Title
A Wireless Vehicle-based mobile network infrastructure designed for smarter cities

Authors
Quer, Giorgio
Aktas, Tugcan
Librino, Federico
et al.

Publication Date
2019-03-01

DOI
10.1016/j.adhoc.2018.10.014

A Wireless Vehicle-based Mobile Network Infrastructure Designed for Smarter Cities

Giorgio Quer, Tugcan Aktas, Federico Librino, Tara Javidi, and Ramesh R. Rao[1]

Abstract

The evolution of smart city services and applications requires a more efficient wireless infrastructure to provide the needed data rate to users in a high-density environment with high mobility, satisfying at the same time the request for high-connectivity and low-energy consumption. To address the challenges in this new network scenario, we propose to opportunistically rely on the increasing number of connected vehicles in densely populated urban areas. The idea is to support the macro base station (BS) with a secondary communication tier composed of a set of smart and connected vehicles that are in movement in the urban area. As a first step towards a comprehensive cost-benefit analysis of this architecture, this paper considers the case where these vehicles are equipped with femto-mobile access points (fmAPs) and constitute a mobile out-of-band relay infrastructure. We first study this network system with a continuous time model, in which three techniques to select an fmAP (if more than one is available) are proposed and the maximal feasible gain in the data rate is characterized as a function of the vehicle density, average vehicle speeds, handoff overhead cost, as well as physical layer parameters. We then introduce a time slotted model, in which we consider a more realistic communication channel, with an exponential path loss model, and we investigate the tradeoff between energy consumption and expected data rate, as a function of the system parameters. The analytical and simulation results, with both the continuous and time slotted models, provide a first benchmark characterizing this architecture and the definition of guidelines for its future realistic study and implementation.

Keywords: Connected vehicles, smart city, wireless networks, mobile access points.

[1]G. Quer was with the University of California, San Diego, and now is with the Scripps Research Translational Institute, 3344 North Torrey Pines Court, La Jolla, CA 92037 (e-mail: gquer@scripps.edu).
T. Aktas was with the University of California, San Diego, and now is with Anketek Elektronik, ODTÜ Teknokent, Ankara, Turkey (e-mail: tugcan.aktas@anketek.com.tr).
F. Librino is with the Italian National Research Council, 56124 Pisa, Italy (e-mail: federico.librino@iit.cnr.it).
T. Javidi and R. Rao are with the University of California, San Diego, University of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093 (e-mail: tjavidi, rrao@ucsd.edu).
A partial and preliminary version of this work [1] has been presented at IEEE GLOBECOM, Dec. 2016.
G. Quer and T. Aktas contributed equally to this work.

1. Introduction

The number of people living in urban areas is ever increasing and is expected to rise from a current estimate of 54% of the world's population to a forecast of 66% by 2050, according to the UN. In an effort to increase the quality of life of citizens through information technologies, various smart city initiatives have been launched [2]. Clearly, smart cities require an ever expanding and evolving wireless infrastructure in terms of data rate, connectivity, and energy efficiency.

In the literature, several works advocate for the expansion of traditional network solutions [3] such as increasing bandwidth [4] or the number of relays and pico and femto cells. The idea of supporting the macro base station (BS) with an additional tier of communication has been widely investigated in the framework of heterogeneous networks (HetNets), see [5] and references therein.

Differently from previous works, we propose to take advantage of yet another trend in urban living: connected vehicles. From Teslas and city buses to Lyft and Uber vehicles, more and more vehicles are equipped with wireless connectivity solutions that allow the exchange of useful information about their status or the implementation of useful applications [6, 7]. The setup and analysis of a multi-hop delay-tolerant vehicular network is discussed in [8], while the integration of cellular and vehicular networks have recently appeared in [9, 10].

In this paper, we propose to exploit connected vehicles to support the macro BS (tier 1, or T1) with a new type of secondary communication tier (T2) composed of a set of connected vehicles as a mobile out-of-band relay infrastructure. Instead of deployment of high-cost pico BSs or relays at fixed positions, we investigate a new network architecture with wireless access points (APs) installed on a selected subset of vehicles that are in constant movement in the urban area, e.g., city buses, taxis, and even car sharing services. We call the AP installed in a vehicle a femto-mobile AP (fmAP). The use of fmAPs bring some significant advantages, including 1) the additional channel diversity offered, as a consequence of the fmAPs movements in the urban scenario, and 2) the possibility of installing small cell APs without incurring in the additional cost of a fixed antenna positioned on the top of existing infrastructures.

In this scenario, the main problem is the management of handoffs within a tier (horizontal handoff) or between the tiers (vertical handoff) [11]. This problem cannot be ignored when we consider user mobility [12]. If the user is using the public transportation system, a promising solution is to install a relay in the public vehicle, as in the case of high speed trains [13], so that the relative speed between the user and the relay is approximately zero. If the user pattern in the urban scenario cannot be predicted, the use of mobile relays can still be effective in terms of capacity by increasing the frequency reuse [14], particularly when the position of the relays can be controlled to optimize the handoff mechanism and balance the backhaul [15].

In our model, the mobility pattern of the fmAPs cannot be controlled, indeed a vehicle equipped with a fmAP continues to travel through the city performing its normal operations. The fmAP is connected to the macro BS through a wireless backhaul link and provides data services to the user equipments (UEs) for the short period in which they are in close proximity. The connection between UE and fmAP is on an orthogonal channel in unlicensed bands (LTE-U) [16]. In order to cope with the Doppler effects due to the high mobility of the fmAP's, we assume that each fmAP is equipped with a large number of antennas and uses beamformed transmission and reception [17]. By maintaining the orthogonality of the transmission in the spatial domain, we do not deal with the limited bandwidth problem at the backhaul [18] in our analysis.

In general, the channel state information (CSI) between the fmAP and the UE is constantly changing due to fmAP mobility. This increased space and time diversity is an opportunity to provide connectivity also to the edge user [19], in the presence of multipath or strong shadowing due to non-line-of-sight conditions in dense urban scenarios. On the other hand, the additional complexity associated with mobility management and the frequent handoffs might introduce significant overhead and additional costs. As a consequence, the value proposition associated with our fmAP infrastructure requires a careful cost-benefit analysis, based on a characterization of the inherent tradeoffs between various network resources.

The first step towards this cost-benefit analysis is to compute the maximal feasible gain in terms of T2 data rate, as a function of the urban scenario, in terms of vehicles' density and average vehicles' speed, as well as physical and data-link layers parameters such as data rate and hand-off transition time. We provide the analytic framework to quantify the value of the proposed two-tier HetNet with mobile fmAPs. Note that our model does not deal with the actual cost of installation, and/or specific implementation issues at the physical (PHY) and data link (DL) layers.

The organization and main contributions of the paper are described in the following.

- In Sec. 2, we propose and conceptualize a new 2-tier communication infrastructure in which vehicles with wireless connectivity act as mobile relays. We consider a general setup in a high density urban scenario by abstracting out the attributes of the PHY and DL layers, and by specifying directly the data rate and handoff transition time (or handoff cost). We model two scenarios: a simpler one, with continuous time, which allows us to derive basic results on the performance of the system in terms of data rate; and a more advanced model, in a time slotted scenario, with an exponential path loss channel and a more realistic representation of the energy consumption of the system.
- In Sec. 3, we detail the strategies for choosing the next fmAP when a handoff occurs, in terms of minimum energy consumption, maximum data rate, or a tradeoff between these two goals. In the time slotted scenario, each choice affects the current performance as well as the performance for the future time slots, thus we model our strategy as a Markov decision process (MDP), which allows us to learn the best strategy by simply observing the parameters of the system.

• In Sec. 4, we derive analytically the network performance in the continuous time scenario. We provide an analytical framework to calculate the performance in terms of handoff rate and additional capacity in T2, as a function of the system parameters. This framework can be used as the basis for a more advanced performance analysis in a more complicated and realistic scenario. The main result in this section is presented in the form of a theorem, and the analytical proof is provided.

• In Sec. 5, we present a proposition which quantifies the average T2 data rate for the time slotted scenario with an exponential path loss model, and we provide the analytical framework for this more realistic scenario. We provide an abstraction of the PHY and DL layers, thus this analytical framework can be adapted to different network characteristics and can be used as a benchmark for the evaluation of future implementations of this network system.

• Then, in Sec. 6, we validate the proposed framework with simulations, presenting a case-of-study for a specific choice of urban scenario (in terms of density and speed of vehicles) and network system (in terms of protocol characteristics and data rate). For the continuous time scenario, we show that the detrimental effects of the frequent handoffs can be alleviated by an opportunistic selection strategy, confirming the insights given by the analytical results. For the time slotted scenario, we consider both average data rate and energy efficiency, showing that the MDP strategy can significantly outperform the other decision strategies when we are looking for a tradeoff between these two goals.

Finally, Sec. 7 concludes the paper and proposes some future research directions.

2. System Model

In our system model, we envision a two tiers network, where T1 is composed of the macro BSs providing service to the UEs, while T2 is a network of fmAPs, supporting T1 on an orthogonal band.

The two tiers differ in terms of 1) cost, and 2) availability. We assume that the cost of a local connection in T2 is significantly lower than in T1, since it allows for frequency reuse in another part of the cell. So, if both connections are available, the UE is always requesting services through T2. Regarding the availability of a low-cost T2 connection, it depends on the dynamics of the fmAPs that randomly arrive in the communication range of the UE and leave it after a certain time interval, thus requiring frequent horizontal handoffs. We assume that a constant time T_H must be spent for each horizontal handoff, during which the communication is interrupted. In cases in which no fmAP is available, the connection with T2 is momentarily suspended until the arrival of a new fmAP.

In this work, we focus on a single UE, deployed within the coverage range of a macro BS. Later on, we will briefly discuss how to consider a multi-user, multi-cell scenario. A graphical representation of our system model is depicted in Fig. 1. The UE, namely U, is assumed to be static and is represented as a cross,

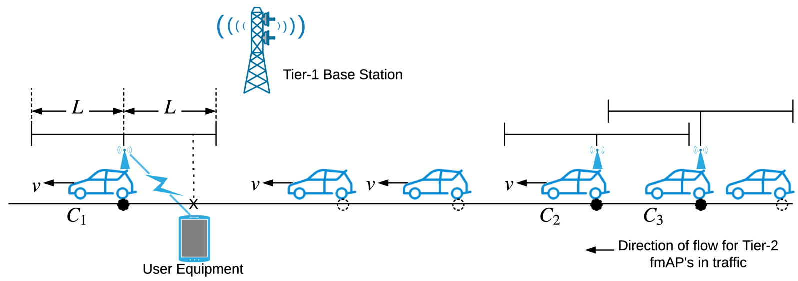

Figure 1: One-way fmAP communication model for T2.

while the macro BS is represented by a cell tower. The vehicles are represented with small circles, and only the black circles are vehicles equipped with an fmAP, namely F_i, with $i = 1, 2, \ldots$. Each vehicle is moving with a constant speed v in one of the W lanes along the street. The macro BS is always connected to the UE, while an fmAP can connect to the UE only if its distance is smaller than L.

2.1. Problem formulation

In this paper, we are interested in providing a first cost-benefit analysis of this new tier composed of fmAPs. We do not deal with the choice between connecting to T1 or T2, which depends on the UE preferences in terms of cost and quality of service. Instead, we focus on T2, and we quantify the provided service for different fmAP selection strategies, requiring different levels of shared information, complexity and additional overhead.

We start with a simplified continuous time scenario. Here, the mathematical properties of Poisson processes allow an elegant yet insightful derivation of the average effective data rate, showing the importance of properly selecting the fmAPs. Energy consumption is not optimized here. Subsequently, we move on to a slotted time scenario. This allows us to relax some of the assumptions needed in the continuous time scenario, and to design an additional strategy able to find a tradeoff between effective data rate and energy consumption. In both scenarios, the UE is located along an urban street of the smart city.

2.2. Continuous time scenario

The dynamics of vehicles' arrivals in an urban setting is a complex process. In general, a stationary model is not suitable to describe the different traffic conditions, where the vehicle flow can vary significantly with the time of the day. By considering only a short time interval, we can indeed model the arrival process of fmAP vehicles with a Poisson process [2] of parameter λ, which can be approximated using one of the existing traffic estimation systems.

We assume that the topology in the proximity of each UE is known, and that a handoff can happen at any time. In this simplified scenario, we assume the presence of a power adaptation mechanism, such that the received power rate is constant if the distance $d(U, F_i)$ between the UE and the fmAP F_i is smaller or equal than $d_{\max}$, otherwise there can not be a connection with F_i. As a consequence, the energy consumption for each transmission increases as a function of the distance, and it is defined as

$$E = \psi d(U, F_i)^{\alpha} \ , \tag{1}$$

where α is the path loss exponent, and ψ is a constant. Since the received power rate is constant, also the received data rate is assumed to be constant and equal to B_r.

2.3. Time slotted scenario

In this case, the time is divided into time slots of duration T, and the length of the road segment within the UE coverage is equal to $L = 2d_{\max}$. We partition this segment into $2K$ sectors, where the length of each sector is $L_s = vT$. We consider the distance $d(U, F_i)$ between the UE U and the fmAP F_i to be approximately constant for the duration of one time slot. If F_i is in sector $k^{(F_i)}$, with $k^{(F_i)} = 1, 2, \dots, 2K$, then the distance to U is approximated by the average distance, which is $d(U, F_i) = |K - k^{(F_i)} + 0.5| L_s$.

We assume the same Poisson process described in Sec. 2.2, so at each time slot the probability that at least one fmAP arrives at the farthest sector that is in the range of connectivity of the UE is given by

$$P_T = 1 - e^{-\lambda T} \ . \tag{2}$$

We will use Eq. (2) in the following, for an easier comparison with the model in Sec. 2.2. We stress the fact that our model and the calculations in the time slotted case will not change if we do not assume a Poisson arrival process. The probability P_T can be a function of time (non stationary case), or it can also be dependent on the state of the system, i.e., on the number of fmAPs already present in the range of connectivity (relaxing the memoryless property). In any case, at each time slot every fmAP already present in the system will deterministically change its sector (in the direction of the traffic flow).

We highlight that this model is fundamentally different from the continuous time case. In this model, the UE needs to make a decision (choice of the fmAP) by observing the current state of the network (at time

[2] If the street is composed of W lanes, vehicles are arriving with a Poisson process of parameter λ_1 in each lane, and ρ is the fraction of vehicles equipped with an fmAP, then $\lambda = \rho \lambda_1 W$.

slot t), while the decision will be put into action in the next time slot $(t+1)$. The complete state at $t+1$ is unknown, since the arrival of a new fmAP is a random event. Furthermore, if we take into account both energy consumption and quality of service (QoS), the choice at time t may affect the performance of the system for more than one time slot in the future, as detailed in Sec. 3.2.

The channel between the fmAP and the UE is modeled by considering an exponential path loss, which is assumed to be constant within a given sector. We assume that the data rate (in bits) at the UE is a fixed fraction of the capacity of the channel, expressed as

$$B_r = \eta \mathcal{C} = \eta B_w \log_2 \left(1 + \frac{P_r}{B_w(I + N_0)}\right) , \quad (3)$$

where $\eta < 1$ is a constant, B_w is the bandwidth, I represents the interference and N_0 represents the noise. The power P_r received by the UE when the fmAP is transmitting with power P_s, is expressed as

$$P_r = \frac{P_s}{d(U, F_i)^\alpha} = \frac{P_s}{(|K - k^{(F_i)} + 0.5|L_s)^\alpha} , \quad (4)$$

where $k^{(F_i)}$ corresponds to the sector of the fmAP F_i. We assume the presence of perfect power adaptation, so the transmission power is expressed as a function of the corresponding sector $k^{(F_i)}$. Thus, when the distance increases, in order to receive the same power P_r, the transmission power $P_s(k^{(F_i)})$ should also increase proportionally, i.e.,

$$P_s(k^{(F_i)}) = P_r d(U, F_i)^\alpha = P_r L_s^\alpha |K - k^{(F_i)} + 0.5|^\alpha . \quad (5)$$

In this scenario, the energy consumption increases with the distance, while the data rate B_r remains constant.

Finally, in order to simplify the notation, in the following we assume that the handoff penalty is $T_H = T$, i.e., in correspondence to each handoff, the UE is disconnected for exactly one time slot, unless otherwise specified.

The strategies and the results are investigated for a single user, single cell scenario, but this simple scenario can be generalized with few modifications. In order to allow multiple users to be helped by the fmAPs in the same cell, it is reasonable to consider that a set of channels is available for the fmAP-UE connections, and that a single fmAP can superimpose signals on more than one of these orthogonal channels if its help is required by more than one UE at the same time. Regarding the inter-cell interference, which is represented by the term I in Eq. (3), it can be considered negligible if a proper channel assignment is made by the core network, following for example the inter-cell interference coordination (ICIC) principle in [20].

3. Strategies to select the fmAPs

In highly dynamic scenarios with high vehicle density, it is possible that two or more fmAPs are available for one UE. Since the distance to each fmAP changes rapidly, it is important to design an effective strategy to choose among the available fmAPs.

In the following, we first propose three simple strategies to be adopted in the continuous time scenario, where we do not take into account the energy consumption. Then, in the time slotted scenario, we discuss the two main parameters to be optimized (energy consumption and maximization of the data rate), and we derive a strategy to optimize the tradeoff between these two.

3.1. Continuous time scenario

We propose three strategies to select the fmAP in this simplified scenario.

i) Select the fmAP with minimum distance: S_m

The strategy S_m selects the fmAP with the best channel, which in our simplified model corresponds to the selection of the closest fmAP. It is possible that a new fmAP is chosen while the previous fmAP is still in the range of communication, thus potentially requiring a handoff before it is strictly necessary.

ii) Select a random fmAP in order to decrease the expected number of handoffs: S_r

According to S_r, the UE remains connected to the fmAP until the fmAP goes out of its connectivity range, even if a new fmAP is available and closer to the UE, thus avoiding unnecessary handoff events. When the previous fmAP goes out of range, the UE randomly selects a new fmAP among the available fmAPs, since it does not have additional information on their residual time of connectivity.

iii) Select the fmAP with a centralized controller: S_c

With S_c, we assume the presence of a centralized controller that is keeping track of the position of each vehicle equipped with an fmAP. Based on this information, the controller can select the optimal fmAP that will guarantee the longest connection time, thus minimizing the handoff rate.

3.2. Time slotted scenario

In this scenario, there are two criteria to select the best fmAP. Looking at a single time slot, in order to minimize both the energy consumption and the interference to the other users, the selection of the closest fmAP is preferable. On the other side, considering the cost in multiple time slots, we should consider that there is a handoff penalty to be paid every time a new fmAP is selected. In order to avoid it, once an fmAP has been selected, a possible strategy may consist in staying connected to that fmAP as long as the distance to that fmAP is $d < L$. In the following, we define a set of strategies[3] to choose the next fmAP based on one of the two criteria. We also describe a third strategy based on an MDP to optimize the tradeoff between these two criteria.

[3]The symbol S_m, as well as S_c, represents two different strategies, in the continuous time and in the slotted time models, respectively. In the following, these models are treated separately, so there is no room for ambiguity.

i) Minimization of the energy consumption: S_m

At each time slot t, if more than one fmAPs, namely F_i, with $i = 1, 2, \ldots$, will remain available also in the next time slot, the strategy $\mathrm{S_m}$ selects the fmAP $F^\star$ which requires the minimum transmission power, i.e.,

$$F^\star = \arg\min_{F_i} P_s(k^{(F_i)}) = \arg\min_{F_i} |K - k^{(F_i)} + 0.5|^\alpha = \arg\min_{F_i} |K - k^{(F_i)} + 0.5| \ , \tag{6}$$

where $k^{(F_i)}$ is the sector corresponding to F_i at time slot $t+1$. We should observe that we can not perfectly forecast the future state based on the current one. Thus, in the implementation of this technique, we assume that there will be no arrival at time slot $t+1$, i.e., the sector $k = 1$ will not have a new fmAP.

ii) Maximization of the time in T2: S_c

The goal of $\mathrm{S_c}$ is to maximize the connection time in T2, which in our scenario is equivalent to minimize the number of handoffs required, since a penalty (a fraction of time T_H without connectivity) is associated with each handoff. In order to do so, the UE remains connected to the fmAP until the fmAP goes out of its connectivity range, even if a new fmAP is available and closer to the UE, thus avoiding unnecessary handoff events. At each time slot t, if the selected fmAP is in the last sector, $k = 2K$, the UE should select a new fmAP among the available fmAPs for time slot $t+1$.

The choice of the fmAP according to $\mathrm{S_c}$ becomes the choice of the fmAP that is expected to remain longer in the connectivity range, i.e.,

$$F^\star = \arg\max_{F_i} |2K - k^{(F_i)}| = \arg\min_{F_i} k^{(F_i)} \ . \tag{7}$$

iii) Tradeoff between minimizing the energy and maximizing the time in T2: S_t

In all the previous cases, the choice of the best fmAP depends only on the current position of all the fmAPs in the connectivity range. If instead we consider both criteria and we want to maximize the performance of the network over an infinite horizon, we should consider that the choice of one fmAP at time t will also influence the future performance, since each handoff in T2 corresponds to a cost for the UE.

In order to find the optimal choice, let's first define the state of this system. The state $\mathbf{s}_t$ at time slot t is completely defined by the pair $\mathbf{s}_t = (\mathbf{z}_t, z_t^\star)$, where $\mathbf{z}_t = [z_t^{(1)}, z_t^{(2)}, \ldots, z_t^{(2K)}]$ is a binary vector of dimension $2K$, such that if at least one fmAP is present in sector k at time t then $z_t^{(k)} = 1$, otherwise $z_t^{(k)} = 0$. We denote a subset of this vector as $\mathbf{z}_t^{[k_1:k_2]} = [z_t^{(k_1)}, z_t^{(k_1+1)}, \ldots, z_t^{(k_2)}]$, with $1 \leq k_1 \leq k_2 \leq 2K$.

The scalar $z_t^\star \in \{0, 1, \ldots, 2K\}$ denotes the position of the fmAP connected to the UE, with $z_t^\star = 0$ if the UE is currently not connected to any fmAP. We note that the state at time t, i.e., $\mathbf{s}_t = (\mathbf{z}_t, z_t^\star)$, is independent on its past once the information about the state at time $t-1$ is known, i.e., $\mathbf{s}_{t-1} = (\mathbf{z}_{t-1}, z_{t-1}^\star)$, thus it is possible to model this system as an MDP.

The UE should choose to which fmAP to connect to in the next time slot. Formally, the set of actions is $\mathcal{A} = \{1, \ldots, 2K\}$, where the action $a_t = k$ indicates that the UE is connecting with the fmAP that will be in sector k in the next time slot. A reward is associated with the action $a_t = k$ if there is a fmAP in sector k in the next time slot.

The transition probability at each time step t depends on the action a_{t-1} taken by the UE in the previous time step, and on the random arrival of a new fmAP in the sector $k = 1$. The probability of arriving in state $\mathbf{s}_t$, conditioned by the previous state $\mathbf{s}_{t-1}$ and the chosen action a_{t-1} is defined as

$$\mathrm{P}_{a_{t-1}}\left[\mathbf{s}_t | \mathbf{s}_{t-1}\right] = \mathrm{P}_{a_{t-1}}\left[\left([z_t^{(1)}, \mathbf{z}_t^{[2:2K]} = \mathbf{z}_{t-1}^{[1:2K-1]}], a_{t-1}\right) | \left(\mathbf{z}_{t-1}, z_{t-1}^{\star}\right)\right] = \begin{cases} P_T & \text{if } z_t^{(1)} = 1 \\ 1 - P_T & \text{if } z_t^{(1)} = 0. \end{cases} \tag{8}$$

Clearly, the transitions probabilities towards states with $z_t^* \neq a_{t-1}$ are equal to 0. The UE remains connected to the same fmAP if action $a_t = z_t^{\star} + 1$ is chosen.

For each time step t, we can now define the efficiency of the transmission as the ratio between the measured data rate $B_r(t)$ and the power consumed, i.e.,

$$\xi_t = \frac{B_r(t)}{B_r} \frac{P_s^{\min}}{P_s(t)} , \tag{9}$$

where B_r is the constant data rate in case there is no handoff, and $P_s^{\min}$ is the minimum transmitting power, in the case in which the fmAP is in one of the two sectors adjacent to the UE. From Eq. (5), we have that $P_s^{\min} = P_r L_s^{\alpha} 0.5^{\alpha}$.

With an abuse of notation, by $P_s(t)$ we mean the transmission power at time t, which depends on the position k of the fmAP, as in Eq. (5). The reward at time step t can be expressed as a function of the states at time $t-1$ and t, and of the action taken at $t-1$ as

$$R_{a_{t-1}}(\mathbf{s}_t, \mathbf{s}_{t-1}) = (\xi_t)^{\nu} = \begin{cases} \frac{|0.5|^{\alpha\nu}}{|K-k+0.5|^{\alpha\nu}} & \text{if } z_t^{\star} = k-1, a_t = k \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

In other words, the reward is greater than zero only if the UE was connected to the same fmAP also in the previous time slot, as detailed in Sec. 2.3. The maximum reward is $R = 1$, assigned in the case the transmitting power $P_s(t) = P_s^{\min}$. The parameter ν is used to tune the metric in Eq. (10): for $\nu \ll 1$ the only goal is to maximize the effective time in T2 (as for $\mathrm{S_c}$), while for $\nu \gg 1$ the only goal is to minimize the energy consumption (as for $\mathrm{S_m}$). For $\nu = 1$ the goal is simply to maximize the efficiency in Eq. (9).

The optimal policy for the MDP, namely $\boldsymbol{\pi}^*$, provides the optimal action $\pi_s \in \mathcal{A}$ for each state $\mathbf{s}$ of the system in order to maximize the expected total reward over a possibly infinite horizon. Even if multiple optimal policies may exist (providing the same expected reward), we can rely on dynamic programming to find one optimal solution.

In order to find $\boldsymbol{\pi}^*$, we define an initial policy $\boldsymbol{\pi}^0$ and an initial vector of values $\boldsymbol{V}^0$, which contains the expected reward $V_{\mathbf{s}}$ from each state $\mathbf{s}$. The length of these vectors corresponds to the number of possible states, that in our case is equal to $n = K2^{2K+1}$. The initial values for these vectors can be chosen arbitrarily, e.g., they can all be set to zero.

The value of $\boldsymbol{\pi}^*$ can be obtained via dynamic programming by iteratively updating these two equations for all $\mathbf{s}$:

$$\pi_{\mathbf{s}}^{i+1} = \arg\max_{a} \left\{ \sum_{\mathbf{s}'} \mathrm{P}_a \left[\mathbf{s}|\mathbf{s}'\right] \left(R_a \left(\mathbf{s}, \mathbf{s}'\right) + \gamma V_{\mathbf{s}'}^{i}\right) \right\}, \tag{11}$$

$$V_{\mathbf{s}}^{i+1} = \sum_{\mathbf{s}'} \mathrm{P}_{\pi_{\mathbf{s}}^{i+1}} (\mathbf{s}|\mathbf{s}') \left(R_{\pi_{\mathbf{s}}^{i+1}} \left(\mathbf{s}, \mathbf{s}'\right) + \gamma V_{\mathbf{s}'}^{i} \right), \tag{12}$$

where $\gamma \in (0, 1)$ is the discount factor, a parameter of the algorithm that influence also the rate of convergence. The iteration will continue for $i = 1, 2, \ldots$ until convergence, i.e., until $\|\boldsymbol{V}^{i+1} - \boldsymbol{V}^{i}\|_1 < \epsilon$, where $\|\cdot\|_1$ indicates the norm-1 of a vector, and ϵ is a small constant, $\epsilon \ll 1$.

4. Analysis of Continuous Time Scenario

In the previous sections, we described the scenarios of interest and the techniques to choose the next fmAP. In this section, we present the first part of the analysis, considering the case of continuous time, detailed in Sec. 2.2. As a first step, we evaluate the length of a T2 connection round, i.e., the contiguous time in which the UE is connected to at least one fmAP. A T2 connection round ends when no fmAP is in the connection range of the UE. It is defined, independently form the strategy chosen, as

$$T_2 \triangleq \sum_{j=1}^{M^{(\mathrm{S})}+1} \tau_j^{(\mathrm{S})}, \tag{13}$$

where $M^{(\mathrm{S})}$ is the number of horizontal handoffs, $M^{(\mathrm{S})} + 1$ is the total number of fmAPs serving the UE, $\tau_j^{(\mathrm{S})}$ is the time interval in which the UE is connected to the j^{th} fmAP, and S is the strategy chosen. At the end of a T2 connection round, a T1 connection round of length T_1 starts, during which the UE is connected to T1. A new T2 connection round starts upon a new fmAP arrival.

In order to take the detrimental effect of the horizontal T2 handoffs into account, we define the effective T2 ratio, $R_2^{(\mathrm{S})}$, which is the ratio between the expected effective time spent in T2 and the sum of the expected connection times of T_1 and T_2, i.e.,

$$R_2^{(\mathrm{S})} \triangleq \frac{E\left[T_2 - M^{(\mathrm{S})} T_H\right]}{E\left[T_1 + T_2\right]}. \tag{14}$$

In (13), $M^{(\mathrm{S})}$, $\tau_j^{(\mathrm{S})}$, and $R_2^{(\mathrm{S})}$ depend on the specific selection strategy adopted, whereas T_2 is the same for the three continuous time strategies considered. For simplicity, in the next section, we adopt the strategy $\mathrm{S_m}$ to derive an expectation expression for T_2.

4.1. Expected duration of a T2 connection round

In terms of the T2 connection time, $\mathrm{S_m}$ is equivalent to a strategy that connects to a new arriving fmAP and maintains this connection until the next arrival. Therefore, the connection time $\tau_j^{(\mathrm{S}_m)}$ is equal to the fmAP interarrival time, i_j, for $j = 1, 2, \ldots, M^{(\mathrm{S}_m)}$. For the last fmAP in a T2 connection round, we have $\tau_{M^{(\mathrm{S}_m)}+1}^{(\mathrm{S}_m)} = T_M$, where $T_M \triangleq \frac{L}{v}$.

After the arrival of the last fmAP, we have a time interval of length T_M without new arrivals, followed by a vertical handoff to T1. Since the arrivals of the fmAPs constitute a Poisson process, the probability that an fmAP is the last one of a connection round can be expressed as

$$P_V \triangleq \mathrm{P}\left\{N\left(t - T_M, t\right] = 0\right\} = e^{-\lambda T_M} \ , \tag{15}$$

where $N\left(t_1, t_2\right]$ is the number of fmAPs that enter the connectivity region of the UE in the time interval $\left(t_1, t_2\right] = \left(t : t_1 < t \leq t_2\right)$. On the other hand, the probability of a horizontal handoff in T2 at the end of a connection to an fmAP is equal to $1 - P_V$. Using this probability, we can identify the probability mass function (pmf) for $M^{(\mathrm{S}_m)}$. Indeed, $M^{(\mathrm{S}_m)}$ is a geometric random variable such that $\mathrm{P}\left\{M^{(\mathrm{S}_m)} = m\right\} = (1 - P_V)^m P_V$, for $m \geq 0$. The expected value of $M^{(\mathrm{S}_m)}$ is simply

$$E\left[M^{(S_m)}\right] = \frac{1 - P_V}{P_V} \ . \tag{16}$$

Using this expectation, we can evaluate the expected time spent in T2 by using the iterated expectation over (13). The duration of the consecutive time interval with a T2 connection, conditioned on the value of $M^{(\mathrm{S}_m)}$, is

$$E\left[T_2 \middle| M^{(\mathrm{S}_m)} = m\right] = E\left[\sum_{j=1}^{m} i_j \middle| i_1 \leq T_M, \ldots, i_m \leq T_M, i_{m+1} > T_M\right] + T_M = T_M + mE\left[i_j | i_j \leq T_M\right]. \tag{17}$$

Since the interarrival time i_j is exponentially distributed with parameter λ, we have

$$E\left[i_j | i_j \leq T_M\right] = \frac{1 - P_V\left(1 + \lambda T_M\right)}{\lambda\left(1 - P_V\right)}. \tag{18}$$

Using (16), (17), and (18), we obtain the following

$$E\left[T_2\right] = E_{M^{(\mathrm{S}_m)}}\left[E\left[T_2 \middle| M^{(\mathrm{S}_m)}\right]\right] = \frac{1 - P_V}{\lambda P_V}. \tag{19}$$

4.2. *Expected effective T2 data rate using S_m*

The data rate is assumed to be constant and equal to B_r, for the duration of the T2 connection round, with the exception of the time spend in handoffs.

The time spent in T1 depends only on the first arrival time after the end of a communication round in T2. Due to the memoryless property of Poisson arrivals, we have $E[T_1] = \lambda^{-1}$, where T_1 is the random arrival time of the next fmAP, counted from the time in which the connection to T2 ends. By using this observation, we state the following proposition.

Proposition 1. *The expected effective data rate with strategy S_m is*

$$\overline{B_r^{(S_m)}} = B_r \frac{E\left[T_2 - M^{(S_m)} T_H\right]}{E\left[T_1 + T_2\right]} = B_r\left((1 - P_V) - T_H(1 - P_V)\lambda\right) , \quad (20)$$

which is given by the data rate B_r multiplied by the effective T2 ratio, calculated using (14), (16), and (19). The numerator in the first equation represents the time spent in T2 minus the time spent into $M^{(S_m)}$ horizontal handoffs, where we assume that the consecutive horizontal penalty intervals are non-intersecting in time, i.e., we assume that T_H is significantly smaller than the average interarrival time for fmAP's. [4]

We observe that the first term in the second equation in (20), $(1 - P_V)$, quickly converges to 1 as λ increases, see (15). The second term instead represents the penalty due to the handoffs, which becomes more significant as λ increases. This result provides an easy to interpret relationship between the system parameter and the effective data rate for strategy S_m in the continuous time scenario.

4.3. *Expected effective T2 data rate using S_c*

The S_c selection strategy makes use of the information from a centralized controller in order to minimize the number of handoffs in T2, while the expected time in T2, given in (19), remains unchanged.

In order to explain the behavior of S_c, we show an example of a T2 connection round in Fig. 2. Among the 7 fmAPs that pass by the UE in this connection round, only four of them are chosen (they are highlighted in dotted circles in the figure), while there is no connection to the other fmAPs. In Fig. 2, we report the time interval of connection to the j^{th} fmAP, $\tau_j^{(S_c)}$. We further define the remaining service time until the next selected fmAP arrival t_j, which is defined as the time interval between the moment in which the UE looses the connection to the $(j-1)^{\text{th}}$ selected fmAP and the arrival time of the $(j+1)^{\text{th}}$ selected fmAP. In this example, the first connected fmAP is C_0. At the end of the connection with C_0, C_3 is selected, since it

[4] This assumption is in compliance with the state-of-the-art wireless communication standards. E.g., in LTE radio interfaces, the upper limit for the control plane latency is 100 ms, so $T_H \leq 100$ ms. In a practical scenario in which the UE observes less than one vehicle equipped with an fmAP per second, $T_H \ll \frac{1}{\lambda}$ is satisfied.

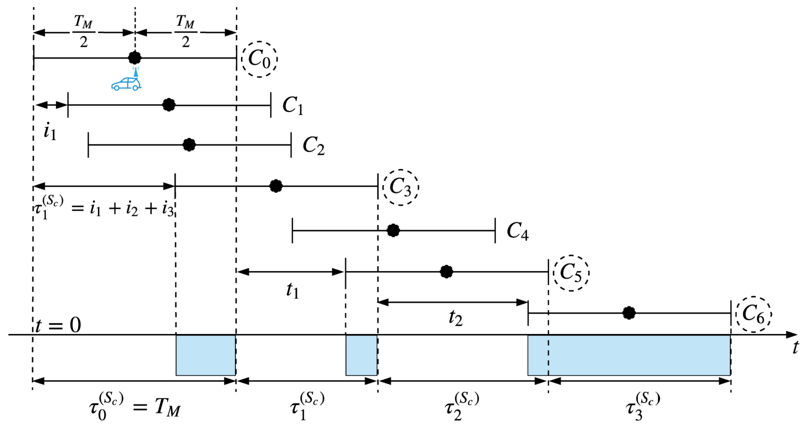

Figure 2: Time of arrivals and S_c operation for a given T2 instance.

is the last arrived fmAP, or, equivalently, the one with the longest residual time of connectivity to the UE. The shaded intervals at the bottom of the figure are the time intervals at the end of the connection time of an fmAP in which no arrivals are observed.

As shown in this example, none of the fmAPs (if any) that arrive in the remaining time t_j are used by the UE, and this translates into a decrease in the handoff rate. We can now approximate the expected effective T2 ratio under S_c, and express the average data rate achieved in T2.

Theorem 1. *The average data rate in T2 with strategy S_c can be approximated by*

$$\overline{B_r^{(S_c)}} \simeq B_r\left(1-P_V\right) - B_r \frac{2\ \left(1-P_V\right) T_H}{E\left[\tau_1^{(S_c)} + t_1 \middle| M^{(S_c)} \geq 1\right] + \frac{2}{\lambda}}. \tag{21}$$

The proof of this theorem is reported in the appendix, in Sec. 8.1.

As in the case of S_m in (20), the term $(1-P_V)$ rapidly converges to 1 as λ increases. The second term, which represents the handoff penalty, converges to T_H/T_M, thus $\overline{B_r^{(S_c)}} > \overline{B_r^{(S_m)}}$ for sufficiently large values of λ. This demonstrates analytically the advantage of strategy S_c over S_m.

5. Analysis of Slotted Time Scenario

The case with continuous time is very informative in terms of the expected coverage provided by the fmAPs, but it does not provide a technique to optimize the tradeoff between time of connectivity and energy consumption, for which we should switch to the time slotted model. In this section, we derive an analytical result on the expected effective time in T2, which is the ratio of time spent in an effective T2 connection. We remind that in our model we should keep into account the time in which there are no fmAPs to connect, and the time spent in handoffs between consecutive fmAPs. This result provides us with important insights to discuss the numerical comparison with the MDP technique.

5.1. Expected duration of a T2 connection round

We start our analysis by calculating the expected duration of a T2 connection round, i.e., the consecutive time during which at least one fmAP is available. As in the continuous time case, this time does not depend on the specific fmAP selection strategy, so the analysis can be simplified. We assume that the UE is always connected to the closest fmAP (as in $\mathrm{S_m}$), and in case there are two fmAPs at the same distance, a handoff is occurring to connect to the new fmAP.

We observe that the first fmAP will be connected for at least K time slots, while also the last fmAP will be connected for at least K time slots. We denote by $M \geq 0$ the number of handoffs in T2 during the connection round, so according to the model in Sec. 2 we have exactly M time slots in a connection round which are dedicated to T2 horizontal handoffs. We also define z_j to be the number of slots with no arrivals between the arrival of the j^{th} and the $j+1^{\text{th}}$ fmAP.

The duration of a T2 connection round can thus be written as

$$T_2 = \left(K + M + \sum_{j=1}^{M} z_j + K \right) T \ , \tag{22}$$

where T is the duration of a time slot. In order to calculate the expectation over T_2, we first observe that the value of M and of all the z_j are independent, once the probability P_T of one arrival in the system in one time slot is given. The expectation over T_2 can thus be obtained by calculating the expectation over all the variables in Eq. (22). The expectation over M can be expressed as

$$E\left[M\right] = \frac{1-(1-P_T)^{2K}}{(1-P_T)^{2K}} \ , \tag{23}$$

by using a similar reasoning as the one used to calculate Eq. (16). The expectation over z_j can be written as

$$E\left[z_j\right] = \frac{(1-P_T)\left(1-(1-P_T)^{2K}\right) - 2K(1-P_T)^{2K} P_T}{P_T\left(1-(1-P_T)^{2K}\right)} \tag{24}$$

by identifying the pmf of z_j and following a series of derivation steps. We can now calculate the expectation over both side of Eq. (22), by using Eq. (23) and Eq. (24), and we obtain

$$E\left[T_2\right] = \frac{1-(1-P_T)^{2K}}{(1-P_T)^{2K} P_T} T \ , \tag{25}$$

which is the expected duration of a T2 connection round in the time slotted case, for all the selection strategies considered.

5.2. *Expected effective T2 data rate using S_m*

We can now calculate the ratio of the expected effective time spent in T2 to the total time of communication, for the strategy S_m. This is given by the ratio of the effective time for one T2 connection round to the sum of the expected time in a T1 and a T2 round.

The effective time for one T2 connection round is simply the T2 connection round, expressed in Eq. (25), minus the time spent for the handoffs between fmAPs, which can be calculated by using $E[M]$ expression given in Eq. (23), and minus the time spent in the 2 handoffs that mark the boundaries of the T2 connection round. It is equal to

$$E\left[T_2^{\mathrm{eff}}\right] = \frac{T}{P_T}\frac{1-(1-P_T)^{2K}}{(1-P_T)^{2K}} - T_H\frac{1+(1-P_T)^{2K}}{(1-P_T)^{2K}} \,. \tag{26}$$

The expected time in a T1 connection round is the time between when the last fmAP of a T2 connection round is leaving the system, and the time in which a new fmAP is entering the system, i.e., $E[T_1] = T/P_T$.

The average data rate is expressed in the following proposition.

Proposition 2. *The average T2 data rate with strategy S_m for the time slotted case is*

$$\overline{B_r^{(S_m)}} = B_r\frac{E\left[T_2^{\mathit{eff}}\right]}{E[T_1]+E[T_2]} = B_r\left(1-(1-P_T)^{2K} - P_T\frac{T_H}{T}\left(1+(1-P_T)^{2K}\right)\right), \tag{27}$$

which is given by the data rate B_r, multiplied by the expected effective ratio of time in T2 and divided by the total communication duration.

We observe that, if the handoff penalty lasts for the duration of one time slot, i.e., $T_H = T$, and if we have one new fmAP in each time slot, i.e., $P_T = 1$, then the average T2 data rate goes to zero, since according to the strategy the fmAP to which the UE is connected changes at each time slot, as expected.

6. Numerical Results

In this section, we show the performance of the proposed system for both the continuous and the time slotted scenario. The simulations aim to showcase the achievable performance of this networking system and the validity of the analytical results.

6.1. *Simulation Setup*

In the simulations, the wireless technology for the connection between the UE and the fmAPs is the LTE-U [16], working at 5 GHz, with 20 MHz bandwidth per stream. The outdoor minimum data rate is $B_r = 7.2$

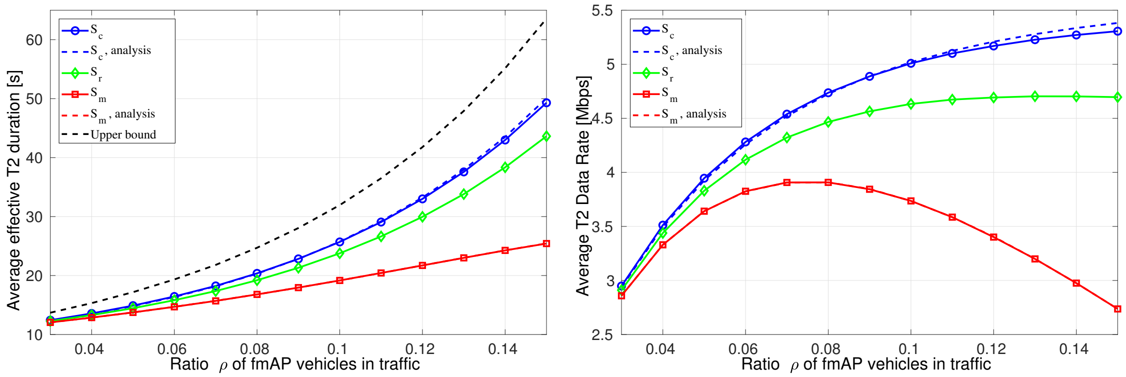

Figure 3: (a) Expected effective time and (b) average data rate in T2 for the continuous time model. The simulation results perfectly match the analytical results.

Mbps at a maximum range of 250 m [21]. In order to take into consideration the detrimental effects of mobility, we assume that $L = 100$ m. The simulations are performed in MATLAB, with a constant data rate B_r if the UE is within a distance L from the corresponding fmAP, and no connection otherwise.

The results are obtained as a function of ρ, the fraction of vehicles equipped with an fmAP. The other simulation parameters are: the speed of vehicles, $v = 20$ m/s, $W = 4$ lanes, and the arrival rates of vehicles in each lane, $\lambda_1 = 0.5$ vehicles/s. In the continuous time scenario, the maximum connection time to an fmAP is $T_M = 10$ s, and the time needed for a horizontal handoff in T2 is $T_H = 2$ s, while a connection to T1 is always available. The total simulated time is 10^8 s for each simulation. In the time slotted scenario, we set the duration of one time slot to $T = 1$ s, which determines $2K = 10$ sectors within the UE's coverage range. Moreover, the path loss exponent is $\alpha = 2.5$ and the total simulated time in this case is 50 hours.

6.2. Continuous Time Scenario

In Fig. 3-(a), we show the expected effective time of a T2 connection for $\mathrm{S_m}$, $\mathrm{S_r}$, and $\mathrm{S_c}$, with the corresponding analytical results for $\mathrm{S_m}$ and $\mathrm{S_c}$. The upper bound in the figure is based on (19), obtained by setting the cost of a horizontal handoff to $T_H = 0$. We observe that, in the case of a high density of fmAPs ($\rho = 0.15$), the effective time in T2 is almost doubled using the $\mathrm{S_c}$ strategy, as compared to the $\mathrm{S_m}$ strategy.

The $\mathrm{S_r}$ selection strategy performs close to $\mathrm{S_c}$, providing a valid alternative in cases for which a centralized controller is unavailable. We also observe that both the exact analysis results given in Sec. 4.2 for $\mathrm{S_m}$ (dashed curve) and the approximate result for $\mathrm{S_c}$ derived from Thm. 1 (dotted curve) closely follow the simulation results.

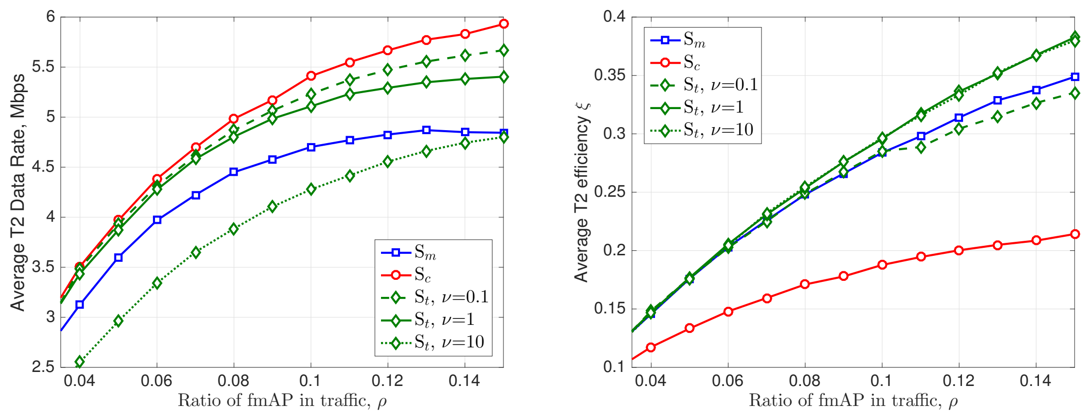

Figure 4: (a) Average data rate and (b) efficiency of the transmissions in T2 for the time slotted model.

In Fig. 3-(b), we evaluate the effective T2 data rate for the strategies considered. We observe that for S_m, as the density of the fmAP increases for $\rho > 0.07$, the T2 data rate decreases. This is due to the fact that this technique does not limit the number of handoffs, but it just connects to the closest fmAP. The cost of all these handoffs results in a significant loss in the average data rate. On the other hand, if we adopt S_r or S_c, the data rate increases also for $\rho > 0.07$.

6.3. Time Slotted Scenario

In Fig. 4-(a) we show the average data rate in T2, as a function of the ratio of fmAPs in the traffic, ρ, for S_m, S_c, and for S_t, where we compare three values for $\nu \in \{0.1, 1, 10\}$. The analytical results have also been compared for S_m, but they are not reported in the figure since they match the simulation results (with an error below 1 %).

We observe that in the case of $\rho = 0.15$, the S_c technique can outperform the S_m technique of about 25 %. As expected, S_t can come closer to S_c, and its data rate is only 5 % less than S_c (for $\nu = 0.1$), or 10 % less (for $\nu = 1$). On the other side, the strategy S_t with $\nu = 10$ performs worse than S_m in terms of average data rate in T2, and it reaches the same performance of S_m only for $\rho = 0.15$, i.e., the case with the highest fmAP arrival rate.

It is indeed interesting to investigate also the tradeoff between the average data rate and the efficiency, which is depicted for the same scenario in Fig. 4-(b). This time the performance of S_m is better than S_c (of about 66 %). S_t, for $\nu = 1$ and $\nu = 10$ can do even better, with a performance increase of about 80 % with respect to S_c. If both metrics are considered together, we obtain the best tradeoff with S_t for $\nu = 1$, which is the strategy that is making each decision in order to maximize the efficiency of the transmissions, described in Eq. (9).

7. Conclusions and Future Works

In this paper we presented a high density network scenario in which smart and connected vehicles are equipped with fmAPs and constitute a mobile out-of-band relay infrastructure to support the macro BS using less costly unlicensed spectrum. We provided the first steps toward a comprehensive cost-benefit analysis of this architecture by investigating the system performance in two fundamentally different system models: a continuous time and slotted time model. For the simpler model, the continuous time model, we designed three techniques to select an fmAP (if more than one is available). In this simplified scenario we computed the maximal feasible gain in the data rate as a function of the vehicle density, average vehicle speeds, handoff overhead cost, as well as physical layer characteristics.

In the slotted time model we introduced a more general fmAP arrival process and a more realistic communication channel model with an exponential path loss. This model allows the study of the tradeoff between minimizing the energy consumption and maximizing the data rate of the system. We propose an optimal solution based on a Markov decision process, and we discussed the possibility of tuning the model depending whether the priority is on reducing the energy consumption or on increasing the data rate. The simulations confirmed the validity of the analytical results. For the continuous time model, we showed that, with a random choice of fmAPs, we can achieve performance close to that observed with a centralized controller, in terms of average data rate. For the time slotted model, instead, we showed that the MDP model is able to achieve a much better tradeoff in terms of average data rate and efficiency of the transmissions, as compared to the other two models.

In a future work, we plan to extend our model to allow simultaneous connections to more than one fmAPs. Furthermore, we aim to evaluate the architecture by implementing a wireless software defined network testbed, as in [22], to separate the two tiers of communication and provide the control information needed by the proposed centralized fmAP selection strategy. Among the possible applications for our proposed model, we plan to investigate the retrieval of measurement data from smart meters in a smart city environment, as well as the transmission and distribution of this data via the secondary tier of connected vehicles. Our approach is particularly attracting for these applications, due to the low cost of the out-of-band data transmissions, as well as the low energy consumption, due to the reduced distance of the mobile fmAP, as compared to the distance to the cellular BS.

8. Appendices

8.1. Proof of Theorem 1

We first present the following two lemmas:

Lemma 1. *The expectation of service time* $\tau_1^{(S_c)}$ *is*

$$E\left[\tau_1^{(S_c)}|M^{(S_c)} \geq 1\right] = \frac{\lambda T_M - \left(1 - e^{-\lambda T_M}\right)}{\lambda\left(1 - e^{-\lambda T_M}\right)}. \tag{28}$$

Lemma 2. *The expectation of* t_1 *is*

$$\begin{aligned} E\left[t_1|M^{(S_c)} \geq 1\right] &= \frac{\lambda}{e^{\lambda T_M} - 1}\int_0^{T_M} \frac{s_1 e^{\lambda s_1}}{1 - e^{\lambda s_1}} ds_1 - \frac{1}{\lambda} \qquad (29) \\ &= \frac{1}{\lambda\left(e^{\lambda T_M} - 1\right)}\left(Li_2\left(e^{\lambda T_M}\right) - \frac{\pi^2}{6} + 1 + + e^{\lambda T_M}\left(\lambda T_M - 1\right) + \lambda T_M \log\left(1 - e^{\lambda T_M}\right)\right) - \frac{1}{\lambda}, \end{aligned}$$

where $\mathrm{Li}_2(x) \triangleq \sum_{k=1}^{\infty} \frac{x^k}{k^2}$ is the polylogarithm function of order 2. The proofs of the these two lemmas are given in our technical report [23].

For a given value $M^{(\mathrm{S}_c)} = n$, the time intervals in which new fmAP arrivals occur are disjoint time intervals of length $\tau_1^{(\mathrm{S}_c)}, t_1, t_2, \ldots, t_{n-1}$, as shown in Fig. 2. With an abuse of notation, we drop the superscript (S_c) in $\tau_i^{(\mathrm{S}_c)}$, and in the following we write $\tau_i = \tau_i^{(\mathrm{S}_c)}$, unless specified.

The number of unserved fmAPs, U, in a T2 round satisfies

$$E\left[U|M^{(\mathrm{S}_c)} = n\right] = E\left[U^{\tau_1} + \sum_{j=1}^{n-1} U^{t_j}\right],$$

where U^{τ_1} and U^{t_j} are the number of unserved fmAPs in the time intervals τ_1 and t_j, respectively. Based on the iterated expectations over these random service times, we obtain

$$\begin{aligned} E\left[U|M^{(\mathrm{S}_c)} = n\right] &= E\left[E_{U|\tau_1, t_j, M^{(\mathrm{S}_c)}=n}\left[U^{\tau_1} + \sum_{j=1}^{n-1} U^{t_j}\right]\right] \\ &= \lambda\left(E\left[\tau_1|M^{(\mathrm{S}_c)} = n\right] + \sum_{j=1}^{n-1} E\left[t_j|M^{(\mathrm{S}_c)} = n\right]\right), \end{aligned} \tag{30}$$

where we use the fact that $E\left[U^{\tau_1}|\tau_1 = s_1\right] = \lambda s_1$. A similar argument is valid for t_j as well.

We can approximate (30) by evaluating $E\left[t_j|M^{(\mathrm{S}_c)} = n\right]$ for only a few values of j. In particular, we use

$$E\Big[U|M^{(\mathrm{S}_c)}{=}n\Big]{\simeq}n\lambda\frac{E\big[\tau_1|M^{(\mathrm{S}_c)}{\geq}1\big]{+}E\big[t_1|M^{(\mathrm{S}_c)}{\geq}1\big]}{2}, \tag{31}$$

where the term $\left(E\left[\tau_1|M^{(\mathrm{S}_c)}\geq 1\right]+E\left[t_1|M^{(\mathrm{S}_c)}\geq 1\right]\right)/2$ is an approximation for the average number of unserved fmAPs between two consecutive horizontal handoffs. We observe that this approximation is asymptotically tight by investigating (28) and (29).

On the other hand, the number of handoffs for $\mathrm{S_m}$ and $\mathrm{S_c}$ in a T2 round satisfy

$$E\left[M^{(\mathrm{S}_m)}\right]=E_{M^{(\mathrm{S}_c)}}\left[E\left[U+M^{(\mathrm{S}_c)}|M^{(\mathrm{S}_c)}\right]\right]\simeq E\Big[M^{(\mathrm{S}_c)}\Big]\left(\lambda\frac{E\big[\tau_1|M^{(\mathrm{S}_c)}{\geq}1\big]{+}E\big[t_1|M^{(\mathrm{S}_c)}{\geq}1\big]}{2}{+}1\right), \tag{32}$$

where we use (31). Solving it for $E\left[M^{(\mathrm{S}_c)}\right]$ in (32) we obtain

$$E\left[M^{(\mathrm{S}_c)}\right]\simeq\frac{2\;E\left[M^{(\mathrm{S}_m)}\right]}{\lambda\left(E\left[\tau_1+t_1|M^{(\mathrm{S}_c)}\geq 1\right]\right)+2}, \tag{33}$$

where the denominator follows from the results of Lemmas 1 and 2. For $\mathrm{S_c}$, the expected effective data rate in T2 is

$$\overline{B_r^{(\mathrm{S}_c)}}=B_r\frac{E\left[T_2\right]-E\left[M^{(\mathrm{S}_c)}\right]T_H}{E\left[T_1\right]+E\left[T_2\right]}, \tag{34}$$

where $E\left[T_1\right]=\lambda^{-1}$ and $E\left[T_2\right]=\frac{1-P_V}{\lambda P_V}$. Finally, we plug (33) into (34).

References

[1] T. Aktas, G. Quer, T. Javidi, R. Rao, From Connected Vehicles to Mobile Relays: Enhanced Wireless Infrastructure for Smarter Cities, in: IEEE GLOBECOM, 2016.

[2] Metrolab Network, `http://metrolabnetwork.org`, Last acc.: May 2018.

[3] A. K. Gupta, X. Zhang, J. G. Andrews, SINR and Throughput Scaling in Ultradense Urban Cellular Networks, IEEE Wireless Communications Letters 4 (6) (2015) 605–608.

[4] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, F. Gutierrez, Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!, IEEE Access 1 (2013) 335–349.

[5] O. Galinina, A. Pyattaev, S. Andreev, M. Dohler, Y. Koucheryavy, 5G Multi-RAT LTE-WiFi Ultra-Dense Small Cells: Performance Dynamics, Architecture, and Trends, IEEE Journal on Selected Areas in Communications 33 (6) (2015) 1224–1240.

[6] F. Librino, M. Renda, P. Santi, Multihop Beaconing Forwarding Strategies in Congested IEEE 802.11p Vehicular Networks, IEEE Trans. on Veh. Technol. 65 (9) (2016) 7515–7528.

[7] T. G. McGiffen, S. Beiker, A. Paulraj, Motivating Network Deployment: Vehicular Communications, IEEE Veh. Technol. Mag. 12 (3) (2017) 22–33.

[8] J. Zhao, G. Cao, VADD: Vehicle-Assisted Data Delivery in Vehicular Ad Hoc Networks, IEEE Trans. on Veh. Technol. 57 (3) (2008) 1910–1922.

[9] X. Zhu, Y. Li, D. Jin, J. Lu, Contact-Aware Optimal Resource Allocation for Mobile Data Offloading in Opportunistic Vehicular Networks, IEEE Trans. on Veh. Technol. 66 (8) (2017) 7384–7399.

[10] G. Aujla, R. Chaundhary, N. Kumar, J. J. P. C. Rodrigues, Data Offloading in 5G-Enabled Software-Defined Vehicular Networks: A Stackelberg-Game-Based Approach, IEEE Commun. Mag. 55 (8) (2017) 100–108.

[11] I. F. Akyildiz, J. Xie, S. Mohanty, A survey of mobility management in next-generation all-IP-based wireless systems, IEEE Wireless Communications 11 (4) (2004) 16–28.

[12] W. Bao, B. Liang, Stochastic Geometric Analysis of User Mobility in Heterogeneous Wireless Networks, IEEE Journal on Selected Areas in Communications 33 (10) (2015) 2212–2225.

[13] W. Li, C. Zhang, X. Duan, S. Jia, Y. Liu, L. Zhang, Performance Evaluation and Analysis on Group Mobility of Mobile Relay for LTE Advanced System, in: IEEE VTC Fall, 2012.

[14] R. Balakrishnan, X. Yang, M. Venkatachalam, I. F. Akyildiz, Mobile relay and group mobility for 4G WiMAX networks, in: IEEE Wireless Communications and Networking (WCNC), 2011.

[15] Z. Liao, X. Zhang, C. Feng, Mobile relay deployment based on Markov chains in WiMAX networks, in: IEEE GLOBECOM, 2014.

[16] Qualcomm Technologies, Inc., LTE in Unlicensed Spectrum: Harmonious Coexistence with Wi-Fi, White paper.

[17] G. Zafari, X. Wang, Cognitive Co-Existence of Unlicensed Wireless Networks through Beamforming, in: IEEE Vehicular Technology Conference (VTC-Fall), 2017.

[18] X. Ge, H. Cheng, M. Guizani, T. Han, 5g wireless backhaul networks: challenges and research advances, IEEE Network 28 (6) (2014) 6–11.

[19] S. Ng, W. Zhang, Y. Zhang, Y. Yang, G. Mao, Analysis of Access and Connectivity Probabilities in Vehicular Relay Networks, IEEE J. Select. Areas Commun. 29 (1) (2011) 140–150.

[20] M. Al-Shalash, F. Khafizov, Z. Chao, Uplink inter-cell interference coordination through soft frequency reuse, in: IEEE GLOBECOM, 2010.

[21] A. Matsumoto, K. Yoshimura, S. Aust, T. Ito, Y. Kondo, Performance evaluation of IEEE 802.11n devices for vehicular networks, in: IEEE Local Computer Networks (LCN), 2009.

[22] M. Danieletto, G. Quer, R. R. Rao, M. Zorzi, CARMEN: a cognitive networking testbed on android OS devices, IEEE Communications Magazine 52 (9) (2014) 98–107.

[23] T. Aktas, G. Quer, T. Javidi, R. R. Rao, Mobile relays for smart cities: mathematical proofs (2016). URL `https://arxiv.org/abs/1609.01322`