Title

From Connected Vehicles to Mobile Relays: Enhanced Wireless Infrastructure for Smarter Cities

ISBN

978-1-5090-1328-9

Authors

Aktas, Tugcan
Quer, Giorgio
Javidi, Tara
et al.

Publication Date

2016-12-01

DOI

10.1109/glocom.2016.7841850

From Connected Vehicles to Mobile Relays: Enhanced Wireless Infrastructure for Smarter Cities

Tugcan Aktas, Giorgio Quer, Tara Javidi, Ramesh R. Rao
University of California San Diego – La Jolla, CA 92093, USA.

***Abstract*—The increasing number of connected vehicles in densely populated urban areas provides an interesting opportunity to counteract the high wireless data demands in high density and highly mobile scenarios. 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. In particular, 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. The analytical and simulation results provide a first benchmark characterizing this architecture and the definition of guidelines for its future realistic study and implementation.

I. 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 technology, various smart city initiatives have been launched [1]. Clearly, smart cities require an ever expanding and evolving wireless infrastructure in terms of data rate, connectivity, and efficiency.

In the literature, several works advocate for the expansion of traditional network solutions [2] such as increasing bandwidth [3] 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) [4].

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. We propose 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 pico BSs or relays at fixed positions, we investigate a new network architecture with fmAPs 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.

In this scenario, the main problem is the management of handoffs within a tier (horizontal handoff) or between the tiers (vertical handoff) [5]. This problem cannot be ignored when we consider user mobility [6]. 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 [7], 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 [8], when the position of the relays can be controlled to optimize the handoff mechanism and balance the backhaul [9]. Differently from the previous works, 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 the UE and fmAP is on an orthogonal channel, possibly in unlicensed bands (LTE-U) [10].

In general, the channel state information (CSI) between the fmAP and the UE is constantly changing due to the fmAP mobility. This increased space and time diversity is an opportunity to provide connectivity also to the edge user, in the presence of multipath or strong shadowing. 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 comprehensive 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. 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. Instead, by incorporating various relevant parameters and characterizing the increased total throughput, our work enables a cost-benefit analysis.

The main contributions of this paper are the following.

- In Sec. II, 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). Our model allows for the inclusion of different strategies to select the next fmAP when a handoff occurs.
- In Sec. III, we provide an analytical framework that lends itself to a close form derivation of the network performance, in terms of handoff rate and additional capacity in T2, as a function of the parameters of the scenario. Due to the generality of

the PHY and DL layers abstraction, this performance analysis can be used as a benchmark for the evaluation of future implementations under various realistic network parameters.

- In Sec. V, 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). In these simulations, we generate the vehicles arrivals with a random process, and we measure the additional data rate provided by T2. 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. Finally, Sec. VI concludes the paper.

II. System Model

In our system model, we envision a network of fmAPs (T2) that supports the macro BS (T1) to provide service to the UEs in a band that is orthogonal to the band of fmAP–macro BS connection. The two tiers differ in terms of 1) availability, 2) cost, and 3) horizontal handoffs. We assume that a connection to T1 is always available and that each UE is not moving, so that there are no horizontal handoffs in T1. On the other hand, the availability of low-cost T2 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. Due to the fmAP mobility, the connection in T2 deals with frequent handoffs.

A graphical representation of our system model is depicted in Fig. 1. The UE is not moving and is represented as a cross, 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. Each vehicle is moving with a constant speed v in one of the W lanes of the street. In each lane, the vehicle arrival rate is described with a Poisson process of parameter λ_1, a common model to represent the arrivals of vehicles in this type of scenario. Furthermore, the fraction of vehicles equipped with an fmAP is ρ, so the arrival rate of vehicles equipped with an fmAP is a Poisson process of rate $\lambda = \rho\lambda_1 W$.

In this scenario, the macro BS is always connected to the UE, while an fmAP is connected with the UE only if its distance is smaller than L. Since we assume a constant speed v, the interval of time in which an fmAP is in the range of connectivity of a UE is $T_M \triangleq 2L/v$. The connection time to each fmAP is thus limited, and frequent handoffs are necessary to maintain the connectivity within T2. 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 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 this paper, we consider three selection strategies, whose performance will be analyzed and compared in the following sections.

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

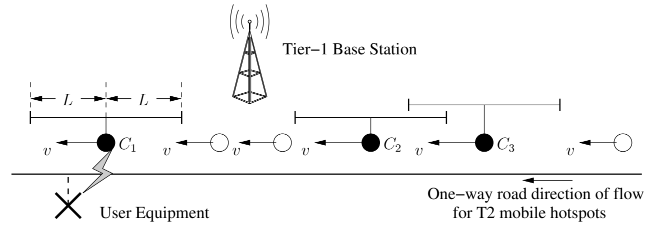

Fig. 1. One-way fmAP communication model for T2.

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. In the presence of a wireless software-defined network architecture (WSDN), the controller can communicate this information to the UE through a control plane (in T1). The UE can then select the fmAP that will guarantee the longer connection time, thus minimizing the handoff rate.

III. Analysis of the Two-tier System

First, we evaluate the expected duration of a T2 connection round, namely T_2, which is the contiguous time of connections to the fmAPs under strategy $S \in \{S_m, S_r, S_c\}$. A T2 connection round ends when no fmAP is in the connection range of the UE. It is defined as

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

where $M^{(S)}$ is the number of horizontal handoffs, $M^{(S)}+1$ is the total number of fmAPs serving the UE, and $\tau_j^{(S)}$ is the time interval in which the UE is connected to the j^{th} fmAP. 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, $R_2^{(S)}$, 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.,

$$R_2^{(S)} \triangleq \frac{\mathrm{E}\left[T_2 - M^{(S)}T_H\right]}{\mathrm{E}\left[T_1 + T_2\right]}. \tag{2}$$

$M^{(S)}$, $\tau_j^{(S)}$, and $R_2^{(S)}$ depend on the specific selection strategy adopted, whereas T_2 is the same for the three strategies considered. For simplicity, in the next section, we adopt the strategy S_m to calculate the value of T_2.

A. 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 $\tau^{(\mathrm{S}_m)}_{M^{(\mathrm{S}_m)}+1} = T_M$.

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:

$$P_V \triangleq \mathrm{P}\left\{N\left(t - T_M, t\right] = 0\right\} = e^{-\lambda T_M}\ , \tag{3}$$

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 $(t_1, t_2]$. 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

$$\mathrm{E}\left[M^{(S_m)}\right] = \frac{1 - P_V}{P_V}\ . \tag{4}$$

Using this expectation, we can evaluate the expected time spent in T2 by using the iterated expectation over (1). The duration of the consecutive time interval with a T2 connection, conditioned on the value of $M^{(\mathrm{S}_m)}$, is

$$\begin{aligned}
&\mathrm{E}\left[T_2 \middle| M^{(\mathrm{S}_m)} = m\right] \\
&= \mathrm{E}\left[\sum_{j=1}^{m} i_j \middle| i_1 \leq T_M, \ldots, i_m \leq T_M, i_{m+1} > T_M\right] + T_M \\
&= T_M + m\mathrm{E}\left[i_j | i_j \leq T_M\right].
\end{aligned} \tag{5}$$

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

$$\mathrm{E}\left[i_j | i_j \leq T_M\right] = \frac{1 - P_V\left(1 + \lambda T_M\right)}{\lambda\left(1 - P_V\right)}. \tag{6}$$

Using (4), (5), and (6), we obtain the following

$$\mathrm{E}\left[T_2\right] = \mathrm{E}_{M^{(\mathrm{S}_m)}}\left[\mathrm{E}\left[T_2 \middle| M^{(\mathrm{S}_m)}\right]\right] = \frac{1 - P_V}{\lambda P_V}. \tag{7}$$

B. Effective T2 ratio using S_m

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

$$\mathrm{E}\left[T_1\right] = \lambda^{-1}, \tag{8}$$

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 T2 ratio with strategy $\mathrm{S_m}$ is

$$R_2^{(\mathrm{S}_m)} = \left(1 - P_V\right) - T_H\left(1 - P_V\right)\lambda, \tag{9}$$

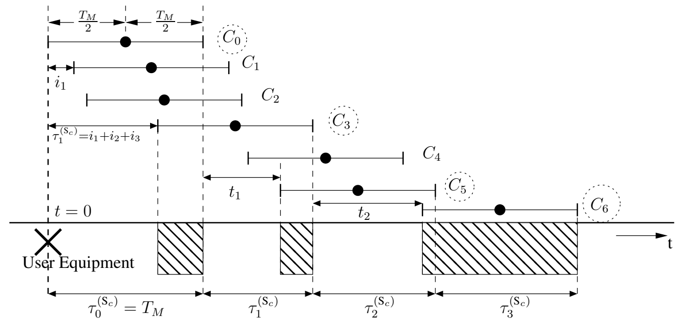

Fig. 2. Time of arrivals and $\mathrm{S_c}$ operation for a given T2 instance.

where we used (2), (4), (7), and (8). The numerator represents the time spent in T2 minus the time spent into $M^{(S_m)}$ horizontal handoffs. We observe that the first term $(1 - P_V)$ quickly converges to 1 as λ increases, see (3). The second term instead represents the penalty due to the handoffs, which becomes more significant as λ increases.

C. Effective T2 ratio using S_c

The $\mathrm{S_c}$ 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 (7), remains unchanged.

In order to explain the behavior of $\mathrm{S_c}$, 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, $\tau_j^{(\mathrm{S}_c)}$. 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 lost 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. Then, at the end of the connection with C_0, the fmAP C_3 is selected, since it 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 $\mathrm{S_c}$.

Theorem 1. *The expected effective T2 ratio with strategy S_c can be approximated by*

$$R_2^{(S_c)} \simeq \left(1 - P_V\right) - \frac{2\ \left(1 - P_V\right) T_H}{E\left[\tau_1^{(S_c)} + t_1 \middle| M^{(S_c)} \geq 1\right] + \frac{2}{\lambda}}. \tag{10}$$

The expression for $\mathrm{E}\left[\tau_1^{(\mathrm{S}_c)} + t_1 | M^{(\mathrm{S}_c)} \geq 1\right]$ is obtained by using (14) and (15), in the proof of Thm.1 in Sec. VII-A.

As in the case of $\mathrm{S_m}$ in (9), 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, demonstrating the advantage of $\mathrm{S_c}$ over $\mathrm{S_m}$. This intuition will be verified through simulation in Sec. V-B.

IV. Extension: Case of Stopping Vehicles

In a realistic urban scenario, commercial vehicles may frequently stop to pick up or drop off passengers. We model this by setting up a probability P_S for a car to stop while in the range of communication of the UE, and we assume that when it stops, the vehicle remains in its position for a time interval T_S. The strategies $\mathrm{S_m}$ and $\mathrm{S_r}$ work as before, with the difference that if an fmAP stops while it is connected to the UE, the UE will remain connected to that fmAP until it goes out of the connectivity range, which translates into a decrease in horizontal handoff rate. The centralized controller in $\mathrm{S_c}$ can also accurately predict if a vehicle will stop, since it knows the vehicle's path and regular stopping pattern, e.g., in the case of public transportation vehicles. $\mathrm{S_c}$ will use this additional information in the selection process.

The analysis of the case of stopping vehicles is non-trivial, thus we provide only the analysis of $\mathrm{S_m}$ for this case. We define the time in T2 and the expected effective T2 ratio for $\mathrm{S_m}$ as $\hat{T}_2$ and $\hat{R}_2^{(S_m)}$, respectively. The definitions follow the same arguments as in (1) and (2), but considering the increased service times due to stopping fmAPs.

First of all, we distinguish between stopping and non-stopping fmAPs. Due to the splitting property of independent Poisson arrivals, we can consider the two groups of stopping and non-stopping fmAPs as two independent Poisson processes with arrival rates $P_S\lambda$ and $(1-P_S)\,\lambda$, respectively, where P_S is the probability of stopping. Let's first consider the interval of connectivity to the first fmAP, after a period without T2 connectivity. With probability P_S, this interval is equal to (T_M+T_S), and with probability $(1-P_S)$, it is T_M. The expected connection time to T2 with stopping fmAPs is

$$\mathrm{E}\left[\hat{T}_2\right] = T_M + P_S T_S + \mathrm{E}\left[\sum_{j=1}^{M^{(S_m)}} \tau_j^{(S_m)}\right], \qquad (11)$$

where $\tau_j^{(S_m)}$ is the service time for the j^{th} connected fmAP, after the first fmAP and $M^{(S_m)}$ is the total number of handoffs. Thm. 2 provides the expected effective T2 ratio for $\mathrm{S_m}$.

Theorem 2. *The expected effective T2 ratio for S_m with stopping fmAPs can be approximated by*

$$\hat{R}_2^{(S_m)} \simeq \frac{\tilde{A}_2 - T_H E\left[M^{(S_m)}\right]}{E\left[\hat{T}_1\right] + \tilde{A}_2}. \qquad (12)$$

In this theorem, we use the approximation $\mathrm{E}\left[\hat{T}_2\right] \simeq \tilde{A}_2$, defined as

$$\tilde{A}_2 = T_M + P_S T_S + \mathrm{E}\left[M^{(S_m)}\right] \frac{\mathrm{E}\left[\tau_1^{(S_m)}\right] + \mathrm{E}\left[\tau_2^{(S_m)}\right]}{2}. \qquad (13)$$

The explicit expressions for $\mathrm{E}\left[\tau_j^{(S_m)}\right]$ and $\mathrm{E}\left[M^{(S_m)}\right]$ are provided in the proof in Sec. VII-B, in (21) and (23), respectively.

V. Numerical Results

In this section, we show the performance of the proposed system as a function of the considered parameters. The simulations aim to showcase the achievable performance of this networking system and the validity of the analytical results.

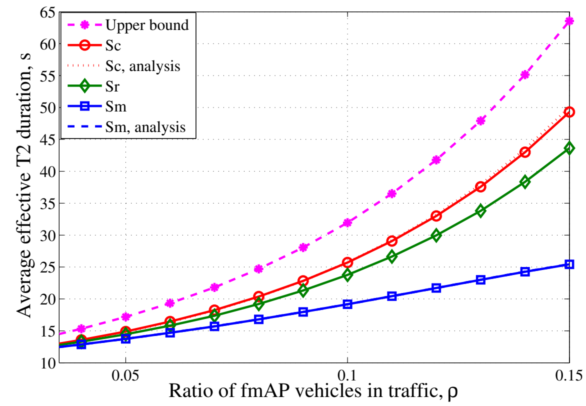

Fig. 3. Expected effective time in T2. The simulation results perfectly match the analytical results.

In particular, we quantify the performance in terms of average effective duration of a connectivity round in T2 and in terms of average data rate that can be obtained via a connection to T2.

A. Simulation Setup

In the simulations, the wireless technology for the connection between the UE and the fmAPs is the IEEE 802.11n, working at 2.4GHz, with 20 MHz bandwidth per stream. The outdoor minimum data rate is $R_b = 7.2$ Mbps at a maximum range of 250 m [11]. 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 R_b 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. As a consequence, the maximum connection time to an fmAP that is not stopping is $T_M = 10$ s. 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.

B. Verification of Analytical Expressions

In Fig. 3, 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 (7), 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 in which a centralized controller is unavailable. We also observe that both the exact analysis results given in Sec. III-B 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.

In Fig. 4, we evaluate the effective T2 data rate for the strategies considered. We observe that for $\mathrm{S_m}$, 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

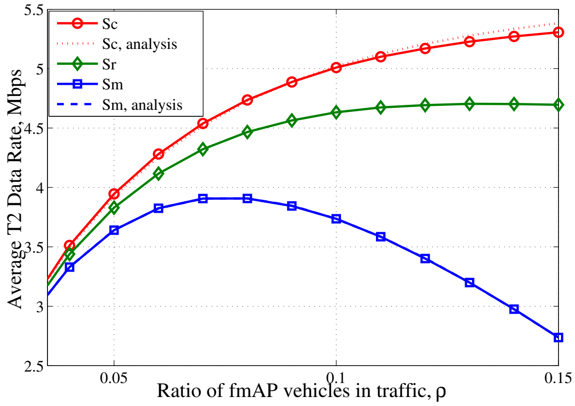

Fig. 4. Average data rate in T2. The simulation results perfectly match the analytical results.

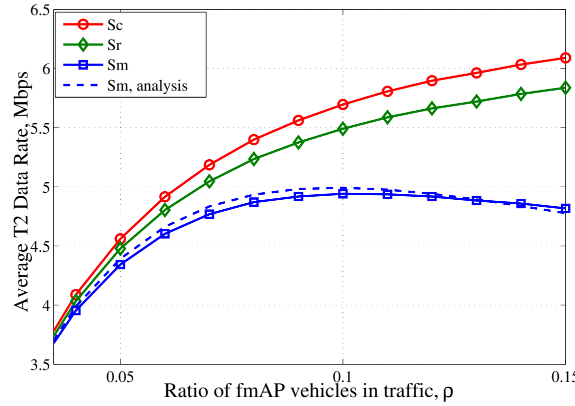

Fig. 5. Average data rate obtained by T2 for a stopping probability of 0.05.

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 $\rho > 0.07$.

Finally, Fig. 5 depicts the expected effective T2 data rate for a probability of stopping $P_S = 0.05$ for the fmAPs. In our simulation, if an fmAP stops, it will remain in that position for $T_S = 45$ s. Since there is the possibility of longer connections with the stopping fmAPs, we observe a significant improvement in the data rate for all the three strategies with respect to Fig. 4. We notice that, for $\rho = 0.15$, the S_c strategy is able to achieve a data rate that is around 85% of the maximum data rate, R_b, achievable in the case of a connection with a fixed fmAP. Also as depicted in Fig. 5, the approximation in Thm. 2 is very close to the simulated results for the considered values of ρ.

VI. Conclusions

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 designing three techniques to select an fmAP (if more than one is available) and by computing 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. The simulations confirmed the validity of the analytical results, showing, in particular, that with a random choice of fmAPs, we can achieve performance close to that observed with a centralized controller. We also showed that the data rate provided by T2 is close to the one provided by a fixed relay in cases in which the arrival rate of fmAPs is high enough.

In a future work, we plan to remove some basic assumptions on the physical and data link layers, investigating the use of fmAPs in a more realistic setting. We plan to evaluate the architecture by implementing a wireless software defined network testbed, as in [12], to separate the two tiers of communication and provide the control information needed by the proposed S_c strategy.

VII. Appendices

A. Proof of Theorem 1

We first present the following two lemmas:

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

$$\mathrm{E}\left[\tau_1^{(S_c)} | M^{(S_c)} \geq 1\right] = \frac{\lambda T_M - \left(1 - e^{-\lambda T_M}\right)}{\lambda\left(1 - e^{-\lambda T_M}\right)}. \quad (14)$$

Lemma 2. The expectation of t_1 is

$$\begin{aligned}\mathrm{E}\left[t_1 | M^{(S_c)} \geq 1\right] &= \frac{\lambda}{e^{\lambda T_M} - 1}\int_0^{T_M} \frac{s_1 e^{\lambda s_1}}{1 - e^{\lambda s_1}} \mathrm{d}s_1 - \frac{1}{\lambda} \\ &= \frac{1}{\lambda\left(e^{\lambda T_M} - 1\right)}\left(\mathrm{Li}_2\left(e^{\lambda T_M}\right) - \frac{\pi^2}{6} + 1\right. \quad (15) \\ &\left. + e^{\lambda T_M}\left(\lambda T_M - 1\right) + \lambda T_M \log\left(1 - e^{\lambda T_M}\right)\right) - \frac{1}{\lambda},\end{aligned}$$

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 [13].

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

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

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

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

$$\begin{aligned}\mathrm{E}\left[U | M^{(S_c)} = n\right] &= \mathrm{E}\left[\mathrm{E}_{U|\tau_1, t_j, M^{(S_c)}=n}\left[U^{\tau_1} + \sum_{j=1}^{n-1} U^{t_j}\right]\right] \\ &= \lambda\left(\mathrm{E}\left[\tau_1 | M^{(S_c)} = n\right] + \sum_{j=1}^{n-1} \mathrm{E}\left[t_j | M^{(S_c)} = n\right]\right), \quad (16)\end{aligned}$$

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

We can approximate (16) by evaluating $\mathrm{E}\left[t_j|M^{(\mathrm{S}_c)}=n\right]$ for only a few values of j. In particular, we use

$$\mathrm{E}\Big[U|M^{(\mathrm{S}_c)}{=}n\Big]{\simeq}n\lambda\frac{\mathrm{E}\big[\tau_1|M^{(\mathrm{S}_c)}{\ge}1\big]{+}\mathrm{E}\big[t_1|M^{(\mathrm{S}_c)}{\ge}1\big]}{2}, \quad (17)$$

where the term $\left(\mathrm{E}\left[\tau_1|M^{(\mathrm{S}_c)}\geq 1\right]+\mathrm{E}\left[t_1|M^{(\mathrm{S}_c)}\geq 1\right]\right)/2$ 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 (14) and (15).

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

$$\mathrm{E}\left[M^{(\mathrm{S}_m)}\right]=\mathrm{E}_{M^{(\mathrm{S}_c)}}\left[\mathrm{E}\left[U+M^{(\mathrm{S}_c)}|M^{(\mathrm{S}_c)}\right]\right]$$
$$\simeq\mathrm{E}\Big[M^{(\mathrm{S}_c)}\Big]\left(\lambda\frac{\mathrm{E}\big[\tau_1|M^{(\mathrm{S}_c)}{\ge}1\big]{+}\mathrm{E}\big[t_1|M^{(\mathrm{S}_c)}{\ge}1\big]}{2}{+}1\right), \quad (18)$$

where we use (17). Solving it for $\mathrm{E}\left[M^{(\mathrm{S}_c)}\right]$ in (18) we obtain

$$\mathrm{E}\left[M^{(\mathrm{S}_c)}\right]\simeq\frac{2\,\mathrm{E}\left[M^{(\mathrm{S}_m)}\right]}{\lambda\left(\mathrm{E}\left[\tau_1+t_1|M^{(\mathrm{S}_c)}\geq 1\right]\right)+2}, \quad (19)$$

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

$$R_2^{(\mathrm{S}_c)}=\frac{\mathrm{E}\left[T_2\right]-\mathrm{E}\left[M^{(\mathrm{S}_c)}\right]T_H}{\mathrm{E}\left[T_1\right]+\mathrm{E}\left[T_2\right]}, \quad (20)$$

where $\mathrm{E}\left[T_1\right]=\lambda^{-1}$ and $\mathrm{E}\left[T_2\right]=\frac{1-P_V}{\lambda P_V}$. The proof is completed by plugging (19) into (20).

B. Proof of Theorem 2

Lemma 3.

$$\mathrm{E}\left[M^{(\mathrm{S}_m)}\right]=\sum_{m=0}^{\infty}\prod_{j=1}^{m+1}\left(1-\hat{P}_V^{(j)}\right), \quad (21)$$

where $\hat{P}_V^{(j)}$ is the probability of a vertical handoff at the end of the service time of the $(j-1)^{\text{th}}$ fmAP. It is expressed as

$$\hat{P}_V^{(j)}=e^{-\lambda T_M}\left[1-P_S\left(1-e^{-P_S\lambda T_S}\right)\frac{1-\Delta^j}{1-\Delta}\right], \quad (22)$$

where $\Delta\triangleq(P_S'-P_S)$, and $P_S'\triangleq\left(1-e^{-P_S\lambda T_S}\right)+e^{-P_S\lambda T_S}\left(1-e^{-\lambda T_M}\right)P_S$ is the probability that an fmAP is a stopping one given that the previous one has stopped.

Proof of Lemma 3 is provided in [13].

The expected service time $\mathrm{E}\left[\tau_j^{(\mathrm{S}_m)}\right]$ can be evaluated by using 4 possible combinations of the stopping property for the $(j-1)^{\text{th}}$ and the jth fmAPs:

$$\mathrm{E}\left[\tau_j^{(\mathrm{S}_m)}\right]=\left(1-P_S'^{(j-1)}\right)\frac{1-e^{-x}\left(1+x\right)}{\lambda\left(1-e^{-x}\right)}\Bigg|_{x=P_S\lambda T_M}$$
$$+\left(1-P_S'^{(j-1)}\frac{P_S'-P_S}{1-P_S}\right)\frac{1-e^{-x}\left(1+x\right)}{\lambda\left(1-e^{-x}\right)}\Bigg|_{x=(1-P_S)\lambda T_M}$$
$$+P_S'^{(j-1)}\frac{P_S'}{P_S}\frac{1-e^{-x}\left(1+x\right)}{\lambda\left(1-e^{-x}\right)}\Bigg|_{x=P_S\lambda(T_M+T_S)}, \quad (23)$$

where $P_S'^{(j)}\triangleq P_S\sum_{k=0}^{j}\left(P_S'-P_S\right)^k$ is the probability that the $(j-1)^{\text{th}}$ fmAP is a stopping one.

Similar to the approximation in the proof of Thm. 1, we can approximate the expected time in T2 by using only a few of the $\mathrm{E}\left[\tau_j^{(\mathrm{S}_m)}\right]$ terms. We notice that we have a tight approximation by utilizing only $E[\tau_1^{(\mathrm{S}_m)}]$ and $E[\tau_2^{(\mathrm{S}_m)}]$, defined in (23). In this case we obtain

$$\mathrm{E}\Big[\hat{T}_2\Big]{\simeq}T_M{+}P_ST_S{+}\mathrm{E}\Big[M^{(\mathrm{S}_m)}\Big]\frac{\mathrm{E}\Big[\tau_1^{(\mathrm{S}_m)}{+}\tau_2^{(\mathrm{S}_m)}\Big]}{2}, \quad (24)$$

where $\mathrm{E}\left[\hat{T}_1\right]=1/\lambda$ as in the case of non-stopping fmAPs. The proof is completed when we replace $\mathrm{E}\left[\hat{T}_2\right]$ in both the numerator and the denominator of (12) with the approximation for $\mathrm{E}\left[\hat{T}_2\right]$ in (24) and use the result on $\mathrm{E}\left[M^{(\mathrm{S}_m)}\right]$ from Lemma 3.

ACKNOWLEDGMENTS

This work was partially supported by the Qualcomm Institute, University of California San Diego, and NSF grant CNS-1329819. Tugcan Aktas was supported by The Scientific and Technological Research Council of Turkey (TUBITAK) 2219 Postdoctoral Research Fellowship Programme.

REFERENCES

[1] "Metrolab Network," http://metrolabnetwork.org, Last acc.: Apr. 2016.

[2] A. K. Gupta, X. Zhang, and J. G. Andrews, "SINR and Throughput Scaling in Ultradense Urban Cellular Networks," *IEEE Wireless Communications Letters*, vol. 4, no. 6, pp. 605–608, Dec. 2015.

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

[4] O. Galinina, A. Pyattaev, S. Andreev, M. Dohler, and Y. Koucheryavy, "5G Multi-RAT LTE-WiFi Ultra-Dense Small Cells: Performance Dynamics, Architecture, and Trends," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 6, pp. 1224–1240, Jun. 2015.

[5] I. F. Akyildiz, J. Xie, and S. Mohanty, "A survey of mobility management in next-generation all-IP-based wireless systems," *IEEE Wireless Communications*, vol. 11, no. 4, pp. 16–28, Aug. 2004.

[6] W. Bao and B. Liang, "Stochastic Geometric Analysis of User Mobility in Heterogeneous Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2212–2225, Oct. 2015.

[7] W. Li, C. Zhang, X. Duan, S. Jia, Y. Liu, and L. Zhang, "Performance Evaluation and Analysis on Group Mobility of Mobile Relay for LTE Advanced System," in *IEEE VTC Fall*, Sep. 2012, pp. 1–5.

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

[9] Z. Liao, X. Zhang, and C. Feng, "Mobile relay deployment based on Markov chains in WiMAX networks," in *2014 IEEE Global Communications Conference*, Dec. 2014, pp. 4508–4513.

[10] Qualcomm Technologies, Inc., "LTE in Unlicensed Spectrum: Harmonious Coexistence with Wi-Fi," *White paper*, Jun. 2014.

[11] A. Matsumoto, K. Yoshimura, S. Aust, T. Ito, and Y. Kondo, "Performance evaluation of IEEE 802.11n devices for vehicular networks," in *IEEE Local Computer Networks (LCN)*, Oct. 2009, pp. 669–670.

[12] M. Danieletto, G. Quer, R. R. Rao, and M. Zorzi, "CARMEN: a cognitive networking testbed on android OS devices," *IEEE Communications Magazine*, vol. 52, no. 9, pp. 98–107, Sep. 2014.

[13] T. Aktas, G. Quer, T. Gavidi, and R. Rao, "Mobile relays for smart cities: mathematical proofs," Tech. Rep., 2016. [Online]. Available: https://arxiv.org/abs/1609.01322.