This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2888968, IEEE Transactions on Mobile Computing

Traffic Flow Control in Vehicular Multi-Hop Networks with Data Caching

Teng Liu, Alhussein A. Abouzeid, and A. Agung Julius
Department of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180-3590, USA
Email: liut7@rpi.edu, abouzeid@ecse.rpi.edu, agung@ecse.rpi.edu

Abstract—Control of conventional transportation networks aims at bringing the state of the network (e.g., the traffic flows in the network) to the *system optimal* (SO) state. This optimum is characterized by the minimality of the social cost function, i.e., the total cost of travel (e.g., travel time) of all drivers. On the other hand, drivers are assumed to be rational and selfish, and make their travel decisions (e.g., route choices) to optimize their own travel costs, bringing the state of the network to a *user equilibrium* (UE). A classic approach to influence users' route choice is using congestion tolls. In this paper we study the SO and UE of future connected vehicular transportation networks, where users consider both the travel cost and the utility from data communication, when making their travel decisions. We leverage the data communication aspect of the decision making to influence the user route choices, driving the UE state to the SO state. We assume the cache-enabled vehicles can communicate with other vehicles via vehicle-to-vehicle (V2V) connections. We propose an algorithm for calculating the values of the data communication utility that drive the UE to the SO. This result provides a guideline on how the system operator can adjust the parameters of the communication network (e.g., data pricing and bandwidth) to achieve the optimal social cost. We discuss the insights that the results shed on a secondary optimization that the operator can conduct to maximize its own utility without deviating the transportation network state from the SO. We validate the proposed communication model via Veins simulation. The simulation results also show that the system cost can be lowered even if the bandwidth allocation does not exactly match the optimal allocation policy under 802.11p protocol.

Index Terms—traffic control; vehicular communication networks; system optimal; user equilibrium

1 Introduction

In transportation systems, the prospect of wide-scale connected autonomous vehicles (CAVs) is approaching its realization, due to the advances in control and communication. In a traditional transportation network, the drivers make travel decisions (e.g., route choices, travel timing) that minimize the transportation related costs, such as travel time, travel distance, etc. With the emergence of CAVs that form vehicular ad-hoc networks (VANETs), data communication network connectivity is not only going to be an important factor for enabling vehicular control, but also going to change the CAV users traveling behavior. Some CAV users will expect the type of data communication service they are accustomed to at their homes and offices. Thus, CAV users may choose routes not only depending on travel time and costs, but also based on the quality of data service that will be provided on the route, since this directly affects their productivity and/or quality of life. CAV users may choose to take a route with longer travel time in order to have a better data communication network connectivity. A similar scenario, where users trade off between different commodities according to their preferences, is that travelers may choose a more expensive hotel, or a less convenient hotel location, if it offers a high speed WiFi connection. Evidence of this behavior has been recently reported in [1], where data connectivity affects the route choice of (human) drivers. Hereafter in this paper, we will refer to the travel decision makers (i.e., drivers or CAV users) as "users."

Travel decision making among users can be analyzed in a game theoretical setting [2], [3]. The travel decision of each user impacts the state of the transportation network, and thereby may also impact the transportation costs of all users. The Nash equilibrium of this game is referred to as the *Wardrop equilibrium* or the *user equilibrium* (UE). Thus, UE occurs if no user can be better off by unilaterally changing his travel decision. In a traditional transportation network, the UE state[1] is achieved if every user tries to minimize his/her travel cost (e.g., travel time).

In contrast to UE, we can also consider the system optimal (SO) state. The system optimal state occurs if the social cost function, i.e., the total of the travel costs of all users, is minimized. In general, assuming that the users are selfish and rational, it is known that UE and SO are not the same. This phenomenon is sometime referred to as the Braess' paradox [4], [5]. The ratio between the social costs at UE and at SO is called the *price of anarchy* (PoA) [6]. The SO is regarded as the ideal state: a closer UE to the SO in terms of the traffic flows results in a lower social cost. Consider a simple network that consists of a single O-D pair connected by two links with traffic flow x_1 and x_2 respectively. For simplicity, we assume that the travel time on the link is $T_1 = 7 + 6x_1 + 4x_2$ and $T_2 = 1 + 2x_1 + 10x_2$ respectively. We will introduce a more realistic travel cost model (Bureau of Public Roads function) in Section 5.3. Fig. 1 shows how the system cost varies with the traffic flow on link 1 under

1. i.e., the state of the transportation network if the game is at UE.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2888968, IEEE Transactions on Mobile Computing

different trip rates in this network. Given the trip rate q, one can solve for the traffic flows at the SO and at the UE. Connecting the SO points (UE points) under all possible trip rates gives the SO trace (UE trace), which is shown by the black dashed line (red dashed line) in Fig. 1. We note that the UE deviates from the SO as the trip rate increases. Traffic control policies, for example congestion tolls, push the UE closer to, and even the same as, the SO, as indicated by the arrow in Fig 1. PoA is eliminated if the UE trace matches the SO trace.

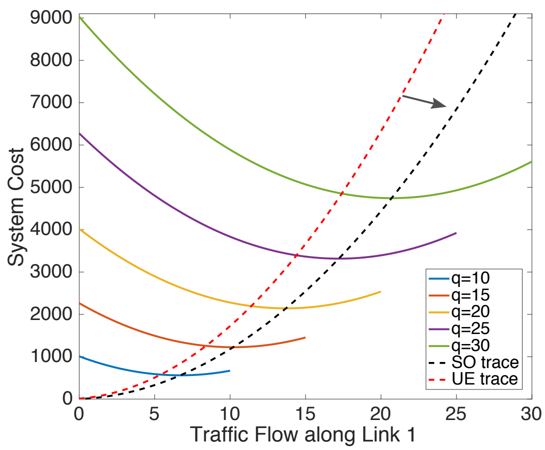

Fig. 1: System cost v.s. traffic flow under different trip rates. Black(red) dashed line connects the SO(UE) states under different trip rates, and is referred to as the SO(UE) trace. The UE trace deviates from the SO trace when the trip rate increases. Our goal is to drive the UE trace closer to, or even the same as, the SO trace, as indicated by the black arrow.

In this paper, we study the UE state and the SO state in the vehicular communication network, where the interdependency between the network condition (including traffic network condition and communication network condition) and the users' valuation of the cost (including travel cost and communication cost) leads to a different UE. We assume that the cache-enabled vehicles can communicate with other vehicles via vehicle-to-vehicle (V2V) connections. We refer to the traffic flows that support data caching and forwarding as the cache-enabled traffic flows hereinafter. Therefore, for a choice of caching, infrastructure, and user connectivity profile, data connectivity depends on the flow density. A dense flow may reduce the quality of service of the V2I connections while benefiting the content users by increasing the cache hit probability, and vice versa. This connectivity dynamics, coupled with the traffic condition, affects users' route planning. For example, a dense traffic flow in a road segment leads to a longer travel time, but can potentially lower the communication cost if the benefit of the V2V caching gain dominates the loss of the V2I QoS declining. On the other hand, as more users choose to use the road segment with low travel cost and low communication cost, congestion may occur in both the traffic network and the communication network, which will discourage other users from using this road segment.

The interaction between the transportation network and the users decisions has been thoroughly studied [2]. However, the effect of network communications, both from a connectivity dynamics point of view, and from a decision point of view, have not been considered. In this work, we study the influence of the traffic condition and the data service on users route planning. We adopt the notion that the system operator can use the communication network parameters as a leverage to push the user equilibrium (UE) to the system optimal (SO). This paper is a substantial extension of our previous conference paper [7]. Specifically, we make the following contributions:

(a) We derive the data throughput in the vehicular communication network that supports content caching and V2V communication, based on which we propose a communication cost model. This model takes into consideration the throughput scaling law in ad-hoc networks, the cache hit probability of the users, and the limitation of V2V bandwidth allocation. Our work makes it possible to incorporate the data communication aspect in traffic flow control.
(b) In order to demonstrate that the proposed communication cost function enables a wide range of applications, we propose a V2V bandwidth allocation scheme with the aim of driving the UE to the SO under the marginal cost pricing framework [2].
(c) We conduct a comprehensive case study on a network in New York State Capital District using the proposed bandwidth allocation scheme, which gives the optimal bandwidth allocation for the main highways in the network.
(d) We validate the proposed communication cost model via simulation. The simulation results also show that the system cost can be lowered under 802.11p protocol.

The remainder of this paper is organized as follows. In Section 2 we review the related work on the data communication and user behavior in vehicular communication networks. In Section 3 we present the model of the transportation network and the communication network, and present the communication cost function and a general trip cost function. In Section 4 we discuss the SO state and the UE state, and the corresponding necessary conditions on the traffic flows. In Section 5 we design a primary optimization technique that steers UE to match SO by leveraging the communication cost, and show the achievability of this UE-SO matching. A secondary optimization is presented in Section 6 with the objective of minimizing the total bandwidth allocation. A comprehensive case study on a network in the New York Capital District is demonstrated in Section 7. In Section 8 we validate the proposed communication cost model via simulation, and show the simulation results when the bandwidth allocation does not exactly match the optimal value under 802.11p protocol, where the bandwidth can only take on 8 possible values. We conclude our work in Section 9.

2 Related Work

In order to decrease the Price of Anarchy (PoA) in transportation networks, congestion tolls have been proposed and is currently adopted in practice [8], [9]. A wealth of research has been done on the design of congestion tolls. Marginal cost pricing is a well-known approach for steering the user equilibrium to the system optimal, which has been proposed in [2]. In marginal cost pricing, the toll of a link is set to the difference between the marginal cost at user

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2888968, IEEE Transactions on Mobile Computing

equilibrium and the marginal cost at system optimal. There exists a number of problems in this marginal cost pricing approach, for example, users may have different sensitivities to the tolls. The recent work by Wang et al. [10] seeks to eliminate the PoA by imposing scaled marginal-cost road pricing on the a transportation network where users have different toll sensitivities. Another problem is that in the marginal cost pricing approach, the trip rate is assumed to be known a-priori. However, the real trip rate may not be the same as the predicted trip rate, thus the marginal cost pricing may result in a high PoA. In [11], the demand-independent tolls have been proposed, which induce the system optimum flow as a Wardrop equilibrium without the prior knowledge of the trip rates if the travel time is a BPR-type cost function [12]. Knowledge of the cost functions is key in characterizing both the system optimal and the user equilibrium. The recent work by Zhang et al. [13] seeks to derive the users travel cost functions from city-wide real traffic data.

It is expected that CAV users needs for, and valuation of, data service vary based on their socioeconomic characteristics and trip-related features. There is wealth of literature on people's behavior in response to transportation service and data communication service. These studies, however, reside in different research fields. Transportation studies typically focus on traveler behavior including mode choice, route choice, departure time choice, etc. For traveler route choice, the main focus ranges from the effects of road pricing [14], fuel costs [15], congestion level [16], reliability [17], land use [18], to advanced traveler information system [19]. User responses to cost and quality of data communication service have been investigated in a wide spectrum of fields including information systems, psychology, and business management. Studies have looked into effects of perceived fee [20], user prior experience and habits [21], social influence [22], perceived monetary value, among others. In a recent literature review, [23] summarized key areas and methods on research related to people's data communication behavior in the past decade. However, no existing study has explored the problem of exploiting the communication aspect to maximize the social welfare when users are faced with the joint choice of transportation and data service , which is the key feature of CAV users, and is the focus of this paper.

Incorporating the communication network in the transportation networks enables a wide range of applications [24], [25], for example, interactive entertainment, urban sensing [26], collision avoidance in platoon formation [27], improving the intersection capacity via platoons [28], etc. A wealth of research focuses on vehicle-to-vehicle (V2V) communication and Vehicle-to-infrastructure (V2I) communication in transportation networks (e.g. [29], [1], [30]). Content caching in vehicular networks has been considered in the past, exploiting the large data storage space of vehicles and the dynamic topology of the networks (e.g. [31], [32], [33]).

In this paper, we use the marginal pricing, which is a classic approach of congestion tolls, to demonstrate a possible application of our proposed communication model. Our novel contribution is in modeling the data throughput of each link in vehicular communication networks that support V2V communication and data caching. Based on this throughput, we derive a communication cost function that takes into consideration the throughput scaling law in ad-hoc networks, the cache hit probability of the users, and the constraints on V2V bandwidth allocation. This communication cost function enables a wide range of applications, and makes it possible to incorporate the data communication aspect in the traffic flow control. For example, it can be used to steer the user equilibrium to system optimal under the aforementioned marginal cost pricing framework. With our proposed model, we analyze if the marginal cost pricing can actually achieve UE-SO matching. This communication cost can be viewed as a type of toll, but this "toll" cannot be controlled by the operator directly, and will be influenced by the traffic flows in real time. Another focus of our work is the development of a simulation environment/tool to simulate both communication network protocols and transportation network dynamics, where the communication quality impacts users' behavior in real time. Up to our knowledge, there does not exist a simulator that captures this joint dynamics of network protocols and transportation networks. We validate the proposed model and the bandwidth allocation scheme via simulation under 802.11p protocol. By using the marginal cost pricing framework, the optimal bandwidth allocation may not be feasible under 802.11p protocol. Therefore, via simulation, we show that the closest possible allocation to the optimal allocation under 802.11p can still lower the system cost.

3 System Model

In this section, we first present the transportation network model and the communication network model in Section 3.1. Then we discuss the costs incurred by traffic and by data communication in Section 3.2. For ease of reference, related notations are shown in Table 1.

3.1 Network Model

The transportation network consists of a number of road segments, which we refer to as links. Infrastructure related parameters, such as the free-flow speed, stay the same throughout a link. Without any loss of generality, we only consider one-way traffic, i.e. all links are directed. A two-way link can be equivalently replaced by two one-way directed links if the traffic on one direction does not have communication overlap with the traffic on the other direction. The set of all links in the transportation network is denoted by A. Each vehicle in this transportation network travels from an origin to a destination via a set of links. We refer to an ordered sequence of links that connects an origin and a destination as a route. The set of all possible origin-destination pairs (O-D pairs) is denoted by N. There are one or more routes between each O-D pair. The set of all possible routes between the O-D pair i is represented by K_i, and the set of all possible routes between all possible O-D pairs is represented by K. We denote the arrival trip rate (trips per unit time) for O-D pair $i \in N$ by q_i. The indicator variable $\delta_{i,k}(a)$ is defined such that $\delta_{i,k}(a) = 1$ if link a is part of route $k \in K_i$. Otherwise $\delta_{i,k}(a) = 0$. Fig. 2 shows an example network that consists of one O-D pair. An O-D pair can potentially be traversed using multiple routes. For

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2888968, IEEE Transactions on Mobile Computing

Notation	Description
A	set of links (road segments)
N	set of all origin-destination (O-D) pairs
K	set of all routes between all possible O-D pairs
K_i	set of all routes connecting O-D pair $i \in N$
$\mathbf{q}$	trip rate vector with entry q_i denoting the trip rate between O-D pair $i \in N$
$\mathbf{x}$	link flow vector with entry x_a denoting the flow on link $a \in A$
$\mathbf{y}$	route flow vector with entry $y_{i,k}$ denoting the flow on route $k \in K_i$ that connects O-D pair $i \in N$
$\mathbf{b}$	bandwidth allocation vector with entry b_a denoting the bandwidth allocated to link $a \in A$
$\mathbf{b}^{max}$	maximum bandwidth vector with entry b_a^{max} denoting the upper bound on the bandwidth along link $a \in A$
$\delta_{i,k}(a)$	=1, if link a is on route k between O-D pair i; =0, otherwise
$\mathbf{T}(\cdot)$	travel cost vector with entry $T_a(\cdot)$ denoting the travel cost of link $a \in A$
$\mathbf{C}(\cdot)$	communication cost vector with entry $C_a(\cdot)$ denoting the communication cost of link $a \in A$
$\mathbf{C}^{max}$	$1 \times \|A\|$ vector with entry C_a^{max} denoting the upper bound of the communication cost on link $a \in A$
$J_{sys}(\cdot)$	system cost
$\mathbf{J}_i(\cdot)$	$1 \times \|K_i\|$ vector with entry $J_{i,k}(\cdot)$ denoting the total trip cost on route $k \in K_i$ that connects O-D pair $i \in N$
$\mathbf{\Lambda}_i$	$1 \times \|K_i\|$ vector with all entries being the same λ
$h_a(x_a)$	one-hop cache hit probability on link $a \in A$
$v_a(x_a)$	average speed on link $a \in A$
l_a	length of link $a \in A$
r	transmission range. $r << l_a, \forall a \in A$
p_a	caching ratio on link $a \in A$
u_a	flow density of link $a \in A$
$\mathbf{0}$	zero vector with every entries being 0

TABLE 1: Table of Notations

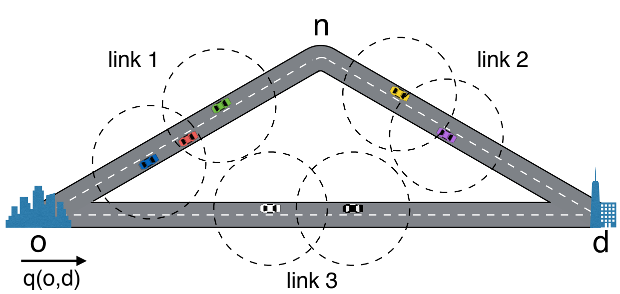

Fig. 2: A transportation network consisting of an origins o and a destination d. Links are indexed by the numbers next to them. Node n is the intersection of link 1, 2. The trip rate from o to d is $q(o, d)$. Communication range is denoted by dotted circle around a vehicle.

example, the O-D pair (o, d) can be traversed using route 1-2 or route 3. The dotted circle represents the communication range of a vehicle.

The vehicles travel along the links and form the link flow vector $\mathbf{x} \in \mathbf{R}^{|A|}$, where the entry x_a represents the traffic flow on link $a \in A$, and $|A|$ represents the cardinality of set A. Similarly, the route flow vector $\mathbf{y} \in \mathbf{R}^{|N| \times |K|}$ represents the number of vehicles that choose certain routes, where the entry $y_{i,k}$ denotes the flow on route $k \in K_i$ that connects the O-D pair $i \in N$.

All routes should satisfy the flow conservation constraints, i.e. the sum of the route flows along all routes that connect an O-D pair equals the O-D trip rate, and the flow on a link equals the sum of the flows that enter the link [3]. The flow conservation constraints are given by:

$$\sum_{k \in K_i} y_{i,k} = q_i, \ \forall i \in N, \tag{1}$$

$$y_{i,k} \geq 0, \ \forall i \in N, k \in K_i, \tag{2}$$

$$x_a = \sum_{i \in N} \sum_{k \in K_i} \delta_{i,k}(a) y_{i,k}, \ \forall a \in A. \tag{3}$$

Note that $\mathbf{x}$ is therefore a linear function of $\mathbf{y}$.

Each user is a participant both in the transportation network and in the communication network. We envision the use of vehicles as nodes with the network interfaces that support V2V communication. Vehicles communicate with each other by broadcasting in an ad-hoc manner, and the system operator can control the V2V communication by limiting the broadcast bandwidth. In practice, the system operator is a certain ”coordinator”, for example the Federal Highway Administration, which coordinates highway transportation programs in cooperation with states and other partners to enhance the country’s safety, economic vitality, quality of life, and the environment. This ”coordinator” will provide incentives to the network operator in order to maximize the social welfare. But how this coordination is done is beyond the scope of the paper. The bandwidth allocation vector is denoted by $\mathbf{b}$, where the entry b_a represents the bandwidth allocated for link a. The concrete shape of the communication cost is discussed in the next subsection.

3.2 Travel & Communication Cost Functions

We associate each route with a cost. As aforementioned, users in the vehicular communication network do not only value travel cost, such as travel time and travel distance, they also need data service for a better travel experience. Therefore, the trip cost consists of two parts: travel cost and communication cost. When a user chooses which route to take, for each route k that connects the O-D pair i, they are presented with the travel cost $T_{i,k}$ and the communication cost $C_{i,k}$. Without specifying how user preference would affect their trade-off between the travel cost and the communication cost, we denote the trip cost of route k that connects the O-D pair i by $J_{i,k}(T_{i,k}, C_{i,k})$.

The travel cost is a measure of the transportation related disutility. We do not restrict the travel cost to be any specific type of disutility. Instead, we represent the travel cost of a link a as a function of the traffic flow vector, i.e. $T_a(\mathbf{x})$. Given the traffic flow, all users experience the same link travel cost on the same link. Note that if the travel time is chosen to be the travel cost, then the average speed can be used to calculate $T_a(\mathbf{x})$. We define the route travel cost $T_{i,k}(\mathbf{x})$ as the sum of the link travel cost along the route.

The communication cost is a measure of the communication network performance, and is a function of the cache-enabled traffic flows and other relevant network parameters. The communication cost involves content downloading delay, data price charged, etc. The network topologies we consider are highway and urban networks where there is negligible communication range overlaps at intersections and parallel roads. The cache-enabled vehicles are envisioned as nodes, all of which send requests according to the following procedure. A query node broadcasts its request to the neighboring nodes that are within its transmission range

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2888968, IEEE Transactions on Mobile Computing

r along the same link. If one of the neighboring nodes has the requested content in its cache, it sends the content to the query node. If none of the neighboring nodes caches the requested content, the request is re-broadcast to the second hop neighbors. After a certain number of hops, the request is dropped if no node has cached the requested content. For ease of discussion, we assume that the request is dropped after one hop, and the neighboring vehicle who has the requested content in the cache sends this content directly to the querying vehicle. This can be readily expanded to multi-hop scenario by incorporating transmission power control policy and routing algorithm, which determine the maximum number of hops, or the maximum searching distance. We also assume that the caching ratio p_a, i.e. the probability that a piece of content can be found in a node along the link a, is the same for all nodes.

We assume the vehicles' location on link a follows a Poisson distribution with the density parameter u_a. The travel time and the average speed on link a given the link flow x_a is denoted by $t_a(x_a)$ and $v_a(x_a)$, respectively. We denote the cache-enabled flow on link a as $\tilde{x}_a$. From [34], the probability that at least one neighboring node has the requested content in cache, i.e. the one-hop cache hit probability, is given by:

$$h_a(\tilde{x}_a, x_a) = 1 - e^{-2rp_a\tilde{u}_a}, \tag{4}$$

$$\text{where } \tilde{u}_a = \frac{\tilde{x}_a}{v_a(x_a)}. \tag{5}$$

The one-hop cache hit probability does not solely determine the network performance. The interference among vehicles traveling in a dense flow may cause a significant access delay even if the one-hop cache hit probability is relatively high. On the other hand, more bandwidth allocation along a link with a high flow density can mitigate the interference among the vehicles, and thus reduce the content access delay. Therefore, we define the communication cost to be inversely proportional to the throughput achieved from cache hit, which is derived as follows. We assume that the traffic network is not highly dense, and the throughput of a road is $\Theta(\frac{1}{r})$ [35]. We further assume that the channel capacity, which is proportional to the bandwidth, is divided equitably between the vehicles on a road and that the throughput scales with $\Theta(\frac{1}{\sqrt{u}})$ [36], then the throughput per node is

$$\Theta(\frac{b_a}{2r\sqrt{\tilde{u}_a}} h_a(\tilde{x}_a, x_a))$$

We assume that the bandwidth allocated by the system operator along a road cannot exceed a maximum value, denoted by $b_a^{max}, a \in A$. Similarly, the user-perceived communication cost will not be unbounded, and we denote the maximum possible communication cost on a road by $C_a^{max}, a \in A$. Let k denote a positive constant, then we can define the communication cost of link a as

$$C_a(\tilde{x}_a, x_a, b_a) = \min\{c_a(\tilde{x}_a, x_a, b_a), C_a^{max}\}, \tag{6}$$

$$\text{where } c_a(\tilde{x}_a, x_a, b_a) = \Theta^{-1}(\frac{b_a}{2r\sqrt{\tilde{u}_a}} h_a(\tilde{x}_a, x_a)), \tag{7}$$

$$\text{and } 0 \le b_a \le b_a^{max}.$$

If all vehicles are cache-enabled, i.e. $\tilde{x}_a = x_a, \forall a \in A$, then

$$\begin{aligned} c_a(\tilde{x}_a, x_a, b_a) &= c_a(x_a, b_a) \\ &= \Theta^{-1}(\frac{b_a\sqrt{l_a}}{2r\sqrt{x_a t_a(x_a)}} h_a(x_a)) \\ &= k(\frac{b_a\sqrt{l_a}}{2r\sqrt{x_a t_a(x_a)}} h_a(x_a))^{-1} \\ &= \frac{\sqrt{x_a t_a(x_a)}}{1 - e^{-2rp_a x_a t_a(x_a)/l_a}} \frac{2r}{\sqrt{l_a}} \frac{k}{b_a}, \end{aligned} \tag{8}$$

We define the route communication cost $C_{i,k}(\tilde{\mathbf{x}}, \mathbf{x}, \mathbf{b})$ as the sum of the link communication costs along the route. Note that the communication model is applicable in highway and urban scenarios where there is negligible communication range overlaps at intersections and parallel roads. Our model is also applicable in two-direction roads where the car densities are the same, as implied in [34]. We justify this model simplification in Section 8.1 by validating that the model's prediction matches the simulation result. The network model for a general transportation topology is an avenue of future work.

4 Problem Formulation

In this section, we first formulate the UE state and the SO state, and discuss the necessary conditions on the traffic flows for both states. Then we present the general formulation with the objective of minimizing the total travel cost at the UE state subject to the constraints on the maximum bandwidths and the maximum possible communication cost perceived by the users.

4.1 System Optimal

From Wardrop's second principle [37], the average travel cost is minimized at system optimal. Therefore, from the system's perspective, a low *total* travel cost improves the social welfare. For example, a low average travel time or average travel distance can alleviate the traffic congestion, reduce air pollution, and be more energy efficient. The state where the total travel cost is minimized is referred to as the System Optimal state (SO). The link flow at the SO is the solution of the following minimization problem:

$$\min \ J_{sys}(\mathbf{x}) = \sum_{a\in A} x_a T_a(\mathbf{x}), \tag{9}$$

which is subject to the flow conservation constraints (1) through (3). We assume that the system cost $J_{sys}(\mathbf{x})$ is a convex function of $\mathbf{x}$. This assumption implies that the SO is the unique and local minimum of J_{sys}. We also assume that the SO occurs at an interior point of the positive orthant of $\mathbf{y}$ (i.e., $\mathbf{y}$ is strictly positive), so that all routes between any O-D pair i are used at the SO. If this is not the case, then routes with zero traveler can be simply removed from K_i without any loss of generality.

The condition of the local optimality of (9) under the flow conservation constraint (1) through (3) can be derived

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2888968, IEEE Transactions on Mobile Computing

using Lagrange multiplier. From [3], the first-order condition for the solution of the above formulation is, for all $i \in N, k \in K_i$,

$$\begin{cases} y_{i,k}(\tilde{T}_{i,k}(\mathbf{x}) - \mu_i) = 0 \\ \tilde{T}_{i,k}(\mathbf{x}) - \mu_i \geq 0 \\ \sum_{k \in K_i} y_{i,k} = q_i \\ y_{i,k} \geq 0 \end{cases}, \tag{10}$$

where μ_i is a positive Lagrange multiplier, and $\tilde{T}_{i,k}(\mathbf{x})$ denotes the marginal travel cost of route k that connects the O-D pair i, which is the sum of the marginal travel cost of all links on the route. The physical meaning of the marginal travel cost of a link is the marginal contribution of an additional user who uses the link to the total travel cost of the network. So we have

$$\frac{\partial J_{sys}}{\partial x_a} \triangleq \tilde{T}_a(\mathbf{x}) = T_a(\mathbf{x}) + \sum_{b \in A} x_b \frac{\partial T_b(\mathbf{x})}{\partial x_a},$$

$$\frac{\partial J_{sys}}{\partial y_{i,k}} \triangleq \tilde{T}_{i,k}(\mathbf{x}) = \sum_{a \in A} \delta_{i,k}(a) \tilde{T}_a(\mathbf{x}).$$

The first-order condition (10) can be interpreted as: at SO, the marginal travel costs on all routes connecting the same O-D pair are the same. Since we assume that all routes are taken at the SO, the first-order condition of the SO can be written as, for all $i \in N, k \in K_i$,

$$\sum_{a \in A} \delta_{i,k}(a) \Big(T_a(\mathbf{x}) + \sum_{b \in A} x_b \frac{\partial T_b(\mathbf{x})}{\partial x_a} \Big) = \mu_i. \tag{11}$$

Solving for $\mathbf{x}$ from (11) and the flow conservation constraints gives us the traffic flows at the SO. Note that there are many papers addressing marginal cost pricing, but they are either in different domains or using different tools (e.g. toll stations). Our main contribution here is not the marginal cost pricing. We use this method as a framework to address the more interesting issues, which include the modeling of user experience considering both the travel cost and the communication cost, and the proposal of exploiting the communication aspect to influence the users' behavior in order to eliminate the price of anarchy.

4.2 User Equilibrium

The users typically behave non-cooperatively in the transportation network, and the system may have one or more user equilibrium (UE) states, where no user can benefit by unilaterally changing routes. We refer to the routes with positive flows at UE as the *used routes*, and the routes with zero flows at UE as the *unused routes*. From the Wardrop's first principle [37], at UE, the costs of all used routes that connect the same O-D pair are the same, and the cost of any unused route is not smaller then these used routes. Therefore, in the traditional transportation network, the necessary condition for UE is given as follows: for any $i \in N$ and $k, l, m \in K_i$, where k and l are used routes and m is an unused route

$$T_{i,k}(\mathbf{x}) = T_{i,l}(\mathbf{x}) \leq T_{i,m}(\mathbf{x}). \tag{12}$$

In the vehicular communication network where the users also take into consideration the communication cost when planning their trips, the UE deviates from that in the traditional transportation network. As described in Section 3.2, the trip cost $J_{i,k}$ is a function of the travel cost and the communication cost. Therefore, (12) becomes: for all $i \in N$, for any used route $k \in K_i$, and for any unused route $m \in K_i$,

$$\begin{cases} J_{i,k}\big(T_{i,k}(\mathbf{x}), C_{i,k}(\tilde{\mathbf{x}}, \mathbf{x}, \mathbf{b})\big) = \lambda_i, \\ J_{i,m}\big(T_{i,m}(\mathbf{x}), C_{i,m}(\tilde{\mathbf{x}}, \mathbf{x}, \mathbf{b})\big) \geq \lambda_i \end{cases} \tag{13}$$

where λ_i is some positive constant for the O-D pair i.

4.3 General Formulation

The objective is to minimize the system cost at the UE state. As discussed in Section 3.2 (and described in (6)), the communication cost is upper-bounded by the maximum possible value of the users' perceived cost, and is affected by the physical layer limitations, such as the bandwidth allocation. The budget for travel cost is intrinsically modeled in the trip cost. If the travel cost of a route is too high, which results in a high route cost, the user may simply choose another route. It is also implied that any user entering the network already has a destination, therefore the user will have to travel no matter what the travel cost is. If any future survey or experiment in the civil engineering field gives a specific user experience model (e.g. budget on travel cost when given the communication cost), similar analysis can be done using the framework presented below.

Under these constraints, we can adjust the communication related parameters with the objective of minimizing the system cost at the UE state, as shown in formulation (14) below.

$$\min_{\mathbf{x},\mathbf{b}} \ \mathbf{x}^T \mathbf{T}(\mathbf{x}) \tag{14a}$$

$$\text{s.t. } \mathbf{J}_i(\mathbf{T}, \mathbf{C}) - \mathbf{\Lambda}_i = \mathbf{0}_{|K_i|}, \forall i \in N \tag{14b}$$

$$\mathbf{\Lambda}_i \geq \mathbf{0}_{|K_i|}, \forall i \in N \tag{14c}$$

$$\mathbf{C}(\tilde{\mathbf{x}}, \mathbf{x}, \mathbf{b}) = \min\{\mathbf{c}(\mathbf{x}^c, \mathbf{x}, \mathbf{b}), \mathbf{C}^{max}\} \tag{14d}$$

$$\mathbf{0} \leq \mathbf{b} \leq \mathbf{b}^{max}, \tag{14e}$$

where (14b) and (14c) guarantee that the objective (14a) is minimized over the flows at the UE state, (14d) caps the communication cost by the maximum possible value of the user's perceived communication cost, and (14e) limits the bandwidth allocation according to the budget of the system operator.

5 Primary Optimization: SO-UE Matching

In this section, we first design a technique that can drive the flows at UE to match the flows at SO based on a necessary condition on the communication cost function, assuming that the trip cost function takes on a specific form. Then we analyze if the UE-SO matching is achievable under the communication constraints (14d) and (14e). We present a case study to demonstrate the achievability of the UE-SO matching. Lastly, we discuss the uniqueness of the UE state in the vehicular communication networks. Note that the parameters in the examples in this section are set in a way such that the corresponding figures are easier to read.

5.1 Necessary Condition of SO-UE Matching

If the communication cost is factored into the route choices, there exists a new dimension in network management, which provides the opportunity for pushing the UE closer to, and even the same as, the SO, as indicated by the arrow in Fig 1. The system operator can thus adjust the communication network related parameters in such a way that the traffic flows UE match the traffic flows at SO. Denote the flows and the cache-enabled flows at SO by $\mathbf{x}^{SO}$ and $\tilde{\mathbf{x}}^{SO}$ respectively. From (11) and (13), in order to drive the flows at the UE state to match the flows at the SO state, it is necessary that for every used route k and every unused route m that connects the O-D pair i,

$$\begin{aligned} J_{i,k}(\mathbf{T}(\mathbf{x}^{SO}), \mathbf{C}(\tilde{\mathbf{x}}^{SO}, \mathbf{x}^{SO}, \mathbf{b})) = \lambda_i, \\ J_{i,m}(\mathbf{T}(\mathbf{x}^{SO}), \mathbf{C}(\tilde{\mathbf{x}}^{SO}, \mathbf{x}^{SO}, \mathbf{b})) \geq \lambda_i, \end{aligned} \tag{15}$$

where λ_i is some positive constant for the O-D pair i.

The necessary condition (15) is also sufficient if there is only one UE state, which can provide a guideline on how the system operator can manage the communication network to achieve the SO. The uniqueness of the UE state is discussed in Section 5.4. The necessary condition (15) decouples the traffic flow control of the transportation network and the management of the communication network. If one can solve the SO using a convex optimization solver, the resulting link traffic flows and travel costs can be substituted into the trip cost $J_{i,k}(T_{i,k}, C_{i,k})$, and the numerical values of the communication costs of every routes are obtained. Then, the operator can adjust the trip cost by tuning the communication network related parameters, such as bandwidth and data price, so that the necessary condition (15) is satisfied. We call such technique as User-System Equilibrium (USE). USE guarantees that the social welfare is maximized at equilibrium if there is only one UE state, even if the users behave non-cooperatively.

We illustrate this technique via an example, where we assume, for illustrative simplicity, that the trip cost is a linear scalarization of the travel cost and the communication cost. Here, we have two objectives: lower the travel cost and lower the communication cost, and linear scalarization is often used in multi-objective optimization problems. In fact, if any future survey or experiment in the civil engineering field gives a different user experience model, similar analysis can be done using the same framework presented in this section. We denote the weight towards the traffic cost by $\alpha \in (0, 1)$, which reflects the tradeoff between the travel cost and the communication cost. If the user's profile can be gathered in real-time, the weight towards travel cost α can be customized to incorporate users preference. If a vehicle does not support data caching and forwarding, α can be set to 1. For ease of discussion, we assume all users have the same preference and α takes on a positive value, which implies all vehicles support data caching and forwarding. Then the route cost is given by:

$$J_{i,k}(\mathbf{T}, \mathbf{C}) = \alpha T_{i,k}(\mathbf{x}) + (1-\alpha) C_{i,k}(\mathbf{x}, \mathbf{b}). \tag{16}$$

Then, for any used route $k \in N_i$ and any unused route $m \in N_i$ the communication cost at SO should satisfy

$$\begin{aligned} C_{i,k}(\mathbf{x}^{SO}, \mathbf{b}) &= \frac{\lambda_i}{1-\alpha} - \frac{1}{1-\alpha} T_{i,k}(\mathbf{x}^{SO}), \\ C_{i,m}(\mathbf{x}^{SO}, \mathbf{b}) &\geq \frac{\lambda_i}{1-\alpha} - \frac{1}{1-\alpha} T_{i,m}(\mathbf{x}^{SO}). \end{aligned} \tag{17}$$

Note that the first term in the RHS of (17) can take on any positive value, and is the same for all routes that connects the O-D pair i. Therefore, only the difference between the routes' communication costs is relevant. For any pair of routes $k, l \in K_i$, define

$$\begin{aligned} \Delta_{k,l} C_i(\mathbf{x}, \mathbf{b}) &:= C_{i,k}(\mathbf{x}, \mathbf{b}) - C_{i,l}(\mathbf{x}, \mathbf{b}), \\ \Delta_{k,l} T_i(\mathbf{x}) &:= T_{i,k}(\mathbf{x}) - T_{i,l}(\mathbf{x}). \end{aligned}$$

The necessary condition for UE-SO matching becomes: for all O-D pair $i \in N$, and all used routes $k, l \in K_i$, and all unused routes $m \in K_i$,

$$\Delta_{k,l} C_i(\mathbf{x}^{SO}, \mathbf{b}) = \frac{\alpha}{1-\alpha} \Delta_{l,k} T_i(\mathbf{x}^{SO}) \tag{18a}$$

$$\Delta_{m,k} C_i(\mathbf{x}^{SO}, \mathbf{b}) \geq \frac{\alpha}{1-\alpha} \Delta_{k,m} T_i(\mathbf{x}^{SO}) \tag{18b}$$

We refer to Eq. (18) as the UE-SO Overlapping condition (USO condition). As aforementioned, we can use the USE technique to adjust the communication related parameters to drive the flows at UE to match the flows at SO. First, $\mathbf{x}^{SO}$ needs to be solved. Then, (18) is applied to solve for the bandwidth $\mathbf{b}$. However, if certain criteria are not satisfied, the USO condition (18) is not achievable, and thus the USE technique is not applicable. In this case, the UE state is impossible to match the SO state under the marginal cost pricing framework, which is discussed in detail in the next subsection. Note that other road pricing framework my be considered in this case, for example, demand-independent tolls proposed in [11].

5.2 Achievability of USO Condition

In this subsection, we study when the USO condition can be achieved under the communication cost constraints (14d) and (14e), assuming that the system operator can change the bandwidth allocation. The USO condition (18) is not always achievable, as the communication cost of route $k \in K_i$ is bounded above by the maximum possible value of users' perceived communication cost $\mathbf{C}_{i,k}^{max}$, and is bounded below by $C_{i,k}(\mathbf{x}^{SO}, \mathbf{b}^{max})$ due to the constraint on the budget of the bandwidth allocation. Without loss of generality, we consider the communication cost of any used route $k \in K_i$, the communication cost of any unused route $m \in K_i$, and the communication cost of a specific used route $k^* \in K_i$. For any route $l \in K_i$, the four points on the $(\mathrm{C_{i,l^*}}, \mathrm{C_{i,l}})$ plane:

$$\begin{cases} \left(C_{i,k^*}(\mathbf{x}^{SO}, \mathbf{b}^{max}),\ C_{i,l}(\mathbf{x}^{SO}, \mathbf{b}^{max})\right) \\ \left(C_{i,k^*}^{max},\ C_{i,l}(\mathbf{x}^{SO}, \mathbf{b}^{max})\right) \\ \left(C_{i,k^*}^{max},\ C_{i,l}^{max}\right) \\ \left(C_{i,k^*}(\mathbf{x}^{SO}, \mathbf{b}^{max}),\ C_{i,l}^{max}\right) \end{cases} \tag{19}$$

form a rectangle, within which the communication costs are feasible. We refer to this region as the feasibility region hereinafter. If the line defined by (18a) on the $(\mathrm{C_{i,k^*}}, \mathrm{C_{i,k}})$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2888968, IEEE Transactions on Mobile Computing

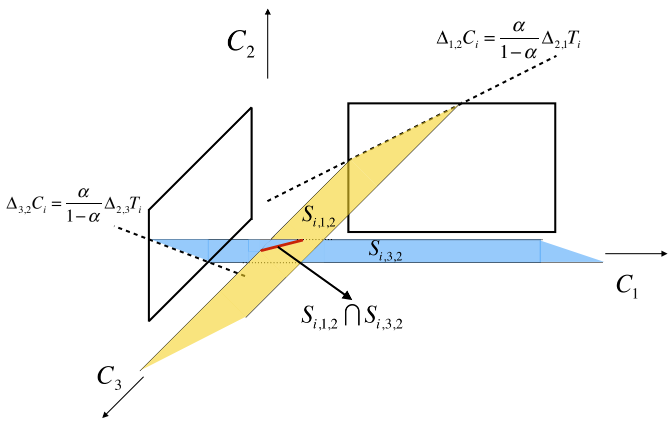

Fig. 3: Example of non-empty intersection of two feasibility sets. There are three routes connecting a single O-D pair, each of which consists of only one link.

plane passes through this feasibility region, and if the plane defined by (18b) on the $(\mathrm{C_{i,k^*}}, \mathrm{C_{i,m}})$ plane intersects with this feasibility region, one can find a bandwidth allocation scheme to satisfy the USO condition for the pair of routes $k, k^* \in K_i$ and $m, k^* \in K_i$. Denote the feasibility set S_{i,l,k^*} as the set of points that satisfy the USO condition and also lie in the feasibility region for the pair of route $l, k^* \in K_i$. One can find the bandwidth allocation scheme for the whole network that achieves the UE-SO matching if and only if the intersection of the feasibility sets of all (l, k^*) pairs is not empty, i.e. there exists a $k^* \in K_i$ such that,

$$\bigcap_{l \in K_i} S_{i,l,k^*} \neq \varnothing, \ \forall i \in N, \tag{20}$$

This is illustrated in the example shown in Fig. 3, where there are three routes (indexed by 1, 2, and 3) connecting a single O-D pair. The USO condition is plotted as the dashed lines. The feasibility regions of the route pair $1, 2$ and $3, 2$ are represented respectively by the rectangles on the $\mathrm{C_1}$-$\mathrm{C_2}$ plane and on the $\mathrm{C_3}$-$\mathrm{C_2}$ plane. The feasibility set for route pair $1, 2$ and $3, 2$ are represented respectively by the yellow plane and the blue plane. The intersection of the feasibility sets is non-empty (the red line in Fig. 3), which means there exists at least one bandwidth allocation scheme that satisfies the USO condition (5.16) and the communication cost constraints (14d) and (14e).

5.3 Case Study

Consider a single O-D pair (with trip rate 6000/h) connected by two routes, each of which consists of only one link (so they can be indexed by 1 and 2 respectively). The first road has length 1000 meters and two lanes. The second road has length 500 meters and one lane. The transmission range, and the caching ratio are 50 meters and 0.01 respectively. We adopt the travel time as the travel cost. The travel time is assumed to follow the Bureau of Public Roads (BPR) function $T(x) = \frac{l}{v_{max}}(1 + \gamma(\frac{x}{P})^{\beta})$, where v_{max} is the speed limit, and P is the capacity of a link. This BPR function is widely used in civil engineering [38]. We will use the same empirical parameters as in [38] hereinafter, unless specified otherwise. Specifically, we set $v_{max} = 35$mph (15.6464m/s), $\gamma = 0.2$, and $\beta = 10$. In order to simulate traffic congestion and avoid corner cases in section 8, we set P = 1500/h per lane. Therefore, the travel cost of route 1 and route 2 are $T_1(x_1) = \frac{1000}{15.6464}(1 + 0.2(\frac{x_1}{3000})^{10})$ and $T_2(x_2) = \frac{500}{15.6464}(1 + 0.2(\frac{x_2}{1500})^{10})$ respectively. For simplicity, we set $k = 1$, thus the communication costs are given by

$$C_1 = \frac{3.1623\sqrt{x_1 T_1(x_1)}}{1 - e^{-x_1 T_1(x_1)/1000}} \frac{1}{b_1}, \tag{21}$$

$$C_2 = \frac{4.4721\sqrt{x_2 T_2(x_2)}}{1 - e^{-x_2 T_2(x_2)/500}} \frac{1}{b_2}. \tag{22}$$

The flows at SO are $x_1 = 3904.3/h, x_2 = 2095.7/h$, which can be computed by any convex programming method. The communication costs at SO are

$$C_1 = \frac{51.2410}{1 - e^{-0.2626}} \frac{1}{b_1}, \tag{23}$$

$$C_2 = \frac{49.8038}{1 - e^{-0.2480}} \frac{1}{b_2}. \tag{24}$$

If we set $\alpha = 0.2$, the USO condition (16) can be written as

$$C_1 - C_2 = -7.2628. \tag{25}$$

Fig. 4 shows how the upper bound on the bandwidth and the upper bound on the communication cost affect the feasibility set. The blue dashed line represents the USO condition. In the three cases shown in Fig. 4, only when $\mathbf{b}^{max} = [4\ 4]^T$ and $\mathbf{C}^{max} = [150\ 150]^T$ (the middle black rectangle, referred to as case 2 hereinafter), the USO condition passes through the feasibility region, so the feasibility set is non-empty. In the red feasibility region (case 1), the bandwidth constraint on route 1 is the bottleneck, and the communication cost cap on route 2 is too low. Although the maximum value of the users' perceived communication cost cannot be modified, the system operator can, if physically possible, increase the maximum bandwidth allocation on route 1 so that the left edge of the red rectangle can intersect with the USO condition. Similar analysis can be applied to the green feasibility region (case 3), where the bandwidth constraint on the route 2 is the bottleneck, and the communication cost cap on route 1 is too low.

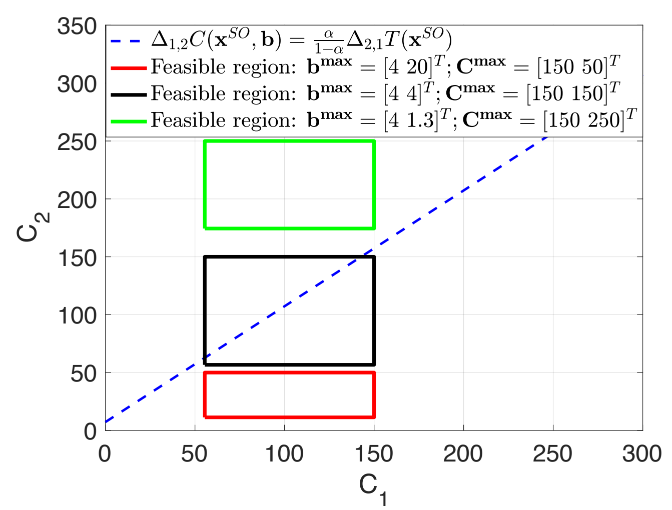

Fig. 4: USO condition and feasibility region.

If we solve the formulation (14) by substituting the parameters of the above three cases, the optimal flow solutions at UE and the corresponding minimum system cost achievable are given in Table 2. Since the feasibility set in the black rectangle is non-empty, one can find a bandwidth allocation scheme to drive the flows at UE to match the flows at SO, as shown in the third column in Table 2, where the system cost is shown to achieve its minimum. The other two cases have empty feasibility sets, therefore the flows at UE cannot be driven to match the flows at SO, and the system cost does not achieve the minimum possible value.

Case	1	2	3
x_1	4038.3	3904.3 (matches SO)	3810.7
x_2	1961.7	2095.7 (matches SO)	2189.3
$J_{sys}(\mathbf{x})$	420.14	386.58 (minimum)	405.57

TABLE 2: Flow solution that minimizes the system cost under corresponding cases.

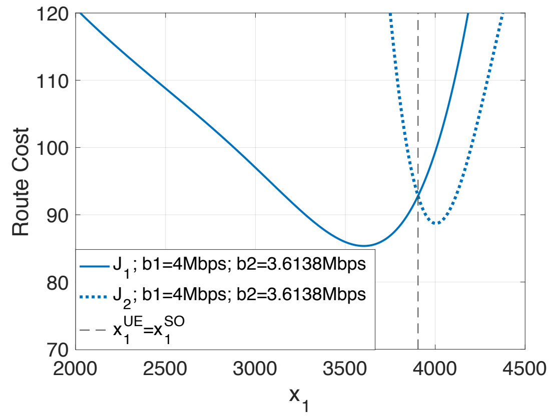

Fig. 5: In case 3, trip costs intersect at a single fixed point. USE technique guarantees that the UE state matches the SO state, as indicated by the black dashed line.

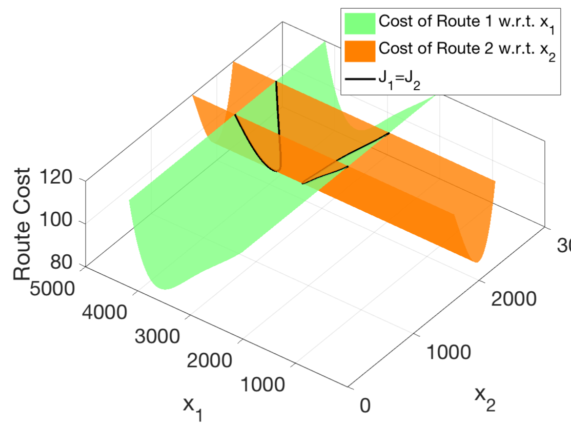

Fig. 6: Route 1 cost w.r.t. x_1 intersects with route 2 cost w.r.t x_2.

5.4 Multiple UE States

It is known that there exists a single UE state in the traditional transportation network if the following assumption holds [3]:

Assumption 1: The travel cost of each link depends on the flow along that link only, and is monotonically increasing w.r.t. the link flow, i.e., $\forall a, b \in A$,

$$\begin{cases} \frac{\partial T_a(x_a)}{\partial x_b} = 0, \text{ if } b \neq a \\ \frac{\partial T_a(x_a)}{\partial x_a} > 0 \end{cases}.$$

However, the additional communication cost in the vehicular communication networks will change the users' behavior. As a result, the system may have multiple UE states. If multiple UE states exist, it is possible that the desired UE state (i.e. the UE state that matches the SO state) is not manifested in the network. In this case, the system operator may need to adjust the network parameters to move the undesired UE state to the desired UE state. On the other hand, if there exists a unique UE state, the necessary condition in Eq. (15) is also sufficient, and the desired UE state is guaranteed to match the SO state if the USE technique is used and the condition (20) holds. Fig. 5 shows the change of the trip costs w.r.t. the flow along link 1 under the configuration of case 2. Note that in Fig. 5, the trip costs intersect at only one fixed point, which corresponds to the unique UE state according to the Wardrop's first principle [37]. The USE technique is used to configure the communication cost, therefore the UE state matches the SO state, as indicated by the black dashed line in Fig. 5. Fig. 6 shows the intersection of route 1 cost w.r.t. x_1 and route 2 cost w.r.t x_2. The projection of the trip cost intersection on the x_1-x_2 plane (black curves in Fig. 7) shows the fixed points under all possible trip rates, where the green points represent the UE states under corresponding trip rates. It is shown in Fig. 7 that the number of UE states depends on the trip rate.

6 Secondary Optimization: Bandwidth Allocation

In this section, we first present a general formulation of the secondary optimization (the system cost minimization in (14) can be regarded as the primary optimization). Then we propose a secondary objective of minimizing the total bandwidth allocation and apply this objective to case 2 discussed in Section 5.3. We present the secondary optimization in order to make the discussion more complete. The solution to the secondary optimization is not obvious, nor is the convexity of the subproblems of the secondary optimization (see Appendix for the proof). Our discussion provides a detailed method to deal with the additional degree of freedom in the system, which is useful from the practical point of view.

6.1 General formulation of the secondary optimization

In practice, we only obtain the relationship among the communication costs by applying the USE technique. Assuming that the trip cost is the weighted sum of the travel cost and the communication cost, the USE technique gives the difference between the communication cost of route pairs, as shown in (16). The numerical values of the communication costs under the UE-SO matching is not yet determined and the method of tuning the communication network related parameters (in this case, the bandwidth allocation) in order to achieve such cost values is not yet specified. In fact, there may exist multiple bandwidth allocation schemes if the system has a non-empty intersection of the feasibility sets. This provides the system operators with the opportunity to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2888968, IEEE Transactions on Mobile Computing

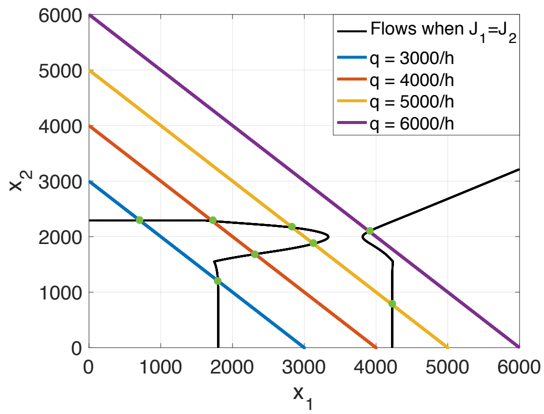

Fig. 7: Projection of the trip cost intersection in Fig. 6 on the x_1-x_2 plane. Each trip rate corresponds to a single UE state indicated by the green points.

conduct certain secondary optimization according to their interests, i.e.

$$\begin{aligned} \text{min./max. } & f(\mathbf{x}^{SO}, \mathbf{b}) \\ \text{s.t. } & \Delta_{k,l} C_i(\mathbf{x}^{SO}, \mathbf{b}) = \frac{\alpha}{1-\alpha} \Delta_{l,k} T_i(\mathbf{x}^{SO}) \\ & \mathbf{C}(\mathbf{x}^{SO}, \mathbf{b}) = min\{\mathbf{c}(\mathbf{x}^{SO}, \mathbf{b}), \mathbf{C}^{max}\} \\ & \mathbf{0}_{|A|} \le \mathbf{b} \le \mathbf{b}^{max}. \end{aligned} \tag{26}$$

6.2 Minimize the total bandwidth allocation

The objective function of the secondary optimization (26) is defined by the system operators according to their interests. Due to the limited budget on the deployment of the vehicular communication network, we propose that the total bandwidth allocated to the network should be minimized. The secondary optimization problem can be formulated as, for all $i \in N$,

$$\text{min. } \sum_{a \in A} b_a \tag{27a}$$

$$\text{s.t. } \Delta_{k,l} C_i(\mathbf{x}^{SO}, \mathbf{b}) = \frac{\alpha}{1-\alpha} \Delta_{l,k} T_i(\mathbf{x}^{SO}),\ \forall k, l \in K_i \tag{27b}$$

$$C_a(x_a^{SO}, b_a) = \min\{c_a(x_a^{SO}, b_a), C_a^{max}\},\ \forall a \in A \tag{27c}$$

$$0 \le b_a \le b_a^{max},\ \forall a \in A. \tag{27d}$$

From (27c), the optimal communication cost of link a can be either of the following two possibilities: (i) $C_a(x_a^{SO}, b_a)$, given that the bandwidth allocation does not push the communication cost to exceed the cap C_a^{max}; (ii) C_a^{max}, if we force the bandwidth allocation of link a to be zero. We can solve the secondary optimization problem for each permutation of these possibilities over all links and choose the best solution. Note that given one such permutation, the secondary optimization (27) can be transformed to an equivalent convex problem, which can be solved efficiently. The equivalent convex formulation is shown in the appendix.

Consider case 2 in Section 5.3 as an example, the secondary optimization can be written as:

$$\begin{aligned} \text{min. } & b_1 + b_2 \\ \text{s.t. } & C_1 - C_2 = -7.2628 \\ & C_1 = \min\{\frac{51.2410}{1 - e^{-0.2626}} b_1^{-1}, 150\} \\ & C_2 = \min\{\frac{49.8038}{1 - e^{-0.2480}} b_2^{-1}, 150\} \\ & 0 \le b_1 \le 4 \\ & 0 \le b_2 \le 4. \end{aligned}$$

According to the analysis in Section 5, there is only one feasibility set which is non-empty. We consider four possible bandwidth allocation schemes, and analyze each of them to obtain the final solution. Note that whenever the communication cost of a link equals the maximum users' perceived communication cost, the bandwidth allocated to this link can be decreased to zero.

Possibility 1: $C_1 = C_1^{max} = 150$ and $C_2 = C_2^{max} = 150$. The bandwidth allocation on route 1 and on route 2 can be both pushed down to zero. Obviously, this is not the solution, because $C_1^{max} - C_2^{max} = 0 > -7.2628$, which violates the USO condition.

Possibility 2: $C_1 = C_1^{max} = 150$ and $C_2 = \frac{49.8038}{1-e^{-0.2480}} b_2^{-1}$. In order to satisfy the USO condition, we need $C_2 = 150 + 7.2628 = 157.2628 > C_2^{max} = 150$. This violates the communication cost cap constraint on route 2.

Possibility 3: $C_1 = \frac{51.2410}{1-e^{-0.2626}} b_1^{-1}$ and $C_2 = C_2^{max} = 150$. In this case, the bandwidth allocation on route 2 can be pushed down to zero. In order to satisfy the USO condition, we need $C_1 = 150 - 7.2628 = 142.7372$, which gives $b_1 = 1.3513$ and $b_2 = 0$.

Possibility 4: $C_1 = \frac{51.2410}{1-e^{-0.2626}} b_1^{-1} < C_1^{max}$ and $C_2 = \frac{49.8038}{1-e^{-0.2480}} b_2^{-1} < C_2^{max}$. In this case, we can bound below the bandwidth allocation on route 1 and on route 2 by $\frac{51.2410}{150(1-e^{-0.2626})}$ and $\frac{49.8038}{150(1-e^{-0.2480})}$ respectively, so the secondary optimization becomes:

$$\begin{aligned} \text{min. } & b_1 + b_2 \\ \text{s.t. } & \frac{51.2410}{1 - e^{-0.2626}} b_1^{-1} - \frac{49.8038}{1 - e^{-0.2480}} b_2^{-1} = -7.2628 \quad (28) \\ & \frac{51.2410}{150(1 - e^{-0.2626})} \le b_1 \le 4 \\ & \frac{49.8038}{150(1 - e^{-0.2480})} \le b_2 \le 4. \end{aligned}$$

Denote the coefficient of b_1^{-1} and b_2^{-1} in (28) by w_1 and w_2 respectively. The secondary optimization is equivalent to

$$\text{min. } b_1 + \frac{w_2 b_1}{w_1 + 7.2628 b_1} \tag{29}$$

$$\text{s.t. } \frac{w_1}{150} \le b_1 \le 4 \tag{30}$$

$$\frac{w_2}{150} \le \frac{w_2 b_1}{w_1 + 7.2628 b_1} \le 4. \tag{31}$$

The first derivative of the objective (29) is $1 + \frac{w_1 w_2}{(w_1 + 7.2628 b_1)^2}$, which is strictly greater than 0. Therefore, the objective function (29) is increasing with b_1. The minimum b_1 that satisfies the constraints (30) and (31) is $b_1 \approx 1.5961$, which gives $b_2 \approx 1.5876$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2888968, IEEE Transactions on Mobile Computing

We note that *possibility 3* gives a lower total bandwidth allocation than *possibility 4*. Therefore, the best bandwidth allocation scheme is $b_1 = 1.3513$ and $b_2 = 0$.

7 CASE STUDY

In this section, we conduct a comprehensive case study, where we apply the USE technique and the secondary optimization on a real world transportation network. We assume the trip cost is the weighted sum of the travel cost and the communication cost. The weight towards the travel cost is set to 0.6. To get the exact weight, comprehensive surveys need to be conducted, which is beyond the scope of our study. The trip cost function is given by

$$J_{i,k} = 0.6T_{i,k} + 0.4C_{i,k}.$$

From (16), it is necessary that for every route k that connects the O-D pair i,

$$\Delta_{k,l}C_i(\mathbf{x}^{SO}, \mathbf{b}) = 0.25\Delta_{l,k}T_i(\mathbf{x}^{SO}). \tag{32}$$

Fig. 8 is obtained from Google Maps, which shows part of the transportation network in the Capital District around Albany, NY. The network under consideration is the grid consisting of the grey and blue links. To simplify the calculation, we assume that there are two O-D pairs in this network: drivers from Latham (node A) either go to Downtown Albany (node C) or Delmar (node E). Therefore, all traffic on link 1 and link 4 is from node A. We also assume that the drivers will only use the links that are indexed in Fig. 8. The links marked as blue (links 1, 2, and 5, denoted by 1-2-5) form a possible route from node A to node E. There are two other routes between the O-D pair (A,E): 4-6, and 4-3-5. Similarly, there are two routes between the O-D pair (A,C): 1-2 and 4-3. The lengths of the links are approximately 6400m, 9700m, 11200m, 10600m, 8400m, and 6400m respectively. The transmission range and the caching ratio are assumed to be 100m and 0.005 respectively.

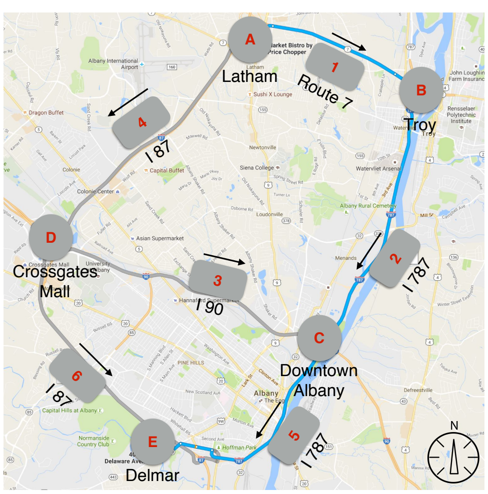

Fig. 8: A sample network in the Capital District, Albany, NY. The link index is in the rectangular box next to each link, and the arrow indicates the direction of each link. The intersections of the links are represented by the grey nodes A through E.

We obtain the traffic flow data from the NYS Traffic Data Viewer [39]. The Traffic Data Viewer (TDV) is a GIS web application for viewing the annual average daily traffic.

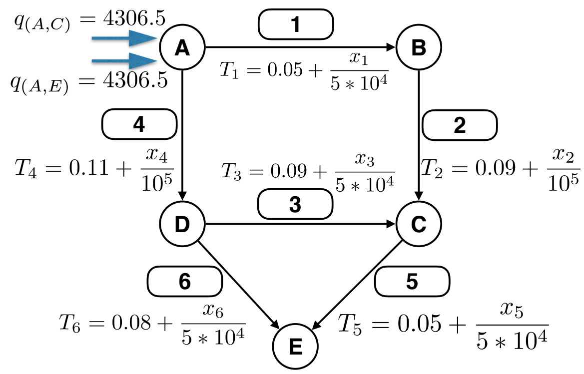

Fig. 9: Graph representation of the network topology of Fig. 8. The link travel costs in terms of the link flows are written next to each link.

According to the data that the TDV averages over several weeks in Spring 2005 and 2006, the traffic flow on link 1 and link 4 are 3786/h, and 4827/h respectively. There are theoretical models and practical methods to estimate the trip rate, but for ease of demonstration, we assume that half of the drivers from node A are traveling to node C, and the other half are traveling to node E. So the trip rates for O-D pair (A,E) and (A,C) are both 4306.5/h. Fig. 9 shows a graph representation of the network topology in Fig. 8. We assume that the link travel cost depends on the traffic flow only on that link.

We combine the first-order condition (11) with the flow conservation constraints to solve for the SO:

$$\begin{cases} (0.05 + \frac{2x_1}{5*10^4}) + (0.09 + \frac{2x_2}{10^5}) = \mu_{(A,C)} \\ (0.11 + \frac{2x_4}{10^5}) + (0.09 + \frac{2x_3}{5*10^4}) = \mu_{(A,C)} \\ (0.11 + \frac{2x_4}{10^5}) + (0.08 + \frac{2x_6}{5*10^4}) = \mu_{(A,E)} \\ (0.11 + \frac{2x_4}{10^5}) + (0.09 + \frac{2x_3}{5*10^4}) + (0.05 + \frac{2x_5}{5*10^4}) = \mu_{(A,E)} \\ (0.05 + \frac{2x_1}{5*10^4}) + (0.09 + \frac{2x_2}{10^5}) + (0.05 + \frac{2x_5}{5*10^4}) = \mu_{(A,E)} \\ x_1 + x_4 = 8613, x_5 + x_6 = 4306.5 \\ x_4 = x_3 + x_6, x_1 = x_2 \end{cases}.$$

Solving the above linear system gives

$$\begin{cases} \mathbf{x} \approx [3605 \;\; 3605 \;\; 1403 \;\; 5008 \;\; 702 \;\; 3605]^T \\ \mathbf{T} \approx [0.12 \;\; 0.13 \;\; 0.12 \;\; 0.16 \;\; 0.06 \;\; 0.15]^T \end{cases}.$$

Substituting the above solution into (32) yields

$$\begin{cases} \Delta_{1,2}C_{(A,C)} \approx 0.045 \\ \Delta_{1,2}C_{(A,E)} \approx 0.045 \\ \Delta_{1,3}C_{(A,E)} \approx 0 \end{cases}, \tag{33}$$

where route 1 and route 2 between O-D pair (A, C) are 1-2 and 4-3 respectively; route 1, route 2, and route 3 between O-D pair (A, E) are 1-2-5, 4-3-5, and 4-6 respectively. The operator can then use (33) to adjust the bandwidth allocation. Suppose the bandwidth allocation cannot exceed 2000MHz for all links and the link communication cost perceived by the users is upper bounded by 10, then the following optimization problem needs to be solved:

$$\begin{aligned}
\min.\ & b_1+b_2+b_3+b_4+b_5\\
\text{s.t. } & \Delta_{1,2}C_{(A,C)}=0.045\\
& \Delta_{1,2}C_{(A,E)}=0.045\\
& \Delta_{1,3}C_{(A,E)}=0\\
& C_{1,(A,C)}=\min\{\frac{6623}{b_1},10\}+\min\{\frac{9929}{b_2},10\}\\
& C_{2,(A,C)}=\min\{\frac{11283}{b_3},10\}+\min\{\frac{11006}{b_4},10\}\\
& C_{1,(A,E)}=C_{1,(A,C)}+\min\{\frac{8423}{b_5},10\}\\
& C_{2,(A,E)}=C_{2,(A,C)}+\min\{\frac{8423}{b_5},10\}\\
& C_{3,(A,E)}=\min\{\frac{11006}{b_4},10\}+\min\{\frac{6678}{b_6},10\}\\
& 0\le b_a\le 2000,\ a=1,2,3,4,5,6.
\end{aligned}$$

Solving the above system gives:

$$\begin{cases} b_2=b_4=b_6=0\\ b_1\approx 1144.2\text{MHz}\\ b_3\approx 1964.5\text{MHz}\\ b_5=2000\text{MHz}\end{cases}. \tag{34}$$

Therefore, allocating 1144.2MHz bandwidth to link 1, 1964.5MHz bandwidth to link 3, and 2000MHz to link 5 is the optimal bandwidth allocation policy that leads to the SO-UE matching in this network.

8 Simulation Results

In this Section, we first validate the proposed communication cost model by comparing the data throughput from the simulation with the throughput from the model. Then we consider a more realistic scenario where the V2V bandwidth can only take on certain values, and show that the system cost decreases if the bandwidth allocation is closer to the optimal allocation policy. We use Veins [40] as the vehicular network simulator. Veins is an open source framework based on the network simulator OMNeT++ [41] and the road traffic simulator SUMO[42]. We use 802.11p protocol in the simulation, which is implemented in Veins.

8.1 Communication Cost

To validate the proposed model of the communication cost, we construct a simple network that has only one road with length 1000m. All vehicles enter the system from one end and leave the system at the other end. We use the empirical BPR function introduced in Section 5.3 in the computation of the speed, which is then used to compute the traffic flow in the model. As discussed in Section 3.2, communication cost is modeled as the inverse of the throughput due to cache hit. Therefore, it is sufficient to validate that the throughput from the proposed communication model matches the throughput from the simulation. In the simulation, we record the throughput under different traffic flows, and compare it with the prediction from the proposed model. The cache hit ratio p is set to 0.05, and the communication range r is set to 360m. Other relevant parameters are shown in Table 3. For each data point (i.e. flow), we average the throughput of 5 runs. In each run, we simulate the system for 500s with a warm-up period of 50s. As shown in the simulation result in Fig. 10, the proposed communication cost model generally matches the simulated throughput. Since we use a Poisson arrival process in the simulation, it is more likely that a vehicle's request cannot reach any vehicle within the communication range when the traffic flow is relatively small. Therefore, the model is more optimistic than the simulated throughput under small traffic flows. Since the proposed model is asymptotic, the predicted throughput fits the simulated throughput better under larger traffic flows.

Parameter	Value	Parameter	Value
Transmission Power	20mW	Sensitivity	-86dBm
Content Size	128KB	Request Interval	1s

TABLE 3: Relevant Parameters

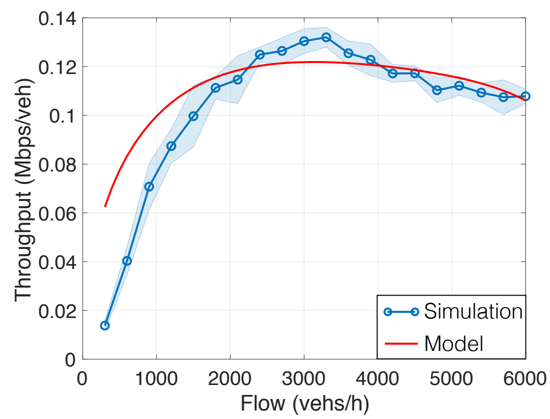

Fig. 10: Throughput from Simulation v.s. throughput from model. The shaded area denotes the standard deviation of each data point.

8.2 V2V Bandwidth Allocation

In practice, the V2V bandwidth may only take certain values in vehicular communication networks. For example, in 802.11p the V2V bandwidth can take eight different values: 3Mbps, 4.5Mbps, 6Mbps, 9Mbps, 12Mbps, 18Mbps, 24Mbps, and 27Mbps. In this section, we show that the system cost can be lowered when we change the V2V bandwidth allocation closer to the optimal value under 802.11p protocol. We use the same transportation network as in Section 5.3 (Fig. 11). When a vehicle enters the network, information on the current throughput and the current travel time of both routes is provided. Then the user chooses the route with the smaller cost. Real time throughput is computed by averaging the throughput recorded in the last 20s, and the current travel time can be obtained directly from the Veins simulator. The trip rate is set to 4000/h, and the V2V bandwidth of route 1 is set to $b_1 = 3Mbps$. After applying the USO condition, we obtain the optimal bandwidth on route 2 $b_2 = 2.89Mbps$. However, the closest value to 2.89Mbps under 802.11p protocol is 3Mbps. We measure the system cost and the traffic flow under three different bandwidth allocation policies: $\mathbf{b} = [3\ 27]^T$, $\mathbf{b} = [3\ 9]^T$, and $\mathbf{b} = [3\ 3]^T$ (with the unit of Mbps). As shown in

Fig. 12, when $b_2 = 3Mbps$, the system cost is the lowest after around 500s. Fig. 13 shows the traffic flow on route 1 under the corresponding bandwidth allocation policies.

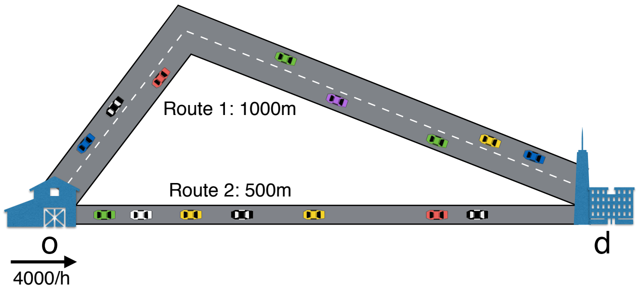

Fig. 11: Topology of the transportation network in the simulation.

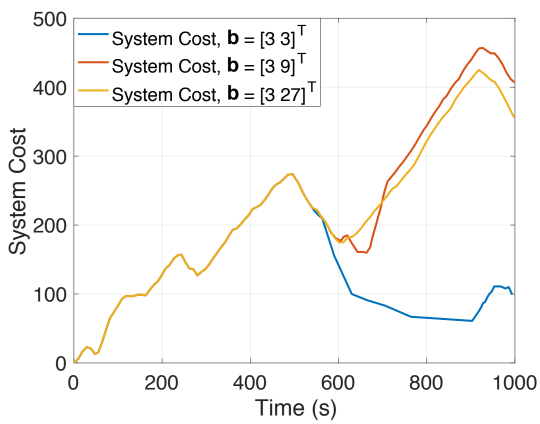

Fig. 12: System cost under different V2V bandwidth allocation policies. Bandwidth has the unit of Mbps, and is specified according to 802.11p protocol.

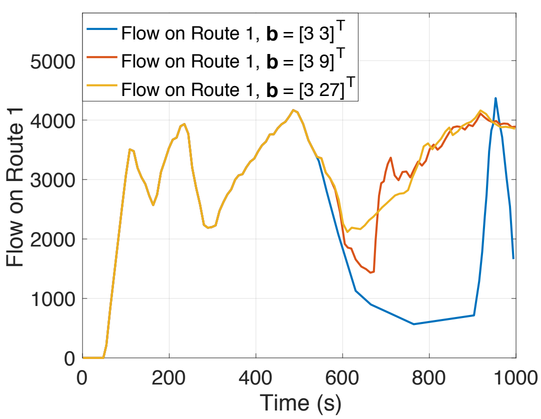

Fig. 13: Traffic flow on route 1 under different V2V bandwidth allocation policies. Bandwidth has the unit of Mbps, and is specified according to 802.11p protocol.

9 Conclusion

In this paper, we model the user trip planning when both the traffic condition and the data communication influence user trip decision. The necessary condition is derived for the SO-UE matching, which provides a guideline on how the system operator can adjust the network parameters to achieve the optimal social welfare even if the users are non-cooperative. The secondary optimization is discussed, which can be utilized according to the system operator's interests. The proposed communication cost model is validated via Veins simulation, and the simulation results show that the system cost can be lowered if the V2V bandwidth allocation is closer to the optimal allocation policy under 802.11p protocol.

References

[1] C. Jiang, Y. Chen, and K. R. Liu, "Data-driven optimal throughput analysis for route selection in cognitive vehicular networks," *Selected Areas in Communications, IEEE Journal on*, vol. 32, no. 11, pp. 2149–2162, 2014.

[2] M. J. Beckman, C. B. McGuire, and C. B. Winsten, *Studies in the Economics of Transportation*. Yale University Press, 1956.

[3] Y. Sheffi, "Urban transportation network," *Pretince Hall*, 1985.

[4] T. Roughgarden and E. Tardos, "How bad is selfish routing?" *Journal of the ACM*, vol. 49, no. 2, pp. 236–259, 2002.

[5] D. Braess, A. Nagurney, and T. Wakolbinger, "On a paradox of traffic planning," *Transportation Science*, vol. 39, no. 4, pp. 446–450, 2005.

[6] H. Youn, M. T. Gastner, and H. Jeong, "Price of anarchy in transportation networks: Efficiency and optimality control," *Physics Review Letters*, vol. 101, no. 12, p. 128701, 2008.

[7] T. Liu, A. A. Abouzeid, and A. A. Julius, "Traffic flow control in vehicular communication networks," in *American Control Conference (ACC)*, 2017, pp. 5513–5518.

[8] J. A. Gomez-Ibanez and K. A. Small, *Road pricing for congestion management: A survey of international practice*. Transportation Research Board, 1994, vol. 210.

[9] K. A. Small and J. A. Gómez-Ibáñez, "Road pricing for congestion management: the transition from theory to policy," *Transport Economics*, pp. 373–403, 1997.

[10] X. Wang, N. Xiao, L. Xie, E. Frazzoli, and D. Rus, "Analysis of price of anarchy in traffic networks with heterogeneous price-sensitivity populations," *IEEE Trans. Control Systems Technology*, vol. 23, no. 6, 2015.

[11] R. Colini-Baldeschi, M. Klimm, and M. Scarsini, "Demand-independent tolls," *arXiv preprint arXiv:1708.02737*, 2017.

[12] T. A. Manual, "Bureau of public roads," *US Department of Commerce*, 1964.

[13] J. Zhang, S. Pourazarm, C. G. Cassandras, and I. C. Paschalidis, "The price of anarchy in transportation networks by estimating user cost functions from actual traffic data," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, Dec 2016, pp. 789–794.

[14] M. W. Burris and R. M. Pendyala, "Discrete choice models of traveler participation in differential time of day pricing programs," *Transport Policy*, vol. 9, no. 3, pp. 241–251, 2002.

[15] C. Weis, K. Axhausen, R. Schlich, and R. Zbinden, "Models of mode choice and mobility tool ownership beyond 2008 fuel prices," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2157, pp. 86–94, 2010.

[16] H. M. Al-Deek, A. J. Khattak, and P. Thananjeyan, "A combined traveler behavior and system performance model with advanced traveler information systems," *Transportation Research Part A: Policy and Practice*, vol. 32, no. 7, pp. 479–493, 1998.

[17] H. X. Liu, W. Recker, and A. Chen, "Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data," *Transportation Research Part A: Policy and Practice*, vol. 38, no. 6, pp. 435–453, 2004.

[18] X. Cao and D. Chatman, "How will smart growth land-use policies affect travel? A theoretical discussion on the importance of residential sorting," *Environment and Planning B: Planning and Design*, vol. 43, no. 1, pp. 58–73, 2016.

[19] D. Levinson, "The value of advanced traveler information systems for route choice," *Transportation Research Part C: Emerging Technologies*, vol. 11, no. 1, pp. 75–87, 2003.

[20] B. Kim, M. Choi, and I. Han, "User behaviors toward mobile data services: The role of perceived fee and prior experience," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8528–8536, 2009.

[21] B. Kim, "The diffusion of mobile data services and applications: Exploring the role of habit and its antecedents," *Telecommunications Policy*, vol. 36, no. 1, pp. 69–81, 2012.

[22] S.-J. Hong, J. Y. Thong, J.-Y. Moon, and K.-Y. Tam, "Understanding the behavior of mobile data services consumers," *Information Systems Frontiers*, vol. 10, no. 4, pp. 431–445, 2008.

[23] A. A. Shaikh and H. Karjaluoto, "Making the most of information technology & systems usage: A literature review, framework and future research agenda," *Computers in Human Behavior*, vol. 49, pp. 541–566, 2015.

[24] K. Kamini and R. Kumar, "Vanet parameters and applications: A review," *Global Journal of Computer Science and Technology*, vol. 10, no. 7, 2010.

[25] J. Yao, S. S. Kanhere, and M. Hassan, "Improving qos in high-speed mobility using bandwidth maps," *IEEE Transactions on Mobile Computing*, vol. 11, no. 4, pp. 603–617, April 2012.

[26] F. D. Da Cunha, A. Boukerche, L. Villas, A. C. Viana, and A. A. Loureiro, "Data communication in vanets: a survey, challenges and applications," Ph.D. dissertation, INRIA Saclay; INRIA, 2014.

[27] S. Biswas, R. Tatchikou, and F. Dion, "Vehicle-to-vehicle wireless communication protocols for enhancing highway traffic safety," *IEEE Communications Magazine*, vol. 44, no. 1, pp. 74–82, 2006.

[28] J. Lioris, R. Pedarsani, F. Y. Tascikaraoglu, and P. Varaiya, "Platoons of connected vehicles can double throughput in urban roads," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 292–305, April 2017.

[29] T. L. Wilke, T. Patcharinee, and N. F. Maxemchuk, "A survey of inter-vehicle communication protocols and their applications," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 2, pp. 3–20, 2009.

[30] F. Malandrino, C. Casetti, C. F. Chiasserini, and M. Fiore, "Optimal content downloading in vehicular networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 7, pp. 1377–1391, July 2013.

[31] S. Lim, S. H. Chae, C. Yu, and C. R. Das, "On cache invalidation for internet-based vehicular ad hoc networks," in *Mobile Ad Hoc and Sensor Systems (MASS), 2008, 5th IEEE International Conference on*, 2008, pp. 712–717.

[32] M. Fiore, F. Mininni, C. Casetti, and C.-F. Chiasserini, "To cache or not to cache?" in *INFOCOM 2009, IEEE*, 2009, pp. 235–243.

[33] O. Attia and T. A. ElBatt, "On the role of vehicular mobility in cooperative content caching," *CoRR*, vol. abs/1203.0657, 2012.

[34] C. Bian, T. Zhao, X. Li, X. Du, and W. Yan, "Theoretical analysis on caching effects in urban vehicular ad hoc networks," *Wireless Communications and Mobile Computing*, vol. 16, pp. 1759–1772, Sept. 2016.

[35] M. Nekoui, A. Eslami, and H. Pishro-Nik, "Scaling laws for distance limited communications in vehicular ad hoc networks," in *Communications, 2008. ICC'08. IEEE International Conference on*, 2008, pp. 2253–2257.

[36] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theor.*, vol. 46, no. 2, pp. 388–404, Sep. 2006.

[37] J. G. Wardrop, "Some theoretical aspects of road traffic research." in *Institution of Civil Engineers*, vol. 1, no. 3, 1952, pp. 325–362.

[38] H. Youn, M. T. Gastner, and H. Jeong, "Price of anarchy in transportation networks: efficiency and optimality control," *Physical review letters*, vol. 101, no. 12, p. 128701, 2008.

[39] "NYS traffic data viewer," available at: https://gis3.dot.ny.gov/tdv, [Accessed Sept. 14, 2016].

[40] "Veins," http://veins.car2x.org, accessed: 2017-12-4.

[41] "Omnet++," https://www.omnetpp.org, accessed: 2017-12-4.

[42] "Sumo," http://sumo.dlr.de/index.html, accessed: 2017-12-4.