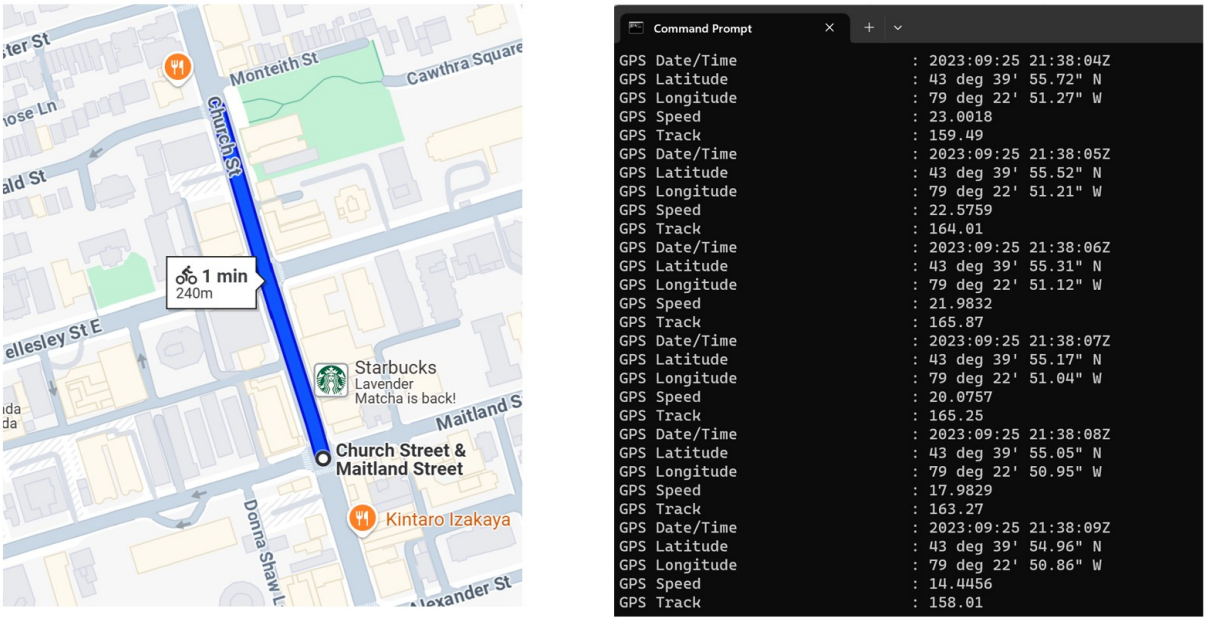
By integrating video content, temporal data, and geospatial coordinates, a comprehensive understanding of on-street parking availability is obtained. The specific requirements of this experiment make selecting a suitable and cost-effective dashcam straightforward. Additionally, since the dashcam is housed within a protective box mounted on the bike, weather conditions such as snow and rain do not affect its performance. The dashcam's power supply is recharged when the bike returns to the station, ensuring it remains operational whenever the bike is in use.

5.4.1 Geofencing

Geofencing creates a virtual boundary around a specific geographical area to trigger actions or alerts when a device crosses this boundary. We use geofencing to establish virtual boundaries around specific areas, such as the streets between two main intersections to monitor parking spots and predict parking availability more accurately. After defining these virtual boundaries, we incorporate the extracted GPS data into our algorithm.

ExifTool is a widely used tool for extracting metadata from media files, including GPS data in images and videos. We use ExifTool to extract embedded GPS data from recorded videos. To facilitate geofencing, we synchronized the video and GPS data chronologically. This synchronization enables our program to accurately determine the precise moments for initiating and ceasing object detection. The vehicles identified within each designated parking area are counted cumulatively, and their respective counts are recorded.

Figure 5.2 panel (b) presents the extracted metadata from the recorded video captured by a dashcam as the biker traverses the path illustrated in panel (a). The extracted metadata includes essential attributes such as date and time, GPS coordinates (latitude and longitude), and the biker's speed. By leveraging this GPS data, we perform a geofencing step by matching the recorded location information with spatial data obtained from Google. This geofencing process enables precise identification of the biker's position within predefined geographic boundaries, facilitating accurate mapping and analysis of parking occupancy on the recorded route.



(a) (b)

Figure 5.2: Overview of the geofencing process using dashcam-recorded data. (a) The recorded path taken by the biker equipped with a dashcam. (b) Extracted metadata from the recorded video.

5.4.2 Video Analysis

We employ the YOLOv8 model by Ultralytics Jocher et al. (2023), trained on the COCO dataset (Lin et al. 2014), for object detection, ensuring accurate and efficient analysis of video data. This model detects various vehicle types and pedestrians, while the open-source Supervision library by Roboflow (Roboflow) is used for annotation tasks, including defining detection zones, drawing bounding boxes, and applying privacy filters by blurring faces.

The dashcam mounted on a bike records video, focusing on the lower half of each frame where parked vehicles are most likely to appear. Detected vehicles are assigned unique IDs, and when the camera enters a geo-fenced parking area, defined using GPS coordinates, the model activates to count the unique vehicle IDs, providing real-time parking occupancy data. In non-parking zones, such as intersections, detection is deactivated, and the count resets to zero.

To ensure accurate detections, the system performs geofencing by matching extracted GPS metadata with Google Maps, activating the model only in designated parking areas. Privacy is preserved by identifying and blurring pedestrians in the video.

Figure 5.3 illustrates this process. The left panel shows a map with geofencing data, including parking availability status, and extracted GPS metadata. The right panel presents a frame from the dashcam footage where the YOLOv8 model has detected and labeled a vehicle with a bounding box, displaying the total parked vehicle count at the bottom. This system enables continuous, real-time parking occupancy monitoring while maintaining privacy considerations.

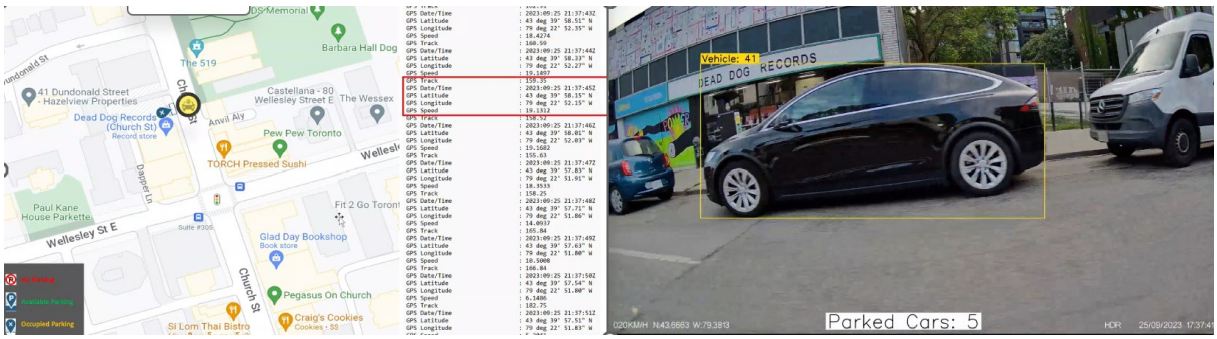


Figure 5.3: Automated Parking Detection System Using Dashcam Video and GPS Data.

5.4.3 Edge Computing

Edge computing involves performing data analysis and computations directly on a local device, rather than transmitting data to a remote cloud server. This method is particularly beneficial in object classification tasks, where processing speed and timely responses are crucial. By running object classification models on an edge device, data can be processed in real time, significantly reducing latency since there is no need to wait for data to travel back and forth between the device and a remote server. Moreover, this approach minimizes bandwidth usage by sending only essential information over the network, if needed, and enhances privacy and security by keeping sensitive data on the device rather than exposing it to potential network vulnerabilities. The resilience of edge computing is especially valuable in environments with limited or unreliable internet connectivity, ensuring that the system remains functional even when off-line. The effectiveness of edge computing for real-time parking occupancy detection using video feeds has been demonstrated by Ke et al. (2020). In this project, we use a Jetson Nano edge processor because its compact design and GPU-accelerated capabilities, provided by NVIDIA's CUDA cores, make it well-suited for these tasks. This device allows for the deployment of complex deep learning models, making it an effective platform for performing high-speed, real-time object classification directly at the source of data capture.

Chapter Summary

This study introduces a scalable and cost-effective approach for real-time parking occupancy detection using mobile sensing agents equipped with dashcams. By integrating the YOLOv8 object detection model with GPS-based geofencing, the system automates vehicle detection and occupancy counting while ensuring privacy through facial blurring. The combination of real-time video analysis and spatial data processing enhances parking monitoring and management in urban environments.

We are actively developing an edge processing framework to enable real-time detection directly on the sensing devices, reducing latency and dependence on cloud-based

computation. Additionally, we aim to incorporate a parking prediction model that leverages historical and real-time data to forecast parking availability. These enhancements will further improve system efficiency, accuracy, and applicability across diverse urban settings, contributing to smarter and more adaptive parking management solutions.

Chapter 6

CHAPTER SIX: CONCLUSION

6.1 Overview of the Research

This dissertation has explored the transformative potential of crowdsensing in addressing key challenges in urban mobility, particularly in dynamic routing, parking management, and real-time urban monitoring. Through four interconnected studies, this work has examined different types of crowdsensing platforms through which user-contributed data can be harnessed to enhance decision-making, optimize system-wide efficiency, and increase social welfare, all while balancing concerns of privacy, participation, and data reliability.

The first study evaluated the relationship between user-generated wait time information and decision-making in service queues. It demonstrated that there exists an optimal ratio of data contributors that maximizes social welfare in high demand service systems. When crowdsensing participation is too low, users make inefficient decisions due to lack of information. Conversely, overly high participation may induce excessive demand and congestion, reducing overall efficiency. The study also highlighted the importance of platform accessibility and the strategic role of voluntary versus mandated data contribution.

The second study investigated the use of crowd-generated travel time data in route choice decisions. It emphasized the role of the contribution ratio and observation window in ensuring accurate and timely information. This study showed that the benefits of such data are sensitive to travelers' risk attitudes and information quality. Moreover, the analysis revealed that excessive access to real-time data can lead to the Informational Braess' Paradox, where more information paradoxically increases congestion. This finding underscores the need for carefully calibrated data-sharing incentives and data access policies.

In the third study, the thesis advanced a resource allocation and route generation framework for urban mobile sensing. This model determined the optimal number and paths of mobile agents tasked with collecting curbside occupancy data. By combining optimization techniques and heuristic algorithms, the study enabled efficient coverage of urban zones while respecting constraints on data collection frequency. The case study on permit parking areas in Toronto validated the proposed approach and demonstrated its

potential to inform the real-world deployment of mobile sensing fleets.

The final study introduced the application of dashcam footage and edge computing for parking occupancy detection. Leveraging modern object detection techniques (e.g., YOLOv8) and geofencing strategies, this study showed that real-time, accurate, and scalable parking information can be collected without relying on expensive fixed infrastructure. This contribution highlights how sensor-equipped vehicles can serve as sensing agents in crowdsensing platforms.

6.2 Summary of Contributions

Collectively, these studies contribute to the theoretical and practical understanding of how crowdsensing can be optimized in complex urban systems. This work:

- Developed analytical and computational models to assess user behavior in response to shared data.
- Quantified the optimal levels of participation and access required for maximizing social benefits.
- Proposed mechanisms for incentivizing participation while addressing privacy and fairness concerns.
- Offered a scalable and implementable approach for curbside data collection using mobile and onboard technologies.

For urban planners, mobility service providers, and policymakers, this dissertation provides strategic guidance on designing crowdsensing platforms that are equitable, efficient, and adaptive to urban growth. From regulating data contribution incentives to allocating mobile sensing resources, the findings present practical insights that can help cities scale smart mobility solutions with minimal infrastructure investment.

6.3 Limitations and Future Research Directions

Despite the contributions, this work is not without limitations. Several simplifying assumptions were made to enable analytical tractability and model interpretability. These include homogeneous user behavior, Poisson-distributed service demand and service times, and the assumption that users act truthfully when contributing data. While useful for theoretical analysis, such assumptions do not fully reflect the complexity of real-world urban systems, where user behavior is heterogeneous, demand patterns are variable, and data quality may be compromised by user manipulation. Addressing these limitations through more flexible behavioral models and robust data validation mechanisms would enhance the realism and applicability of the proposed frameworks. Additionally, the focus of this thesis was limited to parking availability and travel time information. Although these are critical components of urban mobility, broader transportation contexts, such as multimodal trip chains, interdependencies between modes, and system-wide coordination, were beyond the scope of this study.

Building on the findings of this research, several promising directions can be pursued to expand the investigation of crowdsesning platforms. One path is the integration of adaptive incentive mechanisms that dynamically adjust based on system state and user behavior, with the goal of sustaining optimal participation in real-time. Such mechanisms could be grounded in reinforcement learning or game-theoretic frameworks, and tailored to encourage not just data contribution, but also fairness, equity, and long-term engagement.

Another direction is the fusion of passive and active sensing data sources to improve both spatial and temporal coverage. For instance, combining user-generated data with infrastructure-based signals (e.g., traffic signal telemetry, smart curb sensors) may reduce the dependency on participation rates while improving data reliability and system robustness.

6.4 Concluding Remarks

This thesis has demonstrated the potential of crowdsensing to redefine how cities monitor, understand, and manage urban mobility. By transforming ordinary users into distributed sensing agents, crowdsensing creates a participatory ecosystem where data and decision-making are more democratic, adaptive, and efficient. The findings presented here advance the scientific understanding of participation dynamics, resource allocation, and real-time information sharing in complex transportation systems.

Beyond technical contributions, this work explores how emerging technologies can support operational efficiency while addressing sustainability in urban systems. As cities face increasing challenges, ranging from congestion and climate change to infrastructure limitations, solutions that harness collective intelligence will be key to addressing the evolving demands of urban mobility. Advancing crowdsensing at scale will require technological innovation, clear policy frameworks, effective governance, and public trust. This thesis contributes to that effort by informing future research, interdisciplinary collaboration, and real-world implementation in pursuit of smarter, more inclusive urban mobility.

Bibliography

Joseph Abate and Ward Whitt. Transient behavior of the m/m/l queue: Starting at the origin. *Queueing systems*, 2:41–65, 1987.

Kinan Abboud, Hai L. Omar, and Weihua Zhuang. Interworking of dsrc and cellular network technologies for V2X communications: A survey. *IEEE Transactions on Vehicular Technology*, 65(12):9457–9470, 2016. doi: 10.1109/TVT.2016.2532866.

Abdalla Abdelrahman, Amr S El-Wakeel, Aboelmagd Noureldin, and Hossam S Hassanein. Crowdsensing-based personalized dynamic route planning for smart vehicles. *Ieee Network*, 34(3):216–223, 2020.

Hossein Abouee Mehrizi, Mohammad Hossein Eshraghi, and Ming Hu. Efficient sampling: Sample average inference of service capacity in a queue. *Rotman School of Management Working Paper*, (3892899), 2021.

Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar. Informational braess' paradox: The effect of information on traffic congestion. *arXiv preprint arXiv:1601.02039*, 2016.

Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar. Informational braess' paradox: The effect of information on traffic congestion. *Operations Research*, 66 (4):893–917, 2018.

Miguel Aguila, Xu Sun, and Yongxin Wang. A survey on parking management systems: Towards crowdsensing-enabled solutions. In *Proceedings of the 12th ACM Workshop on Vehicular Inter-Networking, Systems, and Applications (VANET)*, pages 12–19, 2019. doi: 10.1145/3349624.3356768.

Hasan Omar Al-Sakran et al. Intelligent traffic information system based on integration of internet of things and agent technology. *International Journal of Advanced Computer*

Science and Applications (IJACSA), 6(2):37–43, 2015.

Akbar Ali, Muhammad Ahsan Qureshi, Muhammad Shiraz, and Azra Shamim. Mobile crowd sensing based dynamic traffic efficiency framework for urban traffic congestion control. *Sustainable Computing: Informatics and Systems*, 32:100608, 2021a.

H. Ali et al. Emerging trends in edge-enabled mobile urban sensing: Blockchain and federated learning for scalability and privacy. *Computers, Environment and Urban Systems*, 88:101–112, 2021b.

Giuseppe Amato, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, and Claudio Vairo. Car parking occupancy detection using smart camera networks and deep learning. In *2016 IEEE Symposium on Computers and Communication (ISCC)*, pages 1212–1217. IEEE, 2016.

Giuseppe Amato, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, Carlo Meghini, and Claudio Vairo. Deep learning for decentralized parking lot occupancy detection. *Expert Systems with Applications*, 72:327–334, 2017.

Khalid Belwafi, Rami Alkadi, Saeed A Alameri, Hadi Al Hamadi, and Giancarlo Fortino. Unmanned aerial vehicles' remote identification: A tutorial and survey. *IEEE Internet of Things Journal*, 10(1):512–536, 2022.

M. L. Betalo, S. Leng, A. M. Seid, and H. N. Abishu. Dynamic charging and path planning for uav-powered rechargeable wsns using multi-agent deep reinforcement learning. *IEEE Transactions on Mobile Computing*, 2025. URL `https://ieeexplore.ieee.org/abstract/document/10962218/`.

Fabian Bock, Sergio Di Martino, and Antonio Origlia. Smart parking: Using a crowd of taxis to sense on-street parking space availability. *IEEE Transactions on Intelligent Transportation Systems*, 21(2):496–508, 2019.

Bosch. Community-based parking, 2016. URL `https://www.bosch.com/stories/community-based-parking/`.

Paul C Box. Curb-parking problems: Overview. *Journal of Transportation Engineering*, 130(1), 2004.

Raj Bridgelall and Denver Tolliver. Accuracy enhancement of anomaly localization with participatory sensing vehicles. *Sensors*, 20(2):409, 2020.

Orhan Bulan, Robert P Loce, Wencheng Wu, Yao Rong Wang, Edgar A Bernal, and Zhigang Fan. Video-based real-time on-street parking occupancy detection system. *Journal of Electronic Imaging*, 22(4):041109, 2013.

Gennaro Cardone, Andrea Cirri, Antonio Corradi, and Luca Foschini. Crowdsensing in urban areas for city-scale mass gathering management: Geofencing and activity recognition. *IEEE Sensors Journal*, 14(12):4185–4195, 2014.

Noppadol Chadil, Apirak Russameesawang, and Phongsak Keeratiwintakorn. Real-time tracking management system using gps, gprs and google earth. In *2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, volume 1, pages 393–396. IEEE, 2008.

S Cheerkoot-Jalim and Z Purahoo. Senseapp: An iot-based mobile crowdsensing application for smart cities. *IEEE International Conference on Emerging Trends in Electrical, Electronic and Communications Engineering (ICETEECE)*, pages 1–6, 2020. doi: 10.1109/ICETEECE49695.2020.9297018.

Alan Chen. How Waze gamifies crowdsourced traffic data. `https://www.waze.com/`, 2019. Accessed: 2023-01-01.

Lun-Chi Chen, Ruey-Kai Sheu, Wen-Yi Peng, Jyh-Horng Wu, and Chien-Hao Tseng. Video-based parking occupancy detection for smart control system. *Applied Sciences*, 10(3):1079, 2020.

Wei Chen, Fang Sun, and Min Yang. Predictive crowdsensing for adaptive routing: A machine learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 19(7):2238–2249, 2018. doi: 10.1109/TITS.2017.2788799.

Z. Chen, H. Wang, and L. Yang. Mobile sensing with reinforcement learning for air quality monitoring. *IEEE Transactions on Mobile Computing*, 2022. doi: 10.1109/TMC.2022.3144700.

J. Cherian, J. Luo, H. Guo, and S.S. Ho. Parkgauge: Gauging the Occupancy of Parking Garages with Crowdsensed Parking Characteristics. In *IEEE International Conference on Mobile Data Management*, 2016. URL `https://personal.ntu.edu.sg/junluo/documents/ParkGauge.pdf`.

City of Toronto. Cycling in toronto – cycling events and programs, 2024a. URL

https://www.toronto.ca/services-payments/streets-parking-transportation/cycling-in-toronto/cycling-events-and-programs/. Accessed: 2024-02-13.

City of Toronto. On-street permit parking area maps. City of Toronto Open Data Portal, 2024b. https://open.toronto.ca/dataset/on-street-permit-parking-area-maps/.

George A. Croes. A method for solving traveling-salesman problems. *Operations Research*, 6 (6):791–812, 1958.

Donatella Darsena, Giuliano Gelli, Ilenia Iudice, and Francesco Verde. Sensing technologies for crowd management, adaptation, and information dissemination in public transportation systems: A review. *IEEE Sensors Journal*, 23(1):44–58, 2023.

Matthew DiChiappari, Nitin Kumar, Ahmed Elhabbash, and Ahmed Kamal. A collaborative crowdsensing platform for urban monitoring using edge and cloud infrastructure. In *2016 IEEE Symposium on Computers and Communication (ISCC)*, pages 1223–1229. IEEE, 2016. doi: 10.1109/ISCC.2016.7543900.

Rong Ding, Zhaoxing Yang, Yifei Wei, Haiming Jin, and Xinbing Wang. Multi-agent reinforcement learning for urban crowd sensing with for-hire vehicles. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*, pages 1–10. IEEE, 2021.

Lili Du and Siyuan Gong. Stochastic poisson game for an online decentralized and coordinated parking mechanism. *Transportation Research Part B: Methodological*, 87:44–63, 2016.

Wei Du, Hui Gao, and Qian Jiang. Incentive mechanisms for crowdsensing: Survey and research challenges. In *Proceedings of the 2017 IEEE International Conference on Communications (ICC)*, pages 1–6, 2017. doi: 10.1109/ICC.2017.7996537.

Noel M Edelson and David K Hilderbrand. Congestion tolls for poisson queuing processes. *Econometrica: Journal of the Econometric Society*, pages 81–92, 1975.

Mustafa A. Fadhel, Ali M. Duhaim, Ali Saihood, and Ahmed Sewify. Comprehensive systematic review of information fusion methods in smart cities and urban environments. *Information Fusion*, 2024. doi: 10.1016/j.inffus.2024.101946.

Albertus Farley, Hanry Ham, et al. Real time ip camera parking occupancy detection using deep learning. *Procedia Computer Science*, 179:606–614, 2021.

Claudio Fiandrino, Fatemeh Anjomshoa, Burak Kantarci, Dzmitry Kliazovich, and Pascal Bouvry. Sociability-driven framework for data acquisition in mobile crowdsensing over fog

computing platforms for smart cities. *IEEE Transactions on Sustainable Computing*, 2(4): 345–358, 2017.

Daniel Freedman, Jihong Chen, and Fang Li. Real-time crowdsourced parking availability estimation using machine learning. In *Proceedings of the 21st IEEE International Conference on Intelligent Transportation Systems (ITSC)*, pages 1081–1086, 2018. doi: 10.1109/ITSC.2018.8569330.

Raghu K Ganti, Fan Ye, and Hui Lei. Mobile crowdsensing: current state and future challenges. *IEEE communications Magazine*, 49(11):32–39, 2011.

Geotab. Data-driven smart city insights, 2017. URL `https://www.geotab.com/white-paper/smart-city-insights/`.

Michele Girolami, Dimitri Belli, Stefano Chessa, and Luca Foschini. How mobility and sociality reshape the context: A decade of experience in mobile crowdsensing. *Sensors*, 21(19):6397, 2021.

A. Gkoumas et al. Deploying micro-mobility units integrated with geospatial satellite data for infrastructure scanning. *Transportation Research Part D: Transport and Environment*, 90: 103–114, 2023.

Thiago S Gomides, E Robson, Fernanda SH Souza, and Daniel L Guidoni. A traffic management system to minimize vehicle congestion in smart cities. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1439–1444. IEEE, 2020.

Giulio Grassi, Kyle Jamieson, Paramvir Bahl, and Giovanni Pau. Parkmaster: An in-vehicle, edge-based video analytics service for detecting open parking spaces in urban environments. In *Proceedings of the Second ACM/IEEE Symposium on Edge Computing*, pages 1–14, 2017.

Radoslav Grbić and Roland Koch. A vision-based system for automatic parking slot detection and occupancy classification. *Computer Vision and Image Understanding*, 226:103653, 2023. doi: 10.1016/j.cviu.2023.103653.

Jin Gu, Zusheng Zhang, Fengqi Yu, and Qun Liu. Design and implementation of a street parking system using wireless sensor networks. In *IEEE 10th International Conference on Industrial Informatics*, pages 1212–1217. IEEE, 2012.

Iftekhar Hasan. *A Framework for Urban Sensing-as-a-Service: Applications in Safety and Mobility Monitoring.* PhD thesis, University of Alabama at Birming-

ham, 2023. URL https://digitalcommons.library.uab.edu/cgi/viewcontent.cgi?article=1420&context=etd-collection.

Refael Hassin and Moshe Haviv. *To queue or not to queue: Equilibrium behavior in queueing systems*, volume 59. Springer Science & Business Media, 2003.

Tianfu He, Jie Bao, Yexin Li, Hui He, and Yu Zheng. Crowd-sensing enhanced parking patrol using sharing bikes' trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 35(4):3589–3602, 2021.

Here. Improve the efficiency of road travel - here, 2020. URL https://www.here.com/sites/g/files/odxslz256/files/2020-09/20_PLA_PMK_One-Pager_Traffic-Product_NEW.pdf.

Here. Traffic analytics speed data, 2022. URL https://www.here.com/learn/blog/traffic-analytics-speed-data.

Takamasa Higuchi and Kentaro Oguchi. Monitoring live parking availability by vision-based vehicular crowdsensing. In *GLOBECOM 2020-2020 IEEE Global Communications Conference*, pages 1–5. IEEE, 2020.

Ming Hu, Yang Li, and Jianfu Wang. Efficient ignorance: Information heterogeneity in a queue. *Management Science*, 64(6):2650–2671, 2018.

Y. Hu, H. Yu, and Z. Shi. Informative path planning for parking occupancy estimation with autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 2023. doi: 10.1109/TITS.2023.3248121.

Longbo Huang and Eytan Modiano. Optimizing age-of-information in a multi-class queueing system. In *2015 IEEE international symposium on information theory (ISIT)*, pages 1681–1685. IEEE, 2015.

Peihuang Huang, Wenxing Zhu, Kewen Liao, Timos Sellis, Zhiyong Yu, and Longkun Guo. Efficient algorithms for flexible sweep coverage in crowdsensing. *IEEE Access*, 6:50055–50065, 2018.

INRIX. INRIX Collaborates with BMW to Provide Real-Time Parking Availability. https://inrix.com/blog/inrix-bmw/. Accessed: 2023-01-01.

INRIX. Customer showcase: Inrix powers bmw's innovative parking service, 2018. URL https://inrix.com/blog/inrix-powers-bmw-ospi-parking-service/.

Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolov8, 2023. URL `https://github.com/ultralytics/ultralytics`.

R. Jones et al. Privacy-preserving techniques in decentralized mobile urban sensing platforms. *IEEE Access*, 12:890–902, 2024.

Sanjit Kaul, Roy Yates, and Marco Gruteser. Real-time status: How often should one update? In *2012 Proceedings IEEE INFOCOM*, pages 2731–2735. IEEE, 2012.

Ruimin Ke, Yifan Zhuang, Ziyuan Pu, and Yinhai Wang. A smart, efficient, and reliable parking surveillance system with edge artificial intelligence on iot devices. *IEEE Transactions on Intelligent Transportation Systems*, 22(8):4962–4974, 2020.

Ben Kim, Sandra M McKay, and Joon Lee. Consumer-grade wearable device for predicting frailty in canadian home care service clients: prospective observational proof-of-concept study. *Journal of Medical Internet Research*, 22(9):e19732, 2020.

Sanghyun Kim, Jihun Park, and Jaehyung Cho. Real-world road network routing via a genetic algorithm enhanced with crowdsensing data. In *Proceedings of the 2019 IEEE International Conference on Big Data (BigData)*, pages 1219–1227, 2019. doi: 10.1109/BigData.2019.9006557.

Mirko Kremer and Laurens Debo. Inferring quality from wait time. *Management Science*, 62 (10):3023–3038, 2016.

Petre Lameski, Elena Mitreska Jovanovska, and Valentin Batz. Methods for urban air pollution measurement and forecasting: Challenges, opportunities, and solutions. *Atmosphere*, 14 (9):1441, 2023. doi: 10.3390/atmos14091441.

Richard C Larson and Amedeo R Odoni. *Urban operations research*. Number Monograph. 1981.

Sangmin Lee, Jihoon Park, and Miyoung Lee. Toward sustainable crowdsensing: Incentive mechanisms, challenges, and future directions. *IEEE Access*, 6:29811–29825, 2018. doi: 10.1109/ACCESS.2018.2841685.

Bin Li and Jia Liu. Can we achieve fresh information with selfish users in mobile crowd-learning? In *2019 International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, pages 1–8. IEEE, 2019.

Dong Li, Hong Zhao, and Zhaohui Wu. Edge computing for crowdsensing in transportation

systems: Architecture and challenges. *Future Generation Computer Systems*, 94:617–624, 2019. doi: 10.1016/j.future.2018.11.001.

Hong Li, Xue Chen, and Rui Wu. Edge-computing-enabled crowdsensing for smart city applications: Architecture and challenges. *IEEE Internet of Things Journal*, 7(8):7067–7079, 2020. doi: 10.1109/JIOT.2020.2982179.

Jun Li, Ying Wang, and Hao Chen. Crowdsensing-based urban mobility monitoring: Challenges and opportunities. *IEEE Transactions on Intelligent Transportation Systems*, 18(3):789–802, 2017. doi: 10.1109/TITS.2017.2654245.

Xin Liang, Licheng Wu, and Hong Cheng. Crowdsensing-based parking guidance with real-time occupancy information. *Transportation Research Part C*, 104:175–188, 2019. doi: 10.1016/j.trc.2019.05.015.

R. Liao, C. Roman, P. Ball, S. Ou, and L. Chen. Crowdsourcing on-street parking space detection. *IEEE Transactions on Intelligent Transportation Systems*, 17(3):721–733, March 2016. doi: 10.1109/TITS.2016.2526785.

Tien-Chun Lin, Yi-Hung Chou, and Wei-Chen Hsieh. A survey on smart parking solutions. *International Journal of Information and Electronics Engineering*, 7(2):59–64, 2017. doi: 10.18178/ijiee.2017.7.2.665.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014.

S. Liniger and B. Stiller. Parking Prediction Techniques in an IoT Environment. Master's Thesis, University of Zurich, 2015. URL `https://capuana.ifi.uzh.ch/publications/PDFs/12259_MT-Samuel_Liniger.pdf`.

Bo Liu, Han Wang, and Xuan Zhu. Deep learning-based parking occupancy prediction with crowdsensing data. *IEEE Transactions on Intelligent Transportation Systems*, 21(11):4625–4635, 2020. doi: 10.1109/TITS.2020.2994185.

X. Liu et al. Integrating satellite imagery with crowd-sourced traffic data for improved urban traffic monitoring. *Transportation Science*, 58(2):89–101, 2024.

Y. Liu et al. Data collection in mobile urban sensing: A review of core architectural models, par-

ticipation strategies, and data quality control mechanisms. *IEEE Transactions on Mobile Computing*, 18(3):560–574, 2019.

Yang Liu and Zhenyu Yang. Information provision and congestion pricing in a risky two-route network with heterogeneous travelers. *Transportation Research Part C: Emerging Technologies*, 128:103083, 2021.

Yang Liu, Yuanyuan Li, and Lu Hu. Departure time and route choices in bottleneck equilibrium under risk and ambiguity. *Transportation Research Part B: Methodological*, 117:774–793, 2018.

Yuxuan Liu, Xiang Zhang, Jiannong Chen, and Yaqian Li. Smartbusaq: Distributed air quality monitoring using sensor-equipped public transportation buses. *IEEE Transactions on Intelligent Transportation Systems*, 22(4):2522–2533, 2021. doi: 10.1109/TITS.2020.2993063.

Antonella Longo, Massimo Zappatore, and Maria Antonietta Bochicchio. Crowd-sourced data collection for urban monitoring via mobile sensors. *ACM Transactions on Sensor Networks*, 2017. URL `https://www.researchgate.net/publication/320640222`.

Qing Lu, Yi Zhang, and Xiao Chen. Multimodal sensing and route optimization for illegal parking detection using bike trajectory data. *Transportation Research Part C: Emerging Technologies*, 150:104005, 2023. doi: 10.1016/j.trc.2023.104005.

Jiaqi Ma, Erin Clausing, and Yimin Liu. Smart on-street parking system to predict parking occupancy and provide a routing strategy using cloud-based analytics. Technical report, SAE Technical Paper, 2017a.

Jun Ma, Wei Zhang, and Li Chen. Vehicle-based crowdsensing for urban traffic monitoring. *Transportation Research Part C*, 77:28–44, 2017b. doi: 10.1016/j.trc.2017.01.004.

Hani S Mahmassani and R Jayakrishnan. System performance and user response under real-time information in a congested traffic corridor. *Transportation Research Part A: General*, 25 (5):293–307, 1991.

G. Marques, R. Pitarma, and B. Ferreira. Air quality monitoring with sensor networks: Optimization approaches. *Journal of Cleaner Production*, 261:121151, 2020. doi: 10.1016/j.jclepro.2020.121151.

Thomas J Matarazzo, Daniel Kondor, Salvatore Milardo, and Carlo Ratti. Crowdsourcing

bridge dynamic monitoring with smartphone vehicle trips. *Communications Engineering*, 1(1):1–10, 2022.

Suhas Mathur, Tong Jin, Nikhil Kasturirangan, Janani Chandrasekaran, Wenzhi Xue, Marco Gruteser, and Wade Trappe. Parknet: drive-by sensing of road-side parking statistics. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pages 123–136, 2010.

Michael F Mcgrath, Donald Gross, and Nozer D Singpurwalla. A subjective bayesian approach to the theory of queues i—modeling. *Queueing Systems*, 1(4):317–333, 1987.

Metrolinx. Costs of road congestion in the gtha (greater toronto and hamilton area), Dec 2008. Accessed: 2023-06-22.

Ricardo Miranda, Carlos Alves, Rui Sousa, Ana Chaves, and Luis Montenegro. Revolutionising the quality of life: The role of real-time sensing in smart cities. *Electronics*, 13(3):550, 2024. doi: 10.3390/electronics13030550.

S. Mohamed et al. Multimodal sensor integration for environmental and mobility monitoring: A case study using drones, vehicles, and wearable devices. *Urban Computing*, 15(3):120–132, 2024.

Douglas C Montgomery and George C Runger. *Applied statistics and probability for engineers*. John wiley & sons, 2010.

A. Montori et al. Collaborative iot framework for urban monitoring using edge computing. *Journal of Urban Technology*, 24(2):134–147, 2017.

Philip McCord Morse et al. Queues, inventories, and maintenance. 1958.

Anandatirtha Nandugudi, Taeyeon Ki, Carl Nuessle, and Geoffrey Challen. Pocketparker: Pocketsourcing parking lot availability. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 963–973, 2014.

Pinhas Naor. The regulation of queue size by levying tolls. *Econometrica: journal of the Econometric Society*, pages 15–24, 1969.

Sarfraz Nawaz, Christos Efstratiou, and Cecilia Mascolo. Parksense: A smartphone based sensing system for on-street parking. In *Proceedings of the 19th annual international conference on Mobile computing & networking*, pages 75–86, 2013.

Elnaz Nemati, Chelsea Batteate, and Michael Jerrett. Opportunistic environmental sensing with

smartphones: A critical review of current literature and applications. *Current Environmental Health Reports*, 4(4):405–414, 2017.

Evdokia Nikolova and Nicolás E Stier-Moses. A mean-risk model for the traffic assignment problem with stochastic travel times. *Operations Research*, 62(2):366–382, 2014.

Bei Pan, Yu Zheng, David Wilkie, and Cyrus Shahabi. Crowd sensing of traffic anomalies based on human mobility and social media. In *Proceedings of the 21st ACM SIGSPATIAL international conference on advances in geographic information systems*, pages 344–353, 2013.

Juan Pan, Iulian Sandu Popa, and Cristian Borcea. Divert: A distributed vehicular traffic rerouting system for congestion avoidance. *IEEE Transactions on Mobile Computing*, 16(1): 58–72, 2016.

Y. Pang, M. Tan, and C. Goh. Minimizing sensing units for real-time on-street parking monitoring using public transit vehicles. *Transportation Research Part C: Emerging Technologies*, 150:104049, 2023. doi: 10.1016/j.trc.2023.104049.

Roma Patni and Gayatri Jain. A survey of traffic congestion detection and management technique using vanet. *Int. J. Emerg. Technol. Innovat. Res*, 2:53–58, 2015.

Jamol Pender, Richard H Rand, and Elizabeth Wesson. An analysis of queues with delayed information and time-varying arrival rates. *Nonlinear Dynamics*, 91(4):2411–2427, 2018.

Virginia Pilloni. How data will transform industrial processes: Crowdsensing, crowdsourcing and big data as pillars of industry 4.0. *Future Internet*, 10(3):24, 2018.

Jesper C Provoost, Andreas Kamilaris, Luc JJ Wismans, Sander J van der Drift, and Maurice van Keulen. Predicting parking occupancy via machine learning in the web of things. *Internet of Things*, 12:100301, 2020.

Mohammad Radmanesh and Hossam Bani Younes. Overview of path planning algorithms for uavs in complex environments. *Journal of Aerospace Technology and Management*, 10:1–12, 2018. doi: 10.5028/jatm.v10.870. URL `https://www.scielo.br/j/jatm/a/xMg6pMksnSSkRWptL5Dxv8F/`.

Sayuti Rahman, Marwan Ramli, Fitri Arnia, Arnes Sembiring, and Rusdha Muharar. Convolutional neural network customization for parking occupancy detection. In *2020 International Conference on Electrical Engineering and Informatics (ICELTICs)*, pages 1–6. IEEE, 2020.

Mohsin Raza, Ali Raza Barket, Ateeq Ur Rehman, Abdul Rehman, and Inam Ullah. Mobile crowdsensing based architecture for intelligent traffic prediction and quickest path selection. In *2020 International Conference on UK-China Emerging Technologies (UCET)*, pages 1–4. IEEE, 2020.

G Revathi and VR Sarma Dhulipala. Smart parking systems and sensors: A survey. In *2012 International Conference on Computing, Communication and Applications*, pages 1–5. IEEE, 2012.

Roboflow. Supervision. URL `https://github.com/roboflow/supervision`.

João G. P. Rodrigues. *Large-scale Geo-referenced Sensing Using Personal Mobile Devices*. PhD thesis, University of Porto, 2020. URL `https://repositorio-aberto.up.pt/bitstream/10216/132045/2/441758.pdf`.

Cristian Roman, Ruizhi Liao, Peter Ball, Shumao Ou, and Martin de Heaver. Detecting on-street parking spaces in smart cities: Performance evaluation of fixed and mobile sensing systems. *IEEE transactions on intelligent transportation systems*, 19(7):2234–2245, 2018.

Alexander Rowley and Osman Karakuş. Predicting air quality via multimodal ai and satellite imagery. *Remote Sensing of Environment*, 293:113659, 2023. doi: 10.1016/j.rse.2023.113659.

A. Santhosh et al. A hybrid framework combining macro-level satellite imagery with micro-level smartphone data to monitor urban heat, air quality, and congestion. *Journal of Urban Mobility*, 12(4):234–245, 2024.

Victor Kathan Sarker, Tuan Nguyen Gia, Imed Ben Dhaou, and Tomi Westerlund. Smart parking system with dynamic pricing, edge-cloud computing and lora. *Sensors*, 20(17): 4669, 2020.

Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Ganesan. Finding a "kneedle" in a haystack: Detecting knee points in system behavior. In *Proceedings of the 31st International Conference on Distributed Computing Systems Workshops*, pages 166–171. IEEE, 2011.

SFpark. Sfpark pilot project evaluation, 2014. URL `https://www.sfmta.com/sites/default/files/reports-and-documents/2018/08/sfpark_pilot_project_evaluation.pdf`.

Meet Mineshbhai Shah, Shubham Kamlesh Sinha, and P Madhavan. Crowdsensing using geofencing for smart parking. *Turkish Journal of Computer and Mathematics Education*, 12 (11):791–796, 2021.

Lin Shi, Minghui Zhang, and Lei Zhao. Privacy-preserving mechanisms in crowdsensing: A comprehensive review. *Computer Communications*, 180:1–12, 2021. doi: 10.1016/j.comcom.2021.08.015.

Donald Shoup. *The High Cost of Free Parking*. Routledge, updated edition, 2011.

Donald C Shoup. Cruising for parking. *Transport policy*, 13(6):479–486, 2006.

Rachel Smith, Tom Anderson, and Jack Miller. Addressing privacy concerns in crowdsensed parking systems. *Sensors*, 20(6):1638, 2020. doi: 10.3390/s20061638.

Marius M. Solomon. Algorithms for the vehicle routing and scheduling problems with time window constraints, 1987. URL `https://www.sintef.no/projectweb/top/vrptw/solomon-benchmark/`. Solomon R1 dataset for the Vehicle Routing Problem with Time Windows.

Stephen Spana and Lili Du. Optimal information perturbation for traffic congestion mitigation: Gaussian process regression and optimization. *Transportation Research Part C: Emerging Technologies*, 138:103647, 2022.

Streets.mn. A history of traffic management technology, 2019. URL `https://streets.mn/2019/10/16/a-history-of-traffic-management-technology/`. Accessed: 2025-02-20.

Lajos Takács. Introduction to the theory of queues. Technical report, 1962.

Prasad P Tasgaonkar, Rajesh D Garg, and Pradeep K Garg. An iot-based framework of vehicle accident detection for smart city. *IETE Journal of Research*, 2024. doi: 10.1080/03772063.2023.2239757.

Hamidreza Tavafoghi and Demosthenis Teneketzis. Strategic information provision in routing games, 2019.

Hamidreza Tavafoghi, Kameshwar Poolla, and Pravin Varaiya. A queuing approach to parking: Modeling, verification, and prediction. *arXiv preprint arXiv:1908.11479*, 2019.

Piyushimita (Vonu) Thakuriah, Katarzyna Sila-Nowicka, and Joon Hong. Integrated multimedia city data (imcd): A composite survey and sensing approach to understanding urban living and mobility. *Computers, Environment and Urban Systems*, 80:101438, 2020. doi: 10.1016/j.compenvurbsys.2019.101438.

Erik T Verhoef, Richard HM Emmerink, Peter Nijkamp, and Piet Rietveld. Information provi-

sion, flat and fine congestion tolling and the efficiency of road usage. *Regional Science and Urban Economics*, 26(5):505–529, 1996.

Eleni I Vlahogianni, Konstantinos Kepaptsoglou, Vassileios Tsetsos, and Matthew G Karlaftis. A real-time parking prediction system for smart cities. *Journal of Intelligent Transportation Systems*, 20(2):192–204, 2016.

J. Wang, M. Zhao, T. Huang, and X. Wang. When and how to sense: A survey of optimization approaches in smart city mobile crowdsensing. *ACM Computing Surveys*, 2024. In press.

Jian Wang, Qun Yang, and Wei Liu. Privacy-aware machine learning for crowdsensing-based transportation applications. *Sensors*, 21(7):2351, 2021. doi: 10.3390/s21072351.

Jianfu Wang and Ming Hu. Efficient inaccuracy: User-generated information sharing in a queue. *Management Science*, 66(10):4648–4666, 2020.

Xiang Wang, Jun Li, and Lei Zhao. A real-time parking guidance system using crowdsensing data. *Transportation Research Part C*, 104:187–197, 2019. doi: 10.1016/j.trc.2019.05.027.

Xiong Wang, Jinbei Zhang, Xiaohua Tian, Xiaoying Gan, Yunfeng Guan, and Xinbing Wang. Crowdsensing-based consensus incident report for road traffic acquisition. *IEEE transactions on intelligent transportation systems*, 19(8):2536–2547, 2017.

Lijun Wei, Yuhan Yang, Jing Wu, Chengnian Long, and Bo Li. Trust management for internet of things: A comprehensive study. *IEEE Internet of Things Journal*, 9(10):7664–7679, 2022.

Ward Whitt. Predicting queueing delays. *Management science*, 45(6):870–888, 1999.

Laurence A. Wolsey. *Integer Programming*. Wiley-Interscience, 1998.

Ming-Che Wu and Mei-Chen Yeh. Early detection of vacant parking spaces using dashcam videos. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 9613–9618, 2019.

Jun Xiao, Yingyan Lou, and Joshua Frisby. How likely am i to find parking?–a practical model-based framework for predicting parking availability. *Transportation Research Part B: Methodological*, 112:19–39, 2018.

Bo Xu, Ouri Wolfson, Jie Yang, Leon Stenneth, S Yu Philip, and Peter C Nelson. Real-time street parking availability estimation. In *2013 IEEE 14th International Conference on Mobile Data Management*, volume 1, pages 16–25. IEEE, 2013.

Tao Xu, Ping Liu, and Lei Wang. A hybrid metaheuristic method for traffic routing optimization

with crowdsensing. *Expert Systems with Applications*, 155:113428, 2020. doi: 10.1016/j.eswa.2020.113428.

Hai Yang. Multiple equilibrium behaviors and advanced traveler information systems with endogenous market penetration. *Transportation Research Part B: Methodological*, 32(3): 205–218, 1998.

Haiyang Yu, Jing Fang, Shuai Liu, Yilong Ren, and Jian Lu. A node optimization model based on the spatiotemporal characteristics of the road network for urban traffic mobile crowd sensing. *Vehicular Communications*, 31:100383, 2021.

Xiaojuan Yu, Vincent AC van den Berg, and Zhi-Chun Li. Congestion pricing and information provision under uncertainty: Responsive versus habitual pricing. *Transportation Research Part E: Logistics and Transportation Review*, 175:103119, 2023.

Marco Zappatore, Antonella Longo, and Maria Angela Bochicchio. Crowd-sensing our smart cities: A platform for noise monitoring and acoustic urban planning. *Journal of Communications Software and Systems*, 13(2):68–76, 2017.

Kenan Zhang and Yu Marco Nie. Mitigating the impact of selfish routing: An optimal-ratio control scheme (orcs) inspired by autonomous driving. *Transportation Research Part C: Emerging Technologies*, 87:75–90, 2018.

L. Zhang et al. Distributed edge computing for real-time environmental sensing in urban monitoring. *Journal of Parallel and Distributed Computing*, 132:112–124, 2023.

Lin Zhang, Hui Zhao, and Jun Li. Adaptive routing with real-time crowdsensed data in urban traffic networks. *Transportation Research Part C*, 110:203–216, 2020a. doi: 10.1016/j.trc.2020.01.005.

Ningyu Zhang, Huajun Chen, Xi Chen, and Jiaoyan Chen. Forecasting public transit use by crowdsensing and semantic trajectory mining: Case studies. *ISPRS International Journal of Geo-Information*, 5(10):180, 2016.

Yifan Zhang, Hongyan Li, and Xiang Wang. Privacy-aware crowdsensing schemes in smart cities: Models and applications. *Sensors*, 20(10):2981, 2020b. doi: 10.3390/s20102981.

Zusheng Zhang, Ming Tao, and Huaqiang Yuan. A parking occupancy detection algorithm based on amr sensor. *IEEE Sensors Journal*, 15(2):1261–1269, 2015. doi: 10.1109/JSEN.2014.2362122.

Yating Zhu, Xiaofei Ye, Jun Chen, Xingchen Yan, and Tao Wang. Impact of cruising for parking on travel time of traffic flow. *Sustainability*, 12(8):3079, 2020.

APPENDICES

Appendix A: Supplementary Material for Chapter Two

A. 1 Transition Probabilities in an $M/M/1/K$ Queue

Takács (1962) defines the transition probability of an $M/M/1/K$ queue, denoted as $P_{\phi,n}(\delta)$, as the summation of two terms as follows

$$P_{\phi,n}(\delta) = P_n + d_{\phi,n}(\delta), 1 \leqslant \phi \leqslant K, 0 \leqslant n \leqslant K, \tag{A.1}$$

where

$$P_n = \begin{cases} \frac{(1-\rho_u)\rho_u^n}{1-\rho_u^{K+1}}, & \text{if } \rho_u \neq 1 \\ \frac{1}{K+1}, & \text{if } \rho_u = 1 \end{cases}, \tag{A.2}$$

and

$$\begin{aligned} d_{n,\phi}(\delta) = \frac{2}{K+1}\rho_u^{(n-\phi)/2} \sum_{m=1}^{K} & \frac{e^{-\left(\lambda_u+\mu-2\sqrt{\lambda_u\mu}\,\cos(\frac{m\pi}{K+1})\right)\delta}}{1-2\sqrt{\rho_u}\,\cos(\frac{m\pi}{K+1})+\rho_u} \\ & \times \left(\sin\frac{\phi m\pi}{K+1} - \sqrt{\rho_u}\,\sin\frac{(\phi+1)m\pi}{K+1}\right) \\ & \times \left(\sin\frac{nm\pi}{K+1} - \sqrt{\rho_u}\,\sin\frac{(n+1)m\pi}{K+1}\right). \end{aligned} \tag{A.3}$$

The first term, P_n presented in (A.2), represents the steady-state probability of having n customers in the queue as δ tends to infinity, where λ_u and $\rho_u = \lambda_u/\mu$ denote the arrival rate of non-contributors to the second stage and the offered load of non-contributors, respectively. Mcgrath et al. (1987) refers to P_n as the steady-state distribution of the

system size. The second term, $d_{\phi,n}(\delta)$, which is provided in (A.3), constitutes a deviation component that depends on the initial state of the queue at the beginning of the transition period, the duration of the transition period, and the final state of the queue at its end.

A. 2 Proof of Proposition 2.1

(i) We can express the partial derivative of $P_{\phi,K}(\delta)$ with respect to β as

$$\frac{\partial P_{\phi,K}(\delta)}{\partial \beta} = \frac{\partial P_{\phi,K}(\delta)}{\partial \rho_u} \times \frac{\partial \rho_u}{\partial \beta}. \tag{A.4}$$

According to the definition of δ, only non-contributors enter the system within the δ time interval. With a higher service request rate for non-contributors λ_u, the probability of having more customers in the system and, consequently, the probability of reaching the maximum queue length within δ also increases. Therefore, given (ϕ, δ), the transition blocking probability, $P_{\phi,K}(\delta)$, increases with the service request rate of the non-contributors, λ_u. Using this property, we have $\partial P_{\phi,K}(\delta)/\partial \lambda_u > 0$. Additionally, $\partial \lambda_u / \partial \rho_u$ is positive for a fixed service rate. Therefore,

$$\frac{\partial P_{\phi,K}(\delta)}{\partial \rho_u} = \frac{\partial P_{\phi,K}(\delta)}{\partial \lambda_u} \times \frac{\partial \lambda_u}{\partial \rho_u} \geqslant 0. \tag{A.5}$$

Considering that $\lambda_u = \lambda(1-\beta)$, the service request rate of non-contributors decreases with β. Therefore, $\partial \lambda_u / \partial \beta$ is negative. Thus, we have

$$\frac{\partial \rho_u}{\partial \beta} = \frac{\partial \rho_u}{\partial \lambda_u} \times \frac{\partial \lambda_u}{\partial \beta} \leqslant 0. \tag{A.6}$$

Finally, using (A.5) and (A.6), we conclude that $\partial P_{\phi,K}(\delta)/\partial \beta$ is negative. □

(ii) If all customers contribute data, $\beta = 1$, the service request rate of non-contributors λ_u is zero, resulting in $\rho_u = 0$. Consequently, for all ϕ values, $d_{\phi,n}(\delta)$ is zero. Furthermore, for $1 \leqslant \phi \leqslant K-1$, the value of P_K is also zero. Based on (A.1), this means that the transition blocking probability $P_{\phi,K}(\delta)$ is zero. □

A. 3 Proof of Lemma 2.2

(i) As δ approaches infinity, the term $d_{\phi,n}(\delta)$ tends to zero. Consequently, the transition probability given in (A.1) equals P_n. $\square$

(ii) In accordance with the definition of the expected congestion, we have

$$E[\Gamma(\phi,\delta)] = \sum_{n=0}^{K} n \times P_{\phi,n}(\delta),\ 1 \leqslant \phi \leqslant K. \tag{A.7}$$

Using Lemma 2.2-(i), we rewrite (A.7) as

$$\lim_{\delta\to\infty} E[\Gamma(\phi,\delta)] = \sum_{n=0}^{K} n \times P_n,\ 1 \leqslant \phi \leqslant K. \tag{A.8}$$

The expression in (A.8) is the definition of the expected number of customers in an $M/M/1/K$ queue. $\square$

A. 4 Proof of Proposition 2.3

Utilizing (2.4), and considering Lemma 2.2, the steady-state entrance utility is

$$U^* = \lim_{\delta\to\infty} U_{enter}(\phi,\delta) = R - h - c\frac{L+1}{\mu} - P_K\Big(R - c\frac{K+1}{\mu}\Big). \tag{A.9}$$

By substituting P_K from (A.2) in (A.9), U^* is further simplified as

$$U^* = \begin{cases} R - h - c\frac{L+1}{\mu} - \frac{(1-\rho_u)\rho_u^K}{1-\rho_u^{K+1}}\Big(R - c\frac{K+1}{\mu}\Big), & \text{if } \rho_u \neq 1 \\ R - h - c\frac{L+1}{\mu} - \frac{1}{K+1}\Big(R - c\frac{K+1}{\mu}\Big), & \text{if } \rho_u = 1 \end{cases}, 1 \leqslant \phi \leqslant K. \square \tag{A.10}$$

A. 5 The Expected Congestion

When a customer requests service long after a QL update, the reported congestion ϕ would no longer represent the system's current state. Therefore, confirmed by Lemma

2.2, the expected congestion is independent of ϕ and equals the steady-state number of customers in an $M/M/1/K$ queue, L. The behaviour of the expected congestion based on δ is illustrated in Figure A.4. In both panels, the service rate μ equals one. Panel (a)

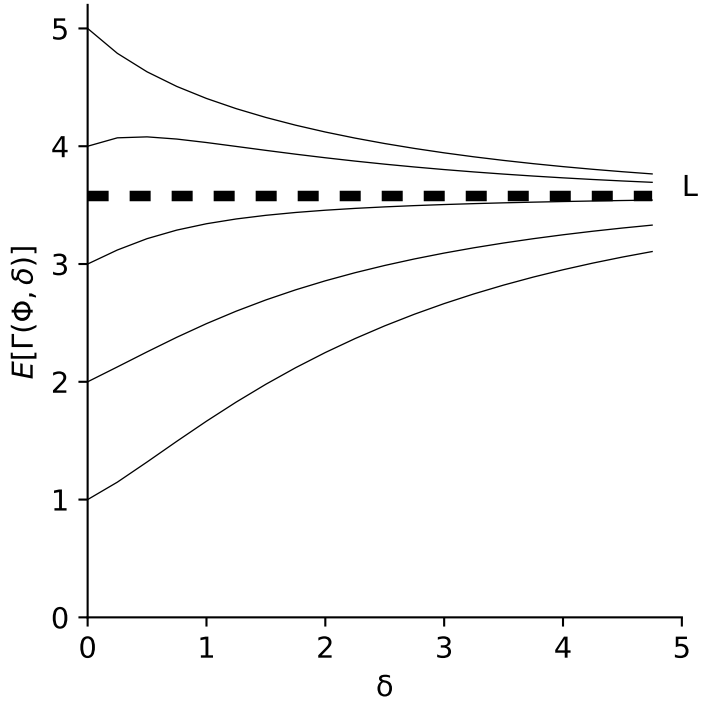


Figure A.1: Expected congestion $E[\Gamma(\phi, \delta)]$ with respect to δ, given the reported QL is $\phi(T) = 1, 2, \ldots, 5$. Parameters: $\mu = 1$, $R = 10$, $c = 2$, $h = 1$, $\beta = 0.5$, and (a) $\lambda = 0.5$, (b) $\lambda = 1.5$.

corresponds to an uncongested system with a service request rate $\lambda = 0.5$, while Panel (b) represents a congested system with $\lambda = 1.5$. As Lemma 2.5 confirms, the expected congestion equals ϕ at $\delta = 0$ in both cases. In Panel (a), all values of $\phi = 1, 2, 3, 4, 5$ exceed the threshold $L = 0.3$, resulting in a decreasing expected congestion with δ. The curve eventually converges to L as δ becomes large enough. In Panel (b), when $\phi = 1$, the expected congestion increases with δ since $L = 1.7$ is greater than one. However, for $\phi = 2, 3, 4, 5$, the expected congestion decreases with δ and eventually converges to L for large values of δ.

Lemma .1. Given (ϕ, δ), the expected congestion $E[\Gamma(\phi, \delta)]$ is a decreasing function of β.

The partial derivative of $E[\Gamma(\delta,\phi)]$ with respect to β is

$$\frac{\partial E[\Gamma(\delta,\phi)]}{\partial \beta} = \frac{\partial E[\Gamma(\delta,\phi)]}{\partial P_{\phi,n}(\delta)} \times \frac{\partial P_{\phi,n}(\delta)}{\partial \lambda_u} \times \frac{\partial \lambda_u}{\partial \beta}. \tag{A.11}$$

Using (A.7), we have

$$\frac{\partial E[\Gamma(\delta,\phi)]}{\partial P_{\phi,n}(\delta)} = \sum_{n=0}^{K} n = \frac{K(K+1)}{2}, \tag{A.12}$$

which is a positive value. As we discussed earlier, $\partial P_{\phi,n}(\delta)/\partial \lambda_u$ is also positive. When β increases, the ratio of non-contributors decreases, resulting in a decrease in the service request rate of non-contributors. Therefore, $\partial \lambda_u/\partial \beta < 0$. Finally, we conclude that $\partial E[\Gamma(\delta,\phi)]/\partial \beta$ is negative. □

Given the total service request rate, λ, it is observed that the service request rate of non-contributors, λ_u, during a given time interval δ, decreases with the contribution ratio β. Furthermore, within the same time interval δ, some customers receive their service and exit the system. Consequently, the expected congestion in the system δ time units after the most recent QL update ϕ, also decreases with the contribution ratio.

We have conducted multiple numerical experiments to support the assumption that δ_d is zero. We present some of these results here to demonstrate that this assumption closely aligns with reality and does not undermine our analysis. Figure A.2 illustrates the expected congestion based on δ for different service request values. In only a few cases does the expected congestion fail to exhibit completely monotone behavior, while in the majority of cases, δ_d is zero, resulting in the expected congestion to be either monotonically increasing or decreasing.

A. 6 Proof of Lemma 2.4

Using (2.10), the partial derivative of the expected entrance utility is

$$\frac{\partial U_{enter}(\phi,\delta)}{\partial \delta} = \frac{\partial P_{\phi,K}(\delta)}{\partial \delta}\Big(\frac{c}{\mu}(K+1) - R\Big) - \frac{c}{\mu} \times \frac{\partial E[\Gamma(\phi,\delta)]}{\partial \delta}. \tag{A.13}$$

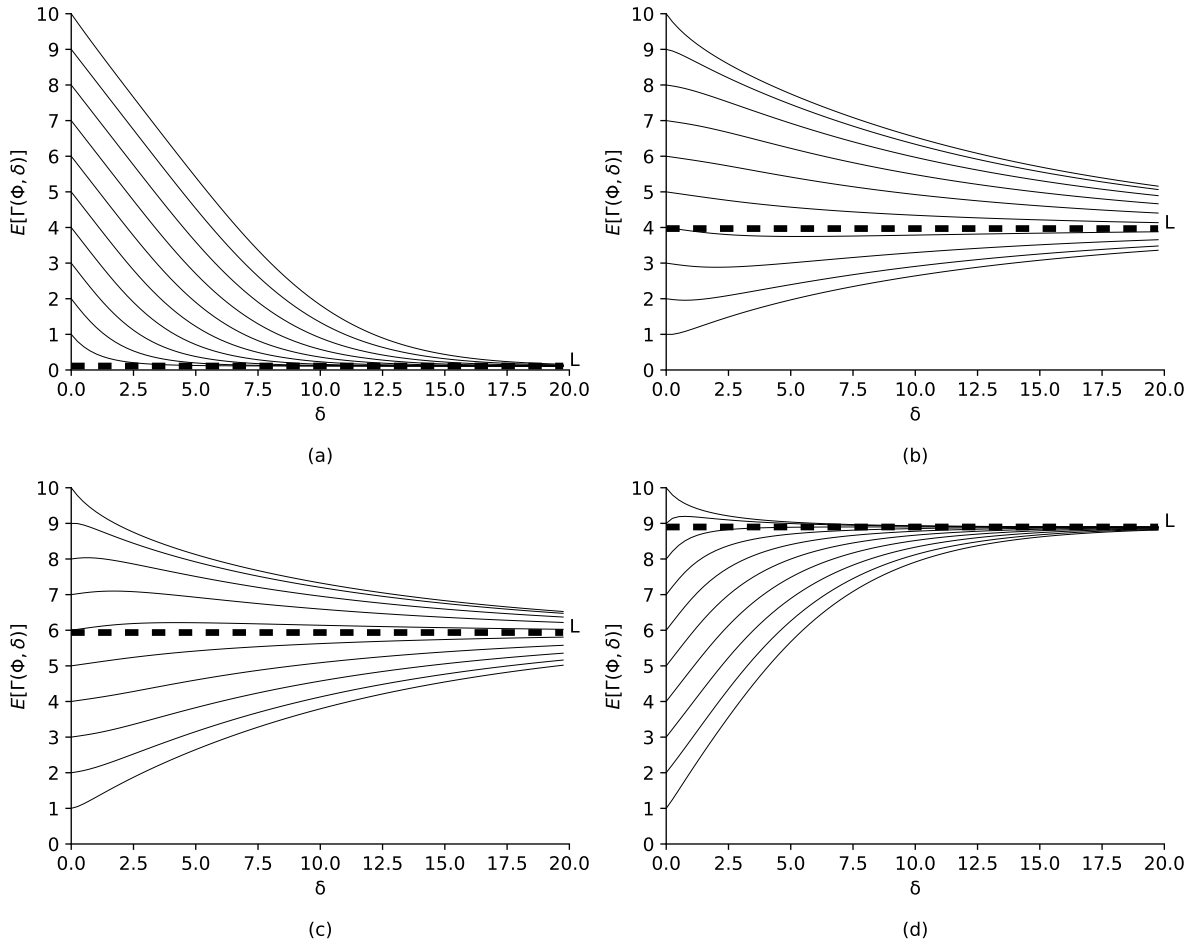


Figure A.2: Expected congestion, $E[\Gamma(\phi,\delta)]$, w.r.t δ, Given the Reported QL is $\phi(T) = 1, 2, \ldots, 10$. (Parameters: $\mu = 1$, $R = 10$, $c = 2$, $h = 1$, and $\lambda =$ (a) 0.1, (b) 0.9, (c) 1.1, (d) 1.9

After substituting K from (2.2) into (A.13), we have

$$\frac{\partial U_{enter}(\phi,\delta)}{\partial\delta} = \frac{c}{\mu}\Big(\frac{\partial P_{\phi,K}(\delta)}{\partial\delta} - \frac{\partial E[\Gamma(\phi,\delta)]}{\partial\delta}\Big). \tag{A.14}$$

Using (A.7), we drive

$$\frac{\partial E[\Gamma(\phi,\delta)]}{\partial\delta} = \sum_{n=0}^{K} n\frac{\partial P_{\phi,n}(\delta)}{\partial\delta} = \sum_{n=0}^{K-1} n\frac{\partial P_{\phi,n}(\delta)}{\partial\delta} + K\frac{\partial P_{\phi,K}(\delta)}{\partial\delta}. \tag{A.15}$$

Therefore, given $K \geqslant 1$, from (A.14) we conclude that

$$\text{Sign}\Big(\frac{\partial U_{enter}(\phi,\delta)}{\partial\delta}\Big) = \text{Sign}\Big(-\frac{\partial E[\Gamma(\phi,\delta)]}{\partial\delta}\Big). \tag{A.16}$$

As discussed in section 2.4.4, given that δ_d goes to zero, when ϕ is less than ϕ_d, $E[\Gamma(\phi, \delta)]$ behaves as a monotone increasing function of δ. This means that $\partial E[\Gamma(\phi, \delta)]/\partial\delta$ is positive. Therefore, we have $\partial U_{enter}(\phi, \delta)/\partial\delta \leqslant 0$, indicating that the entrance utility becomes a monotonically decreasing function of δ. Conversely, suppose ϕ is greater than or equal to ϕ_d. In that case, $\partial E[\Gamma(\phi, \delta)]/\partial\delta$ is negative, which results in $\partial U_{enter}(\phi, \delta)/\partial\delta \geqslant 0$, meaning the expected utility is a monotonically increasing function of δ. □

A. 7 Lemma .2

LEMMA .2. If the time interval δ tends to zero, the probability of change in the number of customers in the system tends to zero, regardless of the contribution ratio β.

$$\lim_{\delta\to 0} P_{\phi,n}(\delta) = \begin{cases} 0 & ,\text{if } n \neq \phi \\ 1 & ,\text{if } n = \phi \end{cases}, \ 1 \leqslant \phi \leqslant K, \ 0 \leqslant n \leqslant K. \tag{A.17}$$

Proof of Lemma .2 By definition, ϕ represents the most recent QL posted by a contributor on the POP. The service request rate of customers follows a Poisson distribution. According to the properties of the Poisson distribution, the probability of two or more events occurring simultaneously is zero. Therefore, if δ tends to zero, the real-time number of customers in the system equals the reported congestion ϕ, regardless of the contribution ratio. □

A. 8 Proof of Lemma 2.5

Using Lemma .2 and the definition of the expected congestion in (A.7), we have

$$\lim_{\delta\to 0} E[\Gamma(\delta, \phi)] = 1 \times 0 + 2 \times 0 + \cdots + \phi \times 1 + \cdots + K \times 0 = \phi. \ \square \tag{A.18}$$

A. 9 Proof of Proposition 3.7

We apply the Intermediate Value Theorem to proof the existence of $\hat{\delta}(\phi)$. Utilizing (2.10), $U_{enter}(\phi, \delta)$ at $\delta = 0$ is calculated. Leveraging Proposition 2.1 and Lemma 2.5, two potential situations are analyzed. If $\phi = K$, then $P_{K,K}(0) = 1$, and $E[\Gamma(K, 0)] = K$. Thus, $U_{enter}(K, 0)$ equals $-h$, which is negative. Otherwise, if $\hat{\phi} < \phi < K$, then $P_{\phi,K}(0) = 0$, and $E[\Gamma(\phi, 0)] = \phi$. Thus, $U_{enter}(\phi, 0) = R - \frac{C}{\mu}(\phi + 1) - h$. Given (2.11), we have $\phi > \frac{\mu}{C}(R-h)-1$. Therefore, $U_{enter}(\phi, 0)$ would be negative for $\hat{\phi} < \phi < K$. On the other hand, Proposition 2.3 states that as δ approaches infinity, for any reported congestion, the entrance utility approaches U^*, which is a positive value. Therefore, $U_{enter}(\phi, \infty) > 0$, for all $1 \leqslant \phi \leqslant K$. In addition, $U_{enter}(\phi, \delta)$ is a continuous function with respect to δ in the interval $[0, \infty)$. Therefore, by applying the Intermediate Value Theorem, for $\hat{\phi} < \phi \leqslant K$, there must exist at least one real number, such as $\hat{\delta}(\phi)$, in the interval $(0, \infty)$ such that $U_{enter}(\phi, \hat{\delta}(\phi)) = 0$. Proving the existence of $\hat{\delta}(\phi)$, the uniqueness of $\hat{\delta}(\phi)$ directly follows from Lemma 2.4. □

A. 10 Proof of Proposition 2.7

Social welfare represents the expected utility that customers receive and is defined as

$$\begin{aligned} SW &= p\left(U_{\text{enter}} \geqslant 0\right)\left[\left(1 - P_{\phi,k}(\delta)\right)\left(R - \frac{c}{\mu}\Big(E\big(\Gamma(\phi,\delta)\big) + 1\Big) - h\right) + P_{\phi,k}(\delta)(-h)\right] \\ &= p\left(U_{\text{enter}} \geqslant 0\right)\left[R - \frac{c}{\mu}\Big(E\big(\Gamma(\phi,\delta)\big) + 1\Big) - h - P_{\phi,k}(\delta)\left(R - \frac{c}{\mu}\Big(E\big(\Gamma(\phi,\delta)\big) + 1\Big)\right)\right]. \end{aligned} \tag{A.19}$$

We define

$$A = R - \frac{c}{\mu}\Big(E\big(\Gamma(\phi,\delta)\big) + 1\Big) - h - P_{\phi,k}(\delta)\left(R - \frac{c}{\mu}\big(E\big(\Gamma(\phi,\delta)\big) + 1\big)\right), \quad A \geqslant 0. \tag{A.20}$$

The partial derivative of A with respect to β is

$$\begin{aligned}\frac{\partial A}{\partial \beta} &= -\frac{c}{\mu}\frac{\partial E\big(\Gamma(\phi,\delta)\big)}{\partial \beta} - \frac{\partial P_{\phi,k}(\delta)}{\partial \beta}\left(R - \frac{c}{\mu}(E\big(\Gamma(\phi,\delta)\big)+1)\right) + \frac{c}{\mu}P_{\phi,k}(\delta)\frac{\partial E\big(\Gamma(\phi,\delta)\big)}{\partial \beta} \\ &= -\frac{\partial E\big(\Gamma(\phi,\delta)\big)}{\partial \beta}\left(\frac{c}{\mu}(1-P_{\phi,k}(\delta))\right) - \frac{\partial P_{\phi,k}(\delta)}{\partial \beta}\left(R - \frac{c}{\mu}(E\big(\Gamma(\phi,\delta)\big)+1)\right).\end{aligned} \tag{A.21}$$

Lemma .4 concludes that $\partial E\big(\Gamma(\phi,\delta)\big)/\partial\beta$ and $\partial P_{\phi,k}(\delta)/\partial\beta$ are both negative. Consequently, $\partial A/\partial\beta$ is positive.

Based on Proposition 3.7,

$$p\left(U_{\text{enter}} \geqslant 0\right) = \begin{cases} 0, & \text{if } \phi > \hat{\phi} \text{ and } \delta < \hat{\delta}(\phi) \\ 1, & \text{if } \begin{cases} \phi \leqslant \hat{\phi} \\ \text{or} \\ \phi > \hat{\phi} \text{ and } \delta \geqslant \hat{\delta}(\phi) \end{cases} \end{cases}. \tag{A.22}$$

We define P_δ as the probability that a customer requests to enter the system δ time units after the most recent QL update. This occurs when no contributor enters the system during δ, and the interarrival between the most recent QL update and the service request of the next customer equals δ. Therefore

$$P_\delta = e^{-\lambda\beta\delta} \times \lambda e^{-\lambda\delta} = \lambda\delta\, e^{-\lambda\delta(\beta+1)}. \tag{A.23}$$

Substituting A and leveraging (A.22) in (A.19) and breaking the sum, social welfare equals

$$SW = \int_{\delta=0}^{\infty}\sum_{\phi=1}^{\phi} A\, P_\delta\, d\delta + \sum_{\phi=\hat{\phi}_{+1}}^{K}\int_{\delta=\hat{\delta}(\phi)}^{\infty} A\, P_\delta\, d\delta. \tag{A.24}$$

Using (A.21) and (A.23), the partial derivative of SW given in (A.24) with respect to

β is as follows

$$\begin{aligned}\frac{\partial SW}{\partial \beta} = &\int_{\delta=0}^{\infty} \sum_{\phi=1}^{\hat{\phi}} \left(-\lambda^2\delta^2\right) e^{-\lambda\delta(1+\beta)} A + \lambda\delta e^{-\lambda\delta(1+\beta)} \frac{\partial A}{\partial \beta}\, d\delta \\ &+ \sum_{\phi=\hat{\phi}+1}^{K} \int_{\delta=\hat{\delta}(\phi)}^{\infty} \left(-\lambda^2\delta^2\right) e^{-\lambda\delta(1+\beta)} A + \lambda\delta e^{-\lambda\delta(1+\beta)} \frac{\partial A}{\partial \beta}\, d\delta\end{aligned} \tag{A.25}$$

Given $\partial A/\partial\beta \geq 0$, the term $\lambda\delta e^{-\lambda\delta(1+\beta)}\frac{\partial A}{\partial\beta}$ is also positive, while the term $\left(-\lambda^2\delta^2\right) e^{-\lambda\delta(1+\beta)} A$ is negative. We introduce the threshold $\hat{\lambda}$ such that when λ is greater than $\hat{\lambda}$, these positive and negative terms cancel each other out, resulting in a local optimum for social welfare. However, when λ is smaller than $\hat{\lambda}$, the positive term dominates the negative one, and $\partial SW/\partial\beta$ becomes positive, indicating that social welfare increases with the contribution ratio. □

A. 11 Proof of Proposition 2.8

The probability of a customer with access to the crowdsensing platform requesting service δ time units after the most recent QL update on the POP, denoted as $P_{\delta\alpha}$, is

$$P_{\delta\alpha} = \alpha\lambda\delta\ e^{-\alpha\lambda\delta(\beta+1)}. \tag{A.26}$$

Applying Proposition 3.7 into (2.15) and using $P_{\delta\alpha}$ from (A.26) results in

$$\begin{aligned}\lambda_{s\alpha} &= (1-\alpha)\lambda + \alpha\lambda\left(\int_{\delta=0}^{\infty}\sum_{\phi=1}^{\hat{\phi}} P_{\delta\alpha}\, d\delta + \int_{\delta=\hat{\delta}(\phi)}^{\infty}\sum_{\phi=\hat{\phi}}^{K} P_{\delta\alpha}\, d\delta\right) \\ &= (1-\alpha)\lambda + \frac{1}{(1+\beta)^2}\left(\hat{\phi} + \sum_{\phi=\hat{\phi}}^{K} e^{-\alpha\lambda\hat{\delta}(\phi)(1+\beta)}\left(1+(1+\beta)\alpha\lambda\hat{\delta}(\phi)\right)\right).\end{aligned} \tag{A.27}$$

Taking the derivative of $\lambda_{s\alpha}$ with respect to α results in

$$\frac{\partial\lambda_{s\alpha}}{\partial\alpha} = -\left(\lambda + \sum_{\phi=\hat{\phi}}^{K} \alpha\lambda^2\hat{\delta}(\phi)^2 e^{-(1+\beta)\alpha\lambda\hat{\delta}(\phi)}\right), \tag{A.28}$$

which is negative. Therefore, $\lambda_{s\alpha}$ is a decreasing function of α. $\square$

A. 12 Proof of Proposition 2.9

The derivative of $SW_{(1-\alpha)}$ with respect to α is

$$\frac{\partial SW_{(1-\alpha)}}{\partial \alpha} = -c(1-P_B)\frac{\partial N}{\partial \alpha} - (R-cN)\frac{\partial P_B}{\partial \alpha}. \tag{A.29}$$

We write the partial derivative of N with respect to α as

$$\frac{\partial N}{\partial \alpha} = \frac{\partial N}{\partial \lambda_{s\alpha}} \times \frac{\partial \lambda_{s\alpha}}{\partial \alpha}. \tag{A.30}$$

As per Proposition 2.8, $\partial\lambda_{s\alpha}/\partial\alpha$ is negative. Given that the expected number of customers in the system rises with $\lambda_{s\alpha}$, we have $\partial N/\partial\lambda_{s\alpha} \geqslant 0$. In summary,

$$\frac{\partial N}{\partial \alpha} \leqslant 0. \tag{A.31}$$

Applying the similar chain rule for P_B, and given that the probability of having K customers in the queue, P_B increases with $\lambda_{s\alpha}$, we conclude

$$\frac{\partial P_B}{\partial \alpha} \leqslant 0. \tag{A.32}$$

Taking into account A.31 and A.32 in (A.29) leads to the conclusion that the social welfare of customers without access to QL information rises with α, $\frac{\partial SW_{(1-\alpha)}}{\partial \alpha} \geqslant 0$.

Substituting λ with $\alpha\lambda$ in (A.24), and taking the derivative of SW_α with respect to α, we have

$$\begin{aligned}\frac{\partial SW_\alpha}{\partial \alpha} = &\int_{\delta=0}^{\infty} \sum_{\phi=1}^{\hat{\phi}} \frac{\partial P_{\delta\alpha}}{\partial \alpha} A + P_{\delta\alpha}\frac{\partial A}{\partial \alpha}\, d\delta \\ &+ \sum_{\phi=\hat{\phi}+1}^{K} \int_{\delta=\hat{\delta}(\phi)}^{\infty} \frac{\partial P_{\delta\alpha}}{\partial \alpha} A + P_{\delta\alpha}\frac{\partial A}{\partial \alpha}\, d\delta,\end{aligned} \tag{A.33}$$

where

$$\frac{\partial P_{\delta\alpha}}{\partial \alpha} = -\delta\lambda e^{-\alpha(1+\beta)\delta\lambda}(-1+\alpha(1+\beta)\delta\lambda) \tag{A.34}$$

is positive. Referring to (A.20), the derivative of A with respect to α is

$$\frac{\partial A}{\partial \alpha} = -\frac{c}{\mu}(1-P_B)\frac{\partial N}{\partial \alpha} - \frac{\partial P_B}{\partial \alpha}\Big(R - \frac{c}{\mu}(N+1)\Big). \tag{A.35}$$

Taking into account A.31 and A.32 in (A.35) results that $\partial A/\partial\alpha$ is positive. Applying this conclusion to (A.33), yields the social welfare of customers with access to QL information increases with α, $\partial SW_\alpha/\partial\alpha \geqslant 0$.

The derivative of the total social welfare given in (2.17) with respect to α is

$$\frac{\partial SW}{\partial \alpha} = \frac{\partial SW_\alpha}{\partial \alpha} + \frac{\partial SW_{(1-\alpha)}}{\partial \alpha}, \tag{A.36}$$

which is also positive. □

A. 13 Changes in U^* with β

Lemma .3. U^* is an increasing function of β.

We can express the partial derivative of U^* with respect to β as

$$\frac{\partial U^*}{\partial \beta} = -\frac{C}{\mu} \times \frac{\partial L}{\partial \beta} - \frac{\partial P_K}{\partial \beta} \times (R - C\frac{K+1}{\mu}). \tag{A.37}$$

Based on Lemma .4 and Proposition 2.1, it is evident that both $\partial L/\partial\beta$ and $\partial P_K/\partial\beta$ are negative. Using (2.2), we find that

$$R - c\frac{K+1}{\mu} \approx -\frac{c}{\mu}, \tag{A.38}$$

which is negative. Therefore, by substituting (A.38) into (A.37), we conclude

$$\frac{\partial U^*}{\partial \beta} = -\frac{c}{\mu} \times (\frac{\partial L}{\partial \beta} - \frac{\partial P_K}{\partial \beta}). \tag{A.39}$$

Using (2.6), we have

$$\frac{\partial L}{\partial \beta} = \sum_{n=0}^{K} n \frac{\partial P_n}{\partial \beta} = K \frac{\partial P_K}{\partial \beta} + \sum_{n=0}^{K-1} n \frac{\partial P_n}{\partial \beta}. \tag{A.40}$$

Therefore,

$$\frac{\partial U^*}{\partial \beta} \approx -\frac{c(K-1)}{\mu} \times \frac{\partial P_K}{\partial \beta}, \tag{A.41}$$

which is positive. □

As shown in (2.7), U^* is a function of the steady-state expected congestion, L, and the steady-state blocking probability, P_K. Proposition 2.1 states that the transition blocking probability decreases with β. According to the definition, a lower blocking probability implies a reduced risk of incurring the access cost without receiving the service. Additionally, given Lemma .4, the expected congestion during a given δ decreases with β, as non-contributors enter the system with a lower rate. A lower expected congestion results in shorter expected waiting times in the queue. Consequently, the steady-state expected utility increases with β due to the less expected waiting time and the reduced risk of wasting the access cost.

Figure A.3 illustrates the expected entrance utility based on β for given δ values of 0.001 and 5. In both panels, the expected utility decreases with the reported congestion

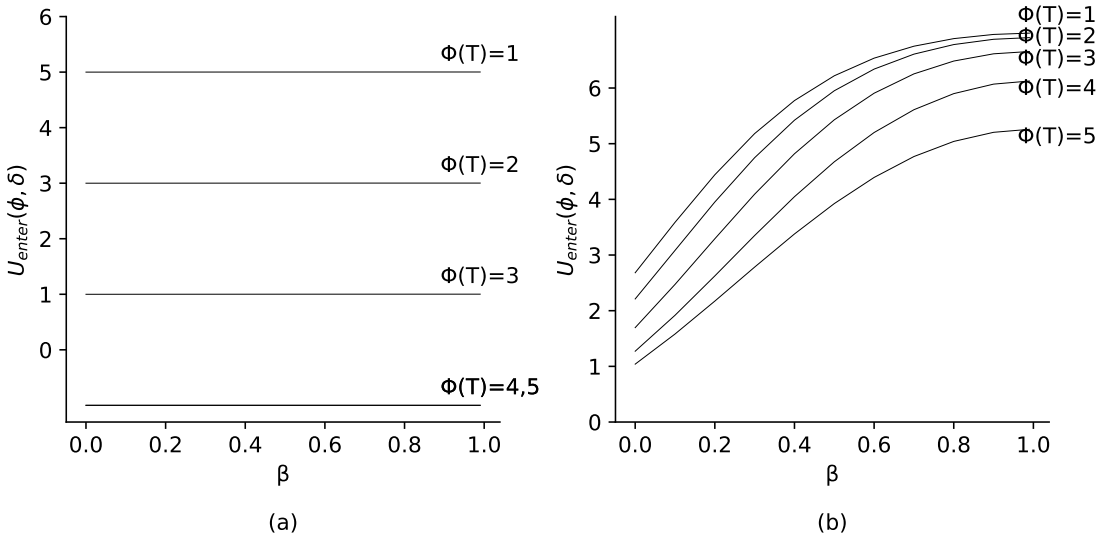


Figure A.3: Entrance Utility, $U_{enter}(\phi, \delta)$, w.r.t. Contribution Ratio, β, Given The Reported QL is $\phi(T) = 1, 2, \ldots, 5$. (Parameters: $\lambda = 1.1$, $\mu = 1$, $R = 10$, $c = 2$, $h = 1$, (a) $\delta = 0.001$, and (b) $\delta = 5$.)

because a higher ϕ results in higher expected congestion, leading to longer expected waiting times. Thus, the largest expected entrance utility occurs when $\phi = 1$. In Panel (a), where $\delta = 0.001$ is negligible, the expected entrance utility does not change with β. Whereas, in Panel (b) with higher δ values of 5, the expected utility clearly increases with β.

A. 14 The Expected Congestion

When a customer requests service long after a QL update, the reported congestion ϕ would no longer represent the system's current state. Therefore, confirmed by Lemma 2.2, the expected congestion is independent of ϕ and equals the steady-state number of customers in an $M/M/1/K$ queue, L. The behaviour of the expected congestion based on δ is illustrated in Figure A.4. In both panels, the service rate μ equals one. Panel (a)

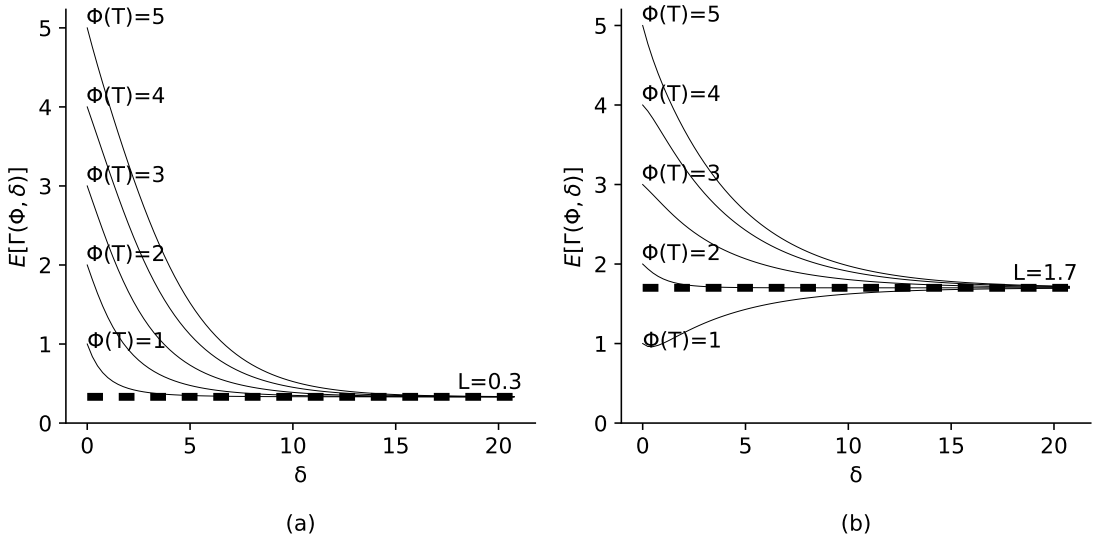


Figure A.4: Expected congestion, $E[\Gamma(\phi, \delta)]$, w.r.t δ, given the reported QL is $\phi(T) = 1, 2, \ldots, 5$. (Parameters: $\mu = 1, R = 10, c = 2, h = 1, \beta = 0.5$, and (a) $\lambda = 0.5$, (b) $\lambda = 1.5$.)

corresponds to an uncongested system with a service request rate $\lambda = 0.5$, while Panel (b) represents a congested system with $\lambda = 1.5$. As Lemma 2.5 confirms, the expected congestion equals ϕ at $\delta = 0$ in both cases. In Panel (a), all values of $\phi = 1, 2, 3, 4, 5$ exceed the threshold $L = 0.3$, resulting in a decreasing expected congestion with δ. The curve eventually converges to L as δ becomes large enough. In Panel (b), when $\phi = 1$,

the expected congestion increases with δ since $L = 1.7$ is greater than one. However, for $\phi = 2, 3, 4, 5$, the expected congestion decreases with δ and eventually converges to L for large values of δ.

LEMMA .4. Given (ϕ, δ), the expected congestion $E[\Gamma(\phi, \delta)]$ is a decreasing function of β.

The partial derivative of $E[\Gamma(\delta, \phi)]$ with respect to β is

$$\frac{\partial E[\Gamma(\delta,\phi)]}{\partial \beta} = \frac{\partial E[\Gamma(\delta,\phi)]}{\partial P_{\phi,n}(\delta)} \times \frac{\partial P_{\phi,n}(\delta)}{\partial \lambda_u} \times \frac{\partial \lambda_u}{\partial \beta}. \tag{A.42}$$

Using (A.7), we have

$$\frac{\partial E[\Gamma(\delta,\phi)]}{\partial P_{\phi,n}(\delta)} = \sum_{n=0}^{K} n = \frac{K(K+1)}{2}, \tag{A.43}$$

which is a positive value. As we discussed earlier, $\partial P_{\phi,n}(\delta)/\partial \lambda_u$ is also positive. When β increases, the ratio of non-contributors decreases, resulting in a decrease in the service request rate of non-contributors. Therefore, $\partial \lambda_u / \partial \beta < 0$. Finally, we conclude that $\partial E[\Gamma(\delta, \phi)]/\partial \beta$ is negative. □

Given the total service request rate, λ, it is observed that the service request rate of non-contributors, λ_u, during a given time interval δ, decreases with the contribution ratio β. Furthermore, within the same time interval δ, some customers receive their service and exit the system. Consequently, the expected congestion in the system δ time units after the most recent QL update ϕ, also decreases with the contribution ratio.

Appendix B: Supplementary Material for Chapter Three

B. 1 Proof of Proposition 3.3

From Equation (3.15) it can easily be observed that $\alpha = 0$ is always an answer. Below we will prove that there is one and only one other solution for $\alpha \in (0, 1]$.

Proof of existence: Equation (3.15) can be rewritten as

$$(1 - \alpha)\Phi(v) = \Phi(u), \tag{B.44}$$

where $v = \sqrt{\alpha\Lambda\beta\delta}$ and $u = \sqrt{\alpha\Lambda\beta\delta}(1 + \alpha\Lambda\tau - \theta\tau)$. Defining

$$f(\alpha) = (1 - \alpha)\Phi(v), \tag{B.45}$$

and

$$g(\alpha) = \Phi(u), \tag{B.46}$$

we have $f(0) = 0.5$ and $f(1) = 0$. On the other hand, for $g(\alpha)$ we have $g(0) = 0.5$, but $g(1) > 0$. To prove the existence of a solution for $\alpha \in (0, 1]$, we need to show

$$\lim_{\alpha \to 0^+} \frac{\partial f(\alpha)}{\partial \alpha} > \lim_{\alpha \to 0^+} \frac{\partial g(\alpha)}{\partial \alpha}, \tag{B.47}$$

and since $g(0) = f(0)$ but $g(1) > f(1)$, it is demonstrated that another solution exists. Simplifying Equation (B.47) we will have

$$\lim_{\alpha\to 0^+} \frac{\beta\delta\Lambda\phi(v)(1-\alpha)}{2\sqrt{\alpha\Lambda\beta\delta}} - \Phi(v) > \lim_{\alpha\to 0^+} \frac{\phi(u)\beta\delta\Lambda(1-\theta\tau+3\alpha\Lambda\tau)}{2\sqrt{\alpha\Lambda\beta\delta}}. \tag{B.48}$$

In the equation above $\phi(\cdot)$ is the standard normal probability distribution function. Denoting $\phi(0) = \frac{1}{\sqrt{2\pi}}$ and $\Phi(0) = \frac{1}{2}$ we have

$$\lim_{\alpha\to 0^+} \frac{\sqrt{\beta\delta\Lambda}\tau\theta}{2\sqrt{2\pi\alpha}} - \frac{1}{2} > 0. \tag{B.49}$$

When evaluating the limit, we observe that the numerator remains positive while the denominator approaches zero. Consequently, the limit of the entire fraction tends toward $+\infty$. This behavior confirms that the inequality (B.49) holds consistently. Thus, we have demonstrated the existence of a positive solution for α.

Proof of uniqueness: For proving the uniqueness of the solution, building upon the previous part we have

$$\frac{\partial^2 f(\alpha)}{\partial\alpha^2} = -2\frac{\partial v}{\partial\alpha}\phi(v) + \frac{\partial^2 v}{\partial\alpha^2}(1-\alpha)\phi(v) + (\frac{\partial v}{\partial\alpha})^2\frac{\partial\phi(v)}{\partial v}(1-\alpha)$$

$$= -\frac{\beta\delta\Lambda}{\sqrt{\alpha\beta\delta\Lambda}}\phi(v) - \frac{\sqrt{\alpha\beta\delta\Lambda}}{4\alpha^2}\phi(v)(1-\alpha) - (\frac{\partial v}{\partial\alpha})^2\sqrt{\frac{\alpha\beta\delta\Lambda}{2\pi}}e^{\frac{-v^2}{2}}(1-\alpha). \tag{B.50}$$

From Equation (B.50) it can be observed $\frac{\partial^2 f(\alpha)}{\partial\alpha^2} \leq 0$ for every $\alpha \in [0,1]$ and $f(\alpha)$ is a decreasing function. On the other hand $g(\alpha) = \Phi(u)$ which is a standard normal CDF and strictly monotonic; therefore, $f(\alpha)$ and $g(\alpha)$ intersect with each other in no more than two α values. Since $\alpha = 0$ is always an answer and we have proven the existence of the other answer in the previous part of the proof, there will only be one unique positive α solution, and Proposition 3.3 is proven. □

B. 2 Proof of Proposition 3.4

Upon taking the derivative of α with respect to β

$$\frac{\partial \alpha}{\partial \beta} = \frac{\partial}{\partial \beta}\big(1 - \frac{1 - \Phi(\sqrt{\alpha\Lambda\beta\delta}(\tau(\theta - \alpha\Lambda) - 1))}{\Phi(\sqrt{\alpha\Lambda\beta\delta})}\big) \tag{B.51}$$

and solving it for $\frac{\partial \alpha}{\partial \beta} = 0$, we have:

$$\frac{\partial \alpha}{\partial \beta} \rightarrow$$
$$- \frac{\delta\Lambda\alpha\left((-1 + \Phi[\sqrt{\beta\delta\Lambda\alpha}(-1 + \theta\tau - \Lambda\alpha\tau)])\phi[\sqrt{\beta\delta\Lambda\alpha}] + \Phi[\sqrt{\beta\delta\Lambda\alpha}](1 - \theta\tau + \Lambda\tau\alpha)\phi[\sqrt{\beta\delta\Lambda\alpha}(-1 + \theta\tau - \Lambda\alpha\tau)]\right)}{2\Phi[\sqrt{\beta\delta\Lambda\alpha}]^2\sqrt{\beta\delta\Lambda\alpha} + \beta\delta\Lambda\left((-1 + \Phi[\sqrt{\beta\delta\Lambda\alpha}(-1 + \theta\tau - \Lambda\alpha\tau)])\phi[\sqrt{\beta\delta\Lambda\alpha}] + \Phi[\sqrt{\beta\delta\Lambda\alpha}](1 - \theta\tau + 3\Lambda\alpha\tau)\phi[\sqrt{\beta\delta\Lambda\alpha}(-1 + \theta\tau - \Lambda\alpha\tau)]\right)} \tag{B.52}$$

In Equation (B.52), $\phi(\cdot)$ represents the probability distribution function of the standard normal distribution, which is always between zero and one. Also, we have

$$\begin{cases} (-1 + \Phi[\sqrt{\beta\delta\Lambda\alpha}(-1 + \theta\tau - \Lambda\alpha\tau)])\phi[\sqrt{\beta\delta\Lambda\alpha}] \leqslant 0 \\ 2\Phi[\sqrt{\beta\delta\Lambda\alpha}]^2\sqrt{\beta\delta\Lambda\alpha} \geqslant 0 \end{cases} \tag{B.53}$$

We simplify the term $(-1 + \theta\tau - \Lambda\tau\alpha)$ as $-1 + \tau/\mu$.

1. When $\mu \geqslant \tau$, equally $-1 + \tau/\mu < 0$, which concludes that $-1 < -1 + \tau/\mu < 0$, as otherwise $\tau/\mu < 0$, which is unacceptable. In this scenario, in the numerator, we have:

$$\Phi[\sqrt{\beta\delta\Lambda\alpha}](1 - \theta\tau + \Lambda\tau\alpha)\phi[\sqrt{\beta\delta\Lambda\alpha}(-1 + \theta\tau - \Lambda\alpha\tau) \geqslant 0. \tag{B.54}$$

   By utilizing (B.53) and (B.54), we can identify a specific $\hat{\beta}$ value where $\partial\alpha/\partial\beta = 0$. If $\beta < \hat{\beta}$, the derivative of α with respect to β will be positive, at $\beta = \hat{\beta}$, the derivative is zero, and for β values greater than $\hat{\beta}$, $\partial\alpha/\partial\beta$ becomes negative.

2. When $\mu < \tau$, equally $-1 + \tau/\mu > 0$, the numerator will be negative. The denominator is positive. (How to show mathematically?). The negative sign in front of the proportion makes $\partial\alpha/\partial\beta$ positive.□

B. 3 Proof of Proposition 3.5

We need to take the derivative of $E(\beta)$ with respect to the β:

$$\frac{\partial E(\beta)}{\partial \beta} = \Lambda \frac{\partial \alpha}{\partial \beta}(\mu - \tau + \alpha(\beta)\Lambda\mu^2). \tag{B.55}$$

According to Proposition 3.4, when $\tau > \mu$, we have $\partial\alpha/\partial\beta > 0$. Furthermore, when $\mu \geqslant \tau$, $\partial\alpha/\partial\beta$ is still positive while $\beta < \beta_I^*$, therefore the sign of $(\mu - \tau + \alpha(\beta)\Lambda\mu^2)$ determines the behavior of the overall travel time. In case $\mu \geqslant \tau$ and $\beta < \beta_I^*$, we have $(\mu - \tau + \alpha(\beta)\Lambda\mu^2) > 0$, meaning the overall travel time increases with β.

For the case of $\tau > \mu$, using Equation (3.3) we can rewrite the parenthesis as $(\frac{1}{(\theta - \alpha\Lambda)} + \frac{\alpha\Lambda}{(\theta-\alpha\Lambda)^2} - \tau)$, then we have

$$(\frac{1}{(\theta - \alpha\Lambda)} + \frac{\alpha\Lambda}{(\theta - \alpha\Lambda)^2} - \tau) \begin{cases} < 0 & \text{if } 0 \leqslant \alpha < \frac{\theta - \sqrt{\frac{\theta}{\tau}}}{\Lambda}, \\ > 0 & \text{if } \frac{\theta - \sqrt{\frac{\theta}{\tau}}}{\Lambda} < \alpha \leqslant 1. \end{cases} \tag{B.56}$$

Note that since $\frac{1}{\theta} < \mu$ is always true and in this case $\tau > \mu$, then $\frac{\theta - \sqrt{\frac{\theta}{\tau}}}{\Lambda} > 0$. Also note that $\alpha = \frac{\theta + \sqrt{\frac{\theta}{\tau}}}{\Lambda}$ will result in $\mu < 0$, which is unacceptable.

In case, $\mu \geqslant \tau$, and $\beta \geqslant \beta_I^*$, we have $\partial\alpha/\partial\beta < 0$, which results in a negative $\frac{\partial E}{\partial \beta}$, indicating that the overall travel time is a decreasing function of β.

From Theorem 3.3 we know for every α there exists a unique β, and from Proposition 3.6 we can see the threshold on α in Equation (B.56) is α^*; therefore, we can rewrite Equation (B.56) based on β^*, the contribution ratio calculated for α^*, as below:

$$\frac{\partial E}{\partial \beta} \begin{cases} < 0 & \text{if } 0 \leqslant \beta < \beta^*, \\ > 0 & \text{if } \beta^* < \beta \leqslant 1. \square \end{cases} \tag{B.57}$$

B. 4 Proof of Proposition 3.6

We seek to obtain a closed-form solution for the mathematical model introduced in (4.1). From the first-order optimality conditions, setting $\frac{\partial E(\beta)}{\partial \beta} = 0$ gives:

$$\Lambda \left[\frac{\partial \alpha}{\partial \beta} M(\lambda) + \alpha \frac{\partial M(\lambda)}{\partial \alpha} \frac{\partial \alpha}{\partial \beta} - \frac{\partial \alpha}{\partial \beta} \tau \right] = 0, \tag{B.58}$$

which will lead to

$$\frac{\partial \alpha}{\partial \beta} [M(\lambda) + \alpha \frac{\partial M(\lambda)}{\partial \alpha} - \tau] = 0. \tag{B.59}$$

Replacing $M(\lambda)$ and $\frac{\partial M(\lambda)}{\partial \alpha}$ with

$$M(\lambda) = \frac{1}{\theta - \lambda} = \frac{1}{\theta - \alpha\Lambda}, \tag{B.60}$$

$$\frac{\partial M}{\partial \alpha} = \frac{\Lambda}{(\theta - \alpha\Lambda)^2}, \tag{B.61}$$

we will have:

$$\frac{\partial \alpha}{\partial \beta} (\frac{1}{(\theta - \alpha\Lambda)} + \frac{\alpha\Lambda}{(\theta - \alpha\Lambda)^2} - \tau) = 0. \tag{B.62}$$

Equation (B.62) holds only if $\partial\alpha/\partial\beta = 0$ or the expression inside the parentheses is equal to 0. Considering the former case, using Equation (3.15) we define $f(\alpha, \beta)$ as

$$f(\alpha, \beta) \; = \; \alpha \; - \; 1 \; + \; \frac{\Phi\big(u(\alpha).v(\alpha, \beta)\big)}{\Phi\big(v(\alpha, \beta)\big)}, \tag{B.63}$$

where $u(\alpha) = 1 - \tau(\theta - \alpha\Lambda)$ and $v(\alpha, \beta) = \sqrt{\alpha\beta\delta\Lambda}$. By the Implicit Function Theorem we have

$$\frac{\partial \alpha}{\partial \beta} \; = \; - \frac{\dfrac{\partial f}{\partial \beta}}{\dfrac{\partial f}{\partial \alpha}}. \tag{B.64}$$

Hence, we must compute $\frac{\partial f}{\partial \alpha}$ treating β as constant, and $\frac{\partial f}{\partial \beta}$ treating α as if it were independent of β. Finally, we take the ratio with a negative sign, and have

$$\frac{\partial \alpha}{\partial \beta} = \frac{-(\frac{1}{2}\sqrt{\frac{\alpha\delta\Lambda}{\beta}})(\phi(v)(\alpha-1)+\phi(u.v)u)}{\Phi(v)+(\frac{1}{2}\sqrt{\frac{\beta\delta\Lambda}{\alpha}})(\phi(v)(\alpha-1)+\phi(u.v)v)+\Lambda\tau\phi(u.v)u} \tag{B.65}$$

where $\phi(\cdot)$ is the probability distribution function of the standard normal distribution.

Since we are interested in $\frac{\partial \alpha}{\partial \beta} = 0$, Equation (B.65) yields

$$\alpha\sqrt{\delta\Lambda}\Big((\alpha-1)\phi(v)+(1-\tau\theta+\alpha\tau\Lambda)\phi(u.v)\Big) = 0. \tag{B.66}$$

One answer for the equation above is

$$\alpha_1 = 0, \tag{B.67}$$

which is not desirable for this problem. Any other solution would come from setting the statement inside the parentheses in Equation (B.66) to zero, and after doing so, we will have

$$\frac{u}{1-\alpha} = \exp(\frac{v^2(u^2-1)}{2}). \tag{B.68}$$

For this equation to be valid and provide an answer for α, $u > 0$ must be true, because otherwise the result of the exponential expression in the right-hand side can never be negative. However, with the given constraints, no answer exists for this expression where all of the parameters remain within their feasible ranges without violating one of the constraints of the original problem or the one pointed out above.

Now moving to the inside of the parenthesis part of Equation (B.62) being equal to zero, we get two other solutions for α as below

$$\alpha_{3,4} = \frac{\theta \pm \sqrt{\frac{\theta}{\tau}}}{\Lambda}. \tag{B.69}$$

Based on the constraints introduced for the mathematical model in (4.1), and boundaries of θ, Λ, and α, only one solution is acceptable, which is:

$$\alpha^* = \frac{\theta - \sqrt{\frac{\theta}{\tau}}}{\Lambda}.\square \tag{B.70}$$

B. 5 Proof of Proposition 3.7

Setting $\alpha = 1$ in Equation (3.15) yields:

$$\Phi\Big(\sqrt{\Lambda\beta\delta}\big(\tau(\theta - \Lambda) - 1\big)\Big) = 1 \tag{B.71}$$

Defining ν such that $\Phi(\nu) = 1$, and conditioning on $\tau > \mu$, we find:

$$\delta_N^* = \frac{1}{\beta\Lambda}(\frac{\nu}{\tau(\theta - \Lambda) - 1})^2.\square \tag{B.72}$$