IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 15, 2016, accepted November 6, 2016, date of publication December 9, 2016, date of current version February 25, 2017.

Digital Object Identifier 10.1109/ACCESS.2016.2638022

Adaptive Resource Management Strategy in Practical Multi-Radio Heterogeneous Networks

MIKHAIL GERASIMENKO[1], DMITRI MOLTCHANOV[1], SERGEY ANDREEV[1], (Member, IEEE), YEVGENI KOUCHERYAVY[1], (Senior Member, IEEE), NAGEEN HIMAYAT[2], (Member, IEEE), SHU-PING YEH[2], (Member, IEEE), AND SHILPA TALWAR[2], (Member, IEEE)

[1]Department of Electronics and Communications Engineering, Tampere University of Technology, 33720 Tampere, Finland
[2]Intel Corporation, Santa Clara, CA 95054-1549, USA

Corresponding author: M. Gerasimenko (mikhail.gerasimenko@tut.fi)

This work was supported by Intel Corporation. The work of S. Andreev was supported in part by the Postdoctoral Researcher Grant by the Academy of Finland and in part by the Jorma Ollila Grant by the Nokia Foundation.

ABSTRACT The ongoing evolution of mobile wireless communications has resulted in the vision of a multi-radio heterogeneous network (HetNet) that comprises cells of different scales controlled by various radio access technologies (RATs). These emerging architectures call for more advanced methods of cross-RAT radio resource allocation, which are the primary focus of this article. In this paper, based on network flow optimization techniques, we adapt the concept of weighted α-fairness for efficient resource management in future HetNets. The corresponding scheme relies on a certain degree of centralized control of the HetNet architecture and allows to achieve the desired balance between the overall system throughput and the fairness of the resulting resource allocations based on a single parameter. Our analytical findings, validated with detailed system-level simulations, are expected to further advance the understanding of feasible resource control strategies in intelligent multi-radio networks, as well as help optimize the performance of next-generation HetNets.

INDEX TERMS 5G, capacity improvement, fairness, heterogeneous networks, multi-RAT, resource allocation.

I. INTRODUCTION

A. GENERAL BACKGROUND ON HetNets

Nowadays, industry and academia are working towards implementing the fifth generation (5G) of cellular networks, targeted for completion by 2020. According to the definition of 5G technologies, they are expected to offer significantly higher capacity (e.g., 1000x increase), universal seamless coverage, and better user experience [1], [2]. In the long run, major industrial players, including infrastructure vendors, device manufacturers, and network operators, are likely to progressively deploy 5G technologies with some of the trends visible already today [3]. Notably, to reach the 1000x capacity, the two key mechanisms are spatial *network densification* and *spectral aggregation* [4]. While the former suggests deploying higher density of increasingly smaller cells in current network architecture, the latter envisions jointly utilizing larger portions of radio spectrum across diverse spectral bands in licensed, unlicensed, and higher frequency spectrum.

The proliferation of inexpensive low-power *small cells* of different sizes (micro, pico, femto, etc.) supporting a mixture of Radio Access Technologies (RATs), overlaid with macro cells for wide area coverage and mobility, together create the vision of a Heterogeneous Network (HetNet) [5]. With further densification towards ultra-dense HetNets [6], based on the network infrastructure that supports multiple air interface technologies (e.g., small cells with multi-radio capabilities), the role of ''anchor'' macro cell base stations to provide wide area coverage becomes even more pronounced. The macro cell can provide central intelligence to create a balanced distribution of users across small ''booster'' base stations, for efficient offload of user data traffic [7], [8].

As the above ''anchor-booster'' architecture develops, macro cell centric intelligence is being increasingly employed to coordinate between multi-RAT small cells [9], such that the degree of network cooperation moves from ''loose'' interworking solutions at the core network level to centralized or distributed schemes offering tighter integration at the Radio Access Network (RAN) layer. Advanced RAN-based interworking mechanisms have recently been addressed in the 3rd Generation Partnership Project (3GPP) standardization, with efficient solutions for 3GPP/WLAN integration captured in

LTE Release 12 [10]. WLAN/LTE aggregation solutions with tighter integration of WLAN into the operator's cellular radio network are further standardized as part of LTE Release 13. Generally, the capacity and quality of service (QoS) gains from aggregation across multiple links depend on how closely RAT selection and management procedures are coordinated at both device and network levels, and increasing levels of coordination result in higher performance gains [11].

Today, with the growing number of alternative RATs (GSM/EDGE, WCDMA/HSPA, LTE-FDD/TDD, WiFi, BT, WiGig, etc.), the emerging RAN-layer cooperation is expected to offer greater flexibility than has been possible before with ''loose'' higher-layer coordination solutions, thus enabling more dynamic radio resource management for improved system and user performance [12]. In particular, RAN-layer integration allows for fine-grained and real-time flow control to split user traffic flexibly between the available radio interfaces when optimizing the overall HetNet performance [13], [14].

B. PROBLEM STATEMENT AND CONTRIBUTIONS

Various aspects of N-tier multi-RAT wireless networks have been subject to extensive investigation over the last several years. Utilizing the tools of stochastic geometry, many authors have characterized capacity and throughput bounds in such systems, see e.g., [15]–[18]. However, this approach has inherent limitations in addressing the practical resource allocation problems that can accommodate different levels of RAN-layer assistance. The related investigations indeed provided useful insights into the ''averaged'' behavior of spatially-uniform N-tier networks, but deliver close to no essential information regarding the instantaneous traffic control procedures in a particular deployment of interest. Hence, whereas probabilistic methods remain crucial for understanding the fundamental capacity limits of such networks, alternative methods are needed at the operational phase of HetNets as user connectivity at any given instant of time could be very different from what is expected on average in the long run.

Using stochastic geometry results as upper bounds, a number of performance optimization frameworks targeting either *max-min* or *proportional fairness* based flow allocations have been developed, see [19] for an extensive tutorial on the matter. As the complexity of radio resource allocation in future HetNets is high, most of the proposed techniques result in NP-hard problems that cannot be solved in polynomial time and thus require heuristic algorithms to achieve sub-optimal performance. To alleviate this problem, game-theoretic approaches have been proposed, see e.g., [20]–[22]. These models enforce a proportional fairness criterion across the allocated flows and allow for simple algorithms to estimate allocations. However, compared to classic optimization methods, game-theoretic approaches do not offer the needed degrees of control between the system throughput and fairness of flow allocations at the air interface. The latter may be of particular interest to mobile network operators. Finally, most optimization studies performed so far do not offer practical procedures for enforcing the computed resource allocations.

In this work, we study different multi-RAT bandwidth allocation techniques. By adopting the concept of weighted α-fairness, which is a generalization of max-min and proportional fairness criteria, we demonstrate that the emerging N-layer HetNet architecture is characterized by a large degree of controllability in terms of the trade-off between system throughput and fairness of allocations. This degree is fully managed by a single parameter, α.

The **contributions** of this work are as follows:

- We comprehensively outline the notion of ''weighted α-fairness'' in heterogeneous wireless networks, which is capable of delivering the appropriate balance between the overall system throughput and the fairness of resulting resource allocations. We also demonstrate that it can be formulated as a special case of max-min problem, hence leading to the optimization tasks of linear programming (LP) type.
- We compare the performance of the proposed criterion against that of the classic max-min and the proportional fairness formulations, as well as consider several heuristic methods indicating that our solution is beneficial in terms of the overall balance between the system throughput and the fairness of resource allocations across the users.
- We investigate the response of our proposed heuristic and the corresponding algorithms to various input parameters of a HetNet, which explains how the capacity at e.g., pico LTE and WiFi HetNet layers can be traded for better coverage across the entire system and vice versa. This, in turn, supplies the network operator with a convenient tool to dynamically control the deployment in question.
- We propose a practical implementation of the proposed weighted α-fairness scheme in HetNets based on the backpressure concept that allows to rely on the directly measurable metrics, such as the current state of traffic buffers, to enforce the computed allocations.

The rest of this text is organized as follows. Section II introduces the considered HetNet architecture, while Section III discusses the choice of an appropriate optimization criterion. Further, Section IV outlines the optimization framework proposed in this work, which is then illustrated by the characteristic numerical results in Section V. Practical implementation of the proposed solution in a HetNet environment is discussed in Section VI. Section VII summarizes the most essential conclusions and take-aways of our study.

II. ENVISIONED HetNet ARCHITECTURE

In this section, we summarize the HetNet concept and provide further details on different HetNet components. In addition, we comment on the current research and development efforts in this area as well as elaborate on how the goals identified in the previous section could be met in practice. Historically, the HetNet paradigm was first introduced in the context of

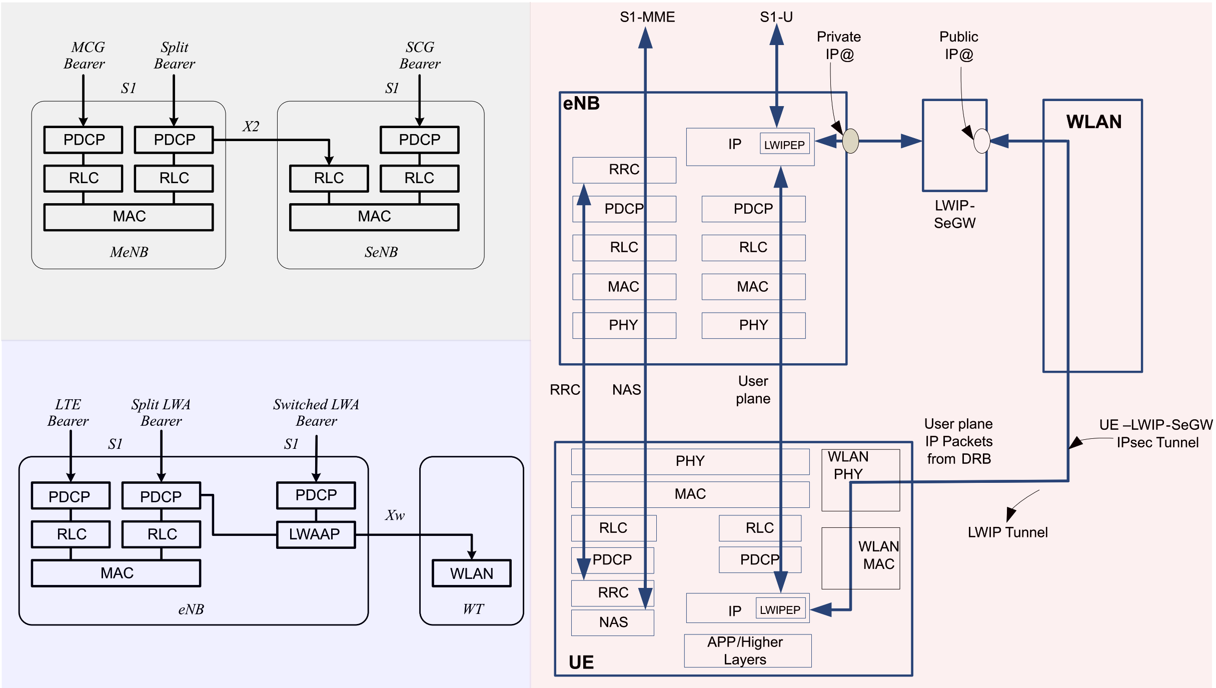


FIGURE 1. HetNet architectures by 3GPP: DC (top left), Non-collocated LWA (bottom left), and LWIP (right), adapted from [24].

cellular systems as part of 2G GSM specifications, with the notion of a small (pico) cell. In subsequent 3G UMTS specifications, this concept did not evolve significantly, as most technical efforts revolved around performance optimization of individual RATs (cellular, WLAN, etc.).

Centralized policy management for multi-radio networks (e.g., WLAN and 3GPP systems) was introduced as part of 4G specifications via a mechanism known as the Access Network Discovery and Selection Function (ANDSF, [23]). This standardized scheme allows the cellular systems to incorporate policies for non-3GPP (e.g., WLAN) network discovery and access. In particular, ANDSF is capable of supplying user equipment (UE) with the relevant information on the neighboring non-3GPP RANs (i.e., the discovery function). It can also notify the UEs regarding the restrictions and preferences of these neighboring RANs through the inter-system mobility policy function. If the UE is allowed to utilize two or more radio links concurrently, the traffic offloading preferences could be specified by the dedicated inter-system routing policy. In general, ANDSF provides policies that apply for longer time scales of operation, but does not offer dynamic control and scheduling of radio resources across a multi-RAT system.

Further extensions to centralized HetNet management were made in LTE Release 12 and 13. In a nutshell, three different architectures were introduced: dual connectivity (DC), LTE-WiFi aggregation (LWA), and LTE/WLAN Radio Level Integration with IPsec Tunnel (LWIP), which are displayed in Fig. 1. In DC, the main focus is set on connecting the UE to two access nodes simultaneously. There are three different bearers that allow to configure different user plane connectivity options. In this research, we concentrate on the split bearer, where Packet Data Convergence Protocol (PDCP) data is transmitted between a Slave eNodeb (SeNB) and a Master eNodeB (MeNB) via X2 interface. In this case, resource scheduling is still performed by each entity separately, but coordination between the MeNB and the SeNB is conducted by the means of X2 interface, which enables support of data splitting controlled by MeNB.

Two other considered possibilities, LWA and LWIP, are in fact different options to support integration of WiFi access point on the RAN level. In LWA (being somewhat closer to DC), the splitting is implemented by using the LTE-WLAN Aggregation Adaptation Protocol (LWAAP), which acts like a gateway between LTE and WiFi. With LWIP, the user plane splitting is performed on the IP layer by creating an IPsec tunnel between the UE and the eNodeB via WiFi. In our research, we adopt LWA as our background architecture, due to the higher flexibility that it delivers to the centralized splitting logic options.

In this work, we concentrate on the conceptual operation of a HetNet [25], having much broader capabilities for radio control than what is possible with e.g., ANDSF, where a multi-radio UE may simultaneously communicate with an arbitrary number of RATs. To outline the inherent properties of such systems, we further discuss the available dimensions of a HetNet "heterogeneity".

RAT-type heterogeneity: Given that each RAT has its specific architecture and signal transmission principles, the efficient simultaneous utilization of multiple RATs is one of the key challenges in HetNets. There are two natural options to enable the related *coordination*. The first alternative is to rely on a centralized coordination entity, which would be in full control of all the associated RANs. However, the necessary signaling procedures making this approach feasible have to be specified separately as part of the corresponding framework (e.g., in future releases of 3GPP LTE). The primary difficulty here is to satisfy the diverse requirements of each involved RAT and construct balanced signaling procedures not affecting the efficiency of individual HetNet components.

Another coordination approach implies the invocation of the interoperability functions within the individual RATs. In this case, potential optimality of the HetNet-wide control might become unavailable due to the lack of complete system information on the accessible network resources from the side of each RAT. Correspondingly, the required cross-RAT coordination information may be exchanged via the appropriate inter-network backhaul links, which might also increase the overall signaling overhead. An alternative option could be to delegate the RAT selection decisions to the UEs themselves, by supplying them with all the necessary information. However, this solution may be highly sub-optimal, especially in dynamic environments, whereas it at the same time offers attractive flexibility to mobile device manufactures and end users.

Application heterogeneity: In addition to distinction based on their architecture, HetNets are also differentiated with respect to the expected use of their individual RAT components. This results in an application-level classification, such as the voice-oriented 2G/3G installations and the data-centric WLAN deployments. Along these lines, end user service provisioning becomes the cornerstone of the respective system optimization, with e.g., machine automation scenarios leading to drastically different design choices than high-rate multimedia use cases.

Importantly, control/user plane separation becomes a fundamental building block for enabling the realization of HetNets, which is adopted as a principal design feature in the next-generation 5G systems [26]. Control/user plane separation offers additional flexibility to future 5G networks as well as enables several architecture design options, such as software defined networks. In the rest of this work, we proceed with considering an important *characteristic* HetNet scenario, which features two RATs (3GPP LTE and IEEE 802.11 WiFi), two types of coverage (macro, as well as pico for LTE and WiFi), and a single application – best effort data transmission. Particularly, we assume that the entire heterogeneous system in question is controlled by a single coordinating node. This node collects all of the relevant information regarding the current user traffic demand, as well as monitors the availability of a particular RAT coverage within the target service area. Once this information is known, the coordinator makes decisions on the best choice of RAT for all the associated users, as well as advises on the actual amount of radio resources that every user may utilize on each accessible RAN. In more detail, we assume that the coordinating node has the full knowledge on the service availability of every RAN together with the current channel conditions of each UE.

Naturally, the discussed coordinator may physically reside on the macro base station side, whereas the relevant control information from the pico cells and the WiFi access points could travel on the fronthaul links. In this case, dedicated gateways between the WLAN access points and the macro LTE base stations would be required. In light of the ongoing RAN-layer 3GPP/WLAN integration, our envisioned architecture becomes structurally similar to the emerging heterogeneous cloud-RAN concept discussed in [27] and [28]. Noteworthy, the proposed control system operates on the packet level (by contrast to the flow level), that is, individual packets belonging to the same flow could be transmitted on the concurrent radio interfaces. The coordinator ensures that the packet stream processed by a certain interface is then assembled in the correct order. In practice, this can be achieved by taking advantage of smart tunneling mechanisms between the UE and the coordinating node. As a proof-of-concept, we have already demonstrated the feasibility of a simple prototype, where the UE is utilizing such smart tunneling (based on the open flow architecture) to split its traffic dynamically between the LTE and the WiFi radio interfaces [29].

In what follows, we elaborate on the available solutions to the aforementioned problem of centralized radio resource allocation in multi-radio HetNets by employing a collection of methods coming from the optimization theory. This corresponds to considering a certain time instant t, when the task of the centralized coordinating node is to decide upon a system-wide resource allocation, which would be optimal with respect to a particular metric of interest. Hence, an important step in optimizing performance of multi-RAT systems is the choice of the appropriate optimization criterion, which could satisfy the expectations of both network operators and end users. We continue with the related discussion below.

III. SELECTION OF OPTIMIZATION CRITERION

From the network perspective, the following optimization metrics are of particular importance: (i) the fairness of resource allocations between the users and (ii) the overall system throughput. As is well known, there typically exists an inherent trade-off between these two criteria. Depending on the effective network operator policies, when allocating resources to the end users, there should remain a possibility to exchange one parameter for another. Therefore, there are currently two well-known fairness criteria resulting in various degrees of this much needed flexibility, which are known as *max-min fairness* and *proportional fairness*.

A. MAX-MIN FAIRNESS

Max-min fairness is one of the most widely known performance criteria introduced originally by Bertsekas and Gallager [30]. Denoting by N the number of user demands to be served in a network and by P_d the number of paths available for these demands d, $d = 1, 2, \ldots, N$ (each demand corresponds to a particular UE throughput), the objective is to maximize the minimum of bandwidth allocations $\sum_{p=1}^{P_d} x_{dp}$, $d = 1, 2, \ldots, N$, subject to the link capacity constraints while at the same time satisfying non-negativity of all allocations. Formally, let $\vec{x} = (x_1, x_2, \ldots, x_N)$ be the allocation vector (bandwidth in Hz) sorted in non-decreasing order. We say that $\vec{x}$ provides the max-min allocation if it is lexicographically maximal among all possible allocation vectors also sorted in non-decreasing order, see [31, Ch. 8]. A vector $\vec{x}$ is said to be lexicographically greater than $\vec{z}$ if there is such index k that $x_i = y_i$, $i = 1, 2, \ldots, k$, and $x_k > z_k$.

Hence, an allocation is called max-min optimal if there is no other way to increase this allocation for a particular demand i at the expense of demands j with a greater allocation. The underlying problem of lexicographic maximization is of LP (linear programming) type allowing for efficient implementations with an array of solutions, such as e.g., simplex method [32]. It is also important to note that in its classical formulation the max-min criterion always exchanges the overall system throughput in favor of fairness of the resource allocations. At the same time, the simplicity of the resulting optimization problem is a very attractive feature that may potentially allow for real-time radio resource management.

B. PROPORTIONAL FAIRNESS

In the course of the last two decades, a number of authors questioned the appropriateness of the max-min fairness criterion. Quoting Massoulie and Roberts [33]: *"In fact, there appears to be no clear economic reason why max-min sharing should be preferred over some other bandwidth allocation. More rational objectives would be to maximize overall utility accounting for costs and perceived value or to minimize the expected response time of any transfer"*. Moreover, Kelly in [34] highlights that one of the most successful window-based rate control procedures — the TCP networking protocol — results in nearly proportional distribution of resources alongside the path of a data flow.

To this end, a commonly used objective criterion for proportional fairness,

$$\sum_{d=1}^{N} \log x_d \rightarrow \max, \qquad (1)$$

has a number of useful features that stem from the mathematical properties of the function $f(x) = \log(x)$. First, when the rate assigned to a certain demand is small, the expression (1) features a very large negative component. Second, although $f(x) = \log(x)$ is a monotonically increasing function, behavior analysis of its derivative $f'(x) = (\log x)' = 1/x \ln(x)$ indicates that the growth rate decreases as x increases. Such behavior naturally precludes from extremely high allocations. From this discussion, one could also conclude that the base of the logarithmic function in (1) is irrelevant.

Compared to the max-min fairness, proportional fairness is known to provide better results in terms of the overall system throughput by preventing from extremely large or small allocations. This, however, comes at the cost of some loss in fairness among the users. Importantly, in its classical formulation and similarly to the max-min criterion, proportional fairness does not provide flexibility in the choice of the throughput–fairness balance point. This crucial consideration can only be introduced by employing alternative objective functions, with the properties similar to $f(x) = \log(x)$. On top of this, the class of optimization problems where the objective function is given by (1) is known as convex programming. Such problems are significantly more difficult to solve compared to simpler LP formulations, thus complicating the efficient implementation of the proportional fairness metric in the real-time resource allocation algorithms, especially for large-scale and ultra-dense HetNet deployments.

C. PROPOSED WEIGHTED α-FAIRNESS CRITERION

As follows from the previous section, the complex HetNet topology together with the unique properties of its individual RAT components suggests to rethink the choice of the fairness criterion. In particular, the said criterion penalizes longer flows (in terms of the number of hops) more heavily than shorter flows, provided that they compete for resources within a certain area of interest. As the users with better channel quality can utilize wireless resources more efficiently, the equivalent of the proportional fairness criterion for HetNets should be based on a certain metric describing the current channel conditions of the users. Further, observe that the max-min criterion attempts to deliver as fair allocations as possible. This feature, while being useful in legacy networks, may become inappropriate for HetNet systems. Indeed, the current distance between a user and the HetNet base station (BS) may significantly affect the choice of the modulation and coding scheme (MCS) and thus the immediate effective service rate provided by a RAT. In a max-min setting, the UEs with worse channel conditions have to be compensated with larger bandwidth allocations. The latter, in turn, implies that attempting to provide an exact fair allocation would dramatically decrease the throughput of the entire network, which may be unacceptable for the service operators (see e.g., [33] for a more detailed discussion).

As an adequate alternative appropriately accounting for the balance between the network throughput and the resource allocation fairness, we propose to employ bandwidth-based max-min fairness. Accordingly, instead of attempting to deliver a fair allocation with respect to the rates provided to the UEs, we first propose to divide the available bandwidth as fairly as possible. Doing so results in a fair allocation with respect to the set of frequencies granted by the overall network to a particular BS. Utilizing these frequencies, users that are closer to their serving BSs are receiving higher data rates than those located farther away due to the use of different MCSs. As one may observe, this advanced criterion still prevents from the infinitesimally small rates since all of the users are provided with the same set of frequencies, whereas the highest rate is upper limited by the fastest MCS that can be achieved on a particular RAT. Therefore, the proposed metric results in a much desired trade-off between the system-wide throughput (benefiting from the dynamic nature of MCS selection) and the fairness of resulting data rate allocations.

In order to further introduce an explicit capability to exchange network throughput for fairness and vice versa, we supplement our baseline objective function with dedicated weighting coefficients. These coefficients are defined by carefully selected functions mindful of the instantaneous UE spectral efficiency and thus allow to flexibly adjust the current operational regime of a HetNet. In summary, the proposed criterion could be thought of as a variation of the proportional fairness metric specifically constructed for HetNets, where the users located at larger distances from the BS are penalized more heavily. An important fact is, as we demonstrate in what follows, that the optimization task still remains in the category of LP problems. The latter allows for effective solution algorithms, whose computational complexity scales well with the dimension of the problem, that is, the size of the HetNet and the number of users.

In summary, the proposed criterion is a type of α-fairness that has been introduced as an extension of proportional fairness, thus allowing to control the trade-off between throughput and fairness in communication networks [35]. Over the last decade, several authors have shifted their attention to this concept within the scope of wireless systems, see e.g., [36], [37]. However, those studies remained primarily theoretical in nature as there have been no standardized mechanisms in 3G, 3G+, and 4G systems to implement the required functionality. In this work, we not only take a careful look at controlling the throughput–fairness trade-off in a simple fashion by using the framework of α-fairness, but also proceed with discussing how it might be implemented in 4G+ and 5G networks. Along these lines, we focus our attention on the architectural aspects of such systems as well as address the problem of realistic scheduling therein by utilizing backpressure algorithms that implement α-fairness in practice.

IV. PROPOSED OPTIMIZATION FRAMEWORK

A. TRAFFIC AND TOPOLOGY MODELING

Specifying the type of user traffic demands is of particular importance in emerging HetNets, as these in turn affect the choice of the objective function and, consequently, the solution algorithm. In this study, we assume that the offered load generated by the users is "greedy elastic" (also known as "full buffer" or "saturated" traffic in 3GPP specifications). Recall that greedy traffic occupies all of the available radio resources, whereas elasticity implies adaptiveness to such allocated resources. In practice, elastic traffic is typical not only for TCP data transfer sessions, but also for contemporary video and voice applications built on top of dynamic rate adaptation algorithms.

Further, we consider a characteristic single-cell scenario in a 3-tier converged HetNet, see Fig. 2.A. The following entities comprise the system of interest: N users

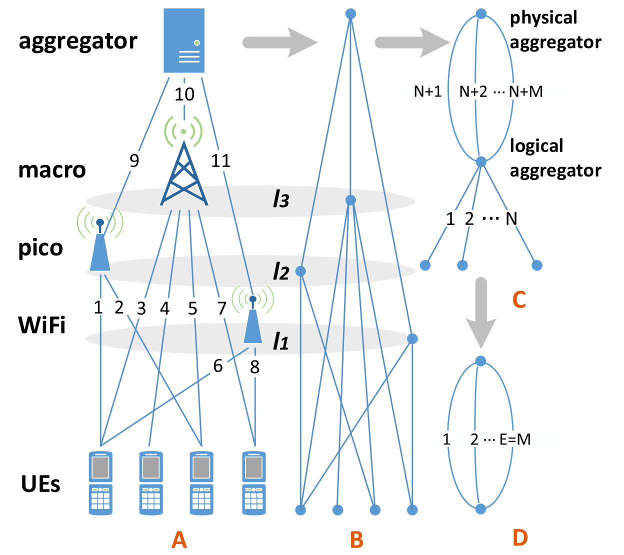


FIGURE 2. Formalization of HetNet system topology.

across 3 separate RAN tiers (called hereinafter *layers*); $M_1 = 1$ BS at layer L1 (e.g., LTE macro cell); M_2 BSs at layer L2 (e.g., LTE pico cells); and M_3 BSs at layer L3 (e.g., WiFi access points). Overall, there are $M = M_1 + M_2 + M_3$ BSs in the considered deployment, as well as one coordinating node (e.g., physical aggregator) assembling traffic after the UEs split it across multiple concurrent RATs.

In summary, we make the following assumptions:

- a user may access one BS at each HetNet layer;
- greedy and elastic traffic demands are considered;
- locations of the users are known to the coordinator;
- achievable data rate depends on the current MCS.

The coordinator, which is a centralized resource allocation module, is physically incorporated into the L1 BS equipment. Hence, the decisions on the appropriate resource allocations are made at the time instants when new users arrive into the system. An example HetNet topology with exactly four users is demonstrated in Fig. 2.B. In what follows, we concentrate on the so-called *bifurcated* resource allocation principles assuming that any user may simultaneously employ two or more radio interfaces and split its traffic arbitrarily between them. For this reason, there has to be an aggregator terminating virtual tunnels over concurrent RANs, as shown in the figure.

Analyzing the HetNet topology demonstrated in Fig. 2.B, we conclude that it is impossible to explicitly define data rates on the links between the users and the BSs. This is due to the fact that those are, in fact, shared links. Furthermore, we notice that the topology in question is redundant as the links connecting the BSs to the aggregator in a properly dimensioned system should have equal or higher capacity than the ones provided by these BSs at the air interface. For example, letting $e = 1, 2, \ldots, E$ be the set of links and using c_e to denote their capacities at the radio layer, we learn from Fig. 2.B that $c_9 \geq c_1 + c_2$, $c_{10} \geq c_3 + c_4 + c_5 + c_7$, and $c_{11} \geq c_6 + c_8$. Hence, links c_e, $e = 9, 10, 11$ yield no additional constraints.

Correspondingly, the simplified network topology is condensed in Fig. 2.C, where the aforementioned "redundant" links have been removed. Here, two additional key system entities are shown: the logical aggregator and the physical aggregator (which corresponds to the aggregator in Fig. 2.A). Within this simplified topology, we do not immediately identify possible paths between the source (S) and the destination (D). However, supplementing it with the set of possible paths for S-D pairs, allows us to indicate such routes unambiguously. Note though that we still cannot explicitly determine the data rates on the links $e = 1, 2, \ldots, N$ to the logical aggregator as they actually constitute flows realizing the user demands.

In summary, our final HetNet modeling topology is illustrated in Fig. 2.D, where all of the demands are concentrated in between two nodes, the logical aggregator and the physical aggregator. The number of links in such a system equals the number of BSs, $M = M_1+M_2+M_3$, whereas their capacities are equal to the effective capacities of the corresponding BSs. Some of these links (but not all) are shared across the demand pairs. Ultimately, the paths for each demand are known to us explicitly and they uniquely identify data flows that must be implemented in the system to satisfy the current user demands.

B. NETWORK FLOW PROBLEM FORMULATION

Here, we first consider simple rate-based resource allocation. In this case, the current spectral efficiency (SE) of a user is not taken into account and all of the user nodes are treated equally in terms of the bandwidth allocation. Further, we extend this formulation by explicitly taking into account the SE as part of the capacity constraints. The latter case is expected to result in better fairness of resource allocations, while the former should potentially enable the trade-off between the fairness of allocation and the overall system throughput. Then, we extend the latter case to the notion of controlled fairness by introducing the weighting coefficients as functions of the user SEs. For all the discussed problems, we are interested in bifurcated solutions, i.e., the schemes where all of the RATs that the UE is associated with can be used simultaneously. Below we begin by specifying the task of bandwidth-based max-min allocation and then extend it to take into account the SE of user nodes.

Denote by N the number of demands that are to be realized in the system. Recall that each user node is associated with exactly one demand. We label the demands with index d, $d = 1, 2, \ldots, N$. The demand volumes are expressed in bits per second (bps) and denoted as

$$h_d, \quad d = 1, 2, \ldots, N. \tag{2}$$

Note that the values of h_d are not known in advance due to the assumption of greedy elastic traffic demands. These values have to be determined such that a certain fairness criterion is satisfied. Let us also denote by Π_d the set of paths for demand d, i.e.,

$$\Pi_d = \{P_{d1}, P_{d2}, \ldots, P_{dP_d}\}, \quad d = 1, 2, \ldots, N, \tag{3}$$

where P_d is the total number of the available paths for the demand d. In practice, P_d is the number of BSs at all layers that a user is associated with. In our target topology, these paths are readily available as all the subsets P_{dp}, $p = 1, 2, \ldots, P_d$, $d = 1, 2, \ldots, N$ consist of exactly one element connecting the aggregators. Here we may avoid using paths as additional variables, hence formulating the task solely in terms of links (there is a one-to-one mapping between those). We prefer using them to avoid confusion with the notation typical for the network flow problems.

Further, let

$$x_{dp}, \qquad d = 1, 2, \ldots, N, \; p = 1, 2, \ldots, P_d, \tag{4}$$

be the flows realizing a part of demand d over the path p. These terms are known as *flow variables*. For bandwidth-based allocation, these flow variables are measured in bps. In the second case, where SEs of the end nodes are taken into account, flow variables x_{dp} are measured in Hz.

The first set of equations that we need to determine is the so-called *demand constraints*. These constraints make sure that all of the demands h_d are fully realized using the flow variables x_{dp}, $p = 1, 2, \ldots, P_d$. The *demand constraints* are conventionally defined as

$$\sum_{p=1}^{P_d} x_{dp} = h_d, \qquad d = 1, 2, \ldots, N. \tag{5}$$

Next, we use the link-path-incidence variables δ_{edp}, as

$$\delta_{edp} = \begin{cases} 1, & e \cap P_{dp} = e, \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

and the *capacity constrains* read as

$$\sum_{d=1}^{N} \sum_{p=1}^{P_d} \delta_{edp} x_{dp} = R_e, \quad e = 1, 2, \ldots, E, \tag{7}$$

where R_e are the rates of the respective BSs measured in bps. The capacity constraints guarantee that there are no overloaded links in our network. Observe that $\delta_{edp} = 1$, if the path p of the demand d uses the link e.

Introducing the allocation vector as

$$\vec{h} = (h_1, h_2, \ldots, h_N), \quad h_d = \sum_{p=1}^{P_d} x_{dp}, \tag{8}$$

we see that the optimization problem is to lexicographically maximize $\vec{h}$ given the constraints (5) and (7). The respective solution yields the sought max-min allocations.

Consider now the second case that explicitly takes into account the impact of SEs of user nodes. This, in turn, requires a modification of both demand and capacity constraints as follows. Let s_{dp} be the current SE of a flow variable x_{dp} realizing a part of the demand d over the path p, measured in bits per second per Hertz (bits/s/Hz). Recall that the SEs specify how efficient a flow allocation x_{dp} is depending on the MCS currently in use by a UE. In order to provide the max-min allocation in the presence of different MCSs at user nodes, we need to allocate more bandwidth to the UEs having MCSs with the lower data rates. Observe that now the flow allocations x_{dp} are measured in Hz.

The demand constraints are now given by

$$h_d = \sum_{p=1}^{P_d} s_{dp} x_{dp}, \quad d = 1, 2, \ldots, N, \tag{9}$$

where the products $x_{dp} s_{dp}$ provide the allocations in bps.

The capacity constrains remain the same as in (7) with the rates R_e replaced by the bandwidth B_e measured in Hz. The task at hand is to lexicographically maximize $\vec{h}$. The problem formulated above (with the SEs taken into account) provides the max-min allocation of rates to the user nodes. In this case, the amount of bandwidth assigned to the UEs will change proportionally to the relative weights of their current MCSs. Since the target task is classified as a *fair network capacitated problem* with fluid capacity, we may replace inequalities in the capacity constraints with equalities. We also note that the resources are scheduled by the LTE BSs based on the discrete resource allocation blocks implying that the actual allocation will slightly deviate from the theoretical prediction.

C. FINE-GRAINED FAIRNESS CONTROL

Consider now the modifications required to the aforementioned task, that is, to enable the possibility of exchanging the overall system throughput for fairness. These modifications are based on several dedicated functions applied to the SE of the UEs. Recall that by default the scheme under discussion will allocate the resources with the maximum possible fairness. This will leave the users with high SE and the best connectivity (e.g., those within coverage of all three layers) with inadequately small amounts of bandwidth in order to improve performance of "unlucky" nodes with poor connectivity and low MCSs. Although this strategy may indeed have a positive effect on the lower percentiles of the cumulative distribution function (CDF) for per-user rate allocations, the total throughput is likely to remain low (see related discussion in Section V). However, we suggest to modify the input SEs, s_{dp}, by an appropriate control function $f(x)$ in such a way that the absolute difference between the UEs with "high" and "low" SEs would be reduced. Hence, the proposed scheme assigns more resources to the UEs with better SEs to trade the fairness of allocation for the overall system throughput.

The key research question for the proposed modification is the choice of the suitable control function $f(x)$. In this work, we consider the two alternatives, $f(x) = x^{\alpha}$ and $f(x) = \alpha^{x}$, where x is the SE of UEs, see Fig. 3. The first option is straightforward and easy to understand. Indeed, the original fair allocation is achieved by setting $\alpha = 1$, whereas the "maximum throughput" solution is obtained by allowing $\alpha \to 0$. The latter is the case for the rate-based allocation,

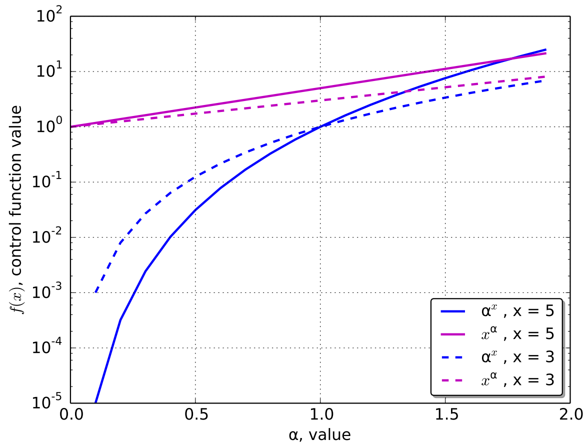


FIGURE 3. Comparison of alternative control functions.

where SEs are not taken into account ($s_{dp} = 1, \forall d, p$). However, the second choice of a function, which is widely used in e.g., range compression of RF power amplifiers, could potentially provide for a better alternative due to the presence of a special balancing point at $\alpha = 1$, see Fig. 3. The values of α lower than 1 lead to better allocations for the users with higher SEs, thus improving the system throughput. Next, choosing $\alpha = 1$ delivers the rate-based resource allocation that still exchanges fairness for the system throughput. For $\alpha > 1$, the scheme tends to improve the fairness of resource allocation eventually reaching the case of bandwidth-based scheduling. Further increase of α, however, does not lead to the corresponding growth in fairness, as we emphasize in what follows. Nevertheless, the use of $f(x) = \alpha^x$ may be more flexible due to its behavior at $\alpha \to 0$, allowing to target the entire controllable range in HetNets. It is important to note that the proposed modifications do not alter the class of the optimization problem in question, as they apply to x_{dp} only in terms of the weighting coefficients.

D. PROPOSED SOLUTION ALGORITHM

The solution to the max-min optimization problem at hand is an extension of the algorithm for the single path problem, see e.g., [38]. In our work, we employ Matlab to implement the corresponding allocation algorithm. This task is classified as the LP problem comprising the initial "water-filling" stage for detecting the maximum allocation that can be assigned to all of the flows simultaneously, and the subsequent refinement of the allocations for selected flows. The first stage may have multiple solutions as opposed to the unique solution for the fixed-path allocation. Due to this ambiguity, we need to additionally perform the so-called non-blocking tests to decide upon the resulting max-min allocation. To have a detailed look at the steps discussed, the reader is encouraged to review Algorithm 1.

The proposed algorithm is to be invoked at the macro-LTE BS whenever a new user arrives into the system or an existing user changes its connectivity state. The reason behind

Algorithm 1 The Weighted Max-Min Algorithm Solving the Multi-RAT Resource Allocation Problem in a Capacitated Fair Network

1: Modify the initial vector of SEs (s_{dp0}) to receive control over the system fairness/throughput. $s_{dp} = f(s_{dp0})$, where $f(s_{dp0})$ is the selected control function (e.g., $f(x) = \alpha^x$)
2: Estimate Δ as a solution to the LP problem and
- set $n = 0$, $\Delta^{(0)} = \Delta$;
- define $Z_0 = \{1, 2, \ldots, N\}$.

3: Set $n = n + 1$ and for each d
4: Check the throughput allocation. The users whose allocations can be increased are defined as a subset $Z_n \subseteq Z_0$
5: **if** $Z_n = \emptyset$ **then**
6: **go to** 13
7: **else**
8: $Z_{n-1} = Z_{n-1} \setminus Z_n$.
9: **end if**
10: Solve the following LP problem

$$\max \Delta,$$

subject to:

$$\begin{aligned} &\sum_{p=1}^{P_d} s_{dp} x_{dp} = h_d, && d = 1, 2, \ldots, N, \\ &\Delta - h_d \leq 0, && d \in Z_n, \\ &\Delta^{(k)} - h_d \leq 0, && k = 0, 1, \ldots, n-1, \\ &\sum_{d=1}^{N} \sum_{p=1}^{P_d} \delta_{edp} x_{dp} = B_e, && e = 1, 2, \ldots, E, \end{aligned}$$

11: Set $\Delta^{(n)} = \Delta$
12: **go to** 3
13: Apply function $f_1(x)$ that will take into account the SE vector modifications considered in step 1. With the said function applied, the final user throughput vector h_{d1} will have the form of $h_{d1} = h_d s_{dp0}/s_{dp}$.

is that the max-min fair allocation for all the users depends on the availability of wireless interfaces for every particular UE. It is important that no ongoing sessions are interrupted due to the new user arrivals as only the available allocations are recomputed and communicated to the UEs. For the same reason, there is no need to solve the underlying allocation problem when no changes in the connectivity of the users are observed. Finally, we emphasize that due to the LP nature of the problem, the computational complexity at a macro-LTE BS is manageable even for ultra-dense scenarios.

E. HEURISTIC APPROACHES

In this subsection, we introduce two simple heuristic resource allocation strategies that may serve as benchmarks for the approach proposed above. The first scheme is a UE-centric mechanism, where the UE independently decides on how much resources will be utilized at each network layer. It is important to note that with this procedure the user is not aware of which resources are generally available at each layer and the UEs locally compete with one another at every accessible radio interface. Furthermore, the assumed greedy elastic type of traffic does not leave room for smarter resource allocation decisions. These two limitations motivate us to implement a simple "greedy" scheme: the UEs consume as much resources as the BS scheduler allows. In the remainder of this text, we refer to this heuristic method as the "max-usage".

The second approach is an example of the network-assisted resource allocation: the wireless system is helping the UE to make more intelligent decisions by providing relevant information on the neighboring service entities. Over the recent years, such algorithms are under active investigation by the 3GPP community for LTE Release 12 and beyond (see e.g., [39]). In our research, we implement a simple example scheme, where the only information provided to the users is the preferred association threshold for each HetNet layer. By appropriately adjusting this threshold, the network may increase/decrease the effective coverage range of pico-LTE stations and WiFi access points in order to better offload the UE traffic from/to the macro LTE cell.

Furthermore, it is also assumed here that the UE can only use a single layer (radio interface) at a time (e.g., WiFi or pico-LTE), which is a common consideration for the ongoing 3GPP standardization efforts. With this constraint in mind, the user first attempts to connect to the WiFi access point, as it typically offers higher data rate. If the network-provided association threshold does not allow for a WiFi connection, a pico-LTE base station is attempted instead. Finally, if none of the previous attempts succeeds the UE remains connected to the macro-LTE base station. Generally, dedicated thresholds might be applied for each individual BS. In this work, we first select the thresholds based on the averaged values of the SE, s_{dp}, of the UEs on a particular layer. Further, we consider different thresholds in order to control the trade-off between the fairness and the overall system throughput. In what follows, we name the discussed heuristic algorithm as "WiFi-preferred".

V. SELECTED NUMERICAL RESULTS

A. PARAMETERS OF THE ENVIRONMENT

The reference HetNet scenario that we consider in this work includes the following system entities: a single macro cell BS (named eNodeB), several pico eNodeBs and WiFi access points (APs), as well as a number of multi-RAT (WiFi and LTE) user nodes. Following 3GPP discussions on the characteristic HetNet use cases [40], [41], pico eNodeBs, WiFi APs, and UEs are deployed uniformly within the coverage area of the macro eNodeB. Each user is assumed to have several active radio connections depending on the availability of RATs and the perceived signal strength (has to exceed the AP/eNodeB association level). The latter is estimated subject to different pathloss models.

According to 3GPP [40], [41], for the macro-cell users the distance-dependent pathloss is given by the ITU Urban Macro (UMa) channel model. For the WiFi APs and pico LTE users, the pathloss is approximated using the ITU Urban Micro (UMi) channel model specified in [42] with the following environmental parameters: the average height of the buildings is $h = 20m$ and the average street width is $W = 20m$. In addition to pathloss, accurate modeling of the propagation environment requires appropriate slow and fast fading models. Slow fading is explicitly taken into account as we consider our network at discrete time instants $n\Delta t$, $n = 0, 1, \ldots$. The effects of fast fading smoothen out over longer periods of time as the performance metrics that we are interested in are the averaged values. Hence, the parameters related to fast and slow fading could be incorporated by employing the constant margins (with the values equal to the standard deviations from [42]), which are added on top of the RAT association threshold. In this work, we also approximate the current levels of interference by following a similar technique. We discuss these important approximations in more detail below. All the basic parameters are summarized in Table 1, where the term LPN (low-power node) refers to both WiFi AP and pico-LTE BS.

TABLE 1. Primary deployment parameters.

Parameter	Value
LTE/WiFi configuration	10 MHz FDD / 20 MHz
Layout	1 macro cell, several LPNs
Macro/LPN-UE pathloss model	ITU UMa/UMi [42]
Macro/LPN antenna gain	17/6 dB
Macro/pico/WiFi max. power	43/23/20 dBm
UE max. power	23/20 (LTE/WiFi) dBm
LTE/WiFi power control	Max power
UE/macro/LPN antenna height	1.5/25/10 m
UE noise figure/feeder loss	5 dB / 0 dB
Traffic model	Full-buffer
LPN/UE deployment type	Uniform [40]
LPN/UE-macro distance	$>$ 75/35 m [41]
LPN/UE-UE distance	$>$ 40/10 m [41]
Trials per experiment	1000

B. CALIBRATING ANALYTICAL ENVIRONMENT

The proposed analytical model incorporates a number of simplifying assumptions. In order to verify whether these assumptions hold with respect to the metrics of interest (including throughput and fairness), the model in question has to be tested against a more realistic HetNet environment. In this research, we employ our own advanced "large-scale" system-level simulator (SLS), which takes into account all the relevant details of HetNet operation and has been thoroughly calibrated in our past publications [43], [44].

Our SLS tool is capable of modeling large-scale multi-RAT environments, including the underlying wireless technologies, such as LTE, WiFi, and mmWave-based RATs. It is based on flexible event-driven architecture, which allows to significantly decrease the computation time for low-loaded scenarios. For all the considered technologies, PHY and MAC layers are implemented in detail, based on the appropriate IEEE and 3GPP specifications, while the higher layers are generally simplified to abstract away the traffic models supported with analytical approximations. Regarding the environment generation, the SLS tool supports 3D geographical models, which take into account time–location based interference characterization, antenna configurations, various UE mobility models, and wrap around. The current open-source version of our SLS may be acquired at [45]. Below we compare the performance of simpler heuristic strategies introduced above with both the analytical performance optimization and the large-scale SLS results. Matching the observed performance for the heuristic schemes targets to confirm that the behavior of the proposed analytical framework essentially repeats that of the SLS tool.

At the calibration phase, the operation of the max-usage scheme implemented in both analytical and SLS environments has been contrasted in a reproducible test scenario with the fixed coverage and capacity settings. The monitored calibration parameters have been the number of users (60 UEs uniformly deployed within the macro cell area), the number of pico-LTE BSs and WiFi APs (5 nodes of either type with uniformly distributed locations across the macro cell area), and transmission power/channel options. One of the key performance indicators discussed below is the fairness of resource allocations. To assess the resulting fairness of a particular set of such allocations, we utilize Jain's index defined as

$$J = \left(\sum_{i=1}^{N} x_i\right)^2 \frac{1}{N \sum\limits_{i=1}^{N} x_i^2}, \tag{10}$$

where x_i is the UE throughput and N is the number of users.

The calibration results averaged over 100 realizations (a real-time run of 4 seconds for each replication in the large-scale SLS) are demonstrated in Fig. 4, where the left subplot outlines the user association statistics (the number of users connected to each RAT), whereas the right subplot details the average per-UE throughput. First, we confirm that the difference in the values of metrics under comparison as obtained with the two approaches is well within 5%, which is sufficiently accurate considering the amount of details accounted for by the SLS tool. The observed smaller deviations are primarily due to the absence of inter-cell interference in the analytical model.

While the deployment of the pico-LTE infrastructure has been implemented according to the 3GPP specifications [40], [41], which should enforce a certain minimum distance between the pico-LTE BSs, the interference between the neighboring nodes is still non-negligible [46]. Observing the throughput CDF, we notice a small step at around 25% for the analytical curve. Such behavior is due to the difference in throughput levels for the users having macro-LTE only connectivity and those with additional RAT associations. Note that in the SLS this effect is smoothened out due to the random interference picture. Similar

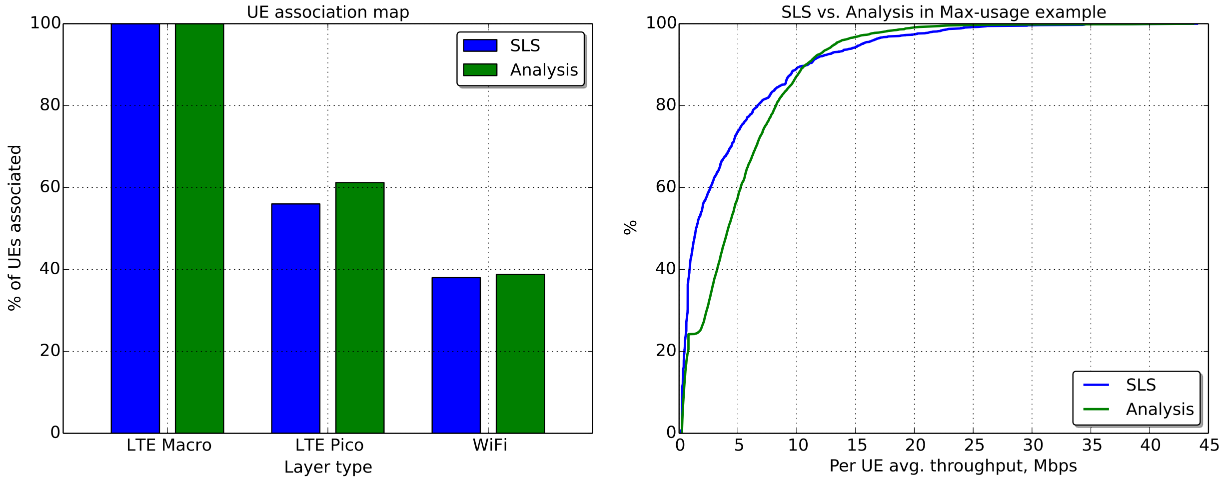


FIGURE 4. Calibration of analytical framework based on connectivity (left) and throughput CDF (right).

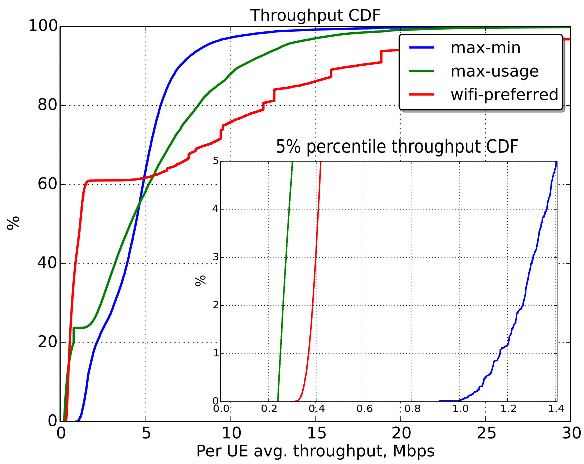


FIGURE 5. Throughput CDF (analytical tool).

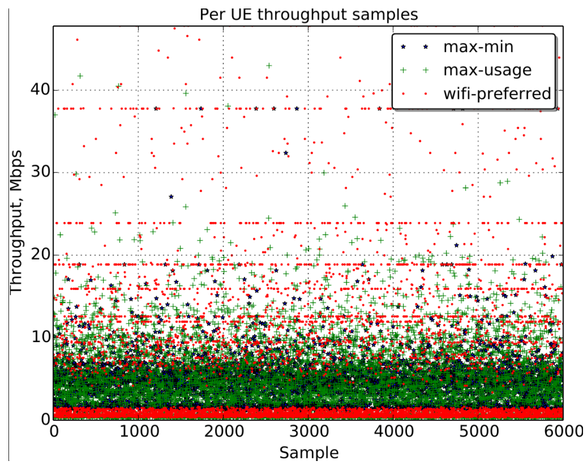


FIGURE 6. Throughput spread diagram (analytical tool).

smoothening effects may be observed in cases when the UE with better connectivity (e.g., associated with all three layers at the same time) does not have an advantage in the throughput as compared to a macro-cell only UE due to the high levels of interference on the WiFi and/or pico layers. In summary, the demonstrated results allow us to conclude that the developed analytical modeling tool is sufficiently accurate to evaluate the effects of the proposed resource allocation strategies.

C. PERFORMANCE ANALYSIS

Let us now compare the performance of the proposed analytical optimization scheme offering the bandwidth-based allocation ($f(x) = s_{dp}^{\alpha}$ with $\alpha = 1$) with that of the simpler heuristic procedures. The results of such comparison are illustrated in Fig. 5 and Fig. 6. In particular, Fig. 5 details the throughput CDF and the 5%-percentile for all the considered strategies. Even though at the first glance the average throughput is nearly equal for all the three alternatives, one may observe that the difference in fairness and the 5%-percentiles is dramatic. For the WiFi-preferred scheme, the throughput figures are exceptional for the UEs having WiFi connectivity, whereas pico- and macro-LTE users are suffering (the step at around 60%). In addition, with this allocation mechanism extra small steps in the CDF curve may be noticed for the WiFi-only users. This effect is caused by the lack of flow splitting (bifurcation) as every UE utilizes a single RAT at a time. Further, the proposed scheme delivers the best 5% throughput percentile performance, which is however achieved at the cost of some degradation in the averaged per-UE throughput.

In more detail, Fig. 6 elaborates on the per-UE throughput values for each user in the form of a spread diagram. This illustration highlights the fact that the red dots (WiFi-preferred scheme) are mostly concentrated in top and bottom parts of the plot resulting in "dotted" lines due to the non-bifurcated nature of resource allocation, as discussed above. By contrast, the green pluses (max-usage scheme) are

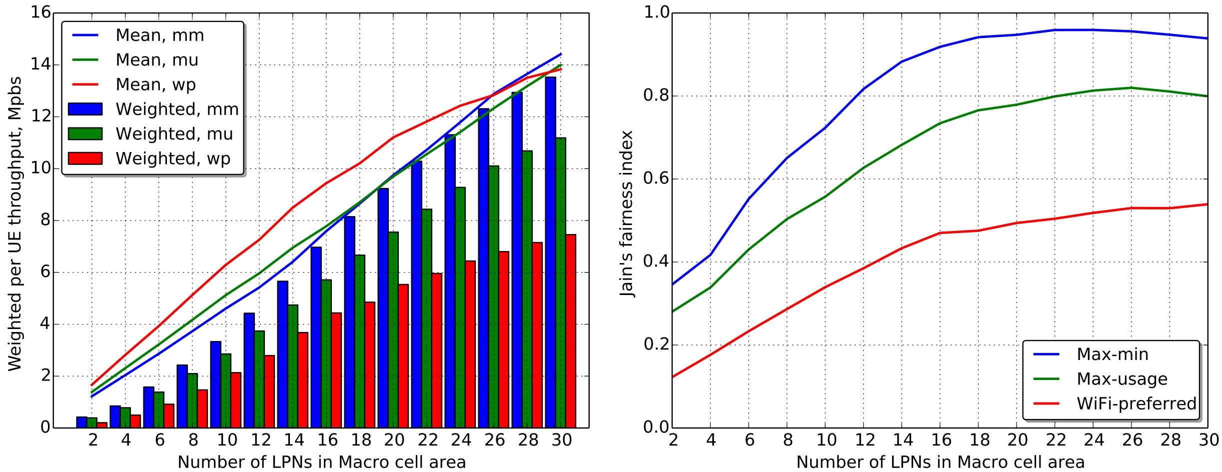


FIGURE 7. Weighted fair and average throughput (left), as well as Jain's index (right) for different deployments.

located essentially everywhere evenly. Notably, the blue stars (max-min scheme, which corresponds to Algorithm 1) tend to group in the center of the plot, thus yielding the best fairness and 5%-percentile performance of the max-min bandwidth-based allocation.

To better understand the impact of capacity and connectivity on the performance of the proposed algorithms, we also assess the behavior of the considered resource allocation schemes for the various densities of infrastructure nodes at different layers. The resulting fairness and the average per-UE throughput statistics for all the three schemes under comparison are highlighted in Fig. 7, where the numbers of LPNs at the pico-LTE and the WiFi layers are varied. Analyzing the obtained results, one may observe that when the connectivity of the users and hence the capacity of the entire network is poor (i.e., almost all of the UEs utilize only the macro LTE cell), the degree of unfairness is large. It is also evident that none of the considered strategies, including the most fair max-min bandwidth-based scheme, can improve over such behavior.

As user connectivity improves, the average per-UE throughput of the WiFi-preferred scheme becomes significantly higher compared to the other strategies. However, this effect is due to a unique position of particular users with WiFi connectivity, who will now enjoy the highest throughput in the system. However, the weighted fair throughput (the throughput multiplied by the Jain's index) of the WiFi-preferred scheme is still lower as the result of much poorer fairness in such a system. If user connectivity improves further, the max-min bandwidth-greedy mechanism begins to perform better in terms of both the average throughput and fairness of resource allocations, eventually reaching and even outperforming the max-usage as well as the WiFi-preferred solutions. Another interesting observation can be made when analyzing the behavior of the Jain's index (the right subplot in Fig. 7). As one may notice, once a certain degree of connectivity is reached no further improvement in fairness can be obtained. We therefore name this effect the "fairness saturation" and note that the fairness plateaus for all the considered schemes. This phenomenon is the consequence of near-ideal connectivity (when almost every UE has access to the BSs on all three layers) and uniform distributions of the UEs and the BSs within the coverage area of a macro cell. The absolute fairness saturation value is also affected by the implementation details of a resource allocation strategy.

D. BALANCING THROUGHPUT AND FAIRNESS

In this subsection, we concentrate on the relative performance of the max-min and max-usage schemes allowing for better control of the trade-off between the system throughput and the fairness of resource allocations. Recall that in case of max-min fairness we may apply two alternative weighting functions to the current SE values of the users, whereas for the max-usage mechanism we can instead control the association threshold for the LPNs. Along these lines, Fig. 8 emphasizes the balance between fairness and per-UE throughput for the max-min allocation strategy, as evaluated for the two different control functions α^x and x^α. As expected, over a certain range of α the performance of the max-min scheme with both control functions remains almost identical. This region is observed for $f(x) = s_{dp}^{\alpha}$ with $\alpha \in (0, 1)$.

However, we also remind that the control function $f(x) = \alpha^x$ has a much wider range of the applicable values of α, thus potentially allowing to achieve higher values of throughput. In addition, Fig. 8 highlights the actual control limits due to the effects of both capacity and coverage. For instance, it could be infeasible to achieve any average per-UE throughput by making α smaller than 0.1 for $f(x) = \alpha^x$, as the scheme in question would attempt to allocate more resources to the UEs with high SE, which may not be possible due to insufficient remaining resources. Therefore, any further decrease in α would only degrade the overall system performance. Similar behavior is observed when the considered functions are applied to improve fairness e.g., making α in

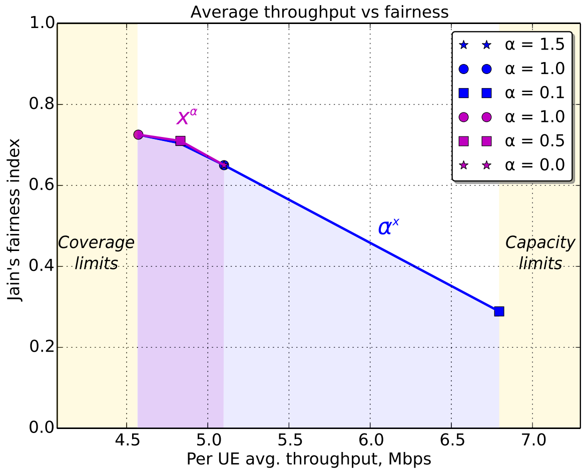


FIGURE 8. **Average per-UE throughput vs. fairness.**

$f(x) = \alpha^x$ higher than 1.5. In that case, the coverage limits are met and further increase in α would lead to excessive resource allocations for the UEs with low connectivity. This, in turn, is wasteful and unfair towards the UEs positioned in "good" locations.

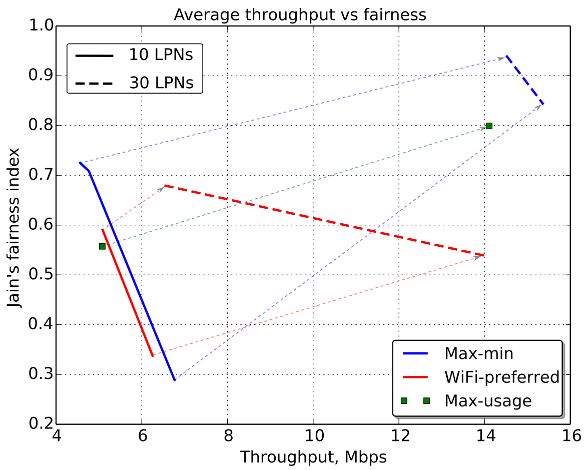


FIGURE 9. **Control regions for different schemes.**

Finally, we compare the control performance of our weighted max-min bandwidth-greedy scheme with that of the max-usage heuristic approach in two different coverage and capacity environments. In fact, in these experiments we target to identify the available operational regimes (control regions) in a multi-RAT HetNet. The corresponding results are shown in Fig. 9, where the control function for the weighted max-min allocation is chosen to be $f(x) = \alpha^x$. In the first scenario, 5 WiFi APs (40% of the macro-LTE cell coverage) and 5 pico-LTE BSs (60% of the macro-LTE cell coverage) are deployed within the macro cell area. The second scenario recreates an ultra-dense deployment with 15 LPNs of each type placed within the same area, thus delivering almost 100% coverage at both pico-LTE and WiFi layers.

In the considered scenarios, in addition to the network-centric max-min solution, it also becomes possible to adjust fairness of the network-assisted WiFi-preferred scheme by altering the association threshold towards very low (higher fairness) vs. very high (more throughput) settings. However, as we learn from Fig. 9, such adjustment possibilities for this heuristic mechanism are significantly more modest than those available for our network-centric weighted max-min algorithm, as the fairness of the former approach is consistently lower. In the ultra-dense case, the max-min control range reduces significantly, as close-to-maximum fairness/throughput ratio has already been reached with the default parameters (bandwidth-based allocation) and no additional adjustment is necessary. Another interesting effect is observed when assessing the behavior of the UE-centric max-usage scheme. In the ultra-dense deployment, this approach performs better than the network-assisted WiFi-preferred strategy. The reason is that in such a scenario even a simple bifurcated scheme outperforms most non-bifurcated solutions.

VI. PRACTICAL IMPLEMENTATION CONSIDERATIONS

In principle, our proposed solution offers network operators a convenient tool to control the throughput–fairness trade-off in a simple and reliable manner. However, in order to actually provide a versatile instrument for the resource allocation and management, the proposed model needs to be formulated in terms of directly controllable network parameters, such as the buffer content at different entities and the actual data rates. In this section, we illustrate how our solution could be mapped onto practical HetNet implementation. The protocol model of the prospective HetNets and the traffic flows between the involved entities are demonstrated in Fig. 10 (left). Generally, the setup follows the DC and LWA logic outlined in [24] and [47] (as described above). Particularly, the aggregation/splitting of the data traffic is enabled at the PDCP layer, even if the serving eNodeB and the WiFi AP are not collocated with the master eNodeB. Note that according to [47], the control messages are still transmitted trough the MeNB hence requiring the UE to be connected to it.

A practical implementation of the weighted α-fairness is known to be provided by the backpressure algorithms [48]. The use of backpressure for generic wireless systems has been addressed in [49] and [50]. Specifically, it has been demonstrated in [35] that by utilizing the Lyapunov function drift the weighted α-fairness can be enabled in the network by computing the difference in the buffer content at the concatenated queues. The latter is of particular importance since it provides a simple way to enforce α-fairness in prospective HetNets by using directly observable variables.

To this end, Fig. 10 (right) demonstrates the abstracted HetNet system model. There are N users assigned to MeNB, SeNB, and WiFi. The task is to decide how much data rate the UE will receive on each interface considering the backlogs in MeNB, SeNB, and WiFi queues. We introduce the following notation:

- x_{i1}, x_{i2} and x_{i3} are the rates of user i, $i = 0, 1 \ldots N$, measured in bit/s, on MeNB, SeNB, and WiFi interfaces,

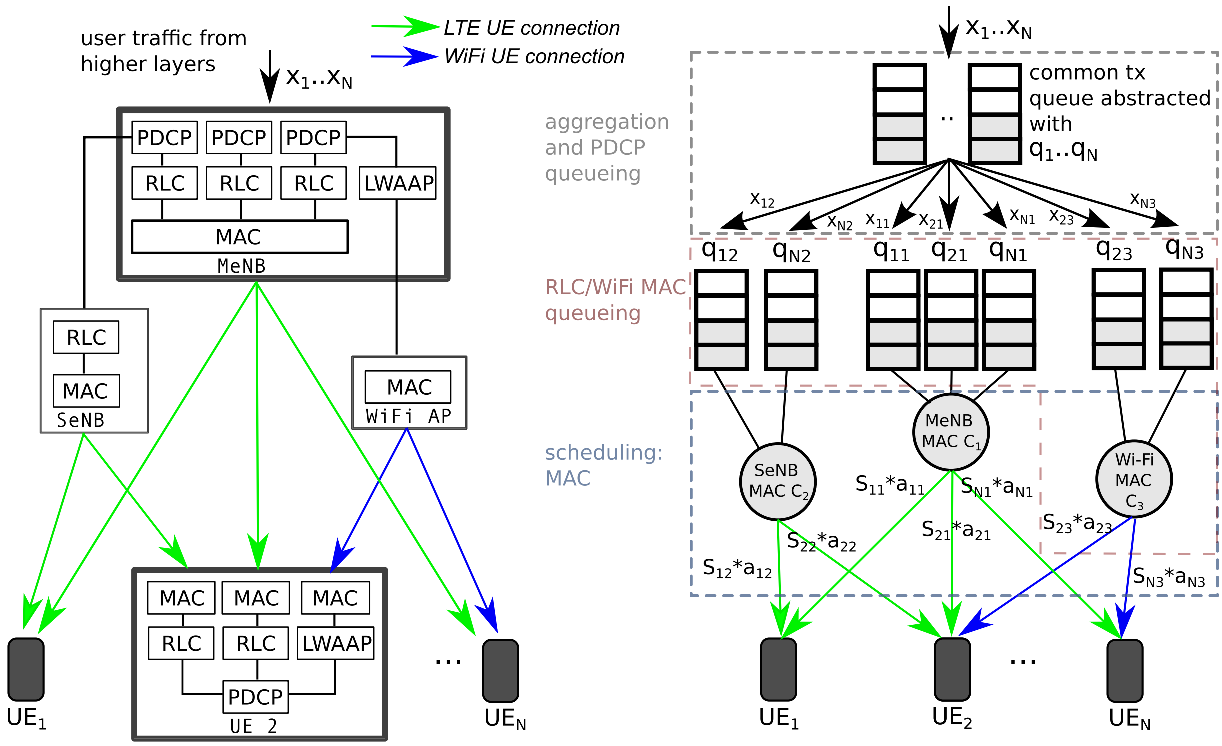


FIGURE 10. Merged LWA and DC architecture (left) and its abstraction (right).

respectively. These values are controlled by the aggregation/splitting logic;

- x_j, $j = 0, 1, 2...N$, are the initial rates defined by the traffic model of user j that will share the MeNB, SeNB, and WiFi interfaces, measured in bit/s;
- q_{i1}, q_{i2}, q_{i3}, $i = 0, 1, 2, 3...N$, are the backlogs in the queue of user i on MeNB, SeNB, and WiFi interfaces, measured in bits. Although in practice there is a common queue for all the UEs, we may sort them based on the packet destination and abstract as a separate transmission queue for each UE;
- q_j, $j = 0, 1, 2...N$, is the backlog in the queue of user j that shares the MeNB, SeNB, and WiFi interfaces, measured in bits;
- C_k, $k = 1, 2, 3$, are the bandwidths of MeNB, SeNB, and WiFi interfaces, respectively, measured in Hz;
- $a_{i1}, a_{i2}, a_{i3}, i = 0, 1...N$, are the shares of the channel on MeNB, SeNB, and WiFi interfaces that user i is provided with, measured in Hz. The latter can be approximated by using the LTE scheduler type and the MAC layer reports;
- S_{i1}, S_{i2}, S_{i3} are the SEs of MeNB, SeNB, and WiFi interfaces, measured in bit/s/Hz.

The α-fairness resource allocations can be formulated in terms of the Lyapunov function drift as

$$\min \sum_{i=1}^{N} U(a_{i1}, a_{i2}, a_{i3}) - \beta \frac{\partial}{\partial t}\Big[\frac{1}{2} \times \sum_{j=1}^{N} \Big[q_{j1}(t)^2 + q_{j2}(t)^2 + q_{j3}(t)^2\Big]\Big], \quad (11)$$

where $U(a_{i1}, a_{i2}, a_{i3})$ is our considered utility function inducing α-fairness, β is a constant that selects the trade-off between the congestion and utility, and the rest of the equation is the drift of the Lyapunov function [35].

Approximating the derivation in (11), we arrive at

$$\begin{aligned}\min \sum_{i=1}^{N} & U(a_{i1}, a_{i2}, a_{i3}) \\ &- \beta\Big[\sum_{j=1}^{N} q_{j1}(x_{j1} - S_{j1}a_{j1}) + q_{j2}(x_{j2} - S_{j2}a_{j2}) \\ &\quad + q_{j3}(x_{j3} - S_{j3}a_{j3}) + q_j(x_j - x_{j1} - x_{j2} - x_{j3})\Big],\end{aligned} \quad (12)$$

subject to

$$\begin{aligned}\sum_{i=1}^{N} a_{ij} &= C_j,\ j = 1..3, \\ \sum_{j=1}^{3} x_{ij} &\leq x_i,\ i = 1..N, \\ x_{ij} &\leq S_{ij}a_{ij},\ i = 1..N, j = 1..3,\end{aligned} \quad (13)$$

thus enforcing the allocations based on the user backlogs. The α-fairness enforcement algorithm in (12) does not modify $U(a_{i1}, a_{i2}, a_{i3})$. Instead, only the radio link control (RLC) offloading rates x_{jk}, $i = 1, 2, ;N$, $k = 1..3$ are managed, while all of the other variables are obtained from different subsystems, including the MAC layer schedulers

(S_{jk}, a_{jk}, C_k), the PDCP layer (x_j, q_j), and the RLC layer (q_{jk}). Every LTE frame (10ms), the decision module located in the MeNB updates these values and solves (12), thus providing x_{jk}. Once the task at hand is solved, the algorithm redistributes the PDCP packets into the RLC queues, according to new x_{jk} values. Finally, the packets are forwarded to lower layers and distributed according to a particular scheduling discipline at each RAT node (e.g., round robin (RR) for the LTE scheduling system and random access (RA) with the theoretical model in [51] for WiFi). Currently, scheduling is modeled with the RR-like bandwidth division, ($Sik * a_{ik}$), which corresponds to the RR, but not to WiFi RA. In this context, the utility function $U(a_{i1}, a_{i2}, a_{i3})$ plays a role of the input values modifier, similar to the previously considered α-fairness. Different options for representing the utility function may improve network fairness or capacity without increasing the complexity of the optimization task.

VII. CONCLUSION

In this work, we addressed the concept of a multi-radio heterogeneous network, which is expected to be the mainstream architecture for emerging 5G wireless systems. To this end, we have applied network optimization theory to assess a range of effective control techniques for intelligent resource allocation in such networks. A key focus of our present study was to address the appropriate balance between the overall system throughput and the fairness of user resource allocations. Our proposed framework is intended to be illustrative of the general set of problems that could be tackled by the offered optimization techniques. We emphasize that due to the linear programming approach of the underlying mathematical formulation, our solution scales well with increasing number of users and infrastructure nodes, even for the most challenging ultra-dense deployments. In addition, the proposed methodology has the potential to consider an arbitrary number of tiers in the system, as well as could be extended in the following important directions: finite traffic demands, alternative objective functions, uplink vs. downlink optimization, and non-bifurcated flows, among others.

The most important features of our proposed solution are the following:

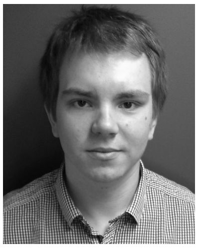
- it may react quickly to the network loading fluctuations, as well as to the changes in the channel properties and/or user densities by adjusting the optimal traffic splitting over the available radio interfaces;
- it can enable efficient resource management for users based on their different traffic patterns;
- it may provide the desired trade-off in terms of the fairness-to-rate ratio, thus potentially offering various value-added features for the network operators (e.g., competitive pricing schemes).

In addition to the above, we demonstrated practical extensions of our model by applying a backpressure algorithm. It should be noted that one of our following steps is the implementation of the proposed backpressure-based model into the 5G-ready SLS tool.

REFERENCES

[1] R. Baldemair et al., "Evolving wireless communications: Addressing the challenges and expectations of the future," *IEEE Veh. Technol. Mag.*, vol. 8, no. 1, pp. 24–30, Mar. 2013.
[2] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu, "5G network capacity: Key elements and technologies," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 71–78, Mar. 2014.
[3] S. Andreev et al., "Intelligent access network selection in converged multi-radio heterogeneous networks," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 86–96, Dec. 2014.
[4] J. G. Andrews et al., "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
[5] J. G. Andrews, "Seven ways that HetNets are a cellular paradigm shift," *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 136–144, Mar. 2013. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6476878
[6] N. Bhushan et al., "Network densification: The dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82–89, Feb. 2014.
[7] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 497–508, Apr. 2012.
[8] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in heterogeneous networks: Modeling, analysis, and design insights," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2484–2497, May 2013.
[9] M. Bennis, M. Simsek, A. Czylwik, W. Saad, S. Valentin, and M. Debbah, "When cellular meets WiFi in wireless small cell networks," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 44–50, Jun. 2013.
[10] D. Astely, E. Dahlman, G. Fodor, S. Parkvall, and J. Sachs, "LTE release 12 and beyond," *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 154–160, Jul. 2013.
[11] J. Andrews, S. Singh, Q. Ye, X. Lin, and H. Dhillon, "An overview of load balancing in HetNets: Old myths and open problems," *IEEE Wireless. Commun.*, vol. 21, no. 2, pp. 18–25, Apr. 2014.
[12] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, "Networks and devices for the 5G era," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 90–96, Feb. 2014.
[13] T. Shuminoski and T. Janevski, "Radio network aggregation for 5G mobile terminals in heterogeneous wireless and mobile networks," *Wireless Pers. Commun.*, vol. 78, no. 2, pp. 1211–1229, Sep. 2014. [Online]. Available: http://link.springer.com/10.1007/s11277-014-1813-0
[14] "Project RAN evolution: Multi-RAT joint radio operation (MRJRO)," NGMN alliance, Reading, MA, USA, Tech. Rep., Mar. 2015.
[15] S. C. Liew and Y. J. Zhang, "Proportional fairness in multi-channel multi-rate wireless networks-Part I: The case of deterministic channels with application to AP association problem in large-scale WLAN," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3446–3456, Sep. 2008.
[16] H. S. Dhillon and J. G. Andrews, "Downlink rate distribution in heterogeneous cellular networks under generalized cell selection," *IEEE Wireless Commun. Lett.*, vol. 3, no. 1, pp. 42–45, 2014.
[17] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of K-tier downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 550–560, Apr. 2012.
[18] A. Damnjanovic et al., "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 10–21, Jun. 2011.
[19] E. Yaacoub and Z. Dawy, "A survey on uplink resource allocation in OFDMA wireless networks," *IEEE Commun. Surv. Tuts.*, vol. 14, no. 2, pp. 322–337, Second 2012.
[20] H. Dai, Y. Huang, and L. Yang, "Game theoretic max-logit learning approaches for joint base station selection and resource allocation in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1068–1081, Jun. 2015.
[21] C. Yang, J. Li, A. Anpalagan, and M. Guizani, "Joint power coordination for spectral-and-energy efficiency in heterogeneous small cell networks: A bargaining game-theoretic perspective," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1364–1376, Feb. 2016.
[22] L. Liang, G. Feng, and Y. Jia, "Game-theoretic hierarchical resource allocation for heterogeneous relay networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1480–1492, Apr. 2015.
[23] 3GPP, "Architecture enhancements for non-3GPP accesses," 3GPP Mobile Competence Centre c/o ETSI, Sophia Antipolis Cedex, France, Tech. Specification TS 23.402, 2015

[24] 3GPP, "Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description; stage 2 (Release 13)," 3GPP Mobile Competence Centre c/o ETSI, Sophia Antipolis Cedex, France, Tech. Specification TS 36.300, 2016.
[25] W. H. Chin, Z. Fan, and R. Haines, "Emerging technologies and research challenges for 5G wireless networks," *IEEE Wireless Commun.*, vol. 21, no. 2, pp. 106–112, Apr. 2014.
[26] NGMN alliance, "NGMN alliance," Reading, MA, USA, 5G White Paper, Feb. 2015.
[27] M. Peng, Y. Li, J. Jiang, J. Li, and C. Wang, "Heterogeneous cloud radio access networks: A new perspective for enhancing spectral and energy efficiencies," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 126–135, Dec. 2014.
[28] C. Ran, S. Wang, and C. Wang, "Balancing backhaul load in heterogeneous cloud radio access networks," *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 42–48, Jun. 2015.
[29] *Offloading Traffic From LTE to WiFi—Quick Start With CMW500*. accessed on Jul. 8, 2014. [Online]. Available: http://winter-group.net/rohde-schwarz-tutorial/
[30] D. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1992.
[31] M. Pioro and D. Medhi, *Routing, Flow, and Capacity Design in Communication and Computer Networks* (The Morgan Kaufmann Series in Networking), 1st ed., 2004.
[32] I. Maros, *Computational Techniques of the Simplex Method*. Springer, 2003.
[33] L. Massoulie and J. Roberts, "Bandwidth sharing: Objectives and algorithms," *IEEE/ACM Trans. Netw.*, vol. 10, no. 3, pp. 320–328, Jun. 2002.
[34] F. P. Kelly, A. K. Malullo, and D. K. H. Tan, "Rate control in communication networks: Shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, pp. 237–252, Mar. 1998.
[35] J. Walrand and A. K. Parekh, "Congestion control, routing and scheduling in communication networks: A tutorial," *IEICE Trans. Commun.*, vol. 96, no. 11, pp. 2714–2723, 2013.
[36] M. J. Neely, E. Modiano, and C.-P. Li, "Fairness and optimal stochastic control for heterogeneous networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 2, pp. 396–409, Apr. 2008.
[37] E. B. Rodrigues and F. Casadevall, "Control of the trade-off between resource efficiency and user fairness in wireless networks using utility-based adaptive resource allocation," *IEEE Commun. Mag.*, vol. 49, no. 9, pp. 90–98, Sep. 2011.
[38] G. Fodor, G. Malicskó, M. Pióro, and T. Szymanski, "Path optimization for elastic traffic under fairness constraints," *Teletraffic Sci. Eng.*, vol. 4, pp. 667–680, Dec. 2001.
[39] 3GPP, "Study on WLAN/3GPP radio interworking," 3GPP Mobile Competence Centre c/o ETSI, Sophia Antipolis Cedex, France, Tech. Rep. TR 37.834, 2013.
[40] 3GPP, "Further advancements for E-UTRA physical layer aspects," 3GPP Mobile Competence Centre c/o ETSI, Sophia Antipolis Cedex, France,, Tech. Rep. TR 36.814, 2010.
[41] 3GPP, "Coordinated multi-point operation for LTE physical layer aspects," 3GPP, Tech. Rep. TR 36.819, 2011.
[42] 3GPP, "Guidelines for evaluation of radio interface technologies for IMT-Advanced," 3GPP Mobile Competence Centre c/o ETSI, Sophia Antipolis Cedex, France, Tech. Rep. ITU-R M.2135, 2009.
[43] M. Gerasimenko, N. Himayat, S.-P. Yeh, S. Talwar, S. Andreev, and Y. Koucheryavy, "Characterizing performance of load-aware network selection in multi-radio (WiFi/LTE) heterogeneous networks," in *Proc. Globecom Workshops*, Dec. 2013, pp. 397–402.
[44] N. Himayat *et al.*, "Multi-radio heterogeneous networks: Architectures and performance," in *Proc. IEEE ICNC*, Feb. 2013, pp. 252–258.
[45] (2016). *WINTERsim System-Level Simulator*. [Online]. Available: http://winter-group.net/downloads/
[46] E. Hossain, M. Rasti, H. Tabassum, and A. Abdelnasser, "Evolution toward 5G multi-tier cellular wireless networks: An interference management perspective," *IEEE Wireless Commun. Mag.*, vol. 21, no. 3, pp. 118–127, Jun. 2014. [Online]. Available: http://arxiv.org/abs/1401.5530
[47] 3GPP, "LTE aggregation & unlicensed spectrum," 3GPP Mobile Competence Centre c/o ETSI, Sophia Antipolis Cedex, France, 4G Americas, White paper, Nov. 2015.
[48] M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synthesis Lectures Commun. Netw.*, vol. 3, no. 1, pp. 1–211, 2010.
[49] M. J. Neely and R. Urgaonkar, "Optimal backpressure routing for wireless networks with multi-receiver diversity," *Ad Hoc Netw.*, vol. 7, no. 5, pp. 862–881, Jul. 2009.
[50] M. J. Neely, "Energy optimal control for time-varying wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 2915–2934, Jul. 2006.
[51] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.



MIKHAIL GERASIMENKO received the Specialist degree from the Saint-Petersburg University of Telecommunications in 2011 and the M.Sc. degree from the Tampere University of Technology in 2013. He started his academic career in 2011. He is currently a Researcher with the Department of Electronics and Communications Engineering, Tampere University of Technology. He has co-authored multiple scientific journal and conference publications. He holds several patents. His main subjects of interest are wireless communications, machine-type communications, and heterogeneous networks. He acted as a Reviewer and participated in educational activities.

DMITRI MOLTCHANOV received the M.Sc. and Cand.Sc. degrees from the Saint-Petersburg State University of Telecommunications, Russia, in 2000 and 2002, respectively, and the Ph.D. degree from the Tampere University of Technology, Finland, in 2006. He is currently a Senior Research Scientist with the Department of Electronics and Communications Engineering, Tampere University of Technology. He authored over 80 publications. His research interests include performance evaluation and optimization issues of wired and wireless IP networks, Internet traffic dynamics, quality of user experience of real-time applications, and traffic localization P2P networks. He serves as TPC Member in a number of international conferences.

SERGEY ANDREEV (M'12) received the Specialist degree in information security and the Cand.Sc. degree in wireless communications from the Saint-Petersburg State University of Aerospace Instrumentation, Saint Petersburg, Russia, in 2006 and 2009, respectively, and the Ph.D. degree in technology from the Tampere University of Technology, Tampere, Finland, in 2012. He is currently a Senior Research Scientist with the Department of Electronics and Communications Engineering, Tampere University of Technology. He has co-authored over 100 published research works. His current research interests include wireless communications, energy efficiency, heterogeneous networking, cooperative communications, and machine-to-machine applications.

YEVGENI KOUCHERYAVY (SM'09) received the Ph.D. degree from the Tampere University of Technology (TUT), Finland, in 2004. He is currently a Full Professor and a Lab Director with the Department of Electronics and Communications Engineering, TUT. He has authored numerous publications in the field of advanced wired and wireless networking and communications. His current research interests include various aspects of heterogeneous wireless communication networks and systems, the Internet of Things and its standardization, and nanocommunications. He is an Associate Technical Editor of the *IEEE Communications Magazine* and an Editor of the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS.

NAGEEN HIMAYAT (M'88) received the B.S.E.E. degree from Rice University, the Ph.D. degree from the University of Pennsylvania, and the MBA degree from the Haas School of Business, University of California at Berkeley, Berkeley. She was with Lucent Technologies and General Instrument Corporation, where she developed standards and systems for both wireless and wire-line broadband access networks. She is currently a Principal Engineer with Intel Labs, where she leads a team conducting research on several aspects of next generation (5G/5G+) of mobile broadband systems. She has authored over 250 technical publications, contributing to several IEEE peer-reviewed publications, 3GPP/IEEE standards, and numerous patents (28 granted, 55 pending). Her research contributions span areas such as multi-radio heterogeneous networks, mm-wave communication, energy-efficient designs, cross layer radio resource management, multi-antenna, and non-linear signal processing techniques.

SHU-PING YEH (M'06) received the B.S. degree from National Taiwan University in 2003, and the M.S. and Ph.D. degrees from Stanford University in 2005 and 2010, respectively, all in electrical engineering. She is currently a Research Scientist with Intel Labs. She specializes in wireless technology development for cellular and local area network access. Her recent research focus includes advanced self-interference cancellation technology, full-duplex PHY/MAC system designs, and multi-tier multi-RAT heterogeneous networks.

SHILPA TALWAR (M'96) received the M.S. degree in electrical engineering and the Ph.D. degree in applied mathematics from Stanford University in 1996. She held several senior technical positions in wireless industry involved in a wide-range of projects, including algorithm design for 3G/4G & WLAN chips, satellite communications, GPS, and others. She is currently a Director of Wireless Multicomm Systems and a Senior Principal Architect with the Wireless Communications Laboratory, Intel, where she leads a research team focused on advancements in network architecture and technology innovations for 5G and contributed to the IEEE and 3GPP standard bodies, including 802.16m and LTE-advanced. She is also coordinating several university collaborations on 5G, and leads Intel Strategic Research Alliance on 5G. She has authored over 50+ technical publications. She was co-edited a book on 5G. She holds 33 patents (42 additional pending). Her research interests include heterogeneous networks, multi-radio interworking, mm wave communications, advanced MIMO, and full-duplex and interference mitigation techniques. She was the Co-Chair of ICC 2014 workshop on 5G Technologies and served as a Co-Editor of a Special Issue on The 5G Revolution for the IEEE Signal Processing Journal, in 2014.