IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 28, 2017, accepted September 6, 2017, date of publication September 11, 2017, date of current version October 25, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2751422

Mobility-Aware User Association for 5G mmWave Networks

ANGELA SARA CACCIAPUOTI, (Senior Member, IEEE)
[1]National Laboratory of Multimedia Communications, National Inter-University Consortium for Telecommunications, 80126 Naples, Italy
[2]Department of Electrical Engineering and Information Technology, University of Naples Federico II, 80138 Naples, Italy

e-mail: angelasara.cacciapuoti@unina.it; angelasara.cacciapuoti@ieee.org

This work was supported by CNIT.

ABSTRACT In this paper, we design a mobility-aware user association strategy for millimeter-wave (mmW) networks to overcome the limitations of the conventional received power (RSS)-based association strategies in a mobile 5G scenario. More in detail, first the design of a mobility-aware strategy for user association in 5G mmW networks is posed as a constrained optimization problem. Then, it is showed that the proposed strategy exhibits several attractive features: a) it is able to track the dynamic changes in the network topology and in the channel conditions induced by the user mobility; b) it takes into account the distribution of the loads among the small base stations (sBSs), thus overcoming to associate an UE to an already congested sBS. This, in turn, affects positively the overall fairness of the network; c) it overcomes overly frequent handovers between sBSs, and thus the need of frequent re-association; d) it takes into account the peculiar aspects of the mmW communications, such as directionality, sensitivity to blockage, and NLoS propagation effects; and e) it is fully distributed, i.e., each mobile user associates to an sBS independently of each other, stemming from local information only. Furthermore, it is showed that the exhaustive search for the solution of the posed optimization problem is computationally unfeasible. Consequently, within this paper, an efficient algorithm exhibiting a polynomial-time complexity is proposed. Finally, the numerical results validate the benefits of adopting the proposed mobility-aware and fully distributed association rule. In particular, it is quantified the very significant performance enhancement of the proposed association with respect to the conventional RSS-based one.

INDEX TERMS mmW communications, 5G networks, user association.

I. INTRODUCTION

The fifth generation (5G) wireless communication paradigm has been proposed to face the bottleneck of current wireless systems in addressing the continuously growing demand for high data rate services [1]–[4]. To this aim, there is a general consensus in both the academia and the industry communities that network densification and the use of millimeter-wave (mmW) are key technologies for enabling the 5G paradigm [1]–[4].

The mmW communications are characterized by higher path-loss and stronger sensitivity to blockage than the microwave (μW) communications [2], [4]–[7]. To face with these impairments, the small wavelengths of mmW frequencies allow for the implementation of a large number of antenna elements in a small space to make narrow beams (pencil-beams) [2], [7]. With this directionality, the links established in the mmW band can be in the range of 200 meters [5]. Furthermore, the network densification generally helps mmW networks, since it increases the probability of serving small Base Station (sBSs) being in Line-of-Sight (LoS). Recent studies have shown that with such highly directional transmissions and sensitivity to blockage, a positive side effect is that the interference is greatly reduced, and so in many or most cases, mmW networks will be noise rather than interference-limited [4], [5], [7].

These unique characteristics make mmW networks completely different from conventional wireless networks, which are instead mainly interference-limited. In particular, the aforementioned mmW channel characteristics, included the NLoS propagation effects [4], have a significant effect on cell coverage, which indicates that the attainable throughput of mmW networks is highly dependent on the user association, i.e., the process that associates a user with a particular serving sBS. In fact, the user association governs the long-term resource allocation policies of a wireless network. Hence, it is urgently needed the design of new association rules for mmW networks as recently surveyed in [4].

More in detail, the current standards for mmW communications, such as the IEEE 802.11ad and IEEE 802.15.3c, adopt the received power based user association rule, which associates a user with the sBS providing the maximum received signal strength (max-RSS) [4]. Such a rule appears rudimentary and ineffective for addressing the unique features of the emerging 5G networks [3], [4], [7], [8], due to the limited size of the sBSs cells and their dense deployment. In particular, the RSS-based association leads to an unbalanced number of users per sBS (unbalanced sBS loads), which limits the available resources per user in highly populated areas, while wasting resources in sparse areas. This in turn affects the network fairness, since overloaded sBSs cannot provide the same resources to their users as the underloaded sBSs. Moreover, RSS-based user association may result in overly frequent handovers between the adjacent sBSs and may increase the overhead/delay of re-association [4].

The above described issues become even more crucial in mmW mobile environments as well-highlighted in [4], to which we refer for a survey on the state-of-the-art about the mmW association rules. In fact, in presence of mobility, the network topology and the channel conditions vary in time. Hence, by associating a mobile user to the closest sBS, without taking into account the mobility patterns, may result in even more frequent handovers than in static environments. To date, to the best of our knowledge, this problem has not yet been investigated.

Hence in this paper, we design a mobility-aware user association strategy for mmW networks, to overcome the limitations of the conventional RSS-based association.

More in detail, first the design of the mobility-aware user association rule is posed as a constrained optimization problem. Then, it is showed that the association resulting by solving such a constrained optimization problem exhibits several attractive features:

a) It is able to track the dynamic changes in the network topology and in the channel conditions induced by the user mobility;
b) It takes into account the sBS loads, thus overcoming to associate a UE to an already congested sBS. This in turn affects positively the overall fairness of the network;
c) It overcomes overly frequent handovers between sBSs and thus the need of frequent re-association;
d) It takes into account the peculiar aspects of the mmW communications, such as directionality, sensitivity to blockage, and NLoS propagation effects.
e) It is fully-distributed, i.e., each mobile user, referred in the following as user equipment (UE) accordingly to the 3GPP terminology, associates to a sBS independently of other UEs and stemming from local information only.

Regarding the feature labeled as *e*), it is widely recognized that distributed solutions alleviate the computational and signaling overhead with respect to centralized solutions, since they need only local and not global information.

Furthermore, it is showed, within the manuscript, that the exhaustive search for the solution of the posed optimization problem is computationally unfeasible. Consequently, for solving the posed optimization problem we also design an efficient algorithm, which exhibits a polynomial-time complexity.

Finally, the numerical results validate the benefits of adopting the proposed mobility-aware and fully-distributed association rule. In particular, it is quantified the very significant performance enhancement of the proposed association with respect to the conventional RSS-based one.

The rest of the paper is organized as follows: Section II describes the system model along with the adopted blockage models. In Section III, the fully-distributed and mobility-aware association strategy is designed. In this same section, the aforementioned properties of the proposed solution are showed along with the design of the computational-feasible algorithm. This section ends with a discussion about the proposed association rule. Section IV presents the numerical results that validate the significant benefits of adopting the proposed strategy. Finally, we conclude the paper in Section V and some proofs are gathered in the appendices.

II. SYSTEM MODEL

In the region $\mathcal{A} \in \mathbb{R}^2$, B sBSs are deployed according to the spatial distribution $f_{\mathbf{x}_B}(\mathbf{x}_B)$.[1] We denote the set of these sBSs as $\mathcal{B} = \{1, 2, \dots, B\}$. In the same region, U mobile UEs move according to some mobility models characterized by a steady-state spatial distribution $f_{\mathbf{x}_u}(\mathbf{x}_u)$ with u belonging to the set of UEs denoted as $\mathcal{U} = \{u_1, u_2, \dots, u_U\}$.

As widely done in literature [7], [9], we assume the sBSs capable of generating multiple pencil-beams, whereas the UEs are able to generate a single pencil-beam, due to the huge complexity of multiple beam generation. This agrees with the recent standards, including IEEE 802.11ad and IEEE 802.15.3c.

Each sBS serves the associated UEs in a time slotted fashion, with time slot duration T_s, counting from time slot $n = 1$, according to a FIFO policy.

A. BLOCKAGE AND CAPACITY MODELS

In this subsection, we collect some definitions that will be used through the paper and Table 1 summarizes the adopted notation.

Definition 1. $P_{u,b}^{\mathrm{LoS}}(d)$ denotes the probability of having a LoS connection between UE u and sBS b, when their relative distance is d. The corresponding achievable capacity is denoted as $\mathcal{C}_{u,b}^{\mathrm{LoS}}(d)$.

Definition 2. $P_{u,b}^{\mathrm{NLoS}}(d)$ denotes the probability of having a NLoS connection between UE u and sBS b, when their relative distance is d. The corresponding achievable capacity is denoted as $\mathcal{C}_{u,b}^{\mathrm{NLoS}}(d)$.

By exploiting the above definitions, it results that the capacity achievable by UE u when it is at distance d from

[1] A widely-adopted assumption in literature is to consider the sBSs spatial distributed accordingly to a Poisson point process [2], [4], [7]. Nevertheless, the results derived in the following do not depend on the adopted sBSs spatial distribution.

TABLE 1. Adopted notation.

Symbol	Definition
B	number of sBSs deployed within region $\mathcal{A} \in \mathbb{R}^2$
U	number of mobile UEs in region $\mathcal{A} \in \mathbb{R}^2$
T_s	sBS time slot duration
T_a^u	association procedure period of UE u
$P_{u,b}^{\text{LoS}}(d)$	probability of having a LoS connection between UE u and sBS b, at distance d
$P_{u,b}^{\text{NLoS}}(d)$	probability of having a NLoS connection between UE u and sBS b, at distance d
$\mathcal{C}_b^u(d)$	achievable capacity of the UE-sBS pair (u, b), at distance d
γ_b	maximum fraction of resources sBS b can allocate to a UE in an arbitrary time slot
ϵ_b	alignment cost to counteract the deafness
$\rho_b(n)$	load of sBS b in the arbitrary time slot n
$t_b(n)$	portion of the n-th time slot of sBS b occupied by the UEs already associated to sBS b
$x_b^u(n)$	association binary variable between UE u and sBS b in the time slot n

sBS b can be evaluated as follows:

$$\mathcal{C}_b^u(d) = \mathcal{C}_{u,b}^{\text{LoS}}(d)P_{u,b}^{\text{LoS}}(d) + \mathcal{C}_{u,b}^{\text{NLoS}}(d)P_{u,b}^{\text{NLoS}}(d) \quad (1)$$

We stress that equation (1) is valid regardless of the adopted models for the channel capacities and the LoS/NLoS probabilities. In particular, the theoretical analysis developed in Section III depends on the adopted blockage and channel model only through the use of the above generic function $\mathcal{C}_b^u(d)$ that in turn depends on: *i*) the probabilities of having LoS/NLoS connections between UE u and sBS b at distance d; *ii*) the corresponding achievable capacities of the UE-sBS pair (u, b). Hence, any arbitrary blockage and capacity model can be used and the analysis continues to hold. In the following, we introduce some specific choices for the sake of clarity.

Regarding the blocking, any model for $P_{u,b}^{\text{LoS}}(d)$ and $P_{u,b}^{\text{NLoS}}(d)$ can be adopted as the ones in [5] and [6] or the ones surveyed in [2]. The common feature of these blockage models, supported by experimental evidences, is to declare the viability of mmW cellular network up to cell radii on the order of 200 m [2]. Hence, the probabilities of having a LoS and NLoS connections go to zero for distances between a UE and a sBS beyond 200m.

Regarding the capacity model, we observe that the assumption of having noise-limited mmW networks has been considered and motivated extensively in literature [4], [5], [7]. In particular, in [4] it has been showed that this assumption holds even for high densities of mmW sBSs. Henceforth, this assumption will be considered within this paper. By taking into account this, the achievable LoS capacity of the UE-sBS pair (u, b) at a given relative distance d can be calculated as [4], [6], [7], [10]:

$$\mathcal{C}_{u,b}^{\text{LoS}}(d) = W_m \log\left(1 + \xi_b^u(d)\right), \quad (2)$$

where $\xi_b^u(d)$ denotes the average SNR observed at the distance d between UE u and sBS b, in the mmW spectrum of width $W_{u,b}$. $\xi_b^u(d)$ can be evaluated as in [4] or as in [7] to account for the directional gain of the mmW antennas that in turn depends on the beamwidth.

Similarly and in agreement with the literature, when a NLoS link is available for the UE-sBS pair (u, b) at a given relative distance d, we adopt the widely-used model that scales the LoS SNR with a factor Δ [6], [10]:

$$\mathcal{C}_{u,b}^{\text{NLoS}}(d) = W_m \log\left(1 + \frac{\xi_b^u(d)}{\Delta}\right) \quad (3)$$

III. MOBILITY-AWARE ASSOCIATION DESIGN

The goal of this paper is to model and analyze a mobility-aware association rule for mmW networks with the objective of overcoming the limitations of the RSS-based association rules. Hence, in a nutshell, the problem is to decide to which sBS the mobile UE u should be associated with.

This problem is addressed in this section, by accounting for the key features *a*) $-$ *e*) described in Section I. To this aim, first some preliminaries are collected in Subsection III-A.

A. PRELIMINARIES

In the following, without loss of generality, we focus on the arbitrary UE u. Let us denote with t_0^u the arbitrary time at which UE u starts its association within region $\mathcal{A}$, belonging, without loss of generality, to the arbitrary n_0-th time slot of the sBSs.[2]

The association is a long-term procedure, repeated every T_a^u corresponding to N_a^u sBSs time slots of duration T_s.

Definition 3. γ_b denotes the maximum fraction of the resources that sBS b can allocate to a UE in an arbitrary time slot of duration T_s.

Definition 4. ϵ_b denotes the time needed by sBS b for establishing a pencil-beam connection with an UE in an arbitrary time slot. Hence, this parameter accounts for the beam alignment overhead required to counteract the deafness in mmW networks. By accounting for Def. 3, it is very reasonable to consider $\epsilon_b < \gamma_b T_s$.

Definition 5. $\rho_b(n)$ denotes the load of sBS b in the arbitrary time slot $n \geq n_0$, generated by the UEs already associated to such a sBS in that time-slot. Consequently, by defining $t_b(n) \triangleq [(n-1) + \rho_b(n)]\, T_s$, we account for the portion of the n-th time slot of sBS b occupied by the UEs already associated to sBS b.

[2]For the sake of simplicity and clarity, the sBSs are assumed synchronized in the time slots. This is a reasonable assumption. In fact, as pointed out in [11], the sBSs can serve only UEs of the same operator. Hence it is reasonable to assume that sBSs of the same operator are synchronized in the time slots. Clearly, for different operators, different sets of sBSs could be individuated and the analysis continues to hold for each of these sets, generalizing so the analysis to the case of un-synchronized sBSs.

Definition 6. Let us denote with $x_b^u(n)$ the binary variable indicating whether UE u is associated in the arbitrary time slot $n \geq n_0$ to sBS $b \in \mathcal{B} = \{1, \ldots, B\}$:

$$x_b^u(n) = \begin{cases} 1, & \text{if } u \text{ is associated to } b \text{ in the time slot } n \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

Definition 7. Let us denote with $1_b^u(n)$ the binary variable indicating whether UE u changes the associated sBS in the arbitrary time slot $n \geq n_0$, with respect to the previous time slot $n-1$, i.e.:

$$1_b^u(n) = \begin{cases} 1, & |x_b^u(n) - x_b^u(n-1)| = 1 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

with $x_b^u(n_0 - 1) = 0$ to account for the initial beam alignment overhead.

Stemming from the above definitions, in Propositions 1 and 2 we derive the achievable data load in an arbitrary time slot $n \geq n_0$, i.e., the maximum number of bits that could be exchanged between UE u and sBS b in the considered time slot by utilizing the mmW spectrum.

Proposition 1. *Given that t_0^u at which UE u starts the association belongs to the n_0-th time slot and that the load of sBS b in n_0 is $\rho_b(n_0)$, the conditional data load achievable by UE u, if it is associated to sBS b in n_0, is given by equation* (6), *reported at the top of the next page. In* (6)*, $C_b^u(d(t))$ is the capacity at distance $d(t)$ evaluated in* (1)*, and ϵ_b and $t_b(n)$ are defined in Defs. 4 and 5, respectively.*

Proof: *See Appendix A.* ■

Proposition 2. *Given that t_0^u at which UE u starts the association belongs to the n_0-th time slot and that the load of sBS b in $n > n_0$ is $\rho_b(n)$, the conditional data load achievable by UE u, if it is associated to sBS b in n, is given by:*

$$\ell_b^u(n) = \int_{\min\{t_b(n)+1_b^u(n)\epsilon_b, nT_s\}}^{\min\{t_b(n)+\gamma_b T_s, nT_s\}} C_b^u(d(t))dt \tag{7}$$

where $C_b^u(d(t))$ is given in (1).

Proof: *See Appendix B.* ■

By accounting for the assumption $\epsilon_b < \gamma_b T_s$ made in Def. 4, the integrals in Propositions 1 and 2 are no-negative. This is in agreement with the definition of data load. In fact, the maximum number of bits that could be exchanged between UE u and sBS b in the considered time slot cannot be negative.

B. DISTRIBUTED OPTIMIZATION PROBLEM

Let $\mathbf{x}_u$ denote the matrix that collects the association indicators $\{x_b^u(n)\}$ of UE u for each sBS and for each time slot within the association period, i.e., $\forall n \in \{n_0, \ldots, n_0 + N_a^u - 1\}$.

We pose the following constrained optimization problem for the association in a mobile mmW network, whose objective function is the overall data load achievable within the temporal horizon N_a^u.

Optimization Problem:

$$\max_{\mathbf{x}_u} \sum_{b=1}^{B} \sum_{n=n_0}^{n_0+N_a^u-1} x_b^u(n)\ell_b^u(n) \tag{8a}$$

$$\text{Subject to: } \sum_{b=1}^{B} x_b^u(n) \leq 1, \quad \forall n \in \{n_0, \ldots, n_0 + N_a^u - 1\} \tag{8b}$$

$$\sum_{b=1}^{B} \sum_{n=n_0}^{n_0+N_a^u-1} 1_b^u(n)\epsilon_b \leq \mathcal{E}_b \tag{8c}$$

$$x_b^u(n) \in \{0, 1\}, b \in \mathcal{B}, \quad n \in \{n_0, \ldots, n_0 + N_a^u - 1\} \tag{8d}$$

with $\ell_b^u(n)$ given in (6) if $n = n_0$, given in (7) otherwise.

The constraint in (8b) assures that in each time slot UE u is associated with at most one sBS, being the UE able to generate a single pencil-beam.

The constraint in (8c) implies that the total alignment cost accumulated in at most N_a^u time slots does not exceed a certain threshold, denoted as $\mathcal{E}_b$. Hence, by accounting for (5), it results that the proposed problem formulation accounts for the cost associated to change the serving sBS between two consecutive time slots. As an example, if $\mathcal{E}_b = \epsilon_b$, UE u can associate to an unique sBS, i.e., it is not possible for UE u to associate to different sBSs in the different time slots belonging to T_a^u.

Remark 1. If the relative distance between UE u and a certain sBS does not fall in the mmW-transmission range in any time slot $n \in \{n_0, \ldots, n_0 + N_a^u - 1\}$, accordingly to the capacity expression (1), the resulting data load $\ell_b^u(n)$ in each time slot is zero. Hence, this sBS does not contribute to maximize the achievable data load in (8a), being so discharged by the UE as possible serving sBS. As a consequence, the dynamic variation of the set of sBSs with which UE u could establish a mmW communication induced by the user mobility is implicitly taken into account within the objective function (8a), through the expression of the achievable data load $\ell_b^u(n)$.

Remark 2. The objective function in (8a), through the expression of the achievable data load $\ell_b^u(n)$ given in Propositions 1 and 2, is able not only to track the dynamics induced by the mobility (Remark 1), but also to take into account the peculiar aspects of the mmW communications: i) directionality, through the alignment cost ϵ_b, ii) mmW channel propagation conditions including blockage and NLoS propagation effects through $C_b^u(d(t))$ in (1).

Remark 3. Finally, the objective function in (8a), through the expression of the achievable data load $\ell_b^u(n)$ given in Propositions 1 and 2, is able to account for the sBSs loads. Specifically, when a sBS is congested in a certain time slot, it provides an achievable data load in that time slot equal to zero. As a consequence, such a sBS will not contribute to maximize the overall achievable data load (8a), being so discharged by UE u as possible serving sBS. This is proved formally in Lemma 1. This property constitutes a further important advance with respect to the RSS-based association rule. In fact, as widely recognized, RSS-based strategies do not account for the sBSs loads, and hence a UE may be associated with a congested sBS.

$$\ell_b^u(n_0) = \begin{cases} \displaystyle\int_{\min\{t_0^u+1_b^u(n_0)\epsilon_b, n_0T_s\}}^{\min\{t_0^u+\gamma_b T_s, n_0T_s\}} C_b^u(d(t))dt, & \text{if } t_b(n_0) < t_0^u \leq n_0T_s \\ \displaystyle\int_{\min\{t_b(n_0)+1_b^u(n_0)\epsilon_b, n_0T_s\}}^{\min\{t_b(n_0)+\gamma_b T_s, n_0T_s\}} C_b^u(d(t))dt, & \text{if } (n_0-1)T_s \leq t_0^u \leq t_b(n_0) \end{cases} \tag{6}$$

Lemma 1. *If in an arbitrary time slot $n \geq n_0$, UE u is associated to sBS b accordingly to the optimization problem* (8)*, then the sBS load $\rho_b(n)$ in the considered time slot is strictly smaller than 1, i.e., $\rho_b(n) < 1$.*

Proof: *See Appendix C.* ■

Lemma 2. *If in an arbitrary time slot $n \geq n_0$, UE u is associated to sBS b accordingly to the optimization problem* (8)*, then the sum of the sBS load $\rho_b(n)$ and the contribution to the load of UE u in the considered time slot cannot exceed 1.*

Proof: *See Appendix D.* ■

Remark 4. The above Lemmas imply that we do not need to insert in the optimization problem a constraint on the load, i.e., we do not need to impose that $\rho_b(n) < 1$ and that the cumulative load accounting also for the UE u contribution is not greater than 1, since this is automatically taken into account in the objective function (8a) through $\ell_b^u(n)$ evaluated in Propositions 1 and 2.

C. FURTHER PROPERTIES

The aim of this section is to analytically disclose the dependence of the proposed optimization problem on some key parameters as the UE mobility patterns, the spatial distributions of the sBSs and the length of the period after which the association procedure is repeated.

To obtain insightful closed-form expressions, we exploit a very light assumption, i.e., we assume that $\epsilon_b = \epsilon, \forall b \in \mathcal{B}$ and $\gamma_b = \gamma, \forall b \in \mathcal{B}$. Indeed, as pointed out in [11], the sBSs can serve only UEs of the same operator. Hence, it is reasonable to assume that sBSs of the same operator devote the same fraction of resources to a UE and spend the same time for the UE alignment. Clearly, for different operators, different sets of sBSs could be individuated and the analysis continues to hold for each of these sets.

Under this assumption, the following results hold.

Proposition 3. *By adopting the proposed association strategy, the communication opportunities of the arbitrary UE u last for an average time T_o^u satisfying the following inequality:*

$$T_o^u \leq N_a^u \gamma T_s - \epsilon = T_a^u \gamma - \epsilon \tag{9}$$

Proof: *See Appendix E* ■

Remark 5. Proposition 3 analytically proves that the communication opportunities of the arbitrary UE u increase linearly as the alignment cost ϵ decreases. This result is reasonable since, as the alignment cost decreases, more time can be devoted to the data transmissions. Furthermore, Proposition 3 analytically proves that the communication opportunities linearly increase with the association period $T_a^u = N_a^u T_s$. This result is reasonable since, by increasing T_a^u, the dynamics of the network topology induced by the mobility and the dynamics of the sBS loads induced by the UE traffic patterns are better exploited. In fact, the higher is N_a^u (equivalently T_a^u), the higher is the dimensionality of the solution space of the optimization problem proposed in (8). However, by increasing N_a^u, the time complexity of the proposed association procedure increases, as highlighted in Insight 2. Hence, there exists a trade-off between communication opportunities and complexity.

Corollary 1. *By adopting the proposed association strategy when the alignment constraint in* (8c) *is the tightest one, i.e., $\epsilon \leq \mathcal{E} < 2\epsilon$, the communication opportunities of the arbitrary UE u last for an average time T_o^u satisfying the following inequality:*

$$T_o^u < \left(\overline{\mathcal{S}}_m^u + T_s\right)\gamma - \epsilon, \tag{10}$$

where $\overline{\mathcal{S}}_m^u$ denotes the average sojourn time UE u spends inside the mmW range of the sBSs.

Proof: *See Appendix F* ■

Remark 6. Corollary 1 analytically proves that the communication opportunities of UE u increase with the average time the UE spends inside the mmW transmission range of the sBSs. The higher is the average sojourn time $\overline{\mathcal{S}}_m^u$, the higher is the time a UE has at its disposition to satisfy the requested traffic demand. The average sojourn time in turn depends on the mobility pattern of UE u along with the spatial distribution of the sBSs inside the network area, as proved in [12], i.e, $\overline{\mathcal{S}}_m^u = \mathcal{F}(f_{\mathbf{x}_u}(\mathbf{x}_u), f_{\mathbf{x}_B}(\mathbf{x}_B))$.

Insight 1. From Proposition 3 and Corollary 1, we obtain the following useful insight: *the association period must be longer than the average sojourn time $\overline{\mathcal{S}}_m^u$:*

$$T_a^u = N_a^u T_s > \overline{\mathcal{S}}_m^u. \tag{11}$$

In fact, as observed, T_a^u (equivalently N_a^u) controls the dimensionality of the solution space of the optimization problem proposed in (8). The higher is T_a^u, the higher are the performance in terms of the achievable throughput, at the cost of an higher time complexity and overhead. With this in mind, we further observe that if we set $T_a^u \leq \overline{\mathcal{S}}_m^u$, it results that the upper bound in (9) is smaller than the upper bound in (10). This constitutes a contradiction, since the result in Corollary 1 has been obtained by imposing the tightest alignment constraint, thus limiting the degrees of freedom with respect to the result in Proposition 3.

D. DISCUSSION

In this section, we discuss some aspects of the proposed association rule.

Specifically, as from (8a), the proposed association strategy needs to evaluate the integrals derived in Propositions 1 and 2. To this aim, it needs the knowledge of: *i*) the path the UE is following within a time frame of duration N_a^u (or equivalently T_a^u) to determine the distances $\{d(t)\}$; *ii*) the loads $\{\rho_b(n)\}$ of the sBSs in each time slot $n \in \{n_0, \ldots, n_0 + N_a^u - 1\}$.

In this regard, we observe that, there is a general consensus in both academia and industry that it is unrealistic to expect universal coverage with mm-waves, for the reasons highlighted in the introduction. Hence, a likely deployment scenario is that a mmW network coexists with a traditional μW network [2], [4], [13], [14], with the μW network used for control signaling and the mmW network for data [2], [13], [14]. As a consequence, we can reasonably assume that μW network can be exploited to get the needed information about distances and sBS loads.

However, the following considerations can be made to limit the volume of the signaling traffic:

i) UE devices are equipped with dedicated positioning systems (e.g. GPS) and users extensively utilize navigation apps running on the UE devices. Hence, it is reasonable to expect that each UE can access to the information about its own path. In this scenario, the μW spectrum could be exploited only to individuate the sBS subset $\subseteq \mathcal{B}$ that contributes to the sum in (8a).

ii) the sBS loads vary on a scale in the order of hours and depend on the UE daily activities also in mmW scenarios [15]. By accounting for the mmW transmission range in the order of 200 meters and the usual UE velocities, also in the case of pedestrian velocities, it is reasonable to assume constant sBSs loads within T_a^u. Hence, a UE can receive through the μW network the information about the sBS loads in the time slot n_0, i.e., $\{\rho_b(n_0)\}$, and then utilizes this information to estimate the loads in the subsequent time slots, by relying on the variability of the loads on a scale in the order of hours, i..e, $\rho_b(n) = \rho_b(n_0)\ \forall n \in \{n_0 + 1, \ldots, n_0 + N_a^u - 1\}$.

We further observe, in agreement with [2], that, since there are not any actual mmW network deployments at the time of writing to compare against, any consideration is speculative, although the above made considerations seem quite reasonable to conclude that the volume of the signaling traffic is limited.

Finally, we observe that the analysis developed in this paper does not depend on the specific strategies adopted to estimate the sBS loads or the distances. If, in the future, more effective models for the loads, for example, will be available, it is enough to utilize these models to evaluate the integrals in Propositions 1 and 2.

Algorithm 1 Polynomial-Time Association Algorithm

1: // input:
2: $\mathcal{N} = \{n_0, \ldots, n_0 + N_a^u - 1\}$
3: $L_b^u(n) = \ell_b^u(n)$ with $1_u^b(n) = 1\ \forall b \in \tilde{\mathcal{B}} \subseteq \mathcal{B},\ n \in \mathcal{N}$
4: $x_b^u(n) = 0\ \forall b \in \tilde{\mathcal{B}},\ n \in \mathcal{N}$
5: $1_u^b(n) = 0\ \forall b \in \tilde{\mathcal{B}},\ n \in \mathcal{N}$
6:
7: // code:
8: // enforcing constraint (8c)
9: **while** $\sum_{b,n} L_b^u(n) > 0$ AND $\sum_{b,n} 1_b^u(n)\epsilon_b < \mathcal{E}_b$ **do**
10: $\quad [\tilde{b}, \tilde{n}] = \max_{b,n} \left\{L_b^u(n)\right\}$
11: $\quad x_{\tilde{b}}^u(\tilde{n}) = 1$
12: $\quad L_b^u(\tilde{n}) = 0\ \forall b \neq \tilde{b}$ // enforcing constraint (8b)
13: $\quad$ // check for a switch between current and next sBS
14: $\quad [b_1, n_1] = \min\left\{n > \tilde{n} : x_b^u(n) = 1\right\}$
15: $\quad$ **if** $b_1 = \tilde{b}$ **then**
16: $\quad\quad 1_{b_1}^u(n_1) = 0$
17: $\quad$ **else**
18: $\quad\quad 1_{b_1}^u(n_1) = 1$
19: $\quad$ **end if**
20: $\quad$ // check for a switch between previous and current sBS
21: $\quad [b_0, n_0] = \max\left\{n < \tilde{n} : x_b^u(n) = 1\right\}$
22: $\quad$ **if** $b_0 \neq \tilde{b}$ OR $b_0 =$ NULL **then**
23: $\quad\quad 1_{\tilde{b}}^u(\tilde{n}) = 1$
24: $\quad$ **end if**
25: **end while**
26: **return** $\left\{x_b^u(n)\right\}_{b \in \tilde{\mathcal{B}}, n \in \mathcal{N}}, \left\{1_b^u(n)\right\}_{b \in \tilde{\mathcal{B}}, n \in \mathcal{N}}$

E. POLYNOMIAL-TIME COMPLEXITY ALGORITHM

Insight 2 (Exhaustive Search Complexity). The exhaustive search for the solution of the optimization problem in (8) exhibits an unfeasible time complexity equal to $\mathcal{O}\left(N_a^{u\tilde{B}}\right)$, exponential with the number $\tilde{B} \leq B$ of sBSs within the mmW transmission range of UE u during T_a^u.

This statement follows directly by observing that the decision variables $\{x_b^u(n)\}$ are integers (actually, binaries) and the presence of the alignment cost ϵ_b within the integrals in Propositions 1 and 2 induces local (not global) maximums. Hence, the time complexity of the problem grows with the number of possible solutions, i.e., $\mathcal{O}\left(N_a^{u\tilde{B}}\right)$.

As a consequence of Insight 2, in Algorithm 1 we design a sub-optimal association algorithm, which exhibits a polynomial-time complexity.

Specifically, the algorithm works by evaluating, for each time slot in T_a^u and for each sBS within the mmW transmission range during T_a^u (this set is denoted in the Algorithm as $\tilde{\mathcal{B}} \subseteq \mathcal{B}$), the achievable data load in the worst-case, i.e., by assuming to pay in each time slot the alignment cost regardless of the change in two consecutive time slots of the serving sBS (line 3). The achievable loads are evaluated accordingly to Propositions 1 and 2. In lines 10-11, as long as

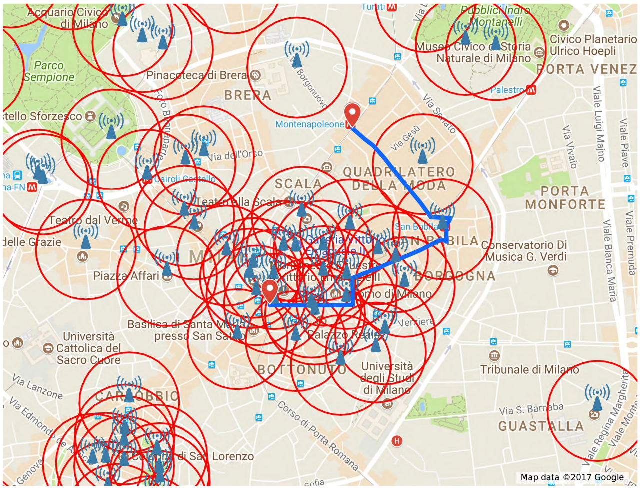

FIGURE 1. Google Maps picture showing the considered UE path (blue line) within *Via Montenapoleone Fashion District*, in the centre of the Italian city of Milan. sBSs placed accordingly to the locations of existing wi-fi hotspots (blue icons).

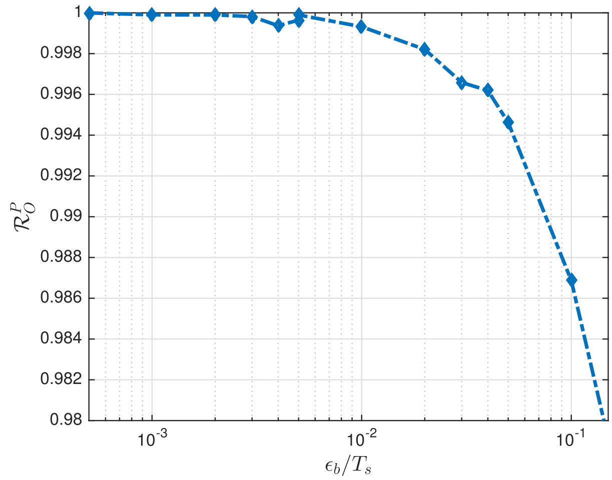

FIGURE 2. Polynomial-time complexity Algorithm 1 versus Exhaustive Search, as function of the ratio between the alignment cost and the duration of the sBS time slot, i.e., ϵ_b/T_s.

a non-null data load exists and the alignment constraint in (8c) is satisfied, the algorithm selects in each time slot the sBS maximizing the data load in the considered time slot. Once this sBS is individuated, the data load achievable by utilizing in that time slot a different sBS is set to zero in order to satisfy the constraint (8b) (line 12). Furthermore, in lines 13-24 the alignment cost is updated, with a cost payed if and only if the serving sBS changes between two consecutive assignments. Finally, the algorithm ends by returning the values of the association variables $\{x_b^u(n)\}$ and of $\left\{1_b^u(n)\right\}$.

Corollary 2. *Algorithm 1 exhibits a feasible time complexity equal to* $\mathcal{O}\left(N_a^{u2}\tilde{B}\right)$, *quadratic with the number of time slots* N_a^u *and linear with the number* $\tilde{B} \leq B$ *of sBSs within the mmW transmission range of UE u during* T_a^u.

Proof: See Appendix G. ■

IV. NUMERICAL PERFORMANCE

In this section, we validate the proposed association procedure by Monte Carlo simulations.

Specifically, we consider the realistic scenario depicted in Fig. 1, showing a possible UE path (blue line) within *Via Montenapoleone Fashion District*, in the centre of the Italian city of Milan. For an accurate and realistic sBS placement, the sBSs are placed at the same location of existing and operating wi-fi hotspots installed by the Municipality of Milan [16]. The red disk-shaped regions are only used for illustrative purposes of the mmW transmission range.

The adopted blockage model for evaluating the LoS/NLoS probabilities (Definitions 1 and 2) is the one proposed in [5] and [6], since this model was validated in a urban environment, as the one considered in this section.

As pointed out in Insight 2, the exhaustive search for the solution of the optimization problem (8) is computational unfeasible. Consequently, we report in Fig. 2 the performance comparison between the exhaustive search and the search obtained with Algorithm 1. Then, in the subsequent performance analysis, we focus only on the polynomial-time complexity Algorithm 1.

Specifically, Fig. 2 reports on the y-axis the ratio between the achievable data load assured by Algorithm 1 and the achievable data load assured through an exhaustive search of the solution for (8). Such ratio is referred in the figure and in the following as $\mathcal{R}_O^P$. On the x-axis, instead, the ratio between the alignment cost and the duration of the sBS time slot, i.e., ϵ_b/T_s, is reported. The results are obtained by setting $T_a^u = 1$s and $T_s = 0.2$s for computational feasibility, whereas the UE moves at a constant velocity equal to 5 Km/h, through the UE path showed in Fig. 1.

As it results from Fig. 2, as long as the alignment cost ϵ_b is shorter than the 1% of the duration of the sBS time slot T_s, the polynomial-time complexity algorithm shows the same performance of the exhaustive search, whereas for longer values a performance degradation is observed. This is reasonable: Algorithm 1 achieves a polynomial time complexity at the price of a sub-optimality in terms of alignment cost minimization. However, the performance degradation is light. Specifically, for ϵ_b equal to $10\% T_s$, Algorithm 1 looses roughly 1.5% with respect to the exhaustive search. Since it is unreasonable to assume an excessive alignment cost with respect to the sBS time slot, we could conclude that the performance assured by Algorithm 1 are comparable with the one assured by the exhaustive search with the advantage of a polynomial time complexity.

Fig. 3 reports on the y-axis with a log-scale, the ratio between the achievable data load assured by adopting Algorithm 1 and the achievable data load assured by adopting the conventional RSS-based association procedure. Such ratio is referred in the figure and in the following as $\mathcal{R}_S^P$. On the x-axis, instead, the duration of the association period, i.e., T_a^u, is reported. The results are obtained for three different values

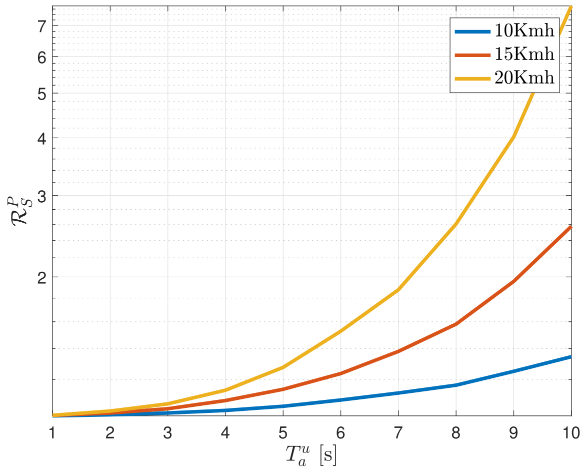

FIGURE 3. Polynomial-time complexity Algorithm 1 versus RSS-based Association, $\mathcal{R}_S^P$, as function of the association period T_a^u for different values of the UE velocity.

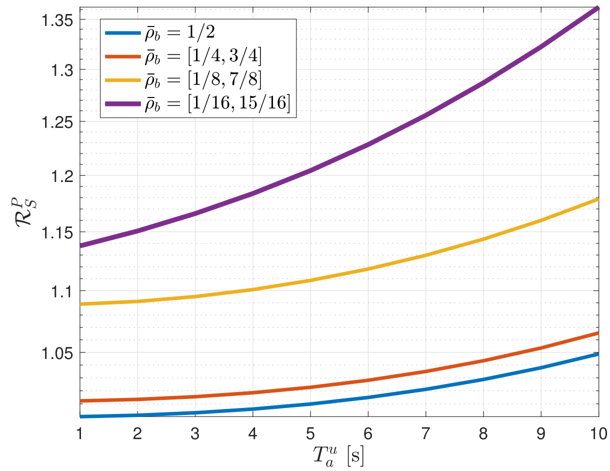

FIGURE 4. Polynomial-time complexity Algorithm 1 versus RSS-based Association, $\mathcal{R}_S^P$, as function of the association period T_a^u for different values of the sBS load distribution, when the UE velocity is 5 Km/h.

of the UE velocities, i.e. 10Km/h, 15Km/h and 20Km/h, when $T_s = 0.2$ s.

As it results from Fig. 3, as long as the association period lasts few sBS time slots, the proposed algorithm does not show an appraisable gain with respect to the conventional RSS-based association procedure. This is expected. In fact, Algorithm 1 cannot exploit its capability of tracking the dynamics induced by UE mobility when N_a^u is very small. This agrees also with the theoretical analysis developed in Section III, since by decreasing N_a^u, the dimensionality of the solution space decreases as well. The aforementioned described effect decreases by increasing the UE velocity. In fact, the higher is the UE velocity, the higher is the loss in term of achievable data load of the conventional RSS-based association procedure, since the variations of the network topology are faster. In particular, we observe that with only $T_a^u = 7$s and 20 Km/h, Algorithm 1 assures a roughly double achievable data load with respect to the RSS-based one. With this same velocity, Algorithm 1 assures an achievable data load seven times higher than the one achievable by adopting the RSS-based association when $T_a^u = 10$s.

We finally, observe that in this simulation we disadvantage the proposed algorithm with respect to the RSS-based one, since we assumed that all the sBSs have the same average load $\bar{\rho}_b = 1/2$, with load uniformly distributed in [0, 1]. As analyzed in Section III, our proposal, instead is able to take advantage from unbalanced load distributed as proved by the next simulation.

Fig. 4 reports the ratio between the achievable data load assured by adopting Algorithm 1 and the achievable data load assured by adopting the conventional RSS-based association procedure, $\mathcal{R}_S^P$, as function of T_a^u. The results are obtained for four different sBS average load values. Specifically, the blue curve refers to the case in which all the sBSs have the same average load $\bar{\rho}_b = 1/2$. Differently, the red, the yellow and the purple curves refer to the case in which the sBSs are split in two equal-size subsets characterized by different average loads, with the constraint that the overall average load, i.e., the load obtained by averaging on all the sBSs, remains equal to $1/2$ for a fair comparison with the blue curve. Specifically, for the red curve, half sBSs have average load $\bar{\rho}_b = 1/4$ and the remaining ones have average load $\bar{\rho}_b = 3/4$. For the yellow curve, half sBSs have average load $\bar{\rho}_b = 1/8$ and the remaining ones have average load $\bar{\rho}_b = 7/8$. Finally, for the purple curve, half sBSs have average load $\bar{\rho}_b = 1/16$ and the remaining ones have average load $\bar{\rho}_b = 15/16$. To separate the effects of the mobility from the ones of the sBS load distributions, and to prove that also in absence of mobility our proposal is able to assure better performance with respect to the RSS-based association, we assume that the UE velocity is equal to 5Km/h, and that T_a^u is within the range [1, 10] s. Hence, in this time horizon, the UE can be considered almost static.

As it results from Fig. 4, as long as the average load is not identical among the sBSs, the gain of the proposed association is 15% with respect to the RSS-based association even for the lowest considered values of T_a^u, and it overcomes 35% for $T_a^u = 10$. This agrees with the theoretical analysis developed in Section III, where it was showed that our proposed association strategy is able to account for the sBS loads. Specifically, when a sBS is congested in a certain time slot, it is efficiently discharged by UE u for the association. Differently the RSS-based association rule does not take into account the sBSs loads and hence it is possible that a UE is associated to a congested sBS.

In Fig. 5 the experiment reported in Fig. 4 is repeated for a different values of the UE velocity, i.e., 10 Km/h. It is very evident that the gain assured by the proposed association strategy increases, for all the considered values of the loads, when the mobility increases. This further validates that the proposal is able to take advantage from the dynamic variation of the network topology induced by the mobility.

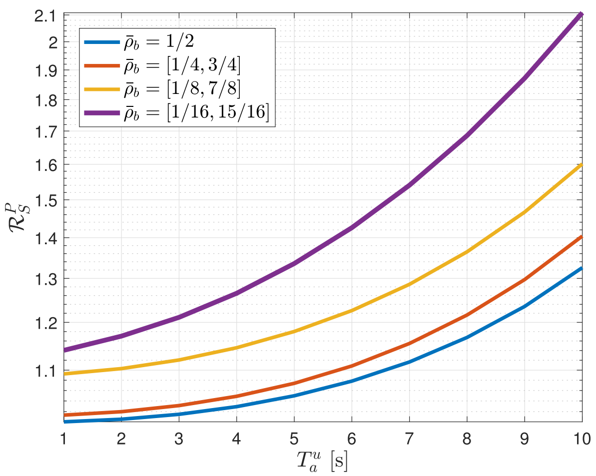

FIGURE 5. Polynomial-time complexity Algorithm 1 versus RSS-based Association, $\mathcal{R}_S^P$, as function of the association period T_a^u for different values of the sBS load distribution, when the UE velocity is 10 Km/h.

V. CONCLUSION

In this paper, we have designed a mobility-aware user association strategy for mmW networks to overcome the limitations of the conventional received power (RSS)-based association strategies. We proved that the proposed strategy exhibits several attractive features: a) it is able to track the dynamic changes in the network topology and in the channel conditions induced by the user mobility; b) it takes into account the distribution of the loads among the small base stations (sBSs), thus overcoming to associate a UE to an already congested sBS; c) it overcomes overly frequent handovers between sBSs, and thus the need of frequent re-association; d) it takes into account the peculiar aspects of the mmW communications, such as directionality, sensitivity to blockage, and NLoS propagation effects; e) it is fully-distributed, i.e., each mobile user associates to a sBS independently of each others, stemming from local information only. Furthermore, since the exhaustive search for the solution of the proposed association problem is computationally unfeasible, we designed an efficient algorithm exhibiting a polynomial-time complexity. Finally, the numerical results validated the significant benefits of adopting the proposed association.

APPENDIX A
PROOF OF PROPOSITION 1

Accordingly to Definition 5, $t_b(n_0) \stackrel{\triangle}{=} [(n_0 - 1) + \rho_b(n_0)]\, T_s$ accounts for the portion of the n_0-th time slot of sBS b occupied by the UEs already associated to b. The random time t_0^u, at which the association procedure starts, could be either greater or smaller than $t_b(n_0)$. These two cases impact on the time the evaluation of the conditional achievable data load in the n_0-th time slot starts.

Specifically, when $t_b(n_0) < t_0^u \leq n_0 T_s$, UE u starts its association procedure in a time in which sBS b has already satisfied the requests of the UEs already associated to it. Hence, UE can evaluate the conditional achievable data load in n_0, starting from t_0^u, by paying the alignment cost ϵ_b, being n_0 the first time slot. In particular, we observe that, since UE u is mobile, the distances between UE u and sBS b vary in time. Consequently the capacity in (1) is a function of the time. Hence, to evaluate the conditional achievable data load, we need to integrate the capacity with respect to the time to account for the variability of the distances with the time. Clearly, since in principle $\epsilon_b + t_0^u$ could be greater than $n_0 T_s$, in the integral one has to consider the minimum between $t_0^u + 1_b^u(n_0)\epsilon_b$ and $n_0 T_s$. Furthermore, a UE has at its disposition at most $\gamma_b T_s$ resources in a time slot. Since in principle $\gamma_b T_s + t_0^u$ could be greater than $n_0 T_s$, in the integral one has to consider the minimum between $\gamma_b T_s + t_0^u$ and $n_0 T_s$. Consequently, the proof follows.

Similar reasonings hold when $(n_0 - 1)T_s \leq t_0^u \leq t_b(n_0)$. The difference is that under this condition, UE u starts its association procedure in a time in which sBS b has not already satisfied the requests of the UEs already associated to it. Consequently, UE u can evaluate the conditional achievable data load starting from $t_b(n_0)$.

APPENDIX B
PROOF OF PROPOSITION 2

The proof follows by adopting similar reasonings as in Appendix A, by observing that since the association procedure started in a previous time slot, i.e., being $n > n_0$, UE u can evaluate the conditional achievable data load starting from $t_b(n)$, determined by the load $\rho_b(n)$ of sBS b in this time slot.

APPENDIX C
PROOF OF LEMMA 1

The proof follows directly from Propositions 1 and 2 by accounting for the definition of $t_b(n)$. Specifically, let us suppose for absurdum that $\rho_b(n) = 1$. In such a case, $t_b(n) = nT_s$ either if $n = n_0$ or $n > n_0$ and hence from (6) or (7) it results $\ell_b^u(n) = 0$. As a consequence, within the objective function (8a) such sBS will not contribute to maximize the overall achievable data load and hence it is discharged by the association procedure. This constitutes a reductio ad absurdum and hence the proof follows.

APPENDIX D
PROOF OF LEMMA 2

The proof follows directly from Lemma 1 and Propositions 1 and 2. In fact, accordingly to Lemma 1, if UE u is associated in a certain time slot to sBS b, $\rho_b(n) < 1$. Since according to Propositions 1 and 2, γ_b is the maximum fraction of T_s that can be allocated to UE u, it results that the sum of the sBS load and the contribution to the load of UE u cannot exceed one.

APPENDIX E
PROOF OF PROPOSITION 3

Let us denote with $\tilde{\mathbf{x}}^u$ the solution of the proposed optimization problem for a given n_0, and with $\{\tilde{1}_b^u(n)\}$ the corresponding values of the binary variables defined in Def. 7. By exploiting the results in Propositions 1 and 2, it results that the conditional time $T_o^u|$ UE u has at its disposition for communication opportunities is bounded by:

$$T_o^u| \leq \sum_{b=1}^{B} \sum_{n=n_0}^{n_0+N_a^u-1} \tilde{x}_b^u(n) \left[\gamma T_s - \tilde{1}_b^u(n)\epsilon\right] \tag{12}$$

By accounting for the constraint (8b), it results, that in each time slot UE u can be associated at most to only one sBS. Plus the maximum number of available time slots is N_a^u. As a consequence, (12) becomes:

$$T_o^u| \leq \sum_{b=1}^{B} \sum_{n=n_0}^{n_0+N_a^u-1} \tilde{x}_b^u(n) \left[\gamma T_s - \tilde{1}_b^u(n)\epsilon\right] \leq N_a^u \gamma T_s - \epsilon \tag{13}$$

where we exploited the inequality $\sum_{b=1}^{B} \sum_{n=n_0}^{n_0+N_a^u-1} \tilde{x}_b^u(n)\tilde{1}_b^u(n)\epsilon \geq \epsilon$, since at least one alignment cost has to be paid. The proof follows, by removing the conditioning and by observing that the upper bound does not depend on the conditioning, since $\epsilon_b = \epsilon, \forall b \in \mathcal{B}$ and $\gamma_b = \gamma$, $\forall b \in \mathcal{B}$.

APPENDIX F
PROOF OF COROLLARY 1

If the alignment constraint is the tightest one, a UE cannot be associated to different sBSs in the different time slots belonging to T_a^u. Consequently, the solution $\tilde{\mathbf{x}}^u$ of the proposed optimization problem for a given n_0 is not a matrix but a vector individuating the sBS with which UE u is associated and in which time slots. Hence, by reasoning as in Proposition 3, it results:

$$\begin{aligned} T_o^u| &\leq \sum_{n=n_0}^{n_0+N_a^u-1} \tilde{x}_b^u(n) \left[\gamma T_s - \tilde{1}_b^u(n)\epsilon\right] \\ &\leq \sum_{n=n_0}^{n_0+\lceil \frac{S_m^u}{T_s} \rceil -1} \gamma T_s - \epsilon = \left\lceil \frac{S_m^u}{T_s} \right\rceil \gamma T_s - \epsilon \end{aligned} \tag{14}$$

where, we utilized the fact that the only time slots contributing to $T_o^u|$ are the ones in which UE u can establish a mmW connection with the considered sBS b. The number of these time slots is determined by the instantaneous value of the sojourn time UE u spends inside the mmW range of sBS b. By exploiting the properties of the ceil function, i.e., $\lceil x \rceil < x + 1$, and by removing the conditioning also with respect to the instantaneous value of the sojourn time UE, the proof follows.

APPENDIX G
PROOF OF COROLLARY 2

The proof follows directly by observing that the *while* loop cycles on the number of time slots N_a^u as a consequence of the *while* condition and line 12. Furthermore, the k-th loop cycle requires $\mathcal{O}(N_a^u B)$ operations for evaluating the *while* condition, $\mathcal{O}(B(k-1))$ operations for line 10, $\mathcal{O}(k-1)$ operations for lines 14-21, and a constant number of operations in the remaining lines.

REFERENCES

[1] I. F. Akyildiz, S. Nie, S.-C. Lin, and M. Chandrasekaran, "5G roadmap: 10 Key enabling technologies," *Comput. Netw.*, vol. 106, pp. 17–48, Sep. 2016.

[2] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, Jr., "Modeling and analyzing millimeter wave cellular systems," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 403–430, Jan. 2017.

[3] D. Liu *et al.*, "User association in 5G networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, 2nd Quart., 2016.

[4] H. Elshaer, M. N. Kulkarni, F. Boccardi, J. G. Andrews, and M. Dohler, "Downlink and uplink cell association with traditional macrocells and millimeter wave small cells," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6244–6258, Sep. 2016.

[5] T. S. Rappaport, G. R. MacCartney, Jr., M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.

[6] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[7] Y. Xu, H. Shokri-Ghadikolaei, and C. Fischione, "Distributed association and relaying with fairness in millimeter wave networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 7955–7970, Dec. 2016.

[8] C. Psomas and I. Krikidis, "Low-complexity base station selection scheme in mmwave cellular networks," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 4049–4064, Sep. 2017.

[9] G. Athanasiou, P. C. Weeraddana, C. Fischione, and L. Tassiulas, "Optimizing client association for load balancing and fairness in millimeter-wave wireless networks," *IEEE/ACM Trans. Netw.*, vol. 23, no. 3, pp. 836–850, Jun. 2015.

[10] A. S. Cacciapuoti, R. Subramanian, M. Caleffi, and K. C. Chowdhury. (Feb. 2017). "Software-defined network controlled switching between millimeter wave and terahertz small cells." [Online]. Available: https://arxiv.org/abs/1702.02775

[11] H. Shokri-Ghadikolaei, F. Boccardi, C. Fischione, G. Fodor, and M. Zorzi, "Spectrum sharing in mmwave cellular networks via cell association, coordination, and beamforming," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 11, pp. 2902–2917, Nov. 2016.

[12] A. S. Cacciapuoti, I. F. Akyildiz, and L. Paura, "Optimal primary-user mobility aware spectrum sensing design for cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 11, pp. 2161–2172, Nov. 2013.

[13] J. Park, S. L. Kim, and J. Zander, "Tractable resource management with uplink decoupled millimeter-wave overlay in ultra-dense cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4362–4379, Jun. 2016.

[14] K. Zheng, L. Zhao, J. Mei, M. Dohler, W. Xiang, and Y. Peng, "10 Gb/s hetsnets with millimeter-wave communications: Access and networking–challenges and protocols," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 222–231, Jan. 2015.

[15] G. K. Tran, H. Shimodaira, R. E. Rezagah, K. Sakaguchi, and K. Araki, "Practical evaluation of on-demand smallcell ON/OFF based on traffic model for 5G cellular networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2016, pp. 1–7.

[16] *Open Wifi Milano*. Accessed: Feb. 2017. [Online]. Available: https://info.openwifimilano.it/EN/index.aspx

ANGELA SARA CACCIAPUOTI (M'10–SM'16) received the Laurea (integrated B.S./M.S. degrees) *(summa cum laude)* degree in telecommunications engineering in 2005 and the Ph.D. degree in electronic and telecommunications engineering in 2009, both from the University of Naples Federico II.

Since 2017, she held the Italian Habilitation as "Associate Professor" in telecommunications engineering.

From 2012 to 2017, she has been a non-tenure-track Assistant Professor at the Department of Electrical Engineering and Information Technology, University of Naples Federico II. Prior to that, she was a Visiting Researcher at the Broadband Wireless Networking Laboratory, Georgia Institute of Technology (USA) and at the NaNoNetworking Center in Catalunya (N3Cat), School of Electrical Engineering, Universitat Politécnica de Catalunya (Spain). Her work in the first tier IEEE journals and she received different awards, including the elevation to the grade of the IEEE Senior Member in 2016, most downloaded article, most cited article awards, and outstanding young faculty/researcher fellowships for conducting research abroad. Her current research interests are in cognitive radio networks, biological communications, 5g networks, mmWave, and THz Communications. She serves as an Editor for the journals IEEE Transactions on Communications, IEEE Communications Letters and *Computer Networks Journal* (Elsevier). She serves as an Associate Editor of IEEE Access.

Dr. Cacciapuoti has been an appointed member of the IEEE ComSoc Young Professionals Standing Committee since 2016. She served as the Award Committee Chair of the 2016 IEEE ComSoc Best YP Awards. Since 2017, she has been appointed as the Award Co-Chair of the N2Women Board. Since 2017, she has been an elected Treasurer of the IEEE Women in Engineering Affinity Group of the IEEE Italy Section.