On the Economic Value of Mobile Caching

Yichen Ruan and Carlee Joe-Wong
Department of Electrical and Computer Engineering, Carnegie Mellon University, USA
Emails: {yichenr, cjoewong}@andrew.cmu.edu

***Abstract*—Recent growth in user demand for mobile data has strained mobile network infrastructure. One possible solution is to use mobile (i.e., moving) devices to supplement existing infrastructure according to users' needs at different times and locations. For instance, vehicles can be used as communication relays or computation points. However, it is unclear how much value these devices add relative to their deployment costs: they may, for instance, interfere with existing network infrastructure, limiting the potential benefits. We take the first step towards quantifying the value of this supplemental infrastructure by examining the use case of mobile caches. We consider a network operator using both mobile (e.g., vehicular) and stationary (small cell) caches, and find the optimal amount of both types of caches under time- and location-varying user demands, as a function of the cache prices. In doing so, we account for interference between users' connections to the different caches, which requires solving a non-convex optimization problem. We show that there exists a threshold price above which no vehicular caches are purchased. Moreover, as the network operator's budget increases, vehicular caching yields little additional value beyond that provided by small cell caches. These results may help network operators and cache providers find conditions under which vehicles add value to existing networks.**

***Index Terms*—caching, economics, vehicle mobility**

I. Introduction

Due to the widespread use of smart devices, recent years have witnessed rapid growth in the volume of mobile data traffic. Cisco [1] predicts that global mobile data traffic will increase sevenfold between 2017 and 2022. At the same time, applications such as augmented reality or 360-degree video streaming increasingly require low network response time. In response, operators have made efforts to move 5G network functionalities to the network edge [2], e.g., caching popular contents on small cells or user devices. These contents can then be retrieved directly from the network edge, without passing through the network core. Such practices have yielded concrete benefits for network operators, e.g., caching in small cells can reduce backhaul traffic by over 45% [3]. Despite these benefits, relying on edge infrastructure also has drawbacks. Small cells have limited range, requiring dense deployments, and cannot adapt to changes in user demand over time [4]. For instance, user demands in a business region of a city are generally much larger during the day than at night, making small cells stay idle for almost half a day. Network infrastructure with more temporal flexibility could then yield further benefits to network operators by adapting to changes in user demands.

We acknowledge funding support from NSF CNS-1751075, the David Barakat and LaVerne Owen-Barakat Fellowship, and the Ann and Martin McGuinn Graduate Fellowship.

Some recent work [5]–[10] has proposed to deploy mobile devices such as drones and vehicles as relay nodes and cache carriers in wireless networks. The physical mobility of these devices allows them to respond to temporal variations in user demands and network conditions at different locations. However, it is not clear whether the benefits of this flexibility outweigh the costs. Using vehicles or other mobile devices, for instance, may be more expensive than simply over-provisioning stationary devices to handle peak user demands at each location, particularly if user demands do not vary enough over time. Users' communication with such devices may also interfere with connectivity to existing network infrastructure. Yet it is hard to say exactly what constitutes "too expensive" or "not varied enough." In this work, we **quantify the value of mobile caching** as a first step towards assessing the value of using vehicles or other mobile devices in network infrastructure. Our results can guide network operators in determining whether and to what extent they should incorporate such devices into their networks.

Mobile devices like vehicles and drones can perform a number of network functions, ranging from caching content [10] to serving as relays [9] to collecting data in urban crowdsensing [11]. We focus on vehicular caching due to caching's known benefits in mobile networks, and vehicles' naturally high density compared to small cells or other infrastructure in an urban environment. One might, for instance, pay public buses, private car owners, or other mobility-as-a-service providers to carry cache servers; some bus systems already carry WiFi access points [12]. Our economic model, however, is general enough to cover any type of cache provider that charges the network operator that uses the caches.

Most prior works on wireless caching focus on finding the optimal caching policies, with little attempt at analyzing the competition between, and relative value of, mobile vehicular caches and other types of caching. We thus encounter several new research challenges in doing so. These involve both *finding the right model* to quantify the value of vehicular caching and *solving the resulting optimization problems*, which we show are in general non-convex. We solve these challenges to make three key research contributions:

Our first contribution lies in **modeling interactions between caching tiers**: Vehicular caching must compete *economically* with existing caching products, such as small cells. The relative costs of these different types of caching will then influence the additional value that vehicular caching brings. We quantify this effect by jointly optimizing the caching purchased from each tier. Vehicular points may also experience *physical interference* from small cells and other network traffic, which affects their ability to serve users and thus their value. Quantifying the effect of such interference between caching tiers is itself a challenging problem, and we use stochastic geometry theories to do so.

Our second research contribution is to develop a **unified framework for different network operator objectives** given the above system model. These include the rate of cache hits and the delivery rate, which quantifies the spectral efficiency of users' connections to different caching tiers. We formulate tractable optimization problems for both objectives in a general framework (*Theorems 1-3*).

Our third major research contribution lies in solving these optimization problems to **concretely quantify the value of vehicular caching**. This value depends on the distribution of user needs over time and space, which is itself hard to quantify given the many variables involved. We show that it also depends on solving for the optimal cache provisioning and provide solutions for this (non-convex) optimization problem in our work (Theorems 4-5). In particular, we will evaluate the value of vehicular caching by answering three questions:

- **Demand functions**: We solve for network operators' optimal caching demands (*Theorems 4 and 5*), and quantify the dependence of vehicle demands at each time on the prices and user needs at other times (*Corollary 1*);
- **Economic viability of vehicular caching**: We quantify the maximum price under which operators would wish to purchase vehicular caching capacities, i.e., under which the value added by vehicular caching exceeds its cost relative to other caching methods (*Corollary 2*);
- **Gain from competition**: We show that as the network operator's budget increases, vehicular caching yields little marginal value above small cell caching (*Corollary 3*), but that the optimal amounts of both vehicular and small cell caching increase linearly with the budget (*Corollary 4*).

We summarize related work in Section II and introduce our system model in Section III. We formulate the problem of optimal cache provisioning, and examine its solution's implications, in Sections IV to VI. Section VII validates and expands on our analytical findings with trace-driven numerical results, and we conclude in Section VIII.

II. Related Work

Many existing works on network caching consider a fixed caching capacity that the network operator can sell to different content providers (CPs) in order to reduce the latency of delivering content to these CPs' users. Economic techniques such as contract theory [13], auction theory [14], and game theory [15] can be used to find the optimal or equilibrium outcomes. Unlike these works, in this paper, we focus on *network operators purchasing cache devices*, whose capacity may later be sold to CPs.

Many works on wireless caching attempt to design optimal content placement schemes. Popular methods include small cell caching [16], [17], device-to-device (D2D) cache sharing [18], [19], mobility-aware caching [19], and content sharing in vehicular ad hoc networks [20]. Comparing these methods reveals that caching in different locations of the wireless network can result in different levels of economic gains [21], but these benefits are not rigorously analyzed. To do so here, we make use of the content placement, device mobility, and wireless connectivity models introduced in these works.

Early works on vehicular caching view vehicles as moving relay nodes that connect user devices to base stations [22], [23]. Later works have considered optimizing user connectivity to vehicles that host caches themselves [10], [24], [25]. In [5], [7], the optimal content placement policy is studied for a 2-tier network with cache-enabled vehicular points and macro cells. Similar models are studied for specific applications, such as the delivery of delay-tolerant contents [6], and video streaming [7]. These works focus only on vehicles and do not consider the existence of other caching methods, e.g., small cells. In contrast, we will explore how vehicular, and more generally mobile, caching adds value to the existing caching methods.

III. System Model

We first show the overall system (Figure 1) and our user and vehicle mobility model in Section III-A, and then describe our models for the content allocation policy (Section III-B) and users' connectivity to the caching devices (Section III-C).

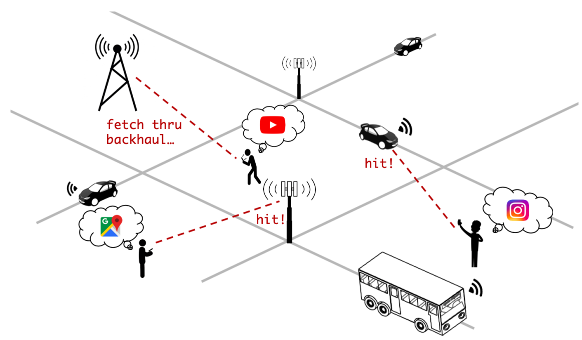

Fig. 1. A heterogeneous network composed of vehicular caching, small cells and macro cells. User requests for contents are served by either the vehicle or small cell caches, or by macro cells that fetch data through the backhaul.

A. Content Caching Market

We consider a network operator, e.g., an Internet service providers (ISP) or content delivery network (CDN) provider, that wishes to utilize caching capacity from a variety of cache providers. We consider two tiers of caching products: small cell caching (which has fixed locations) and vehicular caching. These products can come from different providers. For example, equipment manufacturers may sell cache capacity at small cells, while ride-sharing networks can sell caching capacity in their vehicles. The network operator subscribes to a combination of these two products to maximize its profit, i.e., the benefit from caching, less the payment to the cache providers. To focus on the value that vehicular caching adds to mobile networks and not on the price competition between the small cell and vehicular providers, we assume the prices that the network operator pays for each type of caching are given. For simplicity, we do not consider D2D caching as it is not yet widely deployed. However, our framework can be easily extended to include more caching tiers.

We assume that network operators purchase caching capacity in terms of the expected number of devices in a unit area, i.e., they decide the intensity of small cells L_s and vehicles L_v to be deployed in each region of their networks. We use a stochastic model for the number of vehicles and small cells available in each region and time interval to abstract away from individual device mobility. Indeed, these devices may not be fully controlled by network operators: for instance, taxi vehicles may follow varied mobility patterns. Small cell caches may become inaccessible due to network congestion or other demands. We also assume an existing macro cell tier with fixed, exogenous intensity L_m, leading to a 3-tier heterogeneous network as in [26]. The macro cell can access all content through the network backhaul. Devices of the same tier share the same configuration, i.e., they have the same transmit power p_k, cache capacity N_k and price P_k, where $k \in \mathcal{N} = \{v, s, m\}$ represents vehicles, small cells and macro cells respectively. This price represents the amount that operators need to pay vehicle or small cell owners to cache content for them and may vary with time and location.

To address temporal and spatial dynamics, we divide the decision space into T time slots and D regions. Inside each $(t, d) \in [T] \times [D]$, we assume the locations of all caching points and users follow a homogeneous Poisson point process with constant intensity $L_k(t, d), k \in \mathcal{N} \cup \{u(\text{users})\}$ [19], [27]. Since macro cells and small cells are stationary, their intensity matrices are fixed over time, i.e. $\forall t_1, t_2, L_k(t_1, d) = L_k(t_2, d), k \in \{m, s\}$. The intensity of users $L_u(t, d)$ at each time t and location d is exogenous, and we assume that it is not affected by the cache intensities. The (exogenous) prices $P_k(t, d)$ can also vary with the time t and region d. The mobility of mobile users and vehicles is reflected in the change of intensities across time and space.

B. Content Placement Policies

The value of vehicular caching depends on how content is cached at both the vehcles and small cells. Contents are assumed to be cached as chunks with the same size [7]. Each chunk is associated with an exogenous preference (probability of being requested) $f_i, i = 1, 2, \dots, N$, where N is the size of the content catalog. Without loss of generality, we assume $f_1 \geq f_2 \geq \dots \geq f_N$ following a power-law distribution as in [28]. Chunks from the same content source may have different preferences, e.g., some segments of a video may be more popular than others [7]. The preference, and thus the indices i, can also depend on (t, d), e.g., news videos might be more popular in the morning. For clarity, we generally omit the dependence on (t, d) in our notation.

The content placement policy is modeled in terms of the caching probability. In tier k, the content i is cached with probability $H_{k,i}(t, d)$. All chunks are accessible to the macro cell tier, i.e. $N_m = N$ and $H_{m,i}(t, d) \equiv 1$. By choosing the values of H, we can model different placement policies, provided that the cache does not exceed its capacity, i.e., $\sum_{i=1}^{N} H_{k,i} \leq N_k$. We will consider policies of the form

$$H_{k,i}(t, d) = \begin{cases} u_k \triangleq \frac{N_k}{\mu_k N}, & 1 \leq i \leq \mu_k N \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where μ_k is an exogenous parameter satisfying $\frac{N_k}{N} \leq \mu_k \leq 1$. A larger μ_k means that less popular content (with higher index i) is more likely to be cached. We make the reasonable assumption that $N_v \leq N_s$ and $\mu_v \leq \mu_s$, i.e., vehicles cache fewer chunks than small cells, since they have less space to install storage units. When μ_k takes its extreme values, we obtain two easily interpretable policies:

- **Greedy placement policy** ($\mu_k = N_k/N, u_k = 1$). Devices in each tier cache chunks in descending order of preferences until they reach their capacity N_k.
- **Uniform placement policy** ($\mu_k = 1, u_k = N_k/N$). For each tier, all chunks are cached with the same probability.

We assume that content is replaced when its popularity changes. Since content popularity changes infrequently compared to user requests, we do not include it in our model.

C. Network Interference and Connectivity

We next characterize the network connectivity of each caching tier in order to quantify its benefits to the network operator in the next section. We consider the standard power law propagation model and Rayleigh fading for all tiers [29], i.e., a receiver with distance r from the transmitter receives $p \cdot h \cdot r^{-\alpha}$ of signal power, where p is the transmit power, $h \sim \exp(1)$ represents the fading effect, and $\alpha > 2$ is the path-loss coefficient. Communication inside a region d is assumed to only interfere with devices within that region. Different tiers may or may not interfere with each other, depending on the spectrum allocation scheme. We will consider spectrum sharing, in which all tiers share the same spectrum and may interfere with each other; and an orthogonal partition in which different tiers use different spectrum bands, e.g., cellular connectivity to small cells and WiFi connectivity to vehicles.

Clients can retrieve contents from a caching point only if the requested content is cached and the signal-to-interference ratio (SIR) is greater than some threshold τ_k, which we assume is the same for all tiers ($\tau_k \equiv \tau$). We assume that user requests are always directed to the point with the highest SIR among all points that have the requested content, regardless of its tier. This assumption guarantees that users always benefit from caching, and can be easily relaxed by assigning to each tier a user association preference [30].

IV. Problem Formulation

We propose to use our system model to answer three specific questions regarding the value of vehicular caches:

- **Demand functions.** How do the optimal demands for vehicular and small cell caches change with respect to the time, region, prices, and user intensities?
- **Economic viability.** Can operators make money by subscribing to vehicular caching? How much value does vehicular caching add to existing caching products?

- **Value of competition.** Compared to using only small cells or vehicles, how much gain in utility, or reduction in cost, can the operator obtain when both are available?

The first two questions can be answered by solving the operator's profit maximization problem to find the optimal intensities L_v^*, L_s^* to which it should subscribe.

Problem 1 (Profit Maximization)**.**

$$\begin{aligned} \underset{L_v, L_s \in \mathbb{R}^{T\times D}}{\text{maximize}} \quad & \gamma U(L_v, L_s) - tr(P_v^T L_v) - tr(P_s^T L_s) \\ \text{subject to} \quad & L_v(t,d) \geq 0, L_s(t_1,d) = L_s(t_2,d) \geq 0 \end{aligned}$$

Here $U(L_v, L_s)$ is a utility function, to be defined in Section V, quantifying the benefits of different caching intensities L_v and L_s for the operator, and γ is a scaling coefficient. The remaining terms represent the cost of using each type of cache.

To quantify the added value of vehicular compared to small cell caches, we solve a variant of Problem 1:

Problem 2 (Utility Maximization)**.**

$$\begin{aligned} \underset{L_v, L_s \in \mathbb{R}^{T\times D}}{\text{maximize}} \quad & U(L_v, L_s) \\ \text{subject to} \quad & tr(P_v^T L_v) + tr(P_s^T L_s) \leq P_0 \\ & L_v(t,d) \geq 0, L_s(t_1,d) = L_s(t_2,d) \geq 0 \end{aligned}$$

This problem quantifies the maximum attainable utility subject to a fixed budget constraint. The value of competition is thus the difference of the achieved utilities for the competitive market and single-product markets (i.e., when only vehicular or small cell caches are available). In reality, vehicular and small cell caching providers may decrease their prices in order to better compete with each other, leading to higher utility for the network operator; thus, the solution given by Problem 2 can be interpreted as a lower bound of the real utility gain.

We will solve these optimization problems to answer our research questions analytically and numerically in Sections VI and VII. First, we define the utility function $U(L_v, L_s)$.

V. Defining Operator Utilities

The utility function measures the benefits that an operator obtains from caching. Different types of operators may then have different metrics for utility. We first define a general network operator utility model and then discuss three special cases that are particularly realistic models for operators that are CDN providers that sell cache capacity to CPs or ISPs who use caching to improve their users' quality-of-service.

General utility model. All of our utility models can be written in terms of the probability that users can connect to each caching tier. Thus, we define C_k as an indicator variable of whether a user chooses tier k and connects to it. Formally, $C_k = \mathbb{1}(SIR_k > \tau, Sel = k)$, where Sel is a random variable denoting the tier selected for serving a typical request. We then write the general utility functions

$$U_0 = \mathbb{E}[a(C_v + C_s) + bC_m] \quad (2)$$

where a and b are scalar weights; we use specific values of a and b to define three cases of particular interest below. We also write Sel_i as the tier selected to provide content i. We can now solve for the *tier k connectivity*, defined as probability that users can access tier k, $\mathbb{E}[C_k]$:

Theorem 1 (Expected Tier Connectivity)**.**

$$\mathbb{E}[C_k] = \frac{1}{\|L_u\|_1} \sum_{t,d} L_u(t,d) \sum_{i=1}^{N} f_i(t,d) \mathbb{E}[C_{k,i}(t,d)] \quad (3)$$

where $C_{k,i}(t,d) = \mathbb{1}(SIR > \tau, Sel_i = k | t,d)$ indicates if a typical request for content i is served by tier k at (t,d), and $\|\cdot\|_1$ is the ℓ_1 norm. For a typical user, let $X_{k,i}$ be the distance to the nearest tier-k point that has content i cached and $R_k = \min_i X_{k,i}$. If $R_k|Sel_i = X_{k,i}|Sel_i, \forall i$, then

$$\mathbb{E}[C_{k,i}(t,d)] = \frac{H_{k,i}(t,d) p_k^{\frac{2}{\alpha}} L_k(t,d)}{\sum_{j\in\mathcal{N}} (\rho + H_{j,i}(t,d)) p_j^{\frac{2}{\alpha}} L_j(t,d)} \quad (4)$$

$$\mathbb{E}[C_{k,i}(t,d)] = \frac{H_{k,i}(t,d) p_k^{\frac{2}{\alpha}} L_k(t,d)}{\rho p_k^{\frac{2}{\alpha}} L_k(t,d) + \sum_{j\in\mathcal{N}} H_{j,i}(t,d) p_j^{\frac{2}{\alpha}} L_j(t,d)} \quad (5)$$

for the sharing and orthogonal spectrum schemes respectively, with $\rho = \tau^{2/\alpha} \int_{\tau^{-\frac{2}{\alpha}}}^{\infty} (1 + u^{\frac{\alpha}{2}})^{-1} du$ a function of τ and α.

Proof sketch. For any (t,d), we use the probability density functions (PDFs) of R_k and $X_{k,i}$ and the approximation $R_k|Sel_i = X_{k,i}|Sel_i$ to find the joint probability $\mathbb{P}(R_k > r, Sel_i = k)$, as well as the PDF of $R_k|Sel_i$. Then we use the law of total probability to find $\mathbb{P}(SIR > \tau | Sel_i = k)$, which we simplify by using the Laplace transform of the interference. Reorganizing, we find $\mathbb{E}[C_{k,i}(t,d)]$ as in (4), (5), and we use the law of total expectation to derive (3). □

The approximation $R_k|Sel_i = X_{k,i}|Sel_i$ in Theorem 1, i.e., that content i is served by the closest physical point with that content, holds when all points in the same tier in each region cache the same items, as for the greedy placement policy. In general, it is close to reality if the requested content is cached with a high probability, i.e., when $H_{k,i}$ is close to 1. For less popular contents i, the approximation has limited influence on $\mathbb{E}[C_k]$ in (3) since the corresponding preferences f_i are small.

Utility from cache connectivity. Our first utility model can apply to network operators that are ISPs or CDN providers. Since CDN providers are typically paid for the amount of data directed to their devices, we can model their utility as the probability that a typical request is served by the cache, i.e., the rate of successful cache connectivity, which we define as the sum of the vehicle and small cell connectivity:

$$U_1 = \mathbb{E}[C_v + C_s] \quad (6)$$

where we have set $a = 1, b = 0$ in (2). The γ coefficient in Problem 1's objective then denotes the marginal value of cache connection, which is proportional to the per-byte monetary payoff to the CDN service. Similarly, an ISP might also receive a payoff proportional to the cache connectivity, as each cache connection reduces the operating cost of serving the user's request through the core network.

When the network coverage is good and interference can be ignored, we can assume the SIR threshold $\tau = 0$. In this case the cache connectivity $(C_v + C_s = \mathbb{1}(SIR > \tau, Sel_i = s, v|t,d))$ simply becomes the cache hit rate since we have $SIR_k > \tau = 0$. We thus define the **cache hit utility**, U_2:

$$A_k \triangleq \mathbb{1}(SIR_k > 0, Sel = k) = \mathbb{1}(Sel = k) \quad (7)$$
$$U_2 = \mathbb{E}[A_v + A_s]. \quad (8)$$

By assuming no interference, U_2 represents an optimistic approximation of the cache connectivity; while U_1 can be viewed as the worst-case estimation of cache connectivity as the derivation of C_k in Theorem 1 assumes all stations are transmitting at all time, which is unlikely to happen in reality. The gap between U_1 and U_2 gets larger as we increase the SIR threshold τ. With U_2, we are able to get closed-form solutions for Problems 1 and 2 in the next section. We first derive U_2 and prove its concavity.

Theorem 2 (Expected Cache Hit Rate). $\mathbb{E}[A_k]$ *is given by*

$$\mathbb{E}[A_k] = \frac{1}{\|L_u\|_1} \sum_{t,d} L_u(t,d) \sum_{i=1}^{N} f_i(t,d)\mathbb{E}[A_{k,i}(t,d)] \quad (9)$$

where $A_{k,i} = \mathbb{1}(Sel_i = k)$*, and*

$$\mathbb{E}[A_{k,i}(t,d)] = \frac{H_{k,i}(t,d)p_k^{\frac{2}{\alpha}} L_k(t,d)}{\sum\limits_{j\in\mathcal{N}} H_{j,i}(t,d)p_j^{\frac{2}{\alpha}} L_j(t,d)} \quad (10)$$

Proof. The proof is analogous to the proof of Theorem 1. □

Since the interference has been removed, (9) and (10) are independent of the spectrum allocation scheme, and Theorem 2 does not require that $X_{k,i}|Sel_i = R_k|Sel_i$ as in Theorem 1. In fact, (10) is a special case of (4) and (5) when $\tau = 0$. We find that the cache hit utility U_2 is a concave function:

Theorem 3 (Concavity of Cache Hit Utilities). *The utility function (8) is concave with respect to* L_v *and* L_s*.*

Proof. Plugging (9) and (10) into (8), the utility function is

$$U_2 = \sum_{t,d,i} \frac{L_u(t,d)f_i(t,d)}{\|L_u\|_1}(1 - \mathbb{E}[A_{m,i}(t,d)]) \quad (11)$$

From (10), $\mathbb{E}[A_{m,i}(t,d)]$ is a convex function. Thus $(1 - \mathbb{E}[A_{m,i}(t,d)])$ is concave, as is U_2 in (11). □

We finally define a third type of utility function, which is again a special case of the general utility (2).

Delivery rate utility. If the network operator is an ISP, caching can not only reduce its cost, as discussed above, but also improve user QoS and attract new users. ISPs might then wish to maximize the downlink delivery rate [27], which characterizes the throughput experienced by a typical user and thus the spectrum efficiency. The marginal value of this utility (i.e. the value of γ in Problem 1) can be determined using tools like the discrete choice model [31].

For the utility function to represent the delivery rate, we set $a = \log(1+\tau), b = r_b < \log(1+\tau)$ in the utility expression (2). That is, the typical user is served with constant rate $\log(1+\tau)$ if she is in coverage and served by cache tiers. If she is in coverage yet served by macro cells, the latency is controlled by the processing and queuing rate of the backhaul; thus the user is served with a lower delivery rate $r_b < \log(1+\tau)$. The value of r_b is jointly determined by factors such as the maximum bandwidth of the backhaul, the distribution of macro cells, etc.; we treat it as a known constant. As a result, we can write the **delivery rate utility**, U_3, from (2) as

$$U_3 = \mathbb{E}[\log(1+\tau)(C_v + C_s) + r_b C_m] \quad (12)$$

VI. Optimal Caching Intensities

In this section, we will answer Section IV's research questions on the caching demand functions and economic viability by solving Problem 1, and on the value of competition by solving Problem 2, with Section V's utility functions. With the cache hit utility U_2 in (8), both problems are convex, and can be solved in closed-form. However, the cache connectivity and delivery rate utilities U_1 and U_3 in (6) and (12) are not concave. We design an efficient fractional programming algorithm to find the optimal solution in these cases.

A. Solutions for Profit Maximization (Problem 1)

We first consider the cache hit utility (8) before giving an algorithm to solve Problem 1 with more general objectives.

Theorem 4 (Demand Functions under Cache Hit Utilities). *Under the general placement policy (1), assume* $\mu_v \leq \mu_s$*. With utility function (8), the solution to Problem 1 is:*

$$L_s^*(\cdot,d) = -\frac{1}{u_s}F_{ms}L_m(\cdot,d) + \sqrt{\frac{\gamma F_{ms}L_m(\cdot,d)\sum_t(Q_s(t,d) - \delta_v(t,d)Q_v(t,d))L_u(t,d)}{\|L_u\|_1 u_s \sum_t P_s(t,d) - \delta_v(t,d)\frac{u_s}{u_v}F_{sv}P_v(t,d)}} \quad (13)$$

$$L_v^*(t,d) = \sqrt{\frac{\gamma F_{mv}L_m(\cdot,d)Q_v(t,d)L_u(t,d)}{\|L_u\|_1 u_v P_v(t,d)}} - \delta_s(t,d)\frac{u_s}{u_v}F_{sv}L_s^*(\cdot,d) - \frac{1}{u_v}F_{mv}L_m(\cdot,d) \quad (14)$$

Here $Q_k(t,d) = \sum_{i=1}^{\mu_k N} f_i(t,d)$, $\delta_k(t,d) = \mathbb{1}(L_k^*(t,d) > 0)$, $F_{kj} = (p_k/p_j)^{2/\alpha}$. *If the term inside the square root of (13) is negative, or if (13) or (14) is negative, we set the corresponding intensity to zero.*

Proof. The result follows from the KKT (Karush-Kuhn-Tucker) optimality conditions. □

Demand functions. Theorem 4 allows us to quantify how the optimal demands L_s^* and L_v^* vary with the user intensities L_u. Surprisingly, we observe that the caching intensity L_k^* for $k \in \{s, v\}$ is a *square root function of* $L_u/\|L_u\|_1$, and thus increases at a slower rate than the tier connectivity, which is linear in $L_u(t,d)/\|L_u\|_1$ (Theorem 1), as user requests become more concentrated in a single time and region (t,d) ($L_u(t,d)/\|L_u\|_1$ increases). The optimal small cell demand $L_s^*(t,d)$ depends on the aggregated user intensity $\sum_t L_u(t,d)$, as we would expect given small cells' lack of mobility, while the vehicle demand $L_v^*(t,d)$ only depends on the users at other times through the small cell demand $L_s^*(t,d)$. In fact, *small cells substitute for vehicles:* as $L_s^*(t,d)$ increases, the vehicular demand $L_v^*(t,d)$ decreases linearly. The marginal rate of substitution, $\delta_s(t,d)\frac{u_s}{u_v}F_{sv}$, is independent of $L_u(t,d)/\left\|L_u\right\|_1$ but depends on the ratios of the small cell and vehicle transmit powers p_s/p_v and cache capacities u_s/u_v.

We next investigate the dependence on the prices P_k. The *optimal demand L_k^* decays on the order of $\sqrt{P_k}$ for both tiers $k \in \{s,v\}$*. As with the user intensity, the optimal small cell demand $L_s^*(\cdot,d)$ is a function of the aggregated price $\sum_t P_s(t,d)$ at each location d. However, the optimal vehicle demand $L_v^*(t,d)$ is only affected by the prices P_v at other times through the small cell demand $L_s^*(t,d)$. We can thus find the cross-time price sensitivity of this demand:

Corollary 1 (Cross-Time Sensitivity for Vehicles' Prices)**.** *For cache hit utilities, if the price $P_v(t_1,d)$ increases for some t_1, the optimal demands of vehicles in other time slots $L_v^*(t_2,d)$: $t_2 \neq t_1$ either decrease or stay unchanged.*

Intuitively, if $P_v(t_1,d)$ increases, small cells become more attractive than vehicles in region d. Thus, $L_s^*(t_1,d)$ either increases or stays the same. Since small cells are fixed over time, the number of small cells at (t_2,d) also increases or stays the same, prompting a decrease in $L_v^*(t_2,d)$.

Economic viability. We next use Theorem 4 to find conditions under which vehicular caching is viable, i.e., there is positive demand for some time slots and regions.

Corollary 2 (Economic Viability of Vehicular Caching)**.** *Assume $\mu_v \leq \mu_s$. If at the optimal point all intensities L_v^* and L_s^* are positive, the following relation holds*

$$\sum_t P_s(t,d) > \frac{u_s}{u_v}\left(\frac{p_s}{p_v}\right)^{\frac{2}{\alpha}} \sum_t P_v(t,d) \tag{15}$$

In words, the region-accumulative price $\sum_t P_s(t,d)$ of small cells is lower bounded by that of vehicles, scaled by the ratio of transmit power and caching capacities. We can thus interpret (15) as a lower bound on the small cell prices for which all intensities are positive, i.e., the operator uses both vehicles and small cells. If the small cells are sufficiently less expensive or have sufficiently higher capacity ($u_s \gg u_v$), the operator will not utilize the vehicular caches at all.

Profit maximization with general utility functions. We now turn to the general utility function U_0. In this case, we solve Problem 1 with a quadratic transform [29] as it does not have a closed-form solution. From Theorem 1, we can write

$$\mathbb{E}[C_{k,i}(t,d)] = \frac{W_{k,i}(t,d)}{Z_{k,i}(t,d)} \tag{16}$$

where $W_{k,i}$ and $Z_{k,i}$ are affine functions of L_v and L_s.

Thus, defining $\alpha_v = \alpha_s = a, \alpha_m = b$, (2) can be written as

$$U_0 = \sum_{t,d,i,k} \alpha_k \frac{L_u(t,d)}{\|L_u\|_1} f_i(t,d) \frac{W_{k,i}(t,d)}{Z_{k,i}(t,d)} \tag{17}$$

We then introduce the auxiliary utility function (17):

$$\begin{aligned} V(y,L_v,L_s) = \sum_{t,d,i,k} \alpha_k \frac{L_u(t,d)}{\|L_u\|_1} f_i(t,d) \times \\ \left(2y_{k,i}(t,d)\sqrt{W_{k,i}(t,d)} - y_{k,i}^2(t,d) Z_{k,i}(t,d)\right), \end{aligned} \tag{18}$$

Substituting (17) with (18) in Problem 1, we find the equivalent optimization problem:

Problem 3 (Transformed Profit Maximization)**.**

$$\begin{aligned} \underset{y,L_v,L_s}{\textit{maximize}} \quad & \gamma V(y,L_v,L_s) - tr(P_v^T L_v) - tr(P_s^T L_s) \\ \textit{subject to} \quad & L_v(t,d) \geq 0, L_s(t_1,d) = L_s(t_2,d) \geq 0 \end{aligned}$$

For general content placement policies, the dimension of y is in the order of $\Theta(TDN)$. However, if the value of $H_{k,i}$ is fixed for most contents (e.g. the general placement policy (1)), the summation over i can be greatly simplified, and the number of auxiliary variables y can be reduced to $\Theta(TD)$.

Theorem 5 (Equivalence of Problems 1 and 3)**.** *The variable (L_v^*, L_s^*) maximizes Problem 1 if and only if (L_v^*, L_s^*) together with some y^* maximizes Problem 3. Furthermore, Problems 1 and 3 have the same objective value at the optimal point.*

Proof. See Appendix A of [29]. □

The transformed Problem 3 is not necessarily a convex problem. However, it is easy to see that this problem is convex with respect to (L_v, L_s) when the value of y is fixed. On the other hand, given (L_v, L_s), the optimal y can be written as

$$y_{k,i}(t,d) = \sqrt{W_{k,i}(t,d)}/Z_{k,i}(t,d) \tag{19}$$

This useful property allows the use of a block coordinate ascent algorithm to find the optimum by iteratively solving for the optimal (L_v, L_s) given y and the optimal y given (L_v, L_s). The convergence of this procedure to optimality follows from that of the block coordinate ascent method [32].

B. Solutions for Utility Maximization (Problem 2)

As in Theorem 4, we can also derive a closed-form solution to Problem 2 for the cache hit utility (8). Based on that we can conclude Corollary 3 and 4 below. Given a limited budget, the maximum utility in the competitive market is bound to be no worse than that of single-product markets as the latter are special cases of the two-product optimization. We first consider the **value of competition** with infinite budgets, when the prices of the small cells and vehicles do not matter:

Corollary 3 (Upper Bound of Cache Hit Utilities)**.** *With infinite budget, the maximum utility for the vehicle only and the small cell only market using utility (8) can not exceed*

$$U \leq \sum_{t,d} \frac{L_u(t,d)}{\|L_u\|_1} Q_k(t,d), k \in \{v,s\} \tag{20}$$

Here $Q_k(t,d)$ is given in Theorem 4, and the upper bound for the competitive market is the same as the small cell market, as we assume $\mu_v \leq \mu_s$ in the definition of Q_k.

Corollary 3 suggests that *as the network operator's budget increases, the value of competition approaches zero*: the utilities under the competitive and small cell only markets have the same upper bounds. With infinite budget, either vehicles or small cells can achieve their optimal utilities as their costs are insignificant. We verify this result in Section VII: as the budget increases, the competitive market achieves nearly the same utility as the small cell only market. Corollary 3 also implies that the value of the competitive market is bounded away from 1, i.e., perfect cache hits, due to the limited capacities of the caches. However, despite this lack of competition, both products are purchased at large budgets:

Corollary 4 (Linear Increase of Optimal Demands)**.** *Under cache hit utilities (8), for a sufficiently large value of the budget* $P_0 \gg 0$*, the optimal intensities* L_v^* *and* L_s^* *increase linearly with respect to* P_0 *at every* (t, d)*.*

Thus, for sufficiently large budgets, there is effectively no change in the competition between vehicles and small cells: a budget increase yields constant marginal increases in the demands for both types of caching. This result is consistent with the linear substitution of small cells for vehicles from Theorem 4. Together with Corollary 3, it implies that the marginal value of additional caching intensities is eventually zero, which we verify in Section VII.

To solve Problem 2 for general utility functions, we can still apply the quadratic transform (18) and update rule (19) as for Problem 1. We can then use a block coordinate ascent algorithm to find the optimal solution.

VII. Numerical Simulations

We next verify Section VI's results numerically. We first consider the optimal demands in Section VII-A and the value of competition in Section VII-B. We then investigate the impact of variations in the system model in Section VII-C.

We use the SF311 database [33] to simulate the temporal and spatial dynamics of user requests. The database records the location and time for each mobile 311 service request in San Francisco, California. Figure 2 shows the hourly distribution of user requests for the 30 most popular neighborhoods, averaged over all weekdays in 2017. To facilitate the visualization of our results, we focus on two representative regions: the business/office region Financial District (referred to as R1), and the residential/leisure region Showplace Square (R2). Both regions have about 5000 daily requests and similar land area. We scale their distributions such that the highest hourly intensity equals 50 per squared kilometer (Figure 2). We define the business hour (T1) as 10:00 to 16:00, and the off hour (T2) as 18:00 to 24:00. For both regions, we average the intensities within each time slot, thus yielding four intensity bins (km^{-2}): 39.5 (T1+R1), 11.8 (T1+R2), 27.4 (T2+R1), 25.0 (T2+R2). For both regions, the intensity of macro cells is 0.3 km^{-2}. Unless otherwise noted, we set $P_s \equiv 18, P_v(1, \cdot) = 2, P_v(2, \cdot) = 5$.

We set the transmit power (Watt) as $p_m = 40$, $p_s = 6.3$, and $p_v = 1.0$ [34]. We let the SIR threshold $\tau = 1(0 \text{ dB})$, and the path-loss coefficient $\alpha = 4$. We consider a content catalog with

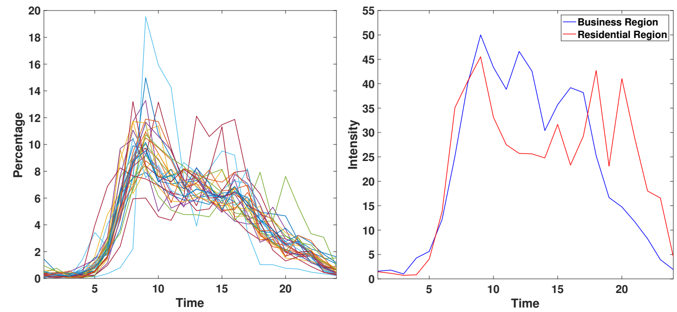

Fig. 2. *Left:* the distribution of user requests in a typical day for the top 30 neighborhoods with highest request counts. *Right:* The scaled hourly user request intensities for the "business" region Financial District and the "residential" region Showplace Square show different temporal patterns.

a total of $N = 10000$ chunks, and let $N_v = 30, N_s = 100$. For simplicity, we assume users' preferences for content chunks do not change for different time slots and regions, and follow the Zipf distribution with shape coefficient equal to 1.2 [35]. For the delivery rate model, let $r_b = 0.5 \log(1 + \tau)$, i.e. the delivery rate when the cache is missed is half that for the cache hit rate. The marginal utility value (i.e. γ) is set to be 500 for the connectivity utility U_1 in (6) and the cache hit utility U_2 in (8), and 1000 for the delivery rate utility U_3 in (12).

A. Demands of Caching

Hourly demand of vehicular caching. An advantage of vehicular caching over small cells is its ability to better respond to the change of user intensities: at any time, more vehicles may be dispatched to regions that currently have higher user intensities. Thus, the hourly dynamics of the optimal vehicle demands are expected to match the patterns of user requests shown in Figure 2.

In Figure 3, we show a heat map of the optimal vehicle intensities for the top 30 neighborhoods under the cache hit utility U_2. In Figure 4, we compare the vehicle intensities in the business and residential regions (R1 and R2) for all three utilities U_1, U_2 and U_3. Comparing Figures 2 and 3, more vehicles are used in the daytime due to higher request intensities, and the vehicle intensity also adapts to variation over different locations. Under all utility functions, the temporal patterns of vehicle demands in R1 and R2 (Figure 4) closely track those of the user requests (Figure 2).

Viability and accumulative demand. The accumulative demand for each caching tier is the sum of the temporally averaged demands across regions. In our setting, it reflects the operators' daily demands for both caching tiers. Thus, vehicular caching is economically viable only if its accumulative demand is positive. We consider the economic viability for two time slots (T1, T2) and two regions (R1, R2) for ease of visualization. For clarity we assume the prices do not change over time and regions (we explore price variation below), and set $P_s \equiv 18$. We then vary the price of vehicles P_v, and find the corresponding demands as per Theorems 4 and 5.

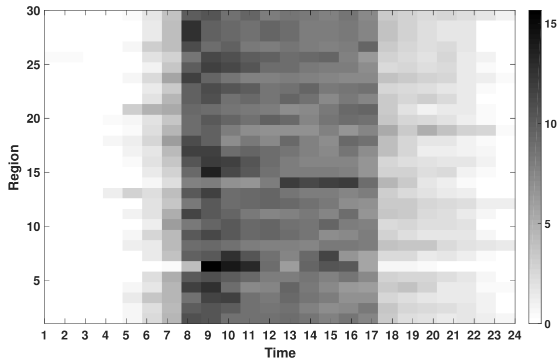

Fig. 3. Heat map of the optimal vehicle demands over 24 hours for the top 30 neighborhoods with the cache hit utility U_2, which adapt to spatiotemporal user request variations.

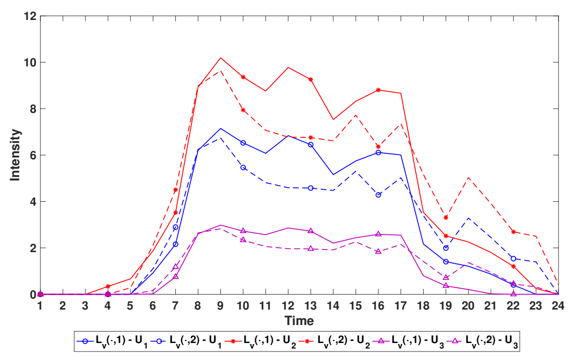

Fig. 4. The optimal vehicle demands over 24 hours in the business region (R1) and the residential region (R2) resemble user request patterns (Figure 2).

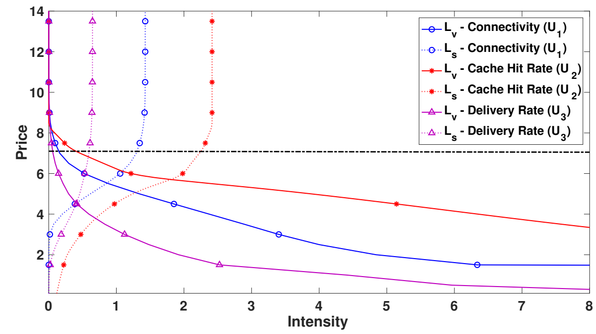

Fig. 5. Accumulative demand curves for vehicular caching as the vehicle price $P_v(\cdot,\cdot)$ varies. Intensities are aggregated over regions and averaged across time. Vehicles exhibit zero demand when the price exceeds a certain threshold. The dash-dot line at 7.17 represents the lower bound of this threshold given by Corollary 2.

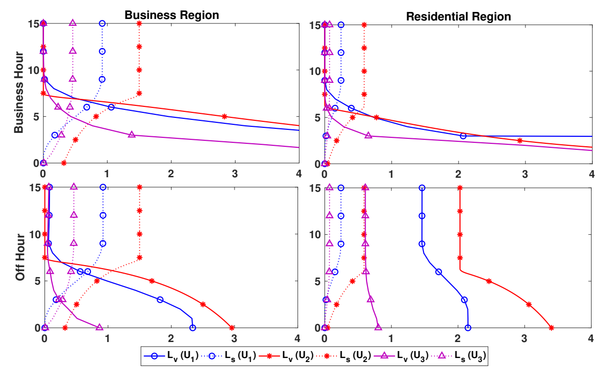

Fig. 6. Spillover effect caused by change of vehicles' prices. X-axes are intensities, Y-axes are price of vehicular caching in business hours $P_v(1,\cdot)$. As the vehicle price in business hours increases, the vehicle (small cell) demand in both the business and off hours decreases (increases).

Figure 5 shows the accumulative demand curves for the three utility functions using the greedy content placement policy and the spectrum sharing scheme. The accumulative demand for vehicles decreases with a diminishing elasticity as the price goes up at approximately a rate of $(\sqrt{P_k})^{-1}$, as expected from Theorem 4. Each of the three utility functions considered exhibit qualitatively similar demand curves. For all three utility functions, the accumulative demand for vehicles drops to zero above a threshold price around 8: vehicular caching is economically viable only if its price is below this threshold. Corollary 2 gives an lower bound of 7.17 on this threshold price, which is very close to the actual threshold.

Spillover effects on time and region. The demands of vehicles and small cells at different times are highly correlated. From Corollary 1, if the price of vehicles increases in some (t_1, d_1), operators will decrease their demands for vehicles $L_v(t_1, d_1)$ and subscribe to more small cells, increasing $L_s(t_1, d_1)$. Since the small cell intensities are fixed over time, $L_s(t_2, d_1)$ also increases, and the operators subscribe to fewer vehicles $L_v(t_2, d_1)$. To reveal this spillover effect, we fix $P_s \equiv 18, P_v(2,\cdot) = 5$, and vary vehicles' price in business hours $P_v(1,\cdot)$, assuming prices do not change across regions. Figure 6 shows the resulting change of demands with greedy placement and spectrum sharing. As $P_v(1,\cdot)$ increases, demand for vehicles (small cells) decreases (increases) at all time slots and regions. The change of vehicle demand in off hours is consistent with Corollary 1: it first decreases, then stays unchanged. However, the decrease is most pronounced during business hours (T1), when the price itself changes.

Effect of user intensities. The dynamics of users' requests can significantly affect the economic performance of vehicular caching. Intuitively, vehicles are more attractive if user intensities are more divergent across time and regions. To illustrate this effect, we set the user intensities as $L_u(1,1) = L_u(2,2) = 30 + d_u, L_u(1,2) = L_u(2,1) = 30 - d_u$. As d_u grows higher, the distribution of users becomes more uneven. We use the cache hit utility U_2.

Figure 7 shows our results. For both regions, as L_u diverges, $L_s^*(\cdot, d)$ first remains constant, then decreases. This is consistent with Theorem 4: since $\sum_t L_u(t,d) \equiv 60$ in our setting, L_s^* changes only when some vehicle demands L_v^* become zero. On the other hand, L_v^* changes significantly at a rate of $\sqrt{d_u}$. When $d_u = 30$, i.e., there are no users in the business (residential) region during off (business) hours, $L_v^*(1,2) = L_v^*(2,1) = 0$. When $d_u = 0$ and the user intensities are uniform, vehicle demand is still positive due to the lower prices P_v compared to small cell prices P_s.

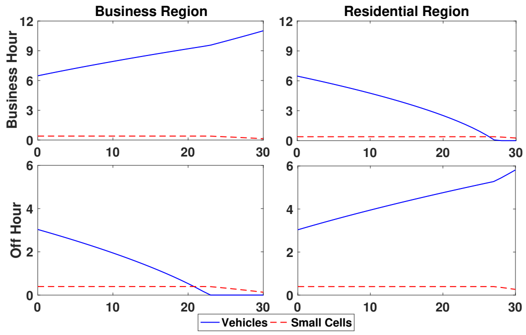

Fig. 7. Evolution of caching demands as user intensities diverge. X axes represent d_u (user divergence), Y axes represent caching demands. While L_v^* visibly changes, L_s^* keeps unchanged before some $L_v^*(t,d)$ becomes zero.

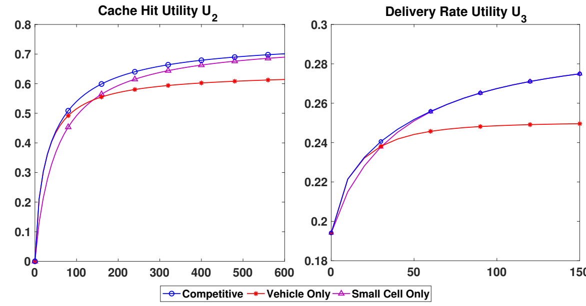

Fig. 8. Comparison of maximum utilities with fixed budget. X-axes represent budgets, Y-axes represent utilities. While initially only vehicles are preferred, as the budget increases the utility in the competitive market approaches that of the market with only small cells.

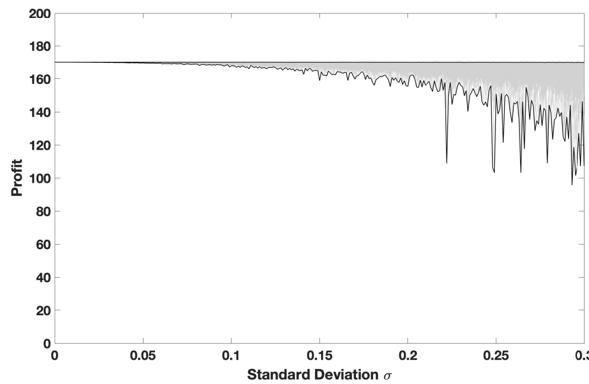

Fig. 9. The change of profits under U_2 when the optimal intensity is offset by a zero mean bias term. The grey lines represent results of 500 trials. The black lines are envelopes, showing little variation in the achieved profit.

B. Value of Competition

As discussed in Section III, we evaluate the value of competition by comparing the solutions to Problem 2 for the competitive and single-product markets.

Figure 8 shows our achieved utilities with greedy placement policy and spectrum sharing. Both plots follow the same pattern. At first, when the budget is small, vehicular caching obtains a higher utility level than small cells, and the optimal choice in the competitive market is to subscribe only to vehicles. In this phase, the competitive curve overlaps with the vehicle curve; the value of competition is thus zero. As the budget increases, we start to observe a positive value of competition - the competitive market achieves a higher utility level than both single-product markets. When the budget is large enough, small cells start to outperform vehicles, and the vehicle curve and the small cell curves intersect. In the meantime, the utility gain from competition continues to rise. Finally, when the budget goes to infinity, the competitive curve converges to the small cell curve, i.e., the value of competition gradually drops to zero. For the cache hit utility, the utility bounds can be calculated by Corollary 3, which states that the curves will eventually converge at $Q_v = 0.639$ for the vehicle only market and $Q_s = 0.751$ for both the competitive market and the small cell only market.

These phase transitions can be explained as follows. When the budget is small, few caching points are deployed in the network. Thus, even requests for popular content chunks can not be served by the cache. In this situation, cached chunks in vehicles can be better utilized due to their mobility, making vehicular caching more valuable. When the budget increases, however, the increase of utility is constrained by the finite vehicle capacity N_v. As a result, the total utility is driven by the larger capacity N_s of the small cells.

C. Variations in the System Model

We finally test the robustness of our results to inaccuracies in our cache intensity models. To do so, we add bias to the optimal intensities $L_k^*, k \in \{v, s\}$, and re-calculate the resulting operator profit. The bias is drawn from a Gaussian distribution with zero mean, and the standard deviation equals $L_k^* \times \sigma$ for some σ. In the case an intensity value is negative, we set it to zero. We run a total of 500 trials, gradually increasing σ from 0 to 0.3. As is illustrated in Figure 9, the profit drops by at most 10 (5.8%) for $\sigma < 0.2$, implying our model is robust to small intensity deviations.

VIII. Conclusion

In this paper, we have proposed a framework to evaluate the economic value of vehicular caching. A system model is developed to capture the economic and physical dynamics of caching tiers. We then use this model to derive the utility functions for different network operators, including both CPs and ISPs. The value of vehicular caching is quantified through its viability, demands, and the gain of competition. These research questions are answered by solving two optimization problems, for which we have provided both analytical and algorithmic solutions. We find that vehicular caching does not add value if its price exceeds a finite threshold or the operator has infinite budget. Simulation results verify the economic value of vehicular caching, and illustrate the dynamics of the system. Future work may build on our model to verify its accuracy under more realistic mobility patterns, and extend our analysis from mobile caching to other network services, e.g., mobile devices that act as edge servers or network relays.

REFERENCES

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2017-2022 white paper," 2019.

[2] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5g wireless networks," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 82–89, 2014.

[3] I. S. C. Trial, "Rethinking the small cell business model," 2011.

[4] X. Zhou and L. Chen, "Demand shaping in cellular networks," *IEEE Transactions on Control of Network Systems*, 2018.

[5] W. Shin, B.-Y. Min, and D. K. Kim, "Vehicaching: Embracing user request on vehicle route with proactive data transportation," in *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st*, pp. 1–5, IEEE, 2015.

[6] B.-Y. Min, W. Shin, and D. K. Kim, "Delay-tolerable contents offloading via vehicular caching overlaid with cellular networks," *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 100, no. 1, pp. 283–293, 2017.

[7] L. Vigneri, *Vehicles as a mobile cloud: modelling, optimization and performance analysis*. PhD thesis, Université Côte d'Azur, 2017.

[8] A. Dhekne, M. Gowda, R. R. Choudhury, and S. Nelakuditi, "If wifi aps could move: A measurement study," *IEEE Transactions on Mobile Computing*, vol. 17, no. 10, pp. 2293–2306, 2018.

[9] S. A. R. Naqvi, S. A. Hassan, H. Pervaiz, and Q. Ni, "Drone-aided communication as a key enabler for 5g and resilient public safety networks," *IEEE Communications Magazine*, vol. 56, no. 1, pp. 36–42, 2018.

[10] K. Zhang, S. Leng, Y. He, S. Maharjan, and Y. Zhang, "Cooperative content caching in 5g networks with mobile edge computing," *IEEE Wireless Communications*, vol. 25, no. 3, pp. 80–87, 2018.

[11] F. Restuccia, S. K. Das, and J. Payton, "Incentive mechanisms for participatory sensing: Survey and research challenges," *ACM Transactions on Sensor Networks (TOSN)*, vol. 12, no. 2, p. 13, 2016.

[12] L. Frayer, "Free wi-fi on buses offers a link to future of 'smart cities'." NPR All Tech Considered, 2015.

[13] K. Hamidouche, W. Saad, and M. Debbah, "Breaking the economic barrier of caching in cellular networks: Incentives and contracts," in *Global Communications Conference (GLOBECOM), 2016 IEEE*, pp. 1–6, IEEE, 2016.

[14] Z. Hu, Z. Zheng, T. Wang, L. Song, and X. Li, "Roadside unit caching: Auction-based storage allocation for multiple content providers," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6321–6334, 2017.

[15] K. Zhao, S. Zhang, N. Zhang, Y. Zhou, Y. Zhang, and X. Shen, "Incentive mechanism for cached-enabled small cell sharing: A stackelberg game approach," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pp. 1–6, IEEE, 2017.

[16] K. Poularakis, G. Iosifidis, L. Tassiulas, *et al.*, "Approximation algorithms for mobile data caching in small cell networks.," *IEEE Trans. Communications*, vol. 62, no. 10, pp. 3665–3677, 2014.

[17] Y. Cui and D. Jiang, "Analysis and optimization of caching and multicasting in large-scale cache-enabled heterogeneous wireless networks," *IEEE transactions on Wireless Communications*, vol. 16, no. 1, pp. 250–264, 2017.

[18] M. Ji, G. Caire, and A. F. Molisch, "Wireless device-to-device caching networks: Basic principles and system performance," *arXiv preprint arXiv:1305.5216*, 2013.

[19] R. Wang, X. Peng, J. Zhang, and K. B. Letaief, "Mobility-aware caching for content-centric wireless networks: Modeling and methodology," *IEEE Communications Magazine*, vol. 54, no. 8, pp. 77–83, 2016.

[20] Y. Zhang, J. Zhao, and G. Cao, "Roadcast: a popularity aware content sharing scheme in vanets," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 13, no. 4, pp. 1–14, 2010.

[21] X. Cai, S. Zhang, and Y. Zhang, "Economic analysis of cache location in mobile network," in *Proc. of IEEE WCNC*, pp. 1243–1248, IEEE, 2013.

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

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

[24] B. Hu, L. Fang, X. Cheng, and L. Yang, "In-vehicle caching (iv-cache) via dynamic distributed storage relay (d ^{2} sr) in vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 843–855, 2018.

[25] Y. Zhang, C. Li, T. H. Luan, Y. Fu, W. Shi, and L. Zhu, "A mobility-aware vehicular caching scheme in content centric networks: Model and optimization," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3100–3112, 2019.

[26] J. G. Andrews, A. K. Gupta, and H. S. Dhillon, "A primer on cellular network analysis using stochastic geometry," *arXiv preprint arXiv:1604.03183*, 2016.

[27] E. Baştu, M. Bennis, M. Kountouris, and M. Debbah, "Cache-enabled small cell networks: Modeling and tradeoffs," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, p. 41, 2015.

[28] L. A. Adamic and B. A. Huberman, "Zipf's law and the internet.," *Glottometrics*, vol. 3, no. 1, pp. 143–150, 2002.

[29] K. Shen and W. Yu, "Fractional programming for communication systemspart i: Power control and beamforming," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2616–2630, 2018.

[30] Y. Lin, W. Bao, W. Yu, and B. Liang, "Optimizing user association and spectrum allocation in hetnets: A utility perspective," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 6, pp. 1025–1039, 2015.

[31] P. Coucheney, P. Maillé, and B. Tuffin, "Impact of competition between isps on the net neutrality debate," *IEEE Transactions on Network and Service Management*, vol. 10, no. 4, pp. 425–433, 2013.

[32] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *Journal of optimization theory and applications*, vol. 109, no. 3, pp. 475–494, 2001.

[33] City and C. of San Francisco, "Sf311 cases." https://sf311.org/, 2018. [Online; accessed 20-December-2018].

[34] M. Di Renzo, A. Guidotti, and G. E. Corazza, "Average rate of downlink heterogeneous cellular networks over generalized fading channels: A stochastic geometry approach," *IEEE Transactions on Communications*, vol. 61, no. 7, pp. 3050–3071, 2013.

[35] Z. Chen and M. Kountouris, "D2d caching vs. small cell caching: Where to cache content in a wireless network?," in *Proc. of the IEEE Workshop on Signal Processing Advances in Wireless Communications*, pp. 1–6, IEEE, 2016.