12-2015

A Robust Data Exchange Framework for Connected Vehicle Technology Supported Dynamic Transit Operations

Yucheng An
Clemson University, yuchena@g.clemson.edu

Recommended Citation

An, Yucheng, "A Robust Data Exchange Framework for Connected Vehicle Technology Supported Dynamic Transit Operations" (2015). *All Theses*. 2250.
https://tigerprints.clemson.edu/all_theses/2250

A ROBUST DATA EXCHANGE FRAMEWORK FOR CONNECTED VEHICLE TECHNOLOGY SUPPORTED DYNAMIC TRANSIT OPERATIONS

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Civil Engineering

---

by
Yucheng An
December 2015

Accepted by:
Dr. Mashrur Chowdhury, Committee Chair
Dr. Yongxi Huang
Dr. Eric Morris

ABSTRACT

Transit systems are an integral part of surface transportation systems. A connected vehicle technology (CVT) supported transit system will assist the users to manage trips both dynamically and efficiently. The primary focus of this research is to develop and evaluate the performance of a secure, scalable, and resilient data exchange framework. In the developed data exchange framework, a new data analytics layer, named Transit Cloud, is used to receive data from different sources, and send it to different users for a Dynamic Transit Operations (DTO) application. The DTO application allows the transit users to request trip information and obtain itineraries, using their personal information devices, (e.g., cell phone), and provides dynamic routing and scheduling information to the transit operators. A case study was conducted to investigate the effectiveness of the developed data exchange framework by comparing the framework with the USDOT recommended data delivery delay requirements. This data exchange framework was simulated in the CloudLab, a distributed cloud infrastructure, in which, the data exchange delay for DTO was examined for different simulation scenarios, utilizing the synthetic data generated from Connected Vehicle Reference Implementation Architecture (CVRIA) and Research Data Exchange (RDE). Security, scalability, and resiliency of the developed data exchange framework are illustrated in this thesis. The results from the simulation network reveal that the data exchange delay satisfies the USDOT data delivery delay requirements. This suggests that the developed secure, scalable, and resilient data exchange framework, which is presented in this study, meets the application performance requirements. Thus, Transit Cloud is a more preferable alternative than the existing framework because of its added benefits in terms of security, scalability, and resiliency.

DEDICATION

I would like to dedicate this thesis to my parents in recognition of their constant and deepest love, inspiration, and support.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation and gratitude to my advisor Dr. Mashrur Chowdhury for his continuous and patient guidance in my Master’s research. In every aspect of this research, Dr. Chowdhury provided a lot of valuable advises and assistance. It is my luck to meet such a nice advisor in my master’s study.

I am also grateful to Dr. Yongxi Huang and Dr. Eric Morris for serving as my thesis committee member. Thank them for squeezing time from their busy schedule to review my thesis and provide insightful comments and suggestions.

I would like to specially thank my friend Mizanur Rahman and Dr. Kakan Dey for their help and inspiration when I was conducting the research. Discussion with them when I was confused gave me both knowledge and mental support. Their positive attitude towards work impresses me a lot.

I would also like to express my gratitude to Katerina Moreland of Parking and Transportation Services for her consultancy and all the staffs in Civil Engineering Department of Clemson University for their hearty help during last two years.

Last but not the least, I would like to thank my parents for their love, support, and encouragement.

TABLE OF CONTENTS

CHAPTER

Table of Contents (Continued) Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Background

Increasing traffic demand and associated traffic congestion cause substantial increase in travel time and contribute to increased fuel consumption. In 2011, due to congestion, 2.9 billion gallons of fuel were wasted, and commuters were stuck for 5.5 billion hours at congested traffic in 498 metropolitan areas in the US, resulting in a total congestion cost of $121 billion (Lomax et al., 2012). In 2014, congestion cost had risen up to $160 billion. Forecasted by Taxes Transportation Institute, this cost will grow to up to $192 billion by 2020 (Lomax et al., 2015).

Over the years, different congestion management strategies were implemented to meet the increasing traffic demand of the US metropolitan areas, and transit service is considered to be one of the most cost-effective strategies to reduce the congestion (Harford, 2006). Transit service is an indispensable part of surface transportation for the passenger's movement, and has been operating in most cities around the world. A study in 85 major urban areas of the U.S. estimated that congestion delay would have increased by 27 percent, and would have cost an additional $18.2 billion to the residents each year in the major urban areas if public transit services were not available (Schrank & Lomax, 2005). Benefits of the transit service include the reduction in fuel consumption, emissions, and improvement of surface transportation mobility efficiency (US Joint Program Office, 2015).

There are several major cities around the world, which have deployed real-time transit information systems using the existing communication technologies, such as Wi-

Fi, Cellular, and satellite. However, the emerging Connected Vehicle (CV) technology could provide a better solution to the traffic problems and contribute to many other improvements, such as reduction in energy consumption and improvement in air quality (FTA, 2002, Ma et al., 2009, Ma et al., 2012, He et al., 2012, and He et al., 2012). Even in smaller cities, real-time transit information services are deployed, e.g., in Clemson, the Tiger Transit, which is the Clemson University's campus shuttle service, uses a mobile application called Transloc to provide real-time bus-tracking service to potential riders. To realize the real-time transit services, an efficient and reliable data exchange between the physical objects, such as vehicles, travelers, infrastructures, is significant (Transloc, 2015). Connected Vehicle environment is a future surface transportation network, which enables wireless communication between the vehicles, infrastructures, and pedestrians, envisioned to improve the transportation safety, mobility, and environmental performance (ITS Joint Program Office, 2015). In the Connected Vehicle Reference Implementation Architecture (CVRIA), an application called "Dynamic Transit Operations (DTO)" is presented, which supports real-time communication between the transit users and the transit service providers. In the CV environment, massive amounts of data will be generated due to data exchange between vehicles, infrastructures, and transit users. Thus, it will be a challenge to enable reliable data exchange between the different physical objects in real-time (US Joint Program Office, 2015). Due to the lack of real-time transit information, trip uncertainty has always been a major problem for planning trips by the transit users, which leads to a long wait time and a decrease in ridership (Mishra et al., 2012). Thus, it is necessary to develop a reliable data exchange framework for DTO in the CV environment, which can provide real-time transit information to the transit users through their mobile devices.

1.2 Overview of DTO

The DTO application provides real-time information to both the transit users and transit operators. It provides the transit users with real-time transit information, and the ability to request trip information via their personal information devices (PID). Meanwhile, transit operators are able to acquire dynamic routing and scheduling information in real-time with this application (ITS Joint Program Office, 2015). Dynamic routing and scheduling information provided to the vehicle operators in real-time will reduce the travel time and trip cost significantly (Taniguchi & Shimamoto, 2004).

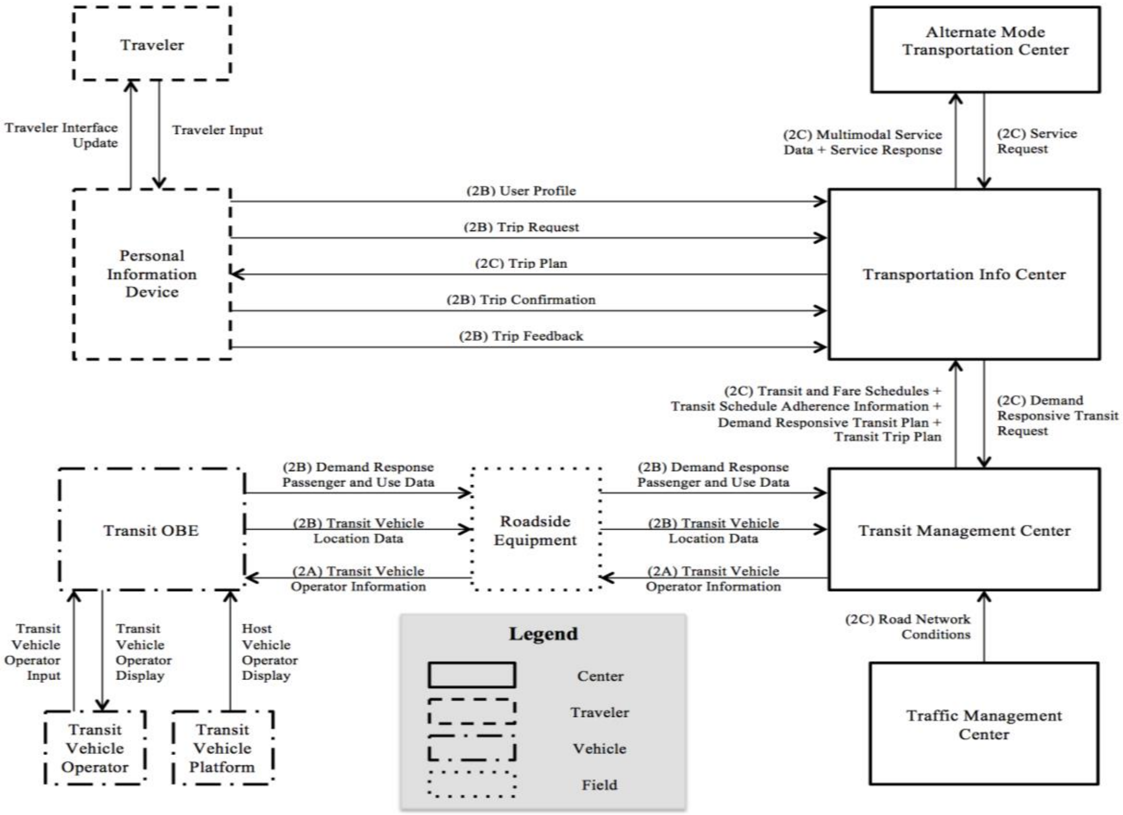

FIGURE 1-1 DTO Application Physical Architecture (Adapted from CVRIA DTO Physical Architecture, ITS Joint Program Office, 2015)

Figure 1-1 illustrates the physical architecture for DTO application, which is adapted from CVRIA. Physical architecture shows the inter-connection and information flows between the different physical objects, which are Center, Traveler, Vehicle, and Field in DTO. The arrow of each information flow indicates flow direction from the data source to the data user. Each information flow has two characteristics: spatial context and time context. As shown in Table 1-1, based on the time context, information flows are grouped into four categories: 1 (Now), 2 (Recent), 3 (Historical), and 4 (Static). On the other hand, based on the spatial context, information flows are categorized into five groups: A (Adjacent), B (Local), C (Regional), D (National), and E (Continental). For example, an information flow with 2B means its time context is recent, and spatial context is local.

Table 1-1 Information Flow Characteristics Defined in CVRIA

Characteristics	Category	Characteristic Value
Time context	1 (Now)	Less than 1 second
	2 (Recent)	1 second -30 minutes
	3 (Historical)	30 minutes – 1 month
	4 (Static)	Greater than 1 month
Spatial context	A (Adjacent)	0-300 meters
	B (Local)	300meters -3 kilometers
	C (Regional)	3 kilometers-30 kilometers
	D (National)	States of U.S.
	E (Continental)	Continental U.S.

In the DTO application, the Transportation Information Center and Transit Management Center receive real-time traffic data from the travelers’ Personal Information Devices (PID), transit vehicles’ On-board Equipment (OBE), and Roadside

Equipment (RSE). Meanwhile, the Traffic Management Center and Alternate Mode Transportation Center provide other information, such as road conditions and weather conditions, to the Transportation Information Center and Transit Management Center (Mishra et al., 2012). After the data is collected from the data sources mentioned above, the Transportation Information Center and Transit Management Center would analyze the transit vehicle schedule, as well as the location status, and provide real-time transit information to the transit users, transit vehicles, and the Alternate Mode Transportation Center.

1.3 Motivation for a Developed Data Exchange Framework

A real-time robust data exchange framework is critical for the proper aggregation, correlation, processing, and distribution of data, which depends on several factors, such as size of data, sending and receiving rate of data, frequency of data collection, and type of collected data. The time requirements of the different information flows are important for modeling the data exchange framework for the DTO application. To provide a real-time service, dynamic data must be reliably exchanged between the different physical objects within a short time. In the CV environment, the massive amount of data generated by vehicles, travelers, and infrastructures make it difficult to redistribute the data reliably while satisfying the application performance requirements. The challenges in designing a robust data exchange framework to support DTO include the following: 1) data exchange security because of the high risk of compromising the privacy of travelers (i.e., security); 2) flexibility of the framework that will have the ability to scale-up (i.e., scalability); and 3) data redundancy plan to recover from any failure (i.e., resiliency of the system).

To support the real-time services, delay, which is caused by data collection, transmission, processing, and analyzing, should satisfy the application requirement. In this research, the developed data exchange framework has an additional layer, named Transit Cloud, which is used to clean raw data, label data into a usable format, and provide information to different data users (e.g., travelers, transportation information center). A Transit Cloud also acts as a data processing and routing medium, so that, each entity of the DTO application can send data to, and receive data from the Transit Cloud.

1.4 Research Objectives

The objectives of this research are: 1) to design a secure, scalable, and resilient data exchange framework, and increase the reliability of DTO application services; and 2) to evaluate the performance of the data exchange framework in terms of the data delivery delay.

A case study for the Clemson Area Transit (CAT) network located in Clemson, South Carolina was conducted to investigate the effectiveness of the developed data exchange framework by comparing with the USDOT recommended data delivery delay requirements. This data exchange framework was simulated in the CloudLab, which is a distributed cloud infrastructure. The data exchange delay for DTO was examined in different simulation scenarios, utilizing synthetic data generated with CVRIA and Research Data Exchange (RDE).

1.5 Research Contributions

The primary contribution of this research is in the development and evaluation of a secure, resilient, and scalable data exchange framework for the DTO application. In this framework, an additional layer, Transit Cloud, is utilized. This framework has the following advantages over the traditional data infrastructure:

- Data users can acquire data without knowing its sources, so that the data privacy of these sources can be protected.
- Each Transit Cloud will replicate the entire data once, thus, the data lost due to any failure will be reduced significantly.
- Different transit applications could be supported simultaneously by increasing the number of Transit Clouds.

1.6 Organization of Thesis

Chapter 2 presents a review of the development history and the benefits of CV, previous research on DTO, and studies related to data exchange reliability. Chapter 3 presents limitations of traditional data exchange framework, the strategy used in this research to develop a robust data exchange framework, and the method employed for the performance evaluation of the data exchange framework. In Chapter 4, an evaluation of the data exchange framework is presented, the results are analyzed, and the potential implementation cost is discussed. Finally, Chapter 5 presents the conclusions, and recommendations.

CHAPTER TWO

LITERATURE REVIEW

2.1 Overview

This chapter presents a review of the literature related to the objectives of this thesis. The literature review includes benefits of CV, DTO application development, and reliability of data exchange framework in terms of security, resiliency, and scalability.

2.2 Connected Vehicle Development

Connected Vehicle is an emerging technology that enables the real-time traffic information sharing between vehicles and vehicles, and between vehicles and infrastructure. The major difference between the CV environment and the traditional transportation environment is that vehicles connected in the CV system can communicate with each other and with transportation infrastructures wirelessly. The wireless communication enables CVs to acquire and disseminate traffic information in real time, which can improve the traffic condition assessment and prediction significantly (Ma et al., 2009, and Ma et al., 2012). Different CV applications including transit application will generate various types and vast amount of data. The application areas of connected vehicle include mobility, safety, environment, and support (ITS Joint Program Office, 2015). These CV applications will improve the traffic mobility and safety, and help relieve the negative impacts of transportation on the environment (Pina, 2015) and energy (He et al., 2012). The support applications are to provide reliable communication service for diverse CV applications.

2.2.1 Mobility

In 2010, the number of vehicle in the world surpassed 1 billion (Sousanis, 2011). According to Statista, in the US, the vehicle number were more than 255 million in 2013, and the new light vehicles registered in 2015 is 1.3 million (Statista, 2015). The rapid increasing number of vehicles and vehicle miles travelled (VMTs) make traffic congestion more severe and decrease trip reliability. As stated by 2015 Urban Mobility Scorecard, 42 hours per commuter and 3 billion gallons of fuel were wasted due to traffic congestion, equivalent to $160 billion in societal cost (Schrank et al., 2015). Numerous Intelligent Transportation Systems (ITS) technologies have been deployed to relieve the traffic congestion. ITS have shown to improve real-time traffic management in response to dynamic traffic conditions (Chwodhury et al., 2006, and Bhavsar et al., 2007).

Sponsored by ITS Joint Program Office, a metropolitan ITS infrastructure deployment tracking system was developed to provide an assessment of level of deployment of the Intelligent Transportation Infrastructure (ITI). Chang conducted a study using various data mining and archiving technologies to compare the incident response time before and after the ITS infrastructures were deployed, and illustrated that ITS system can be used to assess the traffic congestion more effective (Chang, 2004, and Fries et al., 2007).

In 2015, Minelli et al. conducted a research aiming to evaluate the impact of connected vehicles on mode choice and mobility. In the research, dynamic route guidance system is assumed to be equipped on each connected vehicle to select routes automatically based on the real-time information communicated between connected vehicles. Travel time is used to measure the effectiveness of connectivity supported routing in a simulation scenario with connected vehicles and was compared with a base scenario of routing without connected vehicles. This study revealed that connected vehicles would reduce travel time at lower penetration levels, while as the percentage of connected vehicles increases, the travel time increases as well, especially significantly when the percentage is increased from 60% to 100% (Minelli et al., 2015).

2.2.2 Safety

How to improve transportation safety is always a major challenge for transportation agencies. According to the National Highway Traffic Safety Administration (NHTSA), in the U.S., there were 5.6 million vehicle-related crashes in 2013, resulting in 32,719 deaths (NHTSA, 2015). An estimation made by NHTSA shows that 41 to 55 percent of intersection crashes could be reduces with connected vehicle safety applications, and two connected vehicle safety applications, which are to help drivers to negotiate at intersections and turn left at intersections, would reduce 592,000 crashes and 270,000 injuries (NHSTA, 2014).

In 2010, Kattan et al. evaluated the impact of vehicle-to-vehicle communication on random crash scenarios. In the research, APIs are developed to create random crashes based on the collision information, weather information, and the wireless communication between connected vehicles. The results show that under congested condition, to improve traffic safety will increase travel time (Kattan et al., 2010).

In 2015, Genders et al. conducted a research to evaluate the potential safety benefits of CV technology in a work zone. Vehicles with connectivity were able to receive the work zone information via the wireless communication between vehicles to vehicles, so that drivers' awareness increased with information received and reduced vehicle speed. Also, connected vehicles followed the dynamic route guidance to make a detour to avoid the work zone. Time to collision was used to evaluate the safety condition of the network, longer time to collision means better safety condition. The result showed that when the percentage of connected vehicles is less than 40%, CV technology improved the traffic safety, while when the percentage of connected vehicles was over 40%, CV technology decreased traffic safety (Genders and Razavi, 2015). The limitation of this research was that only travel time and work zone information were the input for dynamic route guidance, so the guidance just considered if there is a work zone and which route option was shortest for one connected vehicle, but the impacts of other connected vehicles are not taken into consideration.

2.2.3 Environment

The emission from motorized vehicles is the major contributor to the air pollution in urban areas (Kristensson et al., 2004). Around one third of greenhouse gases and majority of other pollutions, including carbon monoxide, hydrocarbons, oxides of nitrogen, etc., are produced by transportation systems (Jin et al., 2012). Traditionally, signal re-timing/optimization is an effective way to reduce fuel consumption and vehicular emissions on arterials (Stevanovic et al., 2009). Many studies focussed on the fuel efficiency and emission reduction with connected vehicle applications. In 2012, Jin et al. investigated the impact of Advanced Intersection Management System in a connected vehicle environment on vehicle emissions and fuel consumption. This study revealed that the advanced traffic management system with connected vehicles reduces unnecessary stops at intersections as well as vehicle fuel consumption and emissions (Jin et al., 2012).

He et al. showed energy consumption reduction for Plug-in Hybrid Electric Vehicles (PHEVs) with connected vehicle technology (He et al., 2012). In 2015, HomChaudhuri et al. conducted a study on a predictive control strategy to minimize stopping at red lights and reduce fuel consumption for a group of connected vehicles. The signal phase and timing information was collected and provided to individual vehicles using connectivity between vehicles and infrastructure. This study revealed that connected vehicle technology contributed to fuel consumption reduction at signalized intersections (HomChaudhuri et al., 2015).

2.2.4 Connected Transit

Various studies investigated CV technology enabled transit applications to improve transit operations. In 2014, Hao et al. conducted a research on schedule-based coordinated optimization model to improve the transit schedule adherence and transit signal priority. In this research, the wireless communication devices installed on the buses and on roadsides allowed buses to request the signal priority and receive speed guidance. This study revealed that, with connected vehicle technology, the travel delay of the buses between two dedicated bus stops decreased and schedule adherence was improved (Hao et al., 2014). In 2015, another study investigated adjustment of signal timing to accommodate the buses with connectivity with signal controllers. Connected buses are able to request traffic signal priority from intersection controllers via Dedicated Short Range Communications (DSRC), so that buses would stop less times at the signalized intersections and the delay will be reduced. The results show that the average bus delay reduces 19% during peak hour and 49% during off-peak hour (Hsu and Shih, 2015).

2.3 Data Exchange Framework for the DTO Application

Previous studies related to data exchange frameworks for DTO applications in the CV environment are reviewed in the following subsections: DTO application deployment (Section 2.3.1) and the existing data exchange framework for DTO application (Section 2.3.2).

2.3.1 DOT Application Development

The USDOT developed Integrated Dynamic Transit Operations (IDTO) application bundle as a high priority CV mobility application for the CV pilot deployment project (FHWA, 2013). IDTO includes three transit mobility applications: i) Connection Protection (CP), ii) Dynamic Transit Operations (DTO), and iii) Dynamic Ridesharing (DRS). In this research, the author focuses on evaluating a robust data exchange framework to support the DTO application while satisfying USDOT application requirements. The DTO application is an advanced version of the demand responsive transit service that fulfills travelers' requests related to transit service using their destination location and departure time through their personal devices. The DTO application facilitates dynamic scheduling, dispatching and routing services for efficient transit operations (Boenau, and Timcho, 2014). Travelers need real-time information to plan their trip and transit agencies will support these through demand responsive services. The USDOT proposed the concept of operation and system requirements for DTO applications in mid-2012 (Mishra et al., 2012). In October 2012, the USDOT published a report about the test readiness of this application (Schweiger et al., 2012). Prototype development and impact assessment of this application were completed in April 2013, and prototype requirements and architecture were developed in September 2013 (Timcho et al., 2013). IDTO application suites are deployed in two areas: Columbus, Ohio and Central Florida (Boenau et al., 2014). The Central Ohio Transit Authority (COTA) provides a fixed route/fixed schedule on selected routes. The Ohio State University (OSU) Campus Area Bus System (CABS), which provides an on campus central transportation system, operates a shuttle service at the Defense Supply Constriction Center (DSCC) that is connected with COTA routes and supports CP applications. On the other hand, the CP application of an IDTO bundle has also been implemented on the University of Central Florida (UCF) campus along the LYNX routes (i.e., public transit service for Orange, Osceola and Seminole counties in Central Florida). DTO and DRS applications are not implemented in these areas.

2.3.2 Existing Data Exchange Framework

According to the USDOT defined application requirements, data exchange delay must comply with the requirements for applications. The delay for sending data from a vehicle to RSE was measured in the USDOT's vehicle-infrastructure integration (VII) proof-of-concept (POC) test bed in Michigan and the delay range was from 0.5 second to 1.5 seconds. The communication delay depends on the type of communication technology (i.e., wireless, wired) and network congestion (Hamilton, 2009). Large amount of data will lead to network congestion and data packet loss. It will increase the data exchange delay significantly. Dion et al. analyzed the data exchange performance with Intellidrive probe vehicle data (Dion et al., 2011). This study evaluated the interaction between vehicles and RSEs and measured the delay in two data exchange scenarios: 1) exchange data one after another and 2) exchange all data simultaneously.

Average delay was measured to be 65 seconds and 30 seconds, respectively, for these two scenarios.

However, security, scalability and resiliency of a data exchange framework were not considered to develop a data delivery system in the previous literature. In this research, a robust data exchange framework in terms of security, scalability and resiliency is developed. Data exchange delay is also measured to comply with the USDOT system requirements for implementation.

2.4 Data Exchange Reliability

Data reliability means the data is complete and error free to satisfy application requirements (Morgan and Waring, 2004). In connected vehicle applications, as mentioned before, the security, scalability, and resiliency of the data exchange framework must be guaranteed between connected vehicles, infrastructures, and pedestrians. The traditional data exchange framework has several deficiencies that include: 1) higher data exchange delay because of the higher data processing time as they are not separated based on the CVRIA application requirements; 2) the risk of accidental or malicious unauthorized access to the centers' computing systems that could compromise the security of the data processing system at each center; and 3) failure of data processing machines at any center that may shut down the CV application services (i.e., the data exchange framework is not resilient, as there is no back-up infrastructure) (ITS Joint Program Office, 2015).

Originally, the data exchange defined by USDOT for DTO is direct, e.g. the data will be delivered from traveler's personal information devices and received by the Transportation Information Center, and after the data is processed in the Transportation

Center, processed data will be delivered back to personal information devices (ITS Joint Program Office, 2015). In this process, personal information devices and Transportation Information Center will have access to each other, which makes the data exchange unsecure. To improve the security during data exchange, Khadra et al. developed an induced-message cryptosystem to avoid the data transmission across public channels. The encrypted information improved communication security, but it will increase the cost significantly (Khadra et al., 2003). Therefore, a new way to make the data exchange more secure as well as cost effective needs to be developed.

In CVRIA, it is shown that CV data will be aggregated by RSEs, and then delivered to centers via fiber optic cables or other communication options. The limitation is that before the data reaches centers, it may not be replicated. If some of the RSE fails or the network does not work, the data around that area could be lost. Another problem is if other centers need to acquire the data for other applications, or the data from a wider area need to be collected for DTO application, a new wired or wireless communication should be established to connect the RSE and the centers requiring the data. This could create a challenge for scalability of a CV system (ITS Joint Program Office, 2015).

CHAPTER THREE

RESEARCH METHOD

3.1 Overview

In this chapter, a new data exchange framework for DTO application is discussed. At first, an analysis of the limitations of the traditional data exchange framework is presented. Then, a new data exchange framework is developed. To evaluate the performance of the new framework, an evaluation experiment is designed afterwards.

3.2 Limitation of the Traditional Data Exchange Framework

From the traditional perspective of data exchange framework, data will be exchanged between different sources, which are different physical objects that generate data, and the data users that are different physical objects receiving data, directly (as shown in Figure 3-1 and 3-2). In the DTO data exchange framework, data will be exchanged in two phases: 1) from the field (i.e., PID and Transit OBE), and the data-providing centers (i.e., Alternate Mode Transportation Center and Traffic Management Center) to the data-processing centers (i.e., Transit Management Center and Transportation Information Center) (as shown in Figure 3-1); and 2) from the data-processing centers to the field and data-providing centers (as shown in Figure 3-2). Since the data can be transferred in the data exchange framework in two ways, the field and data-providing centers serve as data sources in the first phase, and as data users in the second phase. Similarly, data-processing centers also reverse their roles in two different phases.

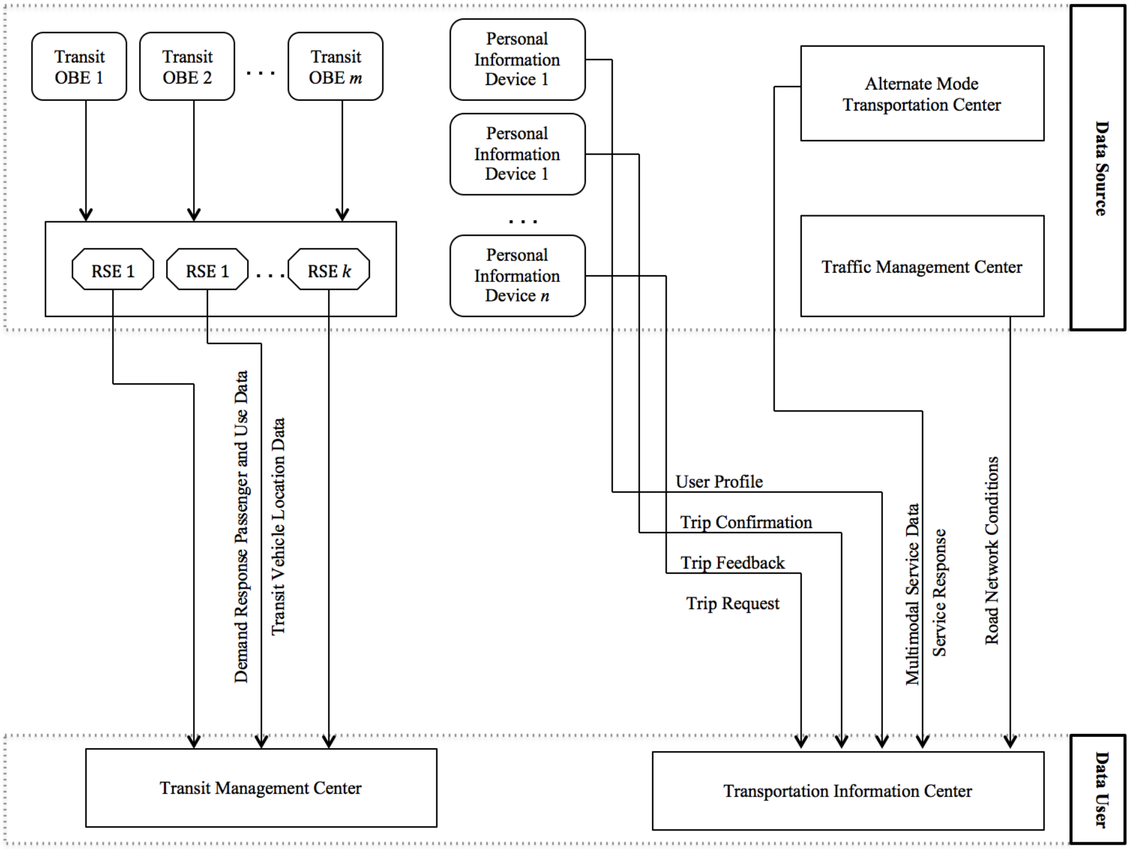

Figure 3-1 First Phase of the Traditional Data Exchange Framework for DTO Application

Figure 3-1 illustrates the first phase of the traditional data exchange framework. In this phase, PID, Transit OBE, Alternate Mode Transportation Center, and Traffic Management Center are data sources, and the Transit Management Center and Transportation Information Center are data users. Data sources send raw data to the data users, and then, the raw data will be processed in the data-processing centers, to be a usable data set in a tabular format, that could be requested/used by different CV applications. Normally, the data users request and acquire data from this table, according to their requirements.

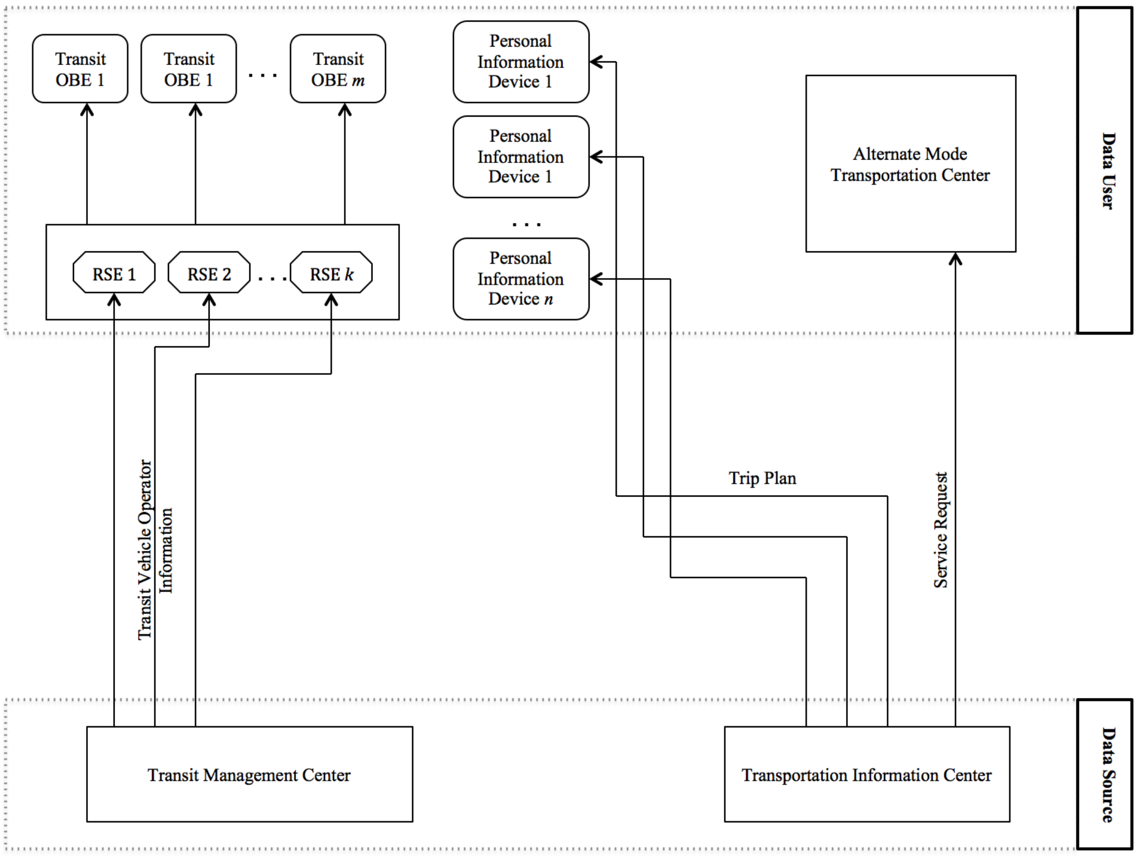

Figure 3-2 Second Phase of the Traditional Data Exchange Framework for DTO Application

Figure 3-2 shows the second phase of traditional DTO data exchange framework, in which, the data sources in the first phase become the data users, and the data users in the first phase serve as the data sources. Transit Management Center provides transit vehicle routing and scheduling information to the transit vehicles via RSE, and Transportation Information Center provides processed data to the transit users and Alternate Mode Transportation Center. Traffic Management Center does not acquire data from the Transportation Information Center in this phase.

The traditional data exchange framework has several deficiencies, which are as follows: 1) the risk of accidental or malicious unauthorized access to the centers' computing systems significantly reduces the security of the data processing system at each center; 2) failure of data-processing machines at any center to shut down the CV application services, due to the lack of back-up infrastructure; and 3) low scalability that makes it more difficult to serve different CV applications with the same data, or to extend the range of the serving area. Thus, it is necessary to design a data exchange framework, which is secure, scalable, and more resilient.

3.3 Developed Robust Data Exchange Framework

The Developed robust data exchange framework contains an additional layer, named Transit Cloud, which supports data collection from different data sources, labels data based on CVRIA information flows, and keeps it available to all the data users. The two phases of this data exchange framework are defined as:

1) From field and data-providing centers to Transit Cloud, and from Transit Cloud to data-processing centers, upward;
2) From data-processing centers to Transit Cloud, and from Transit Cloud to the field and data-providing centers, downward.

Figure 3-3 illustrates the first phase of the developed data exchange framework with Transit Cloud. The Transit Cloud, which consists of data processing resources, can convert raw data into a usable format, and contains a certain number of data, according to the DTO information flow requirements. As shown in Figure 3-3, data sources, including Transit OBE, PID, Alternate Mode Transportation Center, and the Traffic Management Center, will send raw data to the Transit Cloud instead of the data users, which are Transportation Information Center and Transit Management Center. Raw data will be preliminarily converted into a usable format, and then labeled into information flows. In addition, different data users may require the same data at the same time; for example, both the Transportation Information Center and Transit Management Center require dynamic transit information. Thus, data users can receive requested information flows from the Transit Cloud, instead of data sources.

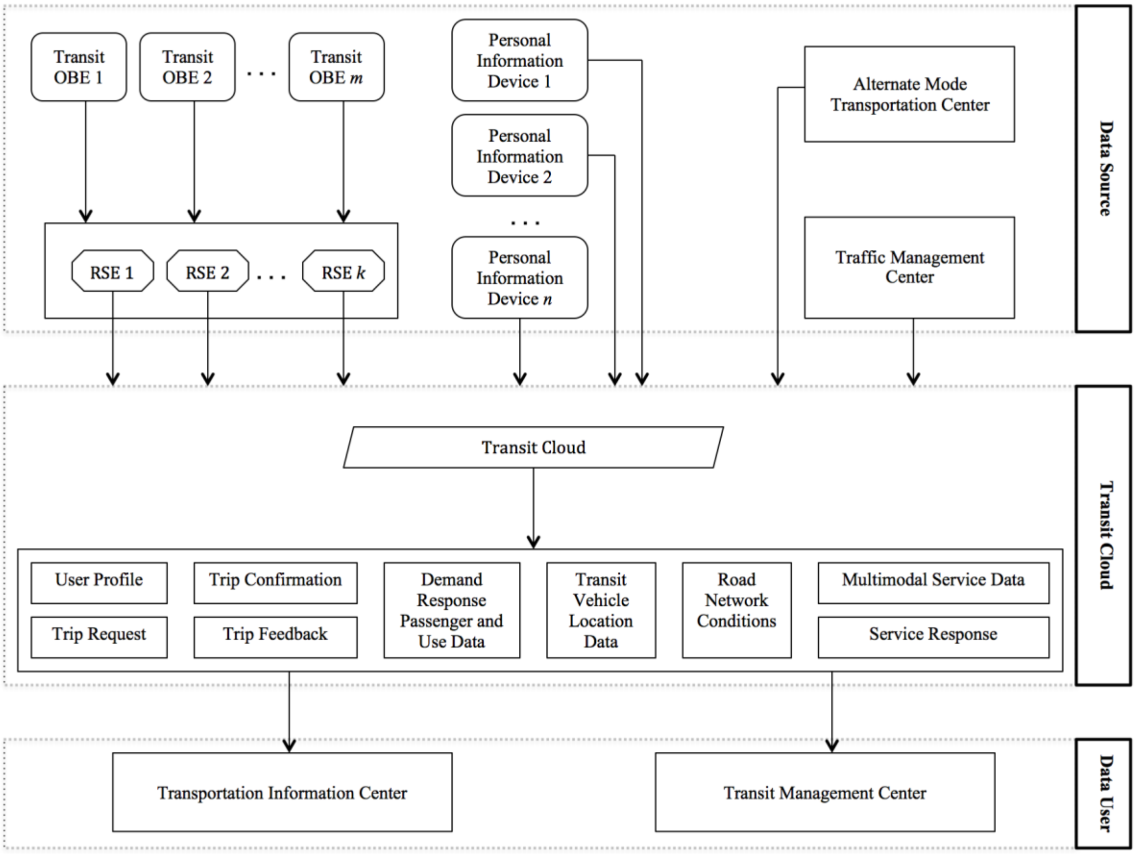

Figure 3-3 First Phase of the Developed Data Exchange Framework for DTO Application

Figure 3-4 shows the second phase of the developed DTO data exchange framework. In this phase, Transportation Information Center and Transit Management Center, which are data users in the first phase, serve as data sources, and provide the processed data to the Transit Cloud. Transit vehicles, transit users, and Alternate Mode Transportation Center will acquire the data from the Transit Cloud afterwards. So, Transit

Cloud will play the same role as it did in the first phase. The security, resiliency and scalability of this data exchange framework are described below.

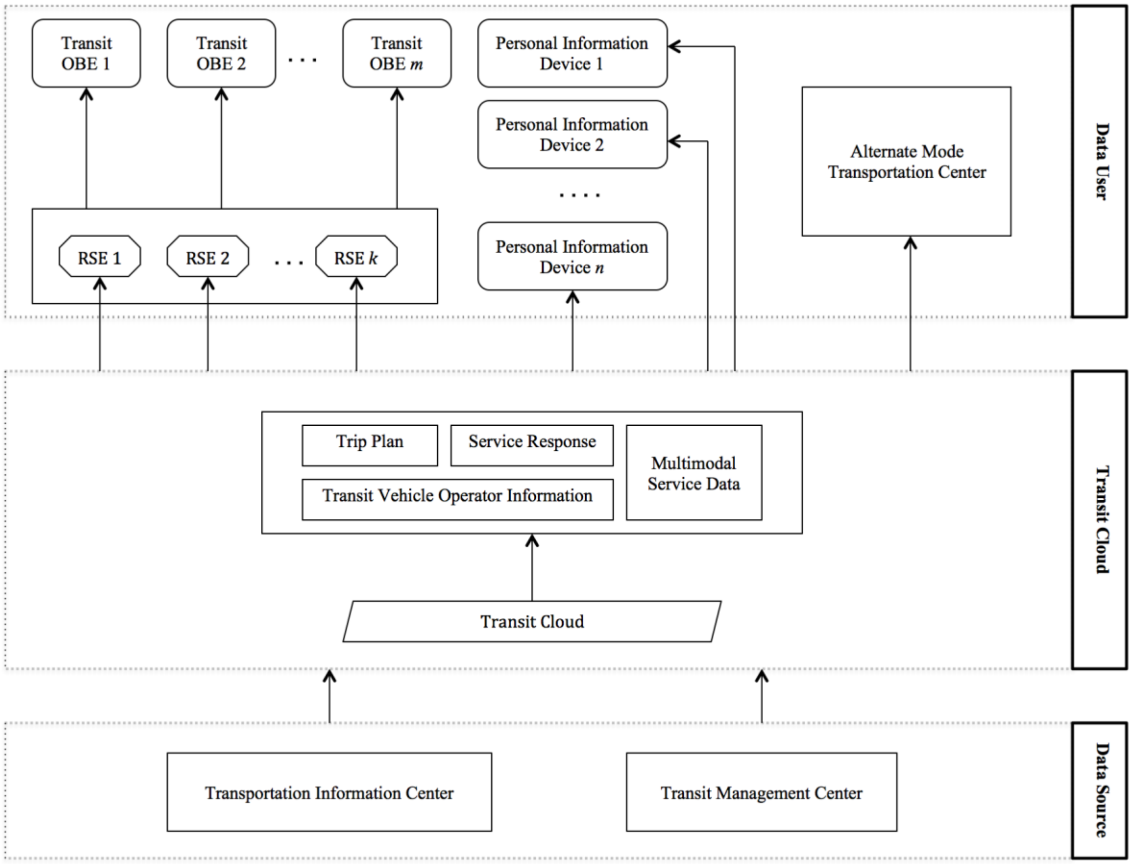

Figure 3-4 Second Phase of the Developed Data Exchange Framework for DTO Application

3.3.1 Security

In this data exchange framework, the data is provided by a variety of data sources. With the Transit Cloud, the data exchange framework makes it possible for the users to acquire data without knowing its source. This improves the privacy and security of the data sources. In the DTO application, the Transportation Information Center and Transit Management Center can acquire dynamic data from the transit users and transit vehicles, without having the contact information of each traveler or transit vehicle (Vilela et al., 2008), so, they may focus more on analyzing the curated data, which makes the data exchange system more robust, in terms of the security perspective (Conzon et al., 2012; Jansen et al., 2011). On the other hand, in the second phase, the transit users can acquire real-time transit information, without having the access to the Transit Management Center and Transportation Information Center, which makes it more secure for the information in these two centers. This enhanced security comes from the fact that data-processing centers and transit users no longer need to provide data access to each other, since the data users can receive what they require directly from the Transit Cloud, and the risk of accidental or malicious unauthorized access to the computing systems is, thus, significantly reduced.

3.3.2 Resiliency

As transportation centers are involved in raw data collection and cleaning in order to put the collected data in a usable format in a traditional data exchange framework, the failure of data processing machines at any center will lead to the failure of CV applications (Ford et al., 2012). The distributed nature of this data exchange framework, using a middle layer (Kreps & Rao, 2011), which is Transit Cloud, will handle machine failure, by duplicating the data into different Transit Clouds. Furthermore, Transit Cloud is an idea inspired by Kafka, where the driving platform of the framework supports automated and graceful transition from the failed components into new components. In Kafka, a middle layer name broker will clean and label the raw data from the data producers, and each broker will replicate the data once, and function similarly to the Transit Cloud. Thus, the impact of the failure recovery process will be minimized.

3.3.3 Scalability

The data exchange framework with the Transit Cloud will support data delivery at a high level of abstraction, and, at a larger scope, include the delivery of data from one source to multiple destinations, across a large geographic area (Manasseh & Sengupta, 2008). The Transit Cloud will act as a routing medium, which may help facilitate data exchange across different data sources and destinations in the connected transportation systems (Kühn et al., 2009). Moreover, other centers of different CV applications may also request this dynamic data from the Transit Cloud, if they also require the same data, in which, the data redundancy is highly reduced, making this framework scalable (Marsh et al., 2008). This entire complex data routing process happens automatically and dynamically, which is not possible, or is quite costly in a single centralized server system. With the popularity of the commercial cloud-computing infrastructure, such as Amazon Web Service (which is a secure cloud services platform providing data storage, computing, and delivery services), it is possible to dynamically scale or reduce the Transit Cloud layer, to support the data exchange demand, according to the actual traffic demand.

3.4 Evaluation of the Developed Data Exchange Framework

The basic function of the new data exchange framework is to deliver data in real time, in order to support the DTO dynamic services. To guarantee that the developed data exchange framework works well in the CV environment, a simulation experiment is conducted, to evaluate the performance. Figure 3-5 illustrates the evaluation procedure steps.

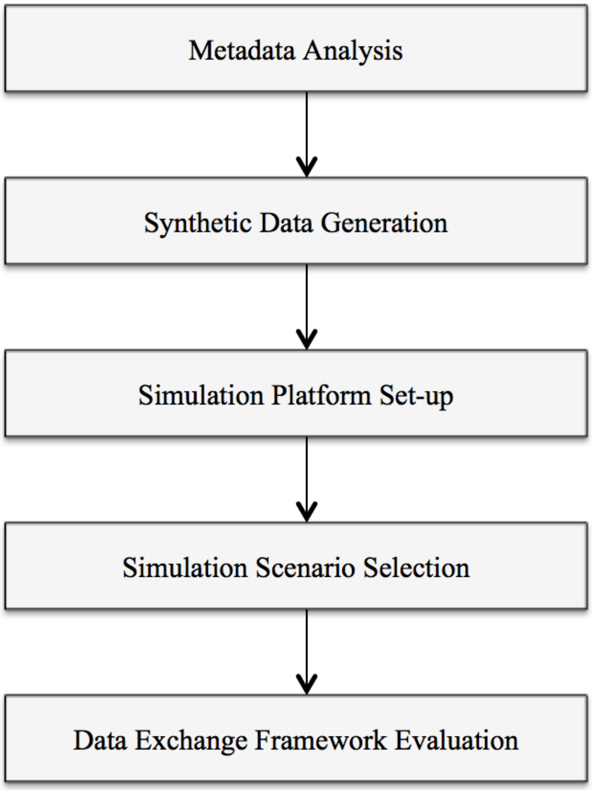

Figure 3-5 Evaluation Procedure of the Developed Data Exchange Framework

3.4.1 Metadata Analysis

According to the physical architecture of DTO, as shown in Figure 1-1, the information flows are delivered between different physical objects, and the information flows are not the basic unit of data delivered in this data exchange framework, but are rather packages of metadata. Defined by CVRIA, information flows can be broken down to a group of primitive elements, which consist of the data delivered in the DTO framework. In different information flows, there will be several same primitive elements, which means, different information flows may contain partially the same data. In the developed data exchange framework, since each Transit Cloud will duplicate all the data once, the repeated data will be cleaned, and unique primitive elements will be left and tagged.

There are three different types of physical objects: 1) Center, 2) Vehicle, and 3) Traveler, in the physical architecture, as proposed in CVRIA for a DTO application (OST-R, 2015). Based on the physical architecture of CVRIA, RSE, which is a field physical object, is utilized in the transit network, to collect data from the transit vehicle's OBE. In a data exchange framework, travelers use their PID to request dynamic transit information from the transportation information centers. Meanwhile, using Dedicated Short-Range Communications (DSRC), the transportation information centers can send travelers the requested information, based on their demand, and also the transit vehicles, which can transmit their location to the nearby RSE, and then, the RSE will send transit vehicle information to the transit management center. Other centers in this physical architecture include the Traffic Management Center and Alternate Mode Transportation Center, which provide road network conditions and service requests to the Transit Management Center and Transportation Information Center, respectively. Considering that Transit Management Center and Transportation Information Center can only get data from certain data sources but not the entire data sources, there will be a data exchange between these two centers after they get data from the field. With all the information, the Transit Management Center and Transportation Information Center process and analyze the data, and provide the travelers with the requested information, such as next available bus arrival time, send back a service request to the Alternate Mode Transportation Center, and demand responsive transit request to the Transit Management Center.

Between the different physical objects, the data is aggregated as information flows, and then sent from one physical object to another. This could be bi-directional (e.g., PID can send an user profile to the Transportation Information Center, and then the

Transportation Information Center will send a trip plan back to the PID). In DTO, there are a total of 17 information flows, which are categorized on the basis of time and spatial context.

In this application, as shown in Table 3-1, based on the different types of data sources and data users, the information flows are classified into five groups, which are center to center, center to vehicle, vehicle to center, center to traveler, and traveler to center, among which, the time and spatial information flows are identified as 2A (recent and adjacent), 2B (recent and local), and 2C (recent and regional), which are summarized in Table 3-1.

Table 3-1 DOT Information Flow Classifications

Center to Center	Road Network Contidions (2C)	Service Request (2C)	Multimodal Service Data (2C)	Service Response (2C)	Demand Responsive Transit Request (2C)	Transit and Fare Schedule (2C)	Transit Schedule Adherence Information (2C)	Demand Responsive Transit Plan (2C)	Transit Trip Plan (2C)
Center to Vehicle	Transit Vehicle Operator Information (2A)	–	–	–	–	–	–	–	–
Vehicle to Center	Demand Response Passenger and Use Data (2B)	Transit Vehicle Location Data (2B)	–	–	–	–	–	–	–
Center to Traveler	Trip Plan 2C	–	–	–	–	–	–	–	–
Traveler to Center	User Profile (2B)	Trip Request (2B)	Trip Confirmation (2B)	Trip Feedback (2B)	–	–	–	–	–

3.4.2 Synthetic Data Generation

As CV technology has not been implemented in the real world at a large scale as yet, it is impossible to get real traffic data. Thus, the experiment was done through a simulated evaluation network. Synthetic data was required to be generated reasonably, to serve as the input of the simulation network.

Due to the limited standards of available CV data format, the format of the existing traffic data, with similar functions, will be used as the CV data format. In CVRIA, there is a description of each primitive element, based on which, CV data functions in DTO can be identify.

Since in the metadata analysis, CV data types have already been decided, and, the format of each data is identified, with the data collection frequency, which is calculated with the case study illustrated in Chapter 4, the synthetic data can be generated with some matrix generation tools, such as MatLab.

3.4.3 Simulation Platform Set-up

In this study, a simulation platform, named CloudLab, which is a distributed cloud infrastructure, was used to evaluate the performance of the simulation scenarios for DTO applications (The University of Utah, 2015). This platform is comprised of 515 machines, which have been used by a consortium of three universities. In the data exchange framework, each machine can work as a data user/ source (e.g., PID, Transit Cloud, or data-processing center). In different scenarios, each machine can work as one component, e.g., one data source, one Transit Cloud, or one data user, and the number of machines will be varied, to model each scenario. In CloudLab, synthetic data was assigned to the data sources or Transit Clouds, and delivered to the machines that work as

Transit Clouds or data users, respectively. For the data source machines, a heterogeneous combination of machines was used, ranging from 16 GB to 64 GB DDR4 RAM, 99 GB to 900 GB hard drives, and 1 Gbps to 10 Gbps Ethernet connections between data sources and data users. The message Transit Cloud machine(s) had 256 GB DDR4 RAM memory, 2 TB 7,200 RPM SATA HDD hard disk, and a 10Gbps Ethernet connection for making data transfer between the nodes. Figure 3-6 is a data delivery interface of CloudLab. In this interface, 6 machines, including 2 data sources, 2 Transit Clouds, and 2 data users, are used to set up the simulation network. Data will be delivered from two data sources to Transit Clouds, and then delivered from Transit Clouds to data users.

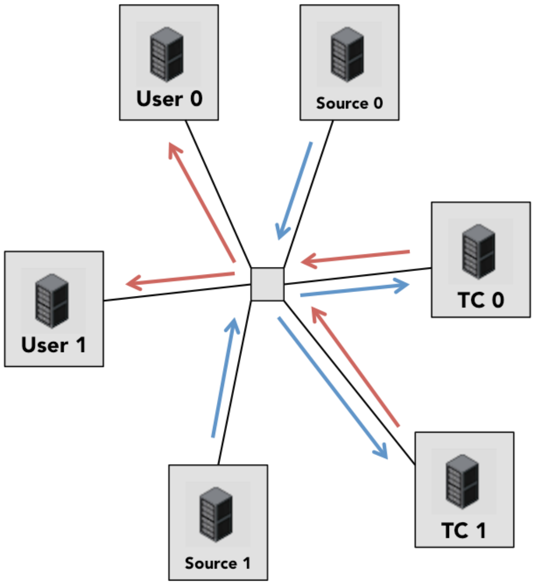

Figure 3-6 CloudLab Data Delivery Interface

In the simulated evaluation network, data was delivered from the machines, which worked as data sources, to the machines that worked as Transit Cloud, and then delivered from the machines that worked as Transit Cloud to the machines, which worked as data users. The time when data was at data sources, at Transit Cloud, and at the data users were recorded, and the time difference when data was at different physical objects was used to be the delay caused by data exchange.

3.4.4 Simulation Scenarios Selection

The security of the developed data exchange framework was discussed in the former section of this chapter. To test the scalability and resiliency, the scenario with different numbers of Transit Clouds were developed. Based on the application requirement of DOT, which is to provide real-time information, the delay caused by data exchange cannot be more than the threshold of delay for CV dynamic mobility applications recommended by USDOT, where the data exchange delay between different components, including data sources, Transit Clouds, and data users, were determined.

3.4.5 Data Exchange Framework Evaluation

To conduct the data exchange framework evaluation, throughput and delay are tested. These two parameters are measured in different scenarios, and the results are compared with the threshold required by USDOT. The simulation evaluation is conducted, to achieve the following objectives:

1) Evaluate throughput and delay from the data sources to Transit Clouds for data duplication, in order to support machine failures;
2) Evaluate the throughput and delay between the data sources and Transit Clouds, and find the worse condition, with the different number of data sources;
3) Evaluate the throughput and delay between the Transit Cloud and data user; and
4) Evaluate the throughput and delay between the data sources, Transit Cloud, and data user.

3.5 Summary

In this chapter, the limitation of the traditional data exchange framework is discussed, and a new data exchange framework for DTO application, with an additional layer, named Transit Cloud, is presented. The new developed data exchange framework would be more secure, resilient, and scalable, than the traditional data exchange framework, by meeting the application requirements. To evaluate the performance of the developed data exchange framework, an evaluation method is designed, which are discussed with a case study in the next chapter.

CHAPTER FOUR

SIMULATION ANALYSIS AND EVALUATION RESULTS

4.1 Overview

This chapter presents an evaluation of the developed data exchange framework, following the evaluation method presented in Chapter 3, based on a case study of CAT bus network. The evaluation is conducted with a simulation platform, named CloudLab, and synthetic data is generated using metadata description of CVRIA and real-world data from RDE (Research Data Exchange) in MATLAB. Evaluation analysis and results of the developed data exchange framework using CloudLab simulation platform in four different scenarios are discussed in this chapter.

4.2 Description of Case Study Area - CAT Bus Network

A case study was conducted to evaluate the performance of the developed data exchange framework following the evaluation experiment steps shown in Figure 3-5. In this case study, a secure, scalable, and resilient data exchange framework was developed for the CAT. The CAT bus network was assumed to be equipped with CV equipment, and it was able to collect real-time data from the transit users and the transit vehicles. Data exchange throughput, data recording rate, average/maximum end-to-acknowledgement delay, and end-to-end delay were, then, measured from the field and data-providing centers to the data-processing centers, to evaluate the performance of the data exchange framework.

CAT is the transit system, which serves the City of Clemson and nearby cities, and provides fare-free transit services. In this network, there are a total of 6 routes (Pendleton Route, Red Route, Seneca Business, Seneca Express, Seneca Residential, and

Clemson University (CU) Campus Routes). Out of these 6 routes, the Seneca Business, Seneca Express, and Seneca Residential routes are partially independent, because they are dedicated to serving Seneca City. This case study includes three routes serving the Clemson area only, which are the Red route, Pendleton Route, and CU Campus Routes. Figure 4-1 illustrates the three CAT routes considered in this research. The length of the Red Route is 14.2 miles with 19 signals along the route, the Pendleton Route is 12.6 miles with 8 signals along the route, and the CU Campus routes are a total of 2.7 miles with 3 signals along the route. All transit vehicles will send data to the RSE. Considering the three routes overlap at one signal, there are total 28 RSEs, which are required at signals. The total number of transit vehicles serving in these three routes is 27, and the average hourly ridership is 241 passengers (KFH, 2014). This information was collected from the CAT bus management center, and used to generate synthetic data, which is discussed in the following subsection.

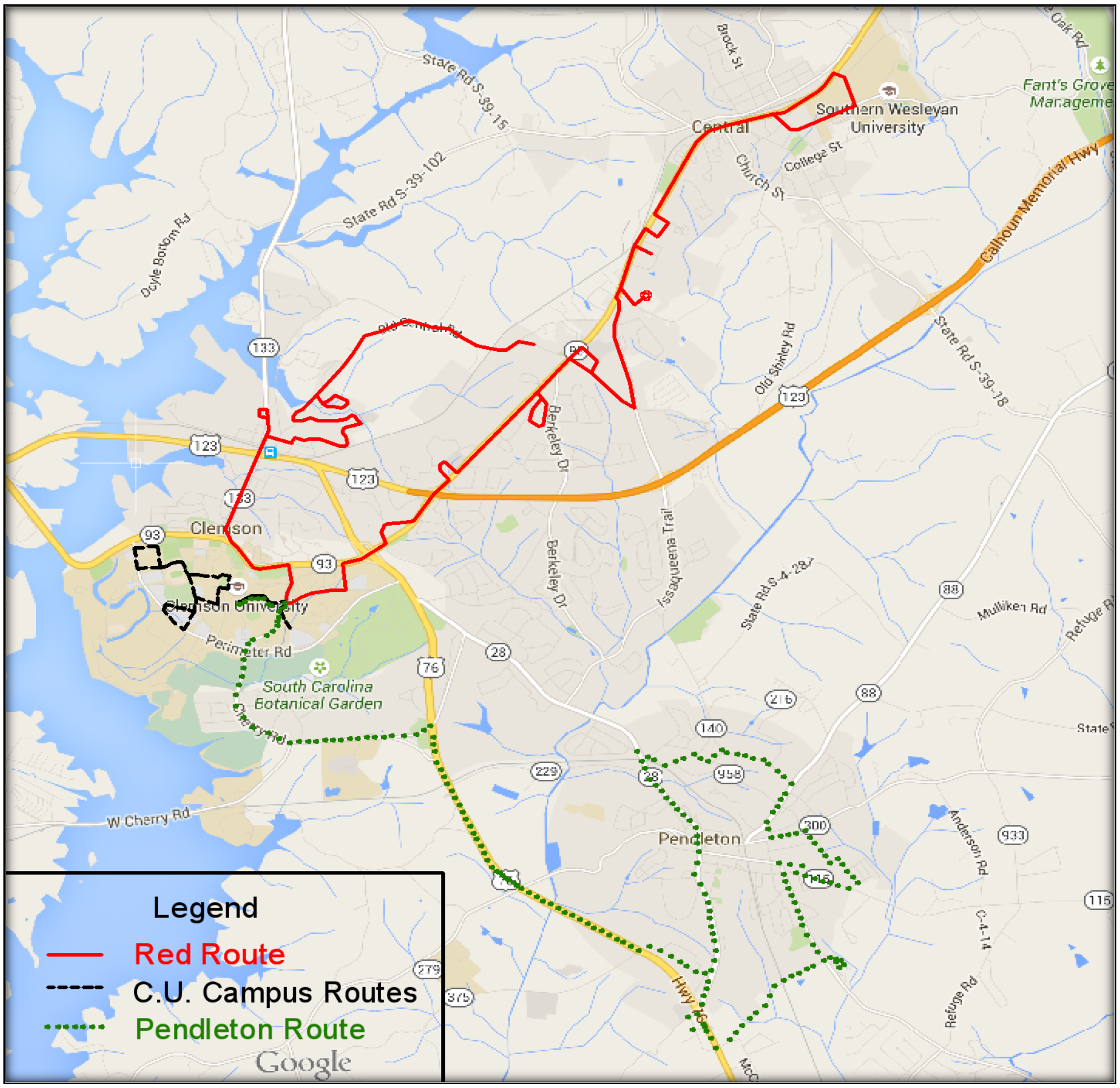

FIGURE 4-1 CAT Bus Routes Map

4.3 Evaluation of DTO Data Exchange Framework

4.3.1 Metadata Analysis

A metadata analysis of the first phase of the developed data exchange framework is conducted to identify the data type. Considering the function similarity of Transit Cloud in two phases, the performance testing of one phase can represent the evaluation of the performance of the entire data exchange framework. In the first phase, there are a total of 9 information flows. From PID, there are 4 information flows, including the user profile, trip request, trip confirmation, and trip feedback. From RSE, there are 2 information flows, which are demand response passenger and use data, and transit vehicle location data. From the data-providing centers, there are 3 information flows, consisting of road network conditions, multimodal service data, and service response. As defined by CVRIA, information flows are composed of a group of data flows, and subsequently, data flows can be broken down into different sub data flows. Sub data flows will continue to be broken down until the primitive elements are obtained. Table 4-1 is an example of the breaking down process of an information flow, trip confirmation, to primitive elements.

Table 4-1 An Example of Information Flow Breaking Down Process

Information Flow	**Data Flow**	**Sub Data Flow 1**	**Sub Data Flow 2**	**Sub Data Flow 3**	**Data (Primitive Element)**
Trip Confirmation	traveler_route_accepted	route_identity			route_identity
	traveler_personal_trip_confirmation	paratransit_service_confirmation	paratransit_service_identity		paratransit_service_identity
			transit_confirmation_flag		transit_confirmation_flag
			traveler_identity		traveler_identity
		traveler_identity	traveler_identity		traveler_identity
		traveler_rideshare_confirmation	credit_identity		credit_identity
			reservation_status	confirmation_flag	confirmation_flag
			rideshare_selection_number		rideshare_selection_number
			traveler_identity		traveler_identity
		traveler_parking_confirmation	traveler_identity		traveler_identity
	traveler_personal_payment_information	credit_identity	credit_identity		credit_identity
		parking_space_details	date		date
			duration		duration
			time		time
		ride_segments	list_size		list_size
			transit_route_segment_number		transit_route_segment_number
		stored_credit	stored_credit		stored_credit
		toll_route_segments	list_size		list_size
			toll_segment_identity	unit_number	unit_number
		traveler_identity			traveler_identity

Each Transit Cloud will replicate all the data once the data is uploaded to it, and the data users will acquire it from the Transit Cloud instead of from the data sources directly. Redundant primitive element, from the same data source will be merged to be one unique data. The unique data will not be delivered multiple times in the data exchange framework, so that, the redundancy will be reduced. The first phase of the developed robust data exchange framework is used to evaluate the performance of the entire framework. Table 4-2 provides the number of unique primitive elements in each information flow. Since, there are some overlapping primitive elements in the different information flows, these overlapping primitive elements are combined to be the same unique data. Finally, there are 105 unique primitive elements, including 51 unique data from PID, 5 unique data from RSE, and 49 unique data from the data-providing centers.

Table 4-2 Number of Unique Primitive Element in each Information Flow

Data Source	**Information Flow**	**Number of Unique Primitive Element**	**Total number of unique data**
PID	User Profile	17	105
	Trip Request	25	
	Trip Confirmation	14	
	Trip Feedback	2	
RSE	Demand Response Passenger and Use Data	2	5
	Transit Vehicle Location Data	5	
Data-providing Center	Road Network Conditions	36	49
	Multimodal Service Data	14	
	Service Response	1	

4.3.2 Synthetic Data Generation

4.3.2.1 Data Format

In this evaluation experiment, the formats of the real data are adopted, using similar data flows from RDE. RDE data is maintained by the Federal Highway Administration, as a transportation data which would share a platform to provide a variety of data sets, which are collected from the field demonstrations, to support the development, testing, and demonstration of multi-modal transportation CV mobility applications (FHWA, 2015). Using the data from RDE, real data format are estimated. For example, the format of the data 'time' from RDE is '14:55:00', and so, the format of the unique data 'time' exchanged in the framework will also be like '14:55:00'.

4.3.2.2 Data Generation

A basic assumption was made, which is, each character or number in one data unit equals to 1 byte, and is estimated as the size of each unique data (e.g., the format of 'time' is 14:55:00, which is 8 digits, so, the unit size of 'time' is 8 bytes). With this assumption, the unit size of each data was estimated. An example of data size estimation for an information flow with synthetic data generation is shown in Table 4-3. The format and size of each piece of data was used to generate synthetic data, using MATLAB later.

Table 4-3 An Example of Data Size Estimation of an Information Flow

Information Flow	Metadata	Format Sample	Data Size (byte)
Trip Confirmation	unit_number	1	1
	traveler_identity	987263516	9
	transit_route_segment_number	1	1
	transit_confirmation_flag	0	1
	time	14:55:00	8
	stored_credit	$9,999.99	9
	route_identity	AMTK_NB	7
	rideshare_selection_number	CL_BH_NB	8
	paratransit_service_identity	WE	2
	list_size	103	3
	duration	749	3
	date	20110705	8
	credit_identity	365	3
	confirmation_flag	0	1

In this case study, the data exchange, throughput, data recording rate, average/maximum end-to-acknowledgement delay, and end-to-end delay from RSE to the Transportation Information Center and Transit Management Center was estimated. The microscopic traffic data collection frequency was assumed to be one per second. Data was collected for four hours of CAT bus operation, to produce sufficient amount of data, and reach the capacity of data transmission bandwidth, which is required to test the data exchange framework. For this case study, synthetic data were generated for all the 27 transit vehicles. Since the number of transit vehicles for the three routes was 27, and hourly ridership was 241 passengers per hour, as described before, the number of data provided by the PID was 3,470,400 (=3,600*4*241), by Transit On-board Equipment (TOBE) was 388,800 (=3,600*4*27), by the Traffic Management Center and Alternate Mode Transportation Center was 14,400 (=3,600*4). With all these estimates, the volume of generated data from PID was 1.38 GB (Gigabyte), the data from TOBE was 16.3 MB (Megabyte), and the data from the Traffic Management Center and Alternate Mode Transportation Center was 4.5 MB.

4.3.3 Simulation Platform Set-up

A distributed cloud infrastructure, named CloudLab, is selected as the simulation platform in this evaluation. Since each machine can work as one component of the data exchange framework, considering the number of components, the maximum number of machines used in the CloudLab is 274, which are used in the fourth simulation scenario, including 28 roadside equipment, 241 transit users, 4 centers, and 1 Transit Cloud. Data is delivered between the different machines working in different roles in the data exchange framework.

4.3.4 Simulation Scenarios Selection

As shown in Table 4-4, four simulation scenarios were designed, to evaluate the data exchange framework. Scenario 1, containing two test rounds, is to evaluate the reliability of the data exchange framework in case of machine failure. In round 1, 271 data sources, including 28 RSEs, 241 PIDs, the Alternate Mode Transportation Center, and the Traffic Management Center, which provided a large volume of dynamic data to 1 Transit Cloud. In the Transit Cloud, the data would be processed into different labeled information flows. Then in round 2, the same data would be transferred to 3 Transit Clouds. In each Transit Cloud, the same information flows labeled would be replicated once, which means, three of the same information flows labels will be created. The reliability of round 2 is higher than round 1 because the replicated labeled information flows are available to the data users, in case of one or two of the Transit Clouds fail.

Table 4-4 Data Exchange Framework Performance Evaluation Scenarios

Scenario	Simulation Category	Objective
1	**Test round 1:** 271 data sources-1 Transit Cloud-1 replication (271 ds-1 tc-1 r) **Test round 2:** 271 data sources-3 Transit Clouds-3 replications (271 ds-3 tc-3 r)	Evaluate the performance from data sources to Transit Cloud for data duplication to support machine failures
2	**Test round 1:** 30 data sources (28 RSE + 2 Center)-1 Transit Cloud (30 ds-1 tc) **Test round 2:** 241 data sources (PID)-1 Transit Cloud (241 ds-1 tc)	Evaluate the performance between data sources and Transit Cloud, and find the worse condition with different number of data sources
3	1 Transit Cloud-2 data users (1 tc-2 du)	Evaluate the performance between Transit Cloud and data user
4	271 data sources-1 Transit Cloud-2 data users (271 ds-1 tc-2 du)	Evaluate the performance between data sources, Transit Cloud and data user

In scenario 2, the performance from the data sources to Transit Cloud, with different number of data sources are tested separately in two rounds. In round 1, 30 data sources, including 28 RSEs and 2 centers, provide data to 1 Transit Cloud. In round 2, the 241 PIDs transfer data to the Transit Cloud. With different number of data sources, the performance of the data exchange framework may be different, and the performance in this scenario for the DTO application will be, therefore, evaluated.

In Scenario 3, objective is to evaluate the performance between the Transit Cloud and the data users. In this scenario, after the processing of data, two data users, the Transportation Information Center and the Transit Management Center, acquire the labeled information flows from the Transit Cloud.

In the last scenario, the delay from the data sources to the data users through Transit Cloud is evaluated. 271 data sources deliver the data messages to 1 Transit Cloud, where the data would be cleaned and labeled. Then, the data users would acquire the topics from the Transit Cloud.

4.3.5 Data Exchange Framework Evaluation

4.3.5.1 Evaluation Parameter

In this research, throughput, data recording rate, average/maximum end-to-acknowledgement delay, and end-to-end delay of the developed data exchange framework were tested. Throughput and data recording rate indicate the data sending capability of the data sources, for transferring data to Transit Cloud. The unit of throughput is megabyte per second (Mb/s), and the unit of data recording rate is records per second (records/sec). During the data exchange, multiple data would be delivered from the data sources to Transit Cloud, and after the data is made available at Transit Cloud, an acknowledgement would be sent back to the data sources, to confirm that the Transit Cloud receives the data. End-to-acknowledgement delay is the delay from the sending out of the data from the data sources, to the time when the, data sources receive the acknowledgement from the Transit Cloud. End-to-acknowledgement delay consists of queuing time at the data sources, data transmission time, and waiting time till the acknowledgement is received by the data sources, while end-to-end delay only consists of transmission delay.

4.3.5.2 Results and Analysis

The evaluation results for the four scenarios are summarized in Table 4-5, 4-6, 4-7, and 4-8. In Scenario 1 and 2, two test rounds, with different set-up were conducted. Scenario 1 demonstrates the performance of the developed data exchange framework with different number of Transit Cloud. Scenario 2 illustrates the performance of the developed data exchange framework when the number of data sources is different. Scenario 3 shows the throughput and delay when the data users acquire the data from the Transit Cloud. Scenario 4 is the result for the entire first phase of the data exchange procedure, which is from field and data-providing centers to the data-processing centers via Transit Cloud.

4.3.5.2.1 Evaluation Results and Analysis in Scenario 1

In Table 4-5, the throughput of 271 data sources (i.e., 28 RSEs, 241 PIDs, the Alternate Mode Transportation Center, as well as the Traffic Management Center are the data sources of Scenario 1), data recording rate, average/maximum end-to-acknowledgement delay, and end-to-end delay of two test rounds are presented. When 1 Transit Cloud is used, the capacity of sending data from each data source is 0.52 Mb/s, or 5,467 records/sec, which is much lower than the test result of round 2, with 3 Transit Clouds (i.e., 1.63 Mb/s or 17,055 records/sec). This is because in the test round 2, all 3 Transit Clouds work at the same time to receive data from the sources, while the data sending capability is not fully used, so, a higher data receiving requirement leads to higher throughput. The average/maximum end-to-acknowledgement delay is 5,804.81 ms (millisecond)/ 10,540.49 ms for test round 1 and 1,613.73 ms/4,187.50 ms for test round 2. As expected, the end-to-acknowledgement delay, with 1 Transit Cloud, is higher than that with 3 Transit Clouds, which means, more Transit Clouds will improve the performance of the data exchange framework. In the condition when the data sending capability is not fully used, higher throughput would decrease the data exchange time. End-to-end delay (i.e., only travel time through a medium (i.e., optical fiber)) is 2 ms, the same for both rounds as it only includes transmission delay.

TABLE 4-5 Simulation Result for Data Exchange Framework (Scenario 1)

Category	271 ds-1 tc-1 r	271 ds-3 tc-3 r
Throughput of Data Source (Mb/s)	0.52	1.63
Recording Rate (records/sec)	5467	17055
Average End to Acknowledgement Delay (ms)	5804.81	1613.73
Maximum End to Acknowledgement Delay (ms)	10540.49	4187.50
End to End Delay (ms)	2.00	2.00

4.3.5.2.2 Evaluation Results and Analysis in Scenario 2

Table 4-6 shows the results for scenario 2. In test round 1, there are 30 data sources (i.e., 28 RSEs and 2 centers were data sources in Scenario 2), sending data to the Transit Cloud, while, in the test round 2, 241 data sources (i.e., PIDs) are sending data. Throughput of the data source, recording rate, average/maximum end-to-acknowledgement delay, and end-to-end delay are reported. The data sending capability of each data source is 1.82 Mb/s, or 19,116 records/sec in test round 1, which is much higher than the data sending capability of each data source in test round 2, 0.61 Mb/s or 6388 records/sec. The average/maximum end-to-acknowledgement delay in test round 1 is 6.67 ms/153.27 ms, which is higher than the average/maximum end-to-acknowledgement delay in test round 2. In test round 1, the end-to-end delay is 1.00 ms, which is the same as the end-to-end delay in test round 2 as it only includes the transmission delay. More data sources will make the data exchange framework more complex, which will decrease the throughput, and increase the data delivery delay. Thus, more data sources will reduce the performance of the data exchange framework.

TABLE 4-6 Simulation Result for Data Exchange Framework (Scenario 2)

Category	30 ds-1 tc	241 ds-1 tc
Throughput Of Data Source (Mb/s)	1.82	0.61
Recording Rate (records /sec)	19116	6388
Average End to Acknowledgement Delay (ms)	6.67	5454.77
Maximum End to Acknowledgement Delay (ms)	153.27	10014.07
End to End Delay (ms)	1.00	1.00

4.3.5.2.3 Evaluation Results and Analysis in Scenario 3

Table 4-7 provides the results of the performance between the data users (i.e. the Transportation Information Center and the Transit Management Center are the data users of Scenario 3) and the Transit Cloud for Scenario 3. Since acknowledgement can only be sent out from the Transit Cloud, the end-to-acknowledgement delay was not tested in Scenario 3. Throughput of data users, recording rate, and end-to-end delay are reported in this table. The data receiving capability of the data users is 43.31Mb/s, or 454,139 records/sec, which is really high. End-to-end latency is 1.00 ms as previous scenario. The result of this scenario indicates that the performance of the developed data exchange framework will not be influenced significantly after the data is available in the Transit Cloud.

TABLE 4-7 Simulation Result for Data Exchange Framework (Scenario 3)

Category	1 tc-2 du
Recording Rate (records/sec)	454139
Throughput of Data Users (Mb/s)	43.31
End to End Delay (ms)	1.00

4.3.5.2.4 Evaluation Results and Analysis in Scenario 4

Table 4-8 shows the results from the data sources (i.e., 28 RSEs, 241 PIDs, the Alternate Mode Transportation Center, and the Traffic Management Center are the data sources of Scenario 4) to the data users (i.e., Transportation Information Center and Transit Management Center). In this table, the throughput at the data user's end is 15.36 Mb/s, which obviously decreased, in comparison to the results in Scenario 3, due to the constraints of the throughput at the data source end, which is 1.72 Mb/s. The average end-to-acknowledgement delay is 3,577.13 ms, and the maximum end-to-acknowledgement delay is 5,656.70 ms., and the reason why end-to-acknowledgement delay in Scenario 4 is shorter than Scenario 1 is because in Scenario 4, no data replication was done in Transit Cloud. End-to-end delay is 3.00 ms.

TABLE 4-8 Simulation Result for Data Exchange Framework (Scenario 4)

Category		271 ds-1 tc-2 du
Data Source	Throughput (Mb/s)	1.72
	Recording Rate (records /sec)	18002
Data User	Throughput (Mb/s)	15.36
	Recording Rate (records /sec)	161098
Average End to Acknowledgement Delay (ms)		3577.13
Maximum End to Acknowledgement Delay (ms)		5656.70
End to End Delay (ms)		3.00

According to the performance requirement report, the Intelligent Network Flow Optimization (INFLO) Prototype, developed by the USDOT, a Traffic Management Entity (TME) is required to have the capability to obtain data from the traffic sensor every 20 seconds for CV-related dynamic mobility applications (FHWA, 2013). Table 4-8 shows that the average/maximum end-to-acknowledgement delay is measured for exchanging the messages from the data source to the data users, using Transit Cloud. The average end-to-acknowledgement delay is 3.58 sec, and the maximum end-to-acknowledgement delay is 5.66 sec. Even with data replication, as shown in Table 4-5, the maximum end-to-acknowledgement delay is 10.54 sec, which is much shorter than the required maximum delay. End-to-end delay from the Transit Cloud to the data users, as shown in Table 4-7, is really short, and can be ignored. Thus, the measured delay in the developed data exchange framework satisfies the USDOT requirement. Figure 4-2 shows the comparison between the average/maximum end-to-acknowledgement delays in different scenarios and the threshold of USDOT requirement, which indicates that all the evaluation requirements satisfy the requirement.

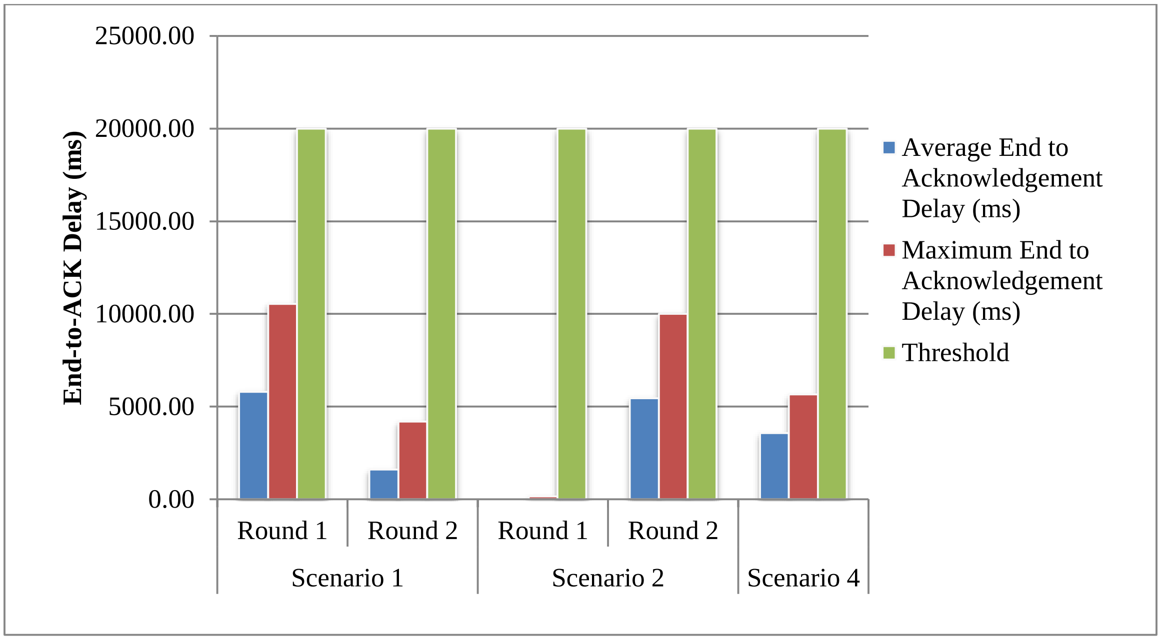

Figure 4-2 The Comparison of End-to-Acknowledgement Delays in Different Scenarios and the USDOT Requirement Threshold

4.4 Potential Implementation Cost

Since the software, which is an open source platform named Kafka, to support the developed data exchange framework is free for use, the required level of investment for developing and deploying the data exchange framework is only limited to the hardware cost (Kreps et al., 2011). Compared to the traditional data exchange framework, the possible hardware cost comes with the addition of the Transit Cloud layer. However, this cost can be reduced by placing the Transit Cloud and its replica on a selected RSE, which can cover an area of a certain number of RSEs. Furthermore, as shown in the simulation results, a single Transit Cloud can support a large number of data sources. Therefore, it is possible to amortize this cost, by combining the Transit Cloud supporting multiple regions, and sharing them via cloud computing resources. With this approach, the cost for human resources can also be reduced.

4.5 Summary

This chapter presented an evaluation of the developed data exchange framework, based on a case study on CAT system. Synthetic data is generated, based on the metadata analysis of the CVRIA and CAT information. Four scenarios are created to evaluate the performance of the developed data exchange framework. Throughput/or recording rate and delay are selected to be the evaluation parameters.

The detailed research findings, based on the simulated evaluation of the developed data exchange framework's performance are discussed. The test was conducted in four scenarios, and from Scenario 1, it shows that more Transit Cloud would increase the data sending capability for its sources, and reduce end-to-acknowledge delay. Scenario 2 shows that more data sources will reduce the data sending capability, and increase the end-to-acknowledge delay. This delay will not be influenced by number of Transit Cloud or data source number. Scenario 3 shows that the throughput and end-to-end delay from Transit Cloud to data users, and it can be seen that the delay is really low. In Scenario 4, the data exchange delay from its sources to the users was evaluated. In comparison with the USDOT’s requirement, it shows that the data exchange delay in the developed data exchange framework is in the required delay range. Finally, the potential implementation cost shows that the implementation of the developed data exchange framework will not be costly.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Overview

This chapter consists of two sections. Section 5.2 presents the conclusions of this research, and Section 5.3 summarizes the recommendations.

5.2 Conclusions

A secured, scalable, and resilient data exchange framework for DTO application was developed through this research. A new layer between the data sources and the data users, called the Transit Cloud, was used to improve the data exchange security, scalability, and resiliency of the framework. This research also investigated the efficiency of the developed data exchange framework, for managing massive transit data for the connected CAT service, by comparing their performance with the USDOT data delivery performance requirements. The DTO metadata from CVRIA and RDE data were analyzed to generate the synthetic data, which was used as the input in the evaluation of the transit network. This data exchange framework was simulated in the Cloud Lab, a distributed cloud infrastructure, in which, the data exchange delay for DTO was examined for different simulation scenarios utilizing the synthetic data.

Data exchange delay, in terms of throughput and delay for different simulation scenarios, was measured to evaluate the performance of the developed data exchange framework. From the simulation results, it is observed that the average data exchange delay for the duplication of data in three Transit Clouds was reduced because the capacity was increased with more Transit Clouds, and the throughput of data users were limited by the throughput of the data sources. A more complex network would potentially reduce the performance of the data exchange framework presented in this thesis. The average and maximum end-to-acknowledgement delay from the data sources to Transit Cloud, and from Transit Cloud to data users were 3.58 seconds and 5.66 seconds, respectively, which satisfy the USDOT requirements. An analysis of four scenarios revealed that the developed data exchange framework with the Transit Cloud framework was more secure, scalable, and resilient, when compared to the existing data analytics framework for supporting the transit operations. Thus, Transit Cloud is a more preferable alternative in comparison to the existing framework because of its added benefits.

5.3 Recommendations

It is necessary to develop and evaluate the performance of potential data exchange framework for CV applications due to the massive amount of data that would be generated in the CV environment. Based on findings of this research, the author presents the following recommendations:

- The case study conducted to evaluate the performance of the data exchange framework presented in this research included a transit network that is not large and complex. An evaluation of the data exchange framework is recommended for a major metropolitan area where a larger and more complex transit network exists. An evaluation with more Transit Clouds is recommended in follow-up research
- Simulation platform and synthetic data are used to evaluate the performance in this study. A real-world evaluation with field data should be conducted in future research.
- A robust data exchange framework is designed for the DTO application in the CV environment. This framework could also be implemented as a data exchange model for other CV applications as well.

REFERENCES

Bacon, J., Eyers, D. M., Singh, J., & Pietzuch, P. R. (2008). Access control in publish/subscribe systems. In Proceedings of the second international conference on Distributed event-based systems, 23-34. ACM, 2008.

Bhavsar, P., Chowdhury, M., Sadek, A., Sarasua, W., and Ogle, J., “Decision Support System for Predicting Traffic Diversion Impact across Transportation Networks using Support Vector Regression,” Transportation Research Record: Journal of the Transportation Research Board, No. 2024, pp 100-106, (2007).

Boenau, R. & Timcho, T. (2014) Integrated Dynamic Transit Operations (IDTO) prototype development -a USDOT connected vehicle research project transit ITS best practices workshop on effective use of transit ITS, Deployment, Data, Development, Orlando, FL.

Chang, E. C.P. (2004). Using ITS data to manage transportation infrastructure, IEEE International Conference on Networking, Sensing and Control (IEEE Cat. No.04EX761), 2004(1), 486.

Chowdhury, M., Sadek, A., Ma, Y., Kanhere, N., and Bhavsar, P., “Applications of Artificial Intelligence Paradigms to Decision Support in Real-time Traffic Management,” Transportation Research Record, Journal of the Transportation Research Board, No. 1968, pp 92-98, (2006).

Conzon, D., Bolognesi, T., Brizzi, P., Lotito, A., Tomasi, R., & Spirito, M. (2012). The virtus middleware: An XMPP based architecture for secure IOT communications.

In Computer Communications and Networks (ICCCN), 21st International Conference on (pp. 1-6). IEEE.

Dion, F., Robinson, R., & Oh, J. S. (2011). Evaluation of usability of Intellidrive probe vehicle data for transportation systems performance analysis. Journal of Transportation Engineering, 137(3).

Federal Transit Administration (FTA) (2002). White Paper on Literature Review of Real-time Transit Information systems. Retrieved from: http://ntl.bts.gov/lib/jpodocs/repts_te/13845.html, 2002

FHWA, Research Data Exchange, available online at https://www.its-rde.net/home, 2015.

The University of Utah, (2015). CloudLab Clemson hardware details. Retrieved from: https://www.cloudlab.us/hardware.php#clemson.

FHWA, (2013). Report on detailed requirements for the INFLO prototype, final report. FHWA-JPO-14-171, US Department of Transportation. Retrieved from: http://ntl.bts.gov/lib/54000/54800/54832/FHWA-JPO-14-171.pdf. 2013.

Ford, D., Labelle, F., Popovici, F. I., Stokely, M., Truong, V. A., Barroso, L., Grimes, C., & Quinlan, S. (2010, October). Availability in globally distributed storage Systems. In OSDI (pp. 61-74), 2012.

Fries, R., Chowdhury, M., and Ma, Y., “Accelerated Incident Detection and Verification: A Benefit to Cost Analysis of Traffic Cameras,” Journal of Intelligent Transportation Systems, No. 11(4), pp 191-203, (2007).

Genders, W. &. Razavi, S.N., (2015). Impact of connected vehicle on work zone network safety through dynamic route guidance, Journal of Computing Civil Engineering, ASCE, ISSN 0887-3801/04015020(11).

Hamilton, B.H. (2009). Vehicle infrastructure integration (VII) Proof of concept (POC) test executive summary—infrastructure. Research and Innovative Technology Administration, U.S. Department of Transportation, Washington DC.

Hao, Y., Wang, Y., & Yang, X., (2014). A schedule-based coordinated optimization model for transit signal priority under connected vehicle environment, IEEE 17th International Conference on Intelligent Transportation Systems (ITSC), October 2014.

Harford, J.D. (2006). Congestion, Pollution, and Benefit-to-cost ratios of US Public Transit Systems, Transportation Research Part D: Transport and Environment, 11(1), 45–58.

He, Y., Chowdhury, M., Pisu, P and Ma, Y., "An Energy Optimization Strategy for Power-split Drivetrain Plug-in Hybrid Electric Vehicles," Transportation Research Part C: Emerging Technologies, Vol. 22, pp 29–41, (2012).

He, Y., Chowdhury, M., Ma, Y., and Pisu, P., "Merging Mobility and Energy Vision with Hybrid Electric Vehicles and Vehicle Infrastructure Integration," Energy Policy, Modeling Transport (Energy) Demand and Policies, Vol. 41, pp 599-609, (2012).

He, Y., Rios, J., Chowdhury, M., Pisu, P., and Bhavsar, P., "Forward Power-Train Energy Management Modeling for Assessing Benefits of Integrating Predictive Traffic Data into Plug-in-Hybrid Electric Vehicles," Transportation Research Part D: Transport and Environment, Vol. 17, Issue 3, pp 201–207, (2012).

HomChaudhuri, B., Vahidi, A., & Pisu, P. (2015). A fuel economic model predictive control strategy for a group of connected vehicles in urban roads, American Control Conference, July 2015.

Hsu, M. & Shih, A. (2015). A traffic signal control mechanism in a connected vehicle environment, 2015 International Conference on Consumer Electronics-Taiwan (ICCE-TW).

ITS Joint Program Office, Federal Highway Administration. Retrieved from: http://www.iteris.com/cvria/html/applications/app77.html#tab-3

Jansen, W., & Grance, T. (2011). Guidelines on security and privacy in public cloud computing. NIST special publication 800, 144.

Jin, Q., Wu, G., Boriboonsomsin, K., & Barth, M., (2012). Advanced intersection management for connected vehicles using a multi-agent systems approach, 2012 Intelligent Vehicles Symposium.

Kattan, L., Moussavi, M., Far, B., Harschnitz, C., Radmanesh, A., & Saidi, S., (2010). Evaluating the potential benefits of vehicle to vehicle communication (V2V) under incident conditions in the PARAMICS model. Proc., 13th Int. IEEE Conf. on Intelligent Transportation, New York.

KFH, Clemson Area Transit Comprehensive Operational Analysis Final Report, August 2014.

Khadra, A., Liu, X., & Shen, X. (2003). Application of Impulsive Synchronization to Communication Security, IEEE Transactions on Circuits and System – I: Fundamental Theory and Applications, 50(3).

Kim, Y. J. Kolesnikov, V. & Thottan, M. (2012). Resilient end-to-end message protection for large-scale cyber-physical system communications. Smart Grid Communications (SmartGridComm), 2012 IEEE Third International Conference on. IEEE, 2012.

Kreps, J., Narkhede, N., & Rao, J. (2011) Kafka: A distributed messaging system for log processing. Proceedings of the NetDB.

Kristensson, A., Johansson, C., Westerhole, R., Swietlicki, E., Gidhagen, L., Wideqvist, U., & Vesely, V., (2004). Real-world traffic emission factors of gases and particles measured in a road tunnel in Stockholm, Sweden, Atmospheric Environment 38, 657–673.

Kühn, E., Mordinyi, R., Keszthelyi, L., Schreiber, C., Bessler, S., & Tomic, S. (2009). Aspect-oriented space containers for efficient publish/subscribe scenarios in intelligent transportation systems. In On the Move to Meaningful Internet Systems: OTM, 432-448. Springer Berlin Heidelberg.

Lomax, T., Schrank, D., & Eisele, B., (2012). TTI's 2012 Urban Mobility Report, Texas A&M Transportation Institute, 2012.

Lomax, T., Schrank, D., Eisele, B., & Bak, J., (2015), "2015 Urban Mobility Scorecard," Texas A&M Transportation Institute, the Texas A&M University System. Retrieved from: http://d2dtl5nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/mobility-scorecard-2015-wappx.pdf

Ma, Y., Chowdhury, M., Sadek, A., and Jeihani, M., "Integration of Vehicle Infrastructure Integration (VII) system and Artificial Intelligence (AI) for online travel time prediction," IEEE Transactions on Intelligent Transportation Systems Vol. 13, Issue: 3, pp 1369 – 1382, (2012).

Ma, Y., Chowdhury, M., Sadek, A., and Jeihani, M., "Real-Time Highway Traffic Condition Assessment Framework Using Vehicle-Infrastructure Integration (VII) with Artificial Intelligence (AI),"IEEE Transactions on Intelligent Transportation Systems, Vol. 10, No. 4, pp 615-627, (2009).

Manasseh, C., & Sengupta, R. (2008) Middleware for cooperative vehicle-infrastructure systems. California Partners for Advanced Transit and Highways (PATH).

Marsh, G., Sampat, A. P., Potluri, S., & Panda, D. K. (2008). Scaling Advanced Message Queuing Protocol (AMQP) architecture with Broker Federation and InfiniBand. Ohio State University, Tech. Rep. OSU-CISRC-5/09-TR17.

Minelli, S., Izadpanah, P., Razavi, S., (2015). Evaluation of connected vehicle impact on mobility and mode choice, Journal of Traffic and Transportation Engineering, 2(5), 301-312.

Mishra, S., Schweiger, C., Abuelhiga, A., Butler, C., Beasley, & Sanchez, R. (2012). Report on functional and performance requirements and high-level data and communication needs for integrated dynamic transit operations (IDTO). Final Report —August 13, 2012, FHWA-JPO-12-085. 2012.

Morgan, S.L., & Waring, C.G., (2004). Guidance on Testing Data Reliability. Retrieved from: http://www.auditorroles.org/files/toolkit/role2/Tool2aAustinCityAud_GuidanceTestingReliability.pdf.

National Highway Traffic Safety Administration (NHTSA), (2014). Vehicle-to-vehicle communications: Readiness of V2V technology for application, DOT HS 812 014. Retrieved from:

http://www.its.dot.gov/cv_basics/pdf/Readiness_V2V_TechnologyApplication.pdf.

National Highway Traffic Safety Administration (NHTSA), (2013). Traffic Safety Facts 2013. Retrieved from: http://www-nrd.nhtsa.dot.gov/Pubs/812139.pdf.

Papadimitratos, P., La Fortelle, A., Evenssen K., Brignolo R., & Cosenza S., (2009). Vehicular communication systems: Enabling technologies, applications, and future outlook on intelligent transportation, Communications Magazine, IEEE, 47(11), 84-95.

Pina, M. (2015). ITS Joint Program Office, How Connected Vehicles Work. Retrieved from: http://www.its.dot.gov/factsheets/pdf/JPO_HowCVWork_v3.pdf.

Report on Detailed Requirements for the INFLO Prototype, Final Report. (2013) FHWA-JPO-14-171, US Department of Transportation. Retrieved from: http://ntl.bts.gov/lib/54000/54800/54832/FHWA-JPO-14-171.pdf.

Schrank, D. & Lomax, T. (2005). The 2005 Urban Mobility Report. College Station: Texas Transportation Institute, Texas A&M University, 2005. Retrieved from: http://mobility.tamu.edu.

Schweiger, C., Mishra, S., Abuelhiga, A., & Sanchez, R. (2012). Test readiness assessment summary for integrated dynamic transit operations (IDTO). Final Report — October 19, 2012, FHWA-JPO-12-086. 2012.

Sousanis, J. (2011), World vehicle population tops 1 billion units, WardsAuto, Retrieved from: http://wardsauto.com/ar/world_vehicle_population_110815, 2015.

Statista, Number of vehicles registered in the United States from 1990 to 2013 (in 1,000s), Retrieved from: http://www.statista.com/statistics/183505/number-of-vehicles-in-the-united-states-since-1990/, 2015.

Stevanovic, A., Stevanovic, J., Zhang, K., Batterman,S., (2009) Optimizing traffic control to reduce fuel consumption and vehicular emissions, Transportation Research Board, ISSN: 0361-1981, 2128, 105–113.

Taniguchi, E. & Shimamoto, H. (2004). Intelligent transportation system based dynamic vehicle routing and scheduling with variable travel times.

The University of Utah, (2015). Cloud Lab, Retrieved from: https://www.cloudlab.us/.

Timcho, T., Burns, M., Mishra, S., Schweiger, C., Rizek, J., & Zink, G. (2013). Report on functional requirements and software architecture for the IDTO prototype. Final Report — August 30, 2013, FHWA-JPO-14-166. 2013.

Transloc. (2015). Retrieved from: http://catbus.transloc.com/.

US Joint Program Office, Federal Highway Administration. Retrieved from: http://www.its.dot.gov/connected_vehicle/connected_vehicles_FAQs.htm.

Vilela, J. P., Lima, L., & Barros, J. (2008). Lightweight security for network coding. In *Communications*, *2008. ICC'08*. IEEE International Conference on (pp. 1750-1754).

Office of the Assistant Secretary for Research and Technology (OST-R) (2015) • U.S. Department of Transportation (US DOT), Connected Research Reference Implementation Architecture (CVRIA). Retrieved from: http://www.iteris.com/cvria/html/applications/applications.html.

Intelligent Transportation Systems -- Joint Program Office, Office of the Assistant Secretary for Research and Technology (OST-R) • U.S. Department of Transportation (US DOT), Southeast Michigan Test Bed (2014) Concept of Operations. Retrieved from: http://www.iteris.com/cvria/docs/SE_Michigan_Test_Bed_2014_ConOps_-_v1_-_2014-Dec_29.pdf, 2014.