This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2752761, IEEE Internet of Things Journal

Using Smart City Data in 5G Self-Organizing Networks

Massimo Dalla Cia, Federico Mason, Davide Peron, Federico Chiariotti, *Student Member, IEEE*, Michele Polese, *Student Member, IEEE*, Toktam Mahmoodi, *Senior Member, IEEE*, Michele Zorzi, *Fellow, IEEE*, Andrea Zanella, *Senior Member, IEEE*

***Abstract*—So far, research on Smart Cities and self-organizing networking techniques for 5G cellular systems has been one-sided: a Smart City relies on 5G to support massive M2M communications, but the actual network is unaware of the information flowing through it. However, a greater synergy between the two would make the relationship mutual, since the insights provided by the massive amount of data gathered by sensors can be exploited to improve the communication performance. In this work, we concentrate on self-organization techniques to improve handover efficiency using vehicular traffic data gathered in London. Our algorithms exploit mobility patterns between cell coverage areas and road traffic congestion levels to optimize the handover bias in HetNets and dynamically manage Mobility Management Entity (MME) loads to reduce handover completion times.**

***Index Terms*—Symbiocity; Traffic for London; handover; HetNets; virtual Mobility Management Entity**

I. Introduction

The fifth generation of mobile networks (5G) is forecasted to rely on virtualization and self-organization techniques to deal with the extreme complexity and heterogeneity of the network and with the massive number of connected devices [2]. The rise of internet-capable sensors and monitoring devices is one of the major drivers of such complexity, due to the volume of information they generate [3]; however, this information can also be a valuable resource in the network decision-making process.

According to the Smart City paradigm, these data can be leveraged to provide innovative services to citizens and to help administrators define smarter policies. However, since they must be transmitted and aggregated by the network in order to be processed [4], there is no reason why the network itself should not benefit from them. For example, traffic data can be used to predict mobility patterns and future cell load with higher accuracy, enabling anticipatory techniques [5]. Cellular network operators would be incentivized to support the deployment of Smart Cities given the possibility of increased efficiency and lower operating costs, improving both the carrier network and the sensors' pervasiveness.

Building upon the "SymbioCity" concept proposed in [6], in this paper we exploit the traffic data from the Transport for London (TfL) Urban Traffic Control (UTC) network [7] in order to dynamically optimize network parameters such as (i) the handover range expansion bias for Heterogeneous Networks (HetNets) and (ii) the number of virtualized Mobility Management Entities (MMEs) deployed city-wide. Since handovers will be one of the major issues in 5G ultra-dense networks, the techniques we propose will reduce the handover completion time and the well-known ping-pong effect [8], [9] without losing the benefits of microcell offloading. The ability to choose the point in the tradeoff between handover frequency and offloading capability is going to be a key element in the design of self-organizing 5G networks.

The rest of this work is organized as follows. Sec. II presents an overview of state of the art techniques in traffic data analysis, self-organizing networks and handover management, while Sec. III describes the London traffic sensor network, the available data and our analysis of the vehicular mobility patterns. We provide the details on the two previously mentioned optimization techniques in Sec. IV, along with an example application of both, using the London traffic data. Finally, in Sec. V we make our final remarks and suggest some possibilities for future research.

II. Related Work

The emerging Smart City paradigm is getting significant attention from researchers, companies and city officials all over the world. A Smart City enables a wide array of services, from environmental monitoring to traffic control and smart parking [10]. These services build upon data generated by a plethora of sensors, and collected by means of possibly different technologies that collectively concur to the shaping of the so-called Internet of Things (IoT) [11]. The data these services need are gathered by millions of distributed sensors [12] and aggregated through a modular event-driven architecture [13]. These devices communicate using either dedicated low power networks (e.g., LoraWAN, SigFox, IEEE 802.15.4) [14] or standard cellular networks. Both these solutions have their advantages and drawbacks; using cellular networks requires no additional infrastructure investment (place & play concept), but the Machine to Machine (M2M) traffic has an impact on traditional human communications [15], [16].

The information that the Smart City generates can be used to make cellular networks aware of the surrounding environment. Although one of the 5G design guidelines is the usage of big-data-driven optimization [17], [18] at various scales (e.g., fog computing [19]), the optimization mostly relies on data generated by the network itself. In our opinion, integrating the

M. Dalla Cia, F. Mason, D. Peron, F. Chiariotti, M. Polese, M. Zorzi and A. Zanella are with the Department of Information Engineering, University of Padova, Via Gradenigo, 6/b, 35131 Padova, Italy. Email: {dallacia, masonfed, perondav, chiariot, polesemi, zanella, zorzi}@dei.unipd.it. T. Mahmoodi is with the Department of Informatics, Kings College London, Strand, London WC2R 2LS, U.K. Email: toktam.mahmoodi@kcl.ac.uk.

Part of this work has been presented at IEEE ISWCS 2017 [1].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2752761, IEEE Internet of Things Journal

Smart City knowledge in this optimization framework would be a significant step towards the cognitive network model [20], as for the "SymbioCity" concept proposed in [6].

Mobility patterns are some of the most valuable pieces of information that the Smart City can provide to the cellular network. Mobility affects ultra-dense network performance significantly, as sub-optimal handover strategies both (i) increase the Radio Access Network (RAN) and the Core Network (CN) signaling and (ii) reduce the overall throughput. Research on mobility models [21], [22] and their integration in communication protocols (e.g., Medium Access [23] or interference coordination [24]) is already ongoing, and using real Smart City data as input for these techniques would reduce the uncertainty compared to purely statistical approaches. To the best of our knowledge, this is the first work to do so, although mobility-aware strategies have been proposed and tested in scenarios with simulated mobility patterns [25], [26].

A. Handover in HetNets

In order to support the ever growing traffic demand within the limits of the available spectrum, cellular networks are becoming denser and denser. Micro-, femto- and picocells have been a hot research topic for the last ten years [27] and are now being deployed all over the world. The main challenges that the network densification is causing are (i) interference coordination and (ii) cell association and mobility management. The Self-Organizing Network (SON) approach is one of the most promising candidates to address these complex issues [28], using tools such as Software-Defined Networking (SDN) [29].

In this paper, we focus on handover management. While handover algorithms are well-studied and several decision criteria have been proposed in the literature [30], the most common ones are based on Received Signal Strength (RSS). 3GPP defines a baseline handover procedure for LTE in [31], and most studies concentrate on optimizing its parameters. The handover is triggered if the difference between the serving and the neighbor cell RSS is larger than a threshold value for at least one Time-to-Trigger (TTT). This parameter is meant to avoid unnecessary handovers due to fluctuations caused by fast fading, but introduces a delay in the association with the optimal evolved Node Base (eNB), whose impact becomes more significant as the UE speed increases [32], [33]. An analytical model to optimize the TTT is introduced in [9].

Using the TTT to reduce ping-pong effects inevitably leads to a higher handover delay. In order to overcome this trade-off, we need to exploit other parameters, such as the hysteresis threshold. Biasing this threshold towards femtocells is already a standard practice to favor offloading from the Macro tier [27], and it is possible to adapt the bias based on the user mobility to reduce the handover delay problem caused by the TTT. In [34], the authors present a heuristic that reacts to late or early handovers and adapts the bias for each pair of neighboring cells. Another work jointly adapts the TTT and bias in a reactive manner [35]. It is even possible to skip handovers entirely, avoiding connections to very small cells while moving at high speed [36].

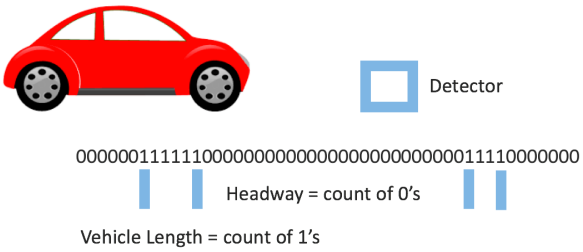


Fig. 1: Scheme of a traffic detector. Source: TfL.

B. Virtual MME

One of the main architectural trends in the evolution towards 5G is Network Function Virtualization (NFV): instead of using specialized and costly hardware in both the core and the access network, most of the processing is virtualized and run on general-purpose machines in the cloud [2]. This allows a larger flexibility and adaptability to the instantaneous load of the control and user planes. The initialization cost of a new Virtual Machine (VM) is orders of magnitude smaller than the cost of the equivalent worst-case dimensioned hardware. A broad overview of the issues and other potential benefits of NFV is presented in [37]. Although this research is still ongoing, preliminary results [38] show that it is possible to increase the energy efficiency of the network without significant performance losses.

In the second part of this work, we focus on handover management in virtualized MMEs. A first model of the performance of the different virtualized CN functions is presented in [39], and the MME is identified as a critical element for scalability of control plane functionalities. Virtualization can also enable distributed MME designs [40].

An optimized design of a virtualized MME is given in [41], where the number of vMME instances is adapted to the traffic load in an M2M scenario, using a traffic model for CN-related events.

III. Data Gathering and Analysis

The TfL UTC network is composed of more than 10000 road sensors, placed at all critical crossings around the city. The Split Cycle Offset Optimization Technique (SCOOT) optimizer uses the traffic flow data from the sensors to adapt the traffic light times to the traffic situation in real time. TfL released the raw sensor data of the first three months of 2015 for the North and Central regions of London, and we use those data in our optimization.

The sensors are actually very basic presence-detectors: every $T_s = 250$ ms, each sensor returns a 1 if it detects a vehicle in close proximity, and a 0 otherwise. The resulting binary signal (see Fig. 1) is packetized and sent to a central collector through different types of technologies.

In this work, we extract two kinds of information from the TfL dataset: (i) the average vehicular speed at any single crossing in London and (ii) the number of handovers between Macro eNBs over the whole city.

These values are not directly provided by TfL. However, they can be roughly estimated using the binary signals generated by the detectors. Indeed, when a vehicle of length L

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2752761, IEEE Internet of Things Journal

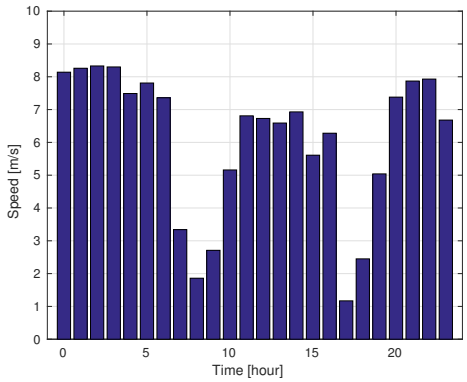


Fig. 2: Hourly average speed for January 23, 2015 at the intersection between Homerton High St. and Daubeney Rd.

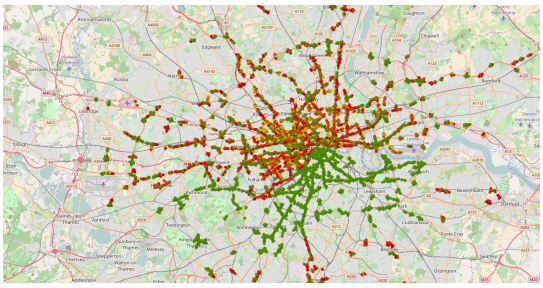


Fig. 3: Map of traffic in London from 12 PM to 1 PM of January 23, 2015. Free intersections are shown in green, heavily congested ones in red.

moving at speed v passes over a sensor, the detector will generate a run of about $n = \frac{L}{vT_s}$ ones, followed by a few zeros corresponding to the inter-vehicle spacing.

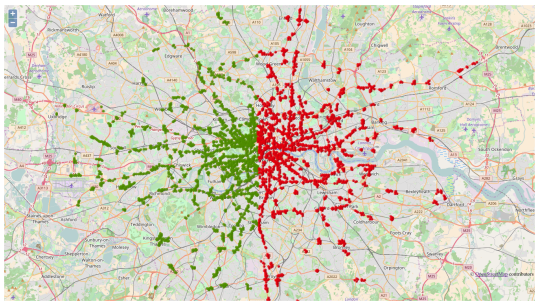
It is then possible to estimate the speed by counting n and assuming a reference vehicle length of $L = 4$ m:

$$v = \frac{L}{nT_s} \tag{1}$$

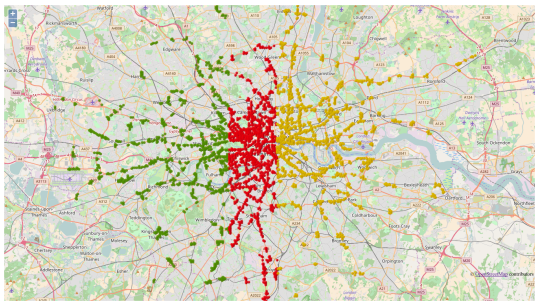
Fig. 2 shows the evolution of the average speed measured by a single sensor over a whole day (namely, January 23, 2015): as expected, the speed of the vehicles is higher at night because of the lighter traffic, while during rush hour (from 8 AM to 9 AM and from 5 PM to 6 PM) the average speed drastically decreases. The spatial distribution of traffic is shown in Fig. 3.

For the second part of our data analysis, we assume that the Macro eNBs are placed using a standard regular hexagonal tiling, with sides of 100 m.

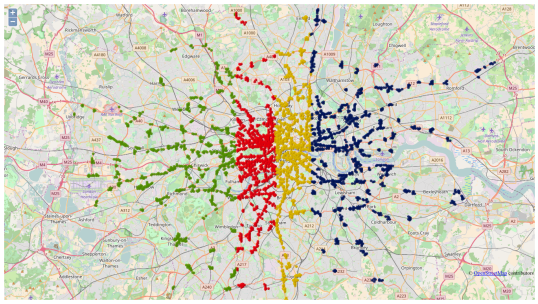
We associate the detection of a car by a sensor in a cell with a handover, and, given a time interval T_{per} equal to 1 hour, we estimate the number of handovers H_m as the total number of detections from the different sensors in cell m during T_{per}. Since the timescale is long and each vehicle is likely detected only once when crossing the area (because of the relatively low density of sensors), the number of vehicles counted in the area in the period T_{per} is roughly equal to the number of cell handovers performed by the vehicles crossing that area in the considered time interval. This assumption is not necessarily realistic for a single cell, but is a valid approximation on the city-wide scale and for timescales of minutes or hours. Moreover, we assume that on average each vehicle carries an LTE device. This is a working assumption based on the available data, and the integration of additional data such as bus position and usage can be easily accommodated by the framework.



(a) $N = 2$



(b) $N = 3$



(c) $N = 4$

Fig. 4: Partition for a different number N of vMMEs. The colors indicate the areas controlled by each vMME.

After computing H_m for all eNBs, the cells are partitioned into N areas, with $N \in \{1, 2, 3, 4\}$, each controlled by a different vMME; given the estimated number of handovers at peak hours, 4 vMMEs should be enough to maintain network stability. The results in Sec. IV-B confirm this hypothesis. These groups are obtained using a clustering algorithm that divides the cells among N vMMEs so that each vMME handles approximately the same number of handovers. An example of this is shown in Fig. 4, which reports the partitions for $N \in \{2, 3, 4\}$.

We define I_i as the total number of handovers for vMME i, and $S_{i,j}$ as the number of handovers from vMME i to vMME j. I_i is given by

$$I_i = \sum_{m \in A_i} H_m \tag{2}$$

where A_i is the set of cells controlled by vMME i. $S_{i,j}$ can be approximated with this formula:

$$S_{i,j} = \sum_{m \in A_i} \sum_{n \in A_j} \frac{H_m}{6} e_{m,n} \tag{3}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2752761, IEEE Internet of Things Journal

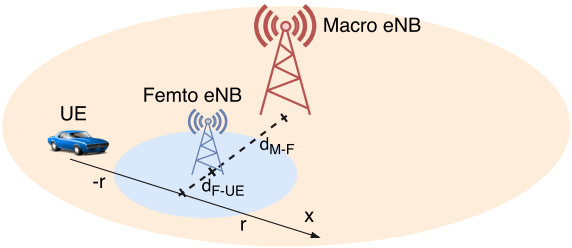


Fig. 5: UE trajectory in the considered scenario.

where the variable $e_{m,n} \in \{0, 1\}$ indicates the number of sides that cells m and n have in common.

IV. Smart City Applications

The information processed as described in Sec. III can be used to perform data-driven optimization of several parameters in a cellular network. In this paper, we use vehicular speed to dimension the handover range expansion bias in a HetNet and the number of handovers over time to find the number of vMME instances that minimize the handover completion time.

A. Asymmetrical Handover Bias Optimization in HetNets

In this simulation we provide a technique to dinamically set the handover range expansion bias of Femto eNBs (FeNBs) in order to improve the capacity provided to the User Equipment (UE) by the only Macro eNB (MeNB). We focus on a scenario consisting of a MeNB with transmission power P_{TX}^{M} and a FeNB with transmission power P_{TX}^{F} placed at a distance d_{MF} from each other. The two tiers transmit at different carrier frequencies (off-band HetNets) to avoid cross-tier interference [42]: f_0^M for the MeNB and f_0^F for the FeNB. Both tiers use the same bandwith B. All the parameters of the simulations are summarized in Table I and are taken from [43].

We consider a channel model with Friis path loss and log-normal shadowing. Let P_{RX}^{H} be the received power at the UE side from the HeNB, with $H \in \{M, F\}$, and P_{TX}^{H} the transmission power of the HeNB. Then

$$P_{RX}^{H}(t) = P_{TX}^{H}(t)\Psi_{SH}\alpha(t)h(f_0, \beta, d), \tag{4}$$

where Ψ_{SH} is the shadowing gain, which is distributed as $\mathcal{N}(0, \sigma)$ when measured in dB, and $\alpha(t)$ is the multipath

Parameter	Value	Description
P_{TX}^{M}	46	MeNB transmission power [dBm]
P_{TX}^{F}	26	FeNB transmission power [dBm]
f_0^M	900	MeNB carrier frequency [MHz]
f_0^F	1800	FeNB carrier frequency [MHz]
B	20	Bandwidth [MHz]
d_{M-F}	40	Distance between MeNB and FeNB [m]
d_{F-UE}	10	Distance between FeNB and UE [m]
σ_M^2	8	MeNB log-normal shadowing variance
σ_F^2	4	FeNB log-normal shadowing variance
β_M^2	4.28	MeNB pathloss exponent
β_F^2	3.76	FeNB pathloss exponent

Table I: Parameters used in the simulation.

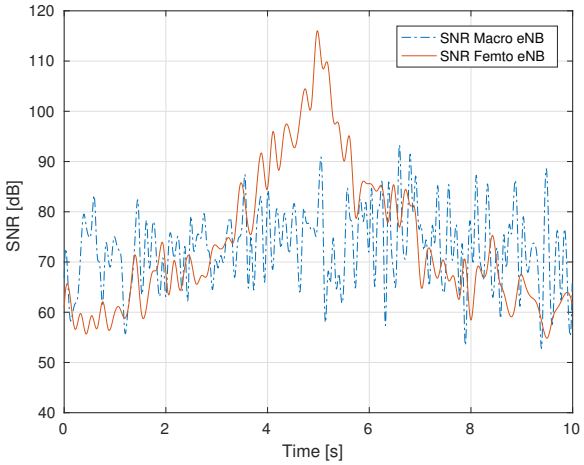


Fig. 6: $\gamma_M(t)$ and $\gamma_F(t)$ with a UE speed of 10 m/s. Multipath fading is not considered in this figure for visual clarity.

fading gain. The channel gain $h(f_0, \beta, d)$ accounts for the path loss attenuation with exponent β, and is given by

$$h(f_0, \beta, d) = A\left(\frac{c}{4\pi f_0}\right)^2\left(\frac{d}{d_0}\right)^{-\beta}, \tag{5}$$

where c is the speed of light, d_0 is the reference distance of the far field model [44], and A is a constant. Finally, $\gamma_H(t)$ denotes the Signal to Noise Ratio (SNR) at time t for the HeNB and is given by

$$\gamma_H(t) = \frac{P_{RX}^{H}}{N_0 B}, \qquad H \in \{M, F\}, \tag{6}$$

where $N_0 = -143.82$ dBW/MHz is the noise power spectral density.

For the sake of simplicity, we assume that one UE is attached in the MeNB, moving as in Fig. 5 with constant speed v. The UE speed at any time is derived from the TfL data as explained in Sec. III; the average speed over the whole day is shown in Fig. 2. We consider the UE to move at the average speed of the traffic around it.

The SNR at the UE while moving depends on its distance from the Macro and Femto eNBs. As we can see in Fig. 6, the SNR from the FeNB is higher than that from the MeNB when the UE is close to the FeNB. The coverage area of the FeNB is defined as the area in which its SNR is higher than that of any other cell.

In this scenario, the UE has to start a handover procedure towards the FeNB when the condition

$$P_{RX}^{F}(t) + \gamma_{th} > P_{RX}^{M}(t) \tag{7}$$

holds for a period of time equal to the TTT, as specified in [45]. Note that in the simulation we have assumed $\gamma_{th} = 0$ for the sake of simplicity. We hence set TTT $= 256$ ms [31], which is large enough to avoid the ping-pong effect but small enough to minimize the handover delay.

This TTT value improves the performance of the system considerably when the traffic is moving slowly, but reduces the Theoretical Spectral Efficiency $\nu = \log_2(1 + \gamma)$ when the UE speed is too high. This is because a fast-moving UE

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2752761, IEEE Internet of Things Journal

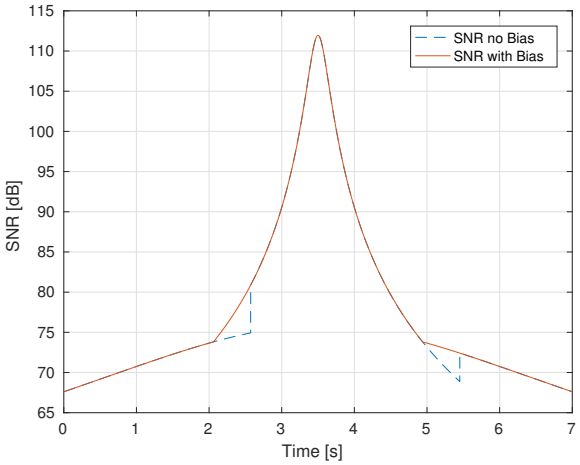


Fig. 7: $\gamma_M(t)$ and $\gamma_F(t)$ with a UE speed of 16 m/s. Multipath fading and shadowing are not considered in this figure for visual clarity.

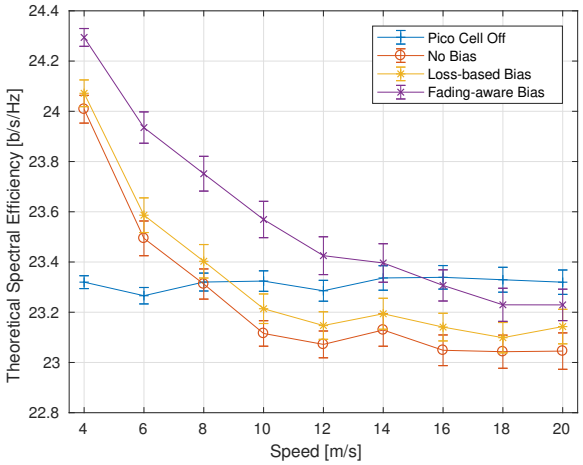


Fig. 8: Theoretical Spectral Efficiency as a function of the vehicular traffic speed v.

exploits the advantages of the FeNB for just a short time, while it remains in the FeNB for TTT seconds after the condition (7) is reversed.

To make sure that the UE starts the handover towards the FeNB as soon as (7) is verified, an asymmetrical handover bias can be applied to P^F_{RX}. When the handover is towards the FeNB, the bias needs to be positive to anticipate the beginning of the procedure, while when the handover is from the FeNB to the MeNB, the bias must be negative. We define the SNR difference in position x along the trajectory as

$$\Delta(x) = \bar{\gamma}_F(x) - \bar{\gamma}_M(x); \tag{8}$$

where $\bar{\gamma}_F(x)$ and $\bar{\gamma}_M(x)$ are the average SNRs from the two eNBs when the UE is in position x. Moreover, the trajectory of the UE draws a chord within the coverage area of the FeNB, with linear coordinates $-r$ and r with respect to the central point of the chord, as shown in Fig. 5. The optimal value of the bias is then given by

$$B_1 = \Delta(-r - vTTT) \tag{9}$$

$$B_2 = -\Delta(r - vTTT). \tag{10}$$

If the FeNB uses the optimal bias, the handover will be performed exactly at the edge of its coverage area.

By applying B_1 and B_2 to P^F_{RX}, (7) becomes

$$P^F_{RX}(t) + B_1 > P^M_{RX}(t) \tag{11}$$

while the condition to leave the FeNB is

$$P^M_{RX}(t) + B_2 > P^F_{RX}(t) \tag{12}$$

The difference between $\bar{\gamma}(x)$ with or without bias can be viewed in Fig. 7. Since the Theoretical Spectral Efficiency ν depends logarithmically on $\bar{\gamma}(x)$, using this asymmetrical handover bias will increase ν, fully exploiting the FeNB.

However, the bias from (9) and (10) does not take shadowing and fading into account: while this is optimal in an ideal situation, real channels often experience deep fading, and a bias value tailored to the path loss difference between the two base stations does not protect the UE from them. In order to avoid resetting the timer every time the fading envelope exceeds the path loss-based bias, we can add an additional bias term B_f, which does not depend on the speed of the UE.

$$B_f = 10\log_{10}\left(\min\left\{B : p\left(\frac{\psi_M \alpha_M}{\psi_F \alpha_F} \geq B\right) \leq 1 - p_{thr}\right\}\right) \tag{13}$$

$$B'_1 = B_1 + B_f \tag{14}$$

$$B'_2 = B_2 + B_f. \tag{15}$$

The parameter p_{thr} in (13) represents the amount of protection against deep fading offered by the extra bias term B_f: a higher value of p_{thr} will reset the TTT timer less often, but a higher bias will lead to stronger ping-pong effects. For this reason, we limit the total handover bias $B'_i, i \in \{1, 2\}$ to a maximum of 7 dB. The value of B_f is shown in Table II for different p_{thr}, which correspond to different percentiles. In the performance evaluation we used $p_{thr} = 0.68$, which is equivalent to one standard deviation in the normal approximation.

The improvement obtained by setting an asymmetric handover bias can be seen in Fig. 8. This figure is obtained calculating the average ν over 100 Monte Carlo simulations with independent shadowing and fading for a UE speed from 4 m/s to 20 m/s.

In the simplest case, in which there is no FeNB and the UE is always attached to the MeNB, ν_{MeNB} is essentially independent of the UE speed. The second case is a legacy handover with no bias: as the plot shows, ν_{noBias} decreases drastically as speed increases, as the delay in the handover caused by the TTT wastes most of the performance improvement from the FeNB. If the UE speed is higher than 6 m/s, the handover is so late that the UE would do better to disregard the existence of the FeNB completely: as soon as the UE finishes the handover process, it has to start it again since it has already moved

p_{thr}	0.5	0.68 (σ)	0.75	0.95 (2σ)	0.99 (3σ)
B_f [dB]	0	3.8	5.4	13.7	21.4

Table II: Values of B_f for different threshold probabilities.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2752761, IEEE Internet of Things Journal

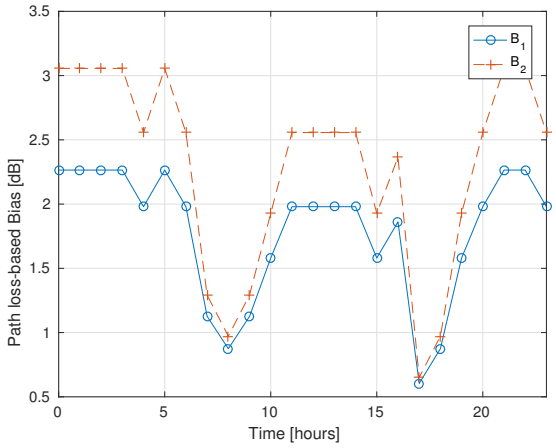


Fig. 9: Optimal handover bias throughout the day for January 23, 2015.

outside of the FeNB coverage area. The improvement given by the fading-aware bias is clear: the FeNB can be exploited if the speed is lower than 16 m/s, and there is a clear performance gain compared to the legacy scheme. The path loss-based bias shows a smaller performance improvement with respect to the legacy scheme, and handing over to the FeNB is already detrimental to the UE at 10 m/s. This is due to the handover happening too late, as even a large bias is not enough to balance the variations of the channel due to fading. In general, ν_{Bias} decreases when the speed increases, since the time in the FeNB coverage area gets shorter, but the FeNB is always fully exploited. Note that the effect of the 7 dB cap is only relevant at a speed of 20 m/s.

The presence of the FeNB is detrimental to vehicular UEs in the legacy scenario (no handover bias) if the speed of traffic exceeds 6 m/s, since $\nu_{\text{noBias}} \leq \nu_{\text{MeNB}}$. However, setting the optimal asymmetrical handover bias allows network operators to keep the FeNB switched on until the speed reaches 16 m/s, benefiting both pedestrian and vehicular UEs, since $\nu_{\text{Bias}} \geq \nu_{\text{MeNB}}$.

We also performed a sensitivity analysis by adding a normally distributed error with standard deviation σ_v to the velocity estimate used to determine the bias and performing multiple independent simulations. The metric we consider is the maximum value $\Delta\nu$ of the difference in the spectral efficiency for all the considered velocities. As shown in Table III, the effects of the errors in the speed estimation are negligible when compared to the randomness of the channel (represented by the standard deviation $\hat{\sigma}$ in the Table). This makes the system robust to small variations of the speed of the flow of traffic, as well as protecting it from imprecisions due to vehicles of different lengths (i.e., the parameter L in (1)).

The optimal asymmetrical handover bias over the course of a day for a specific intersection can be calculated from the TfL data as explained in Sec. III; the speed evolution shown in Fig. 2 results in the bias shown in Fig. 9. As expected, the handover bias is higher at nighttime, as the average speed of traffic is far higher than during the day. For this reason we can fix a threshold for the handover bias beyond which FeNB can be shut down in order to save energy, leaving all traffic to the MeNB. If we fix this threshold to 3 dB, then the FeNB will only turn off in the middle of the night, when the load on the MeNB is very light.

σ_v/v	0.1	0.2	0.3
$\Delta\nu$	$0.05\hat{\sigma}$	$0.07\hat{\sigma}$	$0.09\hat{\sigma}$

Table III: Effect of errors in the speed estimate on the system performance. $\hat{\sigma}$ is the standard deviation of ν across independent channel realizations.

B. Adaptive vMME Allocation

As already mentionded in Sec. II, NFV allows to dynamically allocate the resources needed by a cellular network. In traditional mobile networks a single dedicated MME is typically used to manage millions of end users, such as those in the London metropolitan area [41]. With the NFV approach, instead, it is possible to change the number of vMME instances on the fly, adapting to the number of handovers that are expected to happen in a certain interval.

In this application, we use data processed as in Sec. III to determine the number of handovers that happen in the London area during a typical day. We distinguish between the two kinds of handovers that may happen in LTE networks [46], i.e., intra MME (X2–based) and inter MME (S1–based) handovers, since they require different procedures and different interactions with the MMEs. The X2–based handover happens when the UE remains in an area managed by the same MME and changes the eNB to which it is attached. The S1–based procedure, instead, is used when the UE performs a handover between two eNBs managed by different MMEs. The two procedures are described in detail in [46]. In this paper, we consider the duration of a handover procedure as the interval from the instant in which the source eNB (SeNB) triggers the handover to the instant in which SeNB receives the `RELEASE_RESOURCES` command. During this period the UE first experiences a degraded channel, and then receives packets with an increased latency, thus the Quality of Service perceived by the final user decreases. The goal of this application is to minimize the duration of these intervals, while using as few vMME instances as possible.

In particular, we model the duration of an X2–based handover handled by vMME i as a function of the number of vMMEs N and of the total number of handovers I_i that involve that vMME during an interval T_{per}:

$$t_{HO}^{X2}(N, I_i) = 3t_{Se-Te} + 2t_{Te-SM}(N) + t_{HR} + \tau(I_i) \quad (16)$$

while the time required to complete an S1–based handover that involves vMMEs i and j also depends on the number of handovers I_j that are served by the target vMME j:

$$\begin{aligned} t_{HO}^{S1}(N, I_i, I_j) =& \tau_1(I_i) + 3\tau_2(I_j) + \\ & 4t_{Se-SM}(N) + 4t_{Te-TM}(N) + \\ & 2t_{SM-TM}(N) + \max\{t_{TM-SM}(N) + \\ & t_{SM-Se}(N) + t_{HR}, t_{TM-Se}(N)\} + \\ & \max\{t_{TM-SM}(N) + \tau_1(I_i), t_{TM-Se}\} \end{aligned} \quad (17)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2752761, IEEE Internet of Things Journal

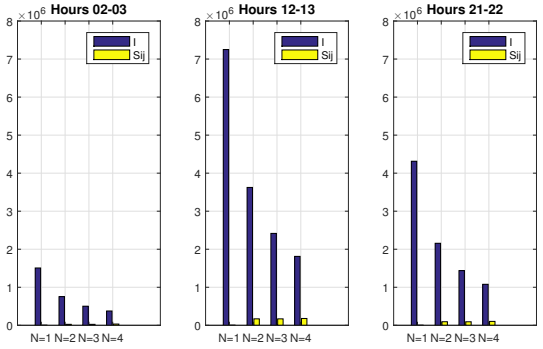


Fig. 10: Average number of X2–based and S1–based handovers per vMME instance, for a different N and different time slots, during January 23, 2015.

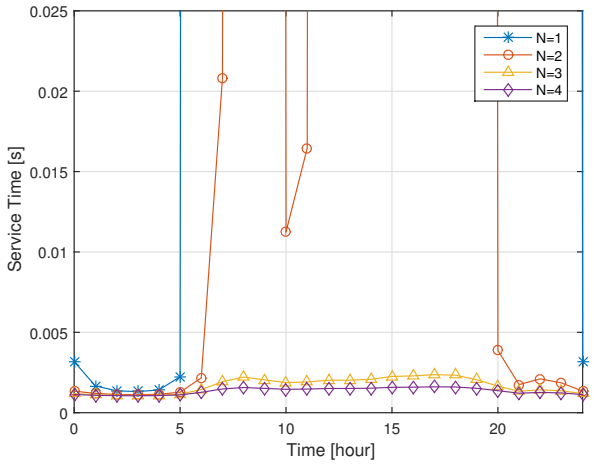


Fig. 11: Average service time τ for different N, during January 23, 2015.

In Eqs. (16) and (17), $t_{A-B}(N)$ with $A, B \in \{Te, Se, SM, TM\}$[1] is the latency between element A and element B of the network. Unless both A and B represent eNBs, we have

$$t_{A-B}(N) = t_{tx} + \frac{d_N(A,B)}{v_f}, \tag{18}$$

where $t_{tx} = 5$ ms is a factor that models the time spent in middleboxes and $t_{PROP} = d_N(A,B)/v_f$ is the propagation delay, given by the ratio of the distance between the two devices and the speed of light inside optical fibers[2] (i.e., $v_f = 2 \cdot 10^8$ m/s). The dependence on the number of vMMEs N is in the distance $d_N(A,B)$ between two network elements, that changes according to the allocation of eNBs to the vMMEs. Instead, t_{Te-Se} is the latency between two adjacent eNBs and does not depend on the relative position between the eNBs and the MMEs, therefore, as in [47], it is modeled as a constant latency $t_{Te-Se} = 2.5$ ms. t_{HR} is the duration of the interval from when the UE actually disconnects from the SeNB to when it connects to the TeNB. In [48], t_{HR} is estimated to be in the order of 50 ms.

Finally, $\tau(I_i)$ is the time that a vMME takes to process the received command. In [41] the process of handover requests is modeled as a Markov process. We adopt the same approach and in particular we model the vMME as an M/D/1 queue, assuming a Poisson arrival process with *arrival rate* $\lambda = I_i/T_{per}$ and a deterministic service time T_s. Given these assumptions, it is possible to compute the value of τ as the *system time* of an M/D/1 queue:

$$\tau = \frac{1}{\mu} + \frac{\rho}{2 \cdot \mu \cdot (1-\rho)}, \tag{19}$$

where $\mu = 1/T_s$ and $\rho = \lambda T$ are the *service rate* and the *loading factor* of the vMME. The study in [39] uses the value $T_s = 110$ μs as *service time* of a vMME, requiring considerable computational resources. Since our work only considers vehicular UEs, and the adaptive nature of our system, overdimensioning each vMME would be a waste of resources: a number of slow vMMEs can provide the same performance as a single powerful vMME during rush hour, and the additional vMMEs can be turned off at less congested times, with a substantial reduction in server management costs and energy requirements. For this reason, we limit the processing power of our vMMEs dedicated to vehicular handovers to the value of $\mu = 1000$ handovers per second.

Since our goal is to find the optimal number of vMMEs N that minimizes the total duration of the handovers, we consider the objective function

$$\begin{aligned} J_{T_{per}}(N) = &\sum_{i=1}^{N}(I_i - \sum_{\substack{j=1 \\ j\neq i}}^{N} S_{i,j}) t_{HO}^{X2}(N, I_i) \\ &+ \sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N} S_{i,j} t_{HO}^{S1}(N, I_i, I_j) + C(N), \end{aligned} \tag{20}$$

where the sums consider all the handovers in a time slot T_{per} of one hour, and $C(N)$ is a penalty function representing the operational cost of N vMMEs. We consider it to be a linear function of the number of vMMEs N, i.e., $C(N) = kN$.

The optimization problem uses the vehicular traffic data processed as in Sec. III to compute the value of I_i, $S_{i,j}$ and $\lambda(I_i) = I_i/T_{per}$ for each vMME $i, j \in \{1, \cdots, N\}$ and computes

$$N_{opt} = \min_N J_{T_{per}}(N) \tag{21}$$

for each interval T_{per} during a certain day.

In the following results we consider the data of January 23, 2015. Fig. 10 shows the average number of handovers inside a single vMME in different time slots. Notice that since we consider only the inter MME handovers for the London area MMEs, then S_{ij} is zero for $N = 1$. The number of handovers in different time slots changes greatly, from $1.5 \cdot 10^6$ per hour during the night to more than $7 \cdot 10^6$ at midday. This justifies a dynamic allocation of resources; a single and dedicated MME that targets the worst case scenario at midday would be wasted during the night. Instead the adaptive approach allows the use of less powerful vMMEs, which are able to

[1] Te stands for Target eNB, Se stands for Source eNB, TM stands for Target MME and SM stands for Source MME

[2] We assume that the backhaul network uses fiber-optic links.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2752761, IEEE Internet of Things Journal

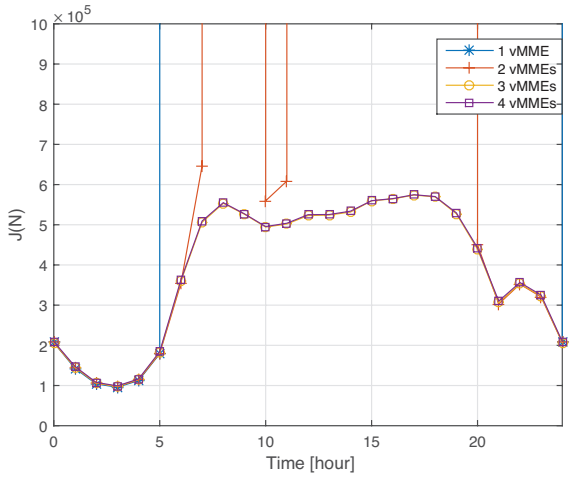


Fig. 12: Objective function $J(N)$, for $N \in \{1, 2, 3, 4\}$, during January 23, 2015.

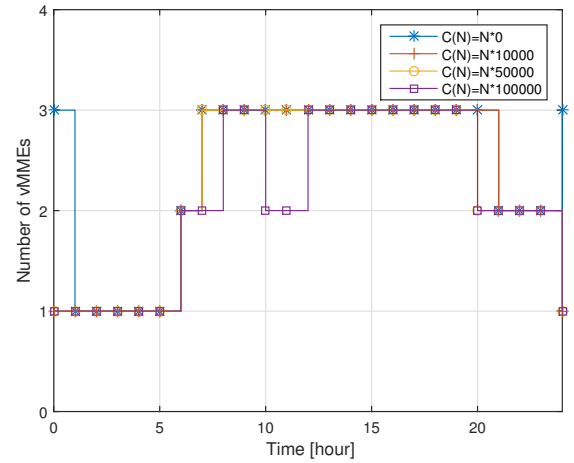


Fig. 13: N_{opt} for different costs $C(N)$, during January 23, 2015.

serve a smaller number of handover requests, and have lower operational expenses than dedicated hardware [37], but can be instantiated on the fly according to the control traffic intensity.

In Fig. 11, the average service time of the vMME instances is shown for different values of N. It can be seen that during the night the values have a small difference, but one or two vMME instances are not enough to handle the load during the day. Fig. 12, instead, shows the value of the objective function $J(N)$ throughout the whole day, assuming a cost factor $k = 0$. In this case, one vMME instance is enough only from midnight to 5 AM, and more instances (up to 3) must be allocated during the day to meet the vehicular handover traffic load.

If we increase the value of k, as shown in Fig. 13, the optimal number of vMMEs changes. At certain times using a lower number of vMMEs becomes more convenient, because of the operational cost which is now accounted for.

The adaptation of the number of vMMEs significantly improves the efficiency of the system: while a worst-case dimensioned system would need 3 vMMEs at all times, the average number of active vMME instances for the most aggressive adaptive system ($k = 0$) is 2.42, while a more conservative system ($k = 100000$) only uses an average of 2.17 vMMEs. This translates into a lower operating cost for the network provider because of a reduced energy consumption and of the need of using fewer virtual functions.

V. Conclusion And Future Work

In this work, we presented two optimization methods that exploit road traffic data to adapt several parameters in a cellular network. Our focus was mostly on handovers, and we showed that a knowledge of the traffic on each road and its speed can help improve the handover performance. A tighter integration between the smart city and the cellular network that serves it might be one of the most promising approaches towards Self-Organizing Networks.

In particular, we exploited our knowledge of the speed of the traffic at any intersection to adapt the femtocell range expansion bias and mitigate the inefficiency caused by the TTT without incurring in the ping pong effect. Since the calculation is simple, this can be easily implemented in real time. We also use the traffic flow data to adaptively provision virtual resources and add or remove virtual MMEs, reducing operating costs without impacting the performance with respect to a worst-case dimensioned system. The performance benefits of the scheme can further increase as the integration of smart city data in the network optimization progresses: for example, data about public transport networks such as buses and the subway system can be exploited to provide a more accurate estimation of the metrics we considered. Moreover, periodic or forecastable events (i.e., holidays and changes in the weather conditions) that impact mobility patterns can be added to the model in order to improve its accuracy.

The two techniques we used in this work are just two examples of the possible benefits that smart city data can provide to cellular networks: in the future, we plan to systematize this approach and integrate existing and new SON techniques, studying and optimizing their interactions using data from both the cellular network itself and the smart city around it. Another challenge for future systems of this kind is the integration with novel technologies such as mmWave, which requires intelligent mobility management.

Acknowledgments

We would like to thank London Smart City and Transport for London (TfL) for providing this research with the road traffic data from the city of London. We also thank Dr. M. Condoluci of King's College of London for his contribution in the definition of the project.

References

[1] M. Dalla Cia, F. Mason, D. Peron, F. Chiariotti, M. Polese, T. Mahmoodi, M. Zorzi, and A. Zanella, "Mobility-aware Handover Strategies in Smart Cities," in *International Symposium on Wireless Communication Systems (ISWCS)*, Aug. 2017.

[2] P. K. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, "Design considerations for a 5G network architecture," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 65–75, Nov. 2014.

[3] Ericsson, "Mobility Report - On The Pulse Of The Networked Society," June 2016. [Online]. Available: https://www.ericsson.com/mobility-report

[4] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Sensing as a Service model for smart cities supported by Internet of Things," *Transactions on Emerging Telecommunications Technologies (ETT)*, vol. 25, no. 1, pp. 81–93, Sep. 2014.

[5] N. Bui, M. Cesana, S. A. Hosseini, Q. Liao, I. Malanchini, and J. Widmer, "A survey of anticipatory mobile networking: Context-based classification, prediction methodologies, and optimization techniques," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1790–1821, July 2017.

[6] F. Chiariotti, M. Condoluci, T. Mahmoodi, and A. Zanella, "Symbiocity: Smart cities for smarter networks," *Transactions on Emerging Telecommunications Technologies (ETT)*, June 2017, e3206 ett.3206. [Online]. Available: http://dx.doi.org/10.1002/ett.3206

[7] P. Hunt, D. Robertson, R. Bretherton, and M. C. Royle, "The SCOOT on-line traffic signal optimisation technique," *Traffic Engineering & Control*, vol. 23, no. 4, Apr. 1982.

[8] P. Muñoz, R. Barco, and I. de la Bandera, "On the potential of handover parameter optimization for self-organizing networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 1895–1905, Feb. 2013.

[9] F. Guidolin, I. Pappalardo, A. Zanella, and M. Zorzi, "Context-aware handover policies in HetNets," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 1895–1906, Mar. 2016.

[10] P. Neirotti, A. De Marco, A. C. Cagliano, G. Mangano, and F. Scorrano, "Current trends in smart city initiatives: Some stylised facts," *Cities*, vol. 38, pp. 25–36, June 2014.

[11] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[12] C. H. Liu, J. Fan, J. W. Branch, and K. K. Leung, "Toward QoI and energy-efficiency in Internet-of-Things sensory environments," *IEEE Transactions on Emerging Topics in Computing*, vol. 2, no. 4, pp. 473–487, Dec. 2014.

[13] L. Filipponi, A. Vitaletti, G. Landi, V. Memeo, G. Laura, and P. Pucci, "Smart city: An event driven architecture for monitoring public spaces with heterogeneous sensors," in *IEEE 4th International Conference on Sensor Technologies and Applications (SENSORCOMM)*, 2010, pp. 281–286.

[14] M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi, "Long-Range Communications in Unlicensed Bands: the Rising Stars in the IoT and Smart City Scenarios," *IEEE Wireless Communications*, vol. 23, Oct. 2016.

[15] A. Biral, M. Centenaro, A. Zanella, L. Vangelista, and M. Zorzi, "The challenges of M2M massive access in wireless cellular networks," *Digital Communications and Networks*, vol. 1, no. 1, pp. 1–19, Feb. 2015.

[16] M. Polese, M. Centenaro, A. Zanella, and M. Zorzi, "M2M massive access in LTE: RACH performance evaluation in a Smart City scenario," in *IEEE International Conference on Communications (ICC)*, May 2016.

[17] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, "Big data-driven optimization for mobile networks toward 5G," *IEEE Network*, vol. 30, no. 1, pp. 44–51, Jan. 2016.

[18] A. Imran, A. Zoha, and A. Abu-Dayya, "Challenges in 5G: how to empower SON with big data for enabling 5G," *IEEE Network*, vol. 28, no. 6, pp. 27–33, Nov. 2014.

[19] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," in *ACM 1st MCC workshop on Mobile cloud computing*, Aug. 2012, pp. 13–16.

[20] M. Zorzi, A. Zanella, A. Testolin, M. D. F. De Grazia, and M. Zorzi, "Cognition-based networks: a new perspective on network optimization using learning and distributed intelligence," *IEEE Access*, vol. 3, pp. 1512–1530, Aug. 2015.

[21] P. I. Bratanov and E. Bonek, "Mobility model of vehicle-borne terminals in urban cellular systems," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 4, pp. 947–952, July 2003.

[22] A. S. Hassani, A. R. Momen, and P. Azmi, "Mobility model of vehicular terminals in cellular networks," in *2nd International Conference on Information Communication Technologies*, vol. 2, 2006, pp. 2434–2437.

[23] Q. Dong and W. Dargie, "A survey on mobility and mobility-aware MAC protocols in wireless sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 88–100, Feb. 2013.

[24] S. Vasudevan, R. N. Pupala, and K. Sivanesan, "Dynamic eICIC – A Proactive Strategy for Improving Spectral Efficiencies of Heterogeneous LTE Cellular Networks by Leveraging User Mobility and Traffic Dynamics," *IEEE Transactions on Wireless Communications*, vol. 12, no. 10, pp. 4956–4969, Oct. 2013.

[25] P.-C. Lin, L. F. Gonzalez Casanova, and B. K. Fatty, "Data-Driven Handover Optimization in Next Generation Mobile Communication Networks," *Mobile Information Systems*, vol. 2016, July 2016.

[26] Z. Yan, H. Zhou, H. Zhang, and S. Zhang, "Speed-Based Probability-Driven Seamless Handover Scheme between WLAN and UMTS," in *2008 The 4th International Conference on Mobile Ad-hoc and Sensor Networks*, Dec. 2008, pp. 110–115.

[27] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 497–508, Mar. 2012.

[28] O. G. Aliu, A. Imran, M. A. Imran, and B. Evans, "A survey of self organisation in future cellular networks," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 336–361, Feb. 2013.

[29] X. Duan and X. Wang, "Authentication handover and privacy protection in 5G hetnets using software-defined networking," *IEEE Communications Magazine*, vol. 53, no. 4, pp. 28–35, Apr. 2015.

[30] D. Xenakis, N. Passas, L. Merakos, and C. Verikoukis, "Mobility management for femtocells in LTE-advanced: key aspects and survey of handover decision algorithms," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 64–91, July 2014.

[31] 3GPP Technical Specification 36.331, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification." [Online]. Available: www.3gpp.org

[32] Y. Lee, B. Shin, J. Lim, and D. Hong, "Effects of time-to-trigger parameter on handover performance in SON-based LTE systems," in *IEEE 16th Asia-Pacific Conference on Communications (APCC)*, 2010, pp. 492–496.

[33] I. Pappalardo, A. Zanella, and M. Zorzi, "Upper bound analysis of the handover performance in hetnets," *IEEE Communications Letters*, vol. 21, no. 2, pp. 418–421, Feb. 2017.

[34] K. Kitagawa, T. Komine, T. Yamamoto, and S. Konishi, "A handover optimization algorithm with mobility robustness for LTE systems," in *IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, 2011, pp. 1647–1651.

[35] S. Nie, D. Wu, M. Zhao, X. Gu, L. Zhang, and L. Lu, "An Enhanced Mobility State Estimation Based Handover Optimization Algorithm in LTE-A Self-organizing Network," *Procedia Computer Science*, vol. 52, pp. 270–277, June 2015.

[36] R. Arshad, H. Elsawy, S. Sorour, T. Y. Al-Naffouri, and M. S. Alouini, "Handover Management in 5G and Beyond: A Topology Aware Skipping Approach," *IEEE Access*, vol. 4, pp. 9073–9081, Dec. 2016.

[37] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, "Network function virtualization: State-of-the-art and research challenges," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 236–262, Sep. 2015.

[38] C. L. I, J. Huang, R. Duan, C. Cui, J. X. Jiang and L. Li, "Recent Progress on C-RAN Centralization and Cloudification," *IEEE Access*, vol. 2, pp. 1030–1039, Aug. 2014.

[39] A. S. Rajan, S. Gobriel, C. Maciocco, K. B. Ramia, S. Kapury, A. Singhy, J. Ermanz, V. Gopalakrishnanz, and R. Janaz, "Understanding the bottlenecks in virtualizing cellular core network functions," in *IEEE 21st International Workshop on Local and Metropolitan Area Networks*. IEEE, 2015, pp. 1–6.

[40] X. An, F. Pianese, I. Widjaja, and U. G. Acer, "dMME: Virtualizing LTE mobility management," in *IEEE 36th Conference on Local Computer Networks (LCN)*, 2011, pp. 528–536.

[41] P. Andres-Maldonado, P. Ameigeiras, J. Prados-Garzon, J. J. Ramos-Munoz, and J. M. Lopez-Soler, "Virtualized MME design for IoT support in 5G systems," *Sensors*, vol. 16, no. 8, p. 1338, 2016.

[42] P. Bhat, S. Nagata, L. Campoy, I. Berberana, T. Derham, G. Liu, X. Shen, P. Zong, and J. Yang, "LTE-advanced: an operator perspective," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 104–114, Feb. 2012.

[43] R. Tanbourgi, S. Singh, J. G. Andrews, and F. K. Jondral, "Analysis of non-coherent joint-transmission cooperation in heterogeneous cellular networks," in *IEEE International Conference on Communications (ICC)*, June 2014, pp. 5160–5165.

[44] A. Goldsmith, *Wireless communications*. Cambridge University Press, 2005.

[45] 3GPP TR 36.839, "Evolved universal terrestrial radio access (E-UTRA); Mobility enhancements in heterogeneous networks," Sep. 2012.

[46] S. Sesia, I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice*. Wiley Publishing, 2009.

[47] Next Generation Mobile Networks Alliance, "Optimised backhaul requirements," 2008.

[48] 3GPP Technical Specification 36.881, "Study on latency reduction techniques for LTE." [Online]. Available: www.3gpp.org