INTELLIGENT ACCESS NETWORK SELECTION IN CONVERGED MULTI-RADIO HETEROGENEOUS NETWORKS

SERGEY ANDREEV, MIKHAIL GERASIMENKO, OLGA GALININA, YEVGENI KOUCHERYAVY, NAGEEN HIMAYAT, SHU-PING YEH, AND SHILPA TALWAR

ABSTRACT

Heterogeneous multi-radio networks are emerging network architectures that comprise hierarchical deployments of increasingly smaller cells. In these deployments, each user device may employ multiple radio access technologies to communicate with network infrastructure. With the growing numbers of such multi-radio consumer devices, mobile network operators seek to leverage spectrum across diverse radio technologies, thus boosting capacity and enhancing quality of service. In this article, we review major challenges in delivering uniform connectivity and service experience to converged multi-radio heterogeneous deployments. We envision that multiple radios and associated device/infrastructure intelligence for their efficient use will become a fundamental characteristic of future 5G technologies, where the distributed unlicensed-band network (e.g., WiFi) may take advantage of the centralized control function residing in the cellular network (e.g., 3GPP LTE). Illustrating several available architectural choices for integrating WiFi and LTE networks, we specifically focus on interworking within the *radio access network* and detail feasible options for intelligent access network selection. Both network- and user-centric approaches are considered, wherein the control rests with the network or the user. In particular, our system-level simulation results indicate that load-aware user-centric schemes, which augment SNR measurements with additional information about network loading, could improve the performance of conventional WiFi-preferred solutions based on minimum SNR threshold. Comparison with more advanced network-controlled schemes has also been completed to confirm attractive practical benefits of distributed user-centric algorithms. Building on extensive system-wide simulation data, we also propose novel analytical space-time methodology for assisted network selection capturing user traffic dynamics together with spatial randomness of multi-radio heterogeneous networks.

Sergey Andreev, Mikhail Gerasimenko, Olga Galinina, and Yevgeni Koucheryavy, are with Tampere University of Technology.

Nageen Himayat, Shu-ping Yeh, and Shilpa Talwar are with Intel Corporation.

RECENT ADVANCES IN MULTI-RADIO NETWORKING

The rapid expansion of wireless communications over the last decades has introduced fundamental changes to "anytime, anywhere" mobile Internet access, as well as posed new challenges for the research community. In 2011, the fourth generation of broadband communication standards was completed and offered aggressive improvements in all aspects of wireless system design, including system capacity, energy efficiency, and user quality of service (QoS). As the respective technologies are being deployed today, the focus of recent research efforts is shifting to what may be referred to as fifth generation (5G) wireless networks.

Given a historical 10-year cycle for every existing generation, it is expected that 5G systems will be deployed sometime around 2020. Whereas there is currently no complete technical definition of what comes after the state-of-the-art networking technology, the anticipated communication requirements may already be understood from the user perspective. Regardless of their current location, human users would like to be connected at all times, taking advantage of the rich set of services provided by contemporary multimedia-over-wireless networks. This creates significant challenges for 5G technology design, as users' connectivity experience should match data rate requirements and be uniform no matter where the user is, to whom the user connects, and what the user's service needs are [1].

Unfortunately, contemporary wireless networks are currently unable to deliver the desired ubiquitous connectivity experience. In the first place, they lack uniformity in data rates and suffer from excessive time delays, or sometimes even service outage due to poor coverage and severe interference conditions. While current technologies have indeed been helpful in coping with some of these challenges, it is commonly believed that they will still be insufficient to meet the anticipated growth in traffic demand

 1536-1284/14/$25.00 © 2014 IEEE

(nearly 11-fold over the following 5 years [2]) aggravated by rapid proliferation in types and numbers of wireless devices. To make matters worse, billions of diverse machine-type devices connect to the network, thus reshaping the Internet as we know it today. All these technological challenges accentuate the need to explore novel solutions within the context of 5G networks.

Major Trends behind 5G Technology

A transformation of mobile user experience requires revolutionary changes in both network infrastructure and device architecture, where the user equipment (UE) is jointly optimized with the surrounding network context [3]. Many believe that the only feasible solution to mitigate the increasing disproportion between the desired QoS and the limited wireless resources is by deploying higher densities of femto- and picocells in current cellular architecture. Due to shorter radio links, smaller cells provide higher data rates and require less energy for uplink transmission, especially in urban environments.

However, introducing an increasing number of serving stations to bridge the capacity gap incurs extra complexity due to more cumbersome interference management, higher rental fees, and increased infrastructure maintenance costs [4]. More importantly, licensed spectrum continues to be scarce and expensive, whereas the traditional methods to improve its efficient use are approaching their theoretical limits. Even when additional spectrum is allocated, these new frequencies are likely to remain fragmented and could require diverse transmission techniques. Consequently, there is a pressing demand to leverage additional capacity across multiple radio access technologies (RATs).

As a result, it becomes crucial to aggregate different radio technologies as part of a common *converged* radio network, in a manner transparent to the end user, and develop techniques that can efficiently utilize the radio resources available across different spectral bands potentially using various RATs [5]. In particular, we expect that the majority of immediate gains will come from advanced architectures and protocols that would leverage the unlicensed spectrum. For example, mobile users with direct device-to-device communication capability may take advantage of their unlicensed-band radios and cooperate with other proximate users to locally improve access in a cost-efficient way [6].

Furthermore, as cell sizes shrink, the footprints of cellular, local, and personal area networks are increasingly overlapping. This creates an attractive opportunity to *simultaneously* utilize multiple RATs for improved wireless connectivity. We thus believe that intelligent multi-RAT coupling will efficiently leverage performance benefits across several dimensions of diversity, including spatial, temporal, frequency, interference, load, and others. In future 5G networks, both short- and long-range technologies may need to work cooperatively and exploit the intricate interactions between the device and the network, as well as between the devices themselves, to realize the desired uniform user experience [7].

Consequently, the incentive to efficiently coordinate between the alternative RATs is growing stronger, and we envision that multiple radios together with the associated device/system intelligence for their efficient use will become a fundamental characteristic of next-generation networks [8]. More specifically, the distributed unlicensed-band network (e.g., wireless local area network, WLAN) may take advantage of the *centralized* control function residing in the cellular network to effectively perform dynamic multi-RAT network association, and hence provide beyond-additive gains in network capacity and user connectivity experience.

Scope and Core Novelty of Current Research

According to the above, there is currently an increasing shift toward tighter interworking between different RATs. To this end, our research campaign is targeting joint RAT assignment, selection, and scheduling algorithms, which provide significant improvement in overall system performance. In what follows, our focus is set on integration between multiple RATs within heterogeneous network architecture. As our case study, we consider convergence of WLAN-based small cells with operator-managed cellular deployment to illustrate feasible architectural options for integration and their associated performance benefits. Consequently, we seek to explore the potential of a diverse range of devices requiring connectivity at different scales to augment system capacity and improve user connectivity experience.

We emphasize that interworking between WLAN and cellular networks has already been considered in the past, but largely from the perspective of internetwork (vertical) handoff [9]. The cellular standards community, represented by the Third Generation Partnership Project (3GPP), has also been involved in developing specifications that address cellular/WLAN interworking for a number of years. Several new study and work items have recently emerged to develop specifications toward tighter integration of WLAN with cellular networks. The areas of investigation range from solutions for trusted access to 3GPP services with WLAN devices, seamless mobility between 3GPP and WLAN technologies, and support for the access network discovery and selection function (ANDSF). While much of this effort has focused on loose interworking solutions only requiring changes within the core network, there has been a recent shift in 3GPP Release 12 to address interworking within the radio access network (RAN) [10].

This emerging trend is driven by the need to support better QoS on unlicensed spectrum as demanded by a consortium of network operators who have introduced stringent requirements for carrier-grade WiFi. The WLAN community has also responded with new initiatives such as Hot Spot 2.0, as well as a novel high-efficiency WLAN standardization effort by the IEEE 802.11 working group. Hence, it is timely to investigate RAN-based integration solutions, which assume increased cooperation between 3GPP Long Term Evolution (LTE) and WiFi radio technologies. Along these lines, our work details several intelligent network selection

Even when additional spectrum is allocated, these new frequencies are likely to remain fragmented and could require diverse transmission techniques. Consequently, there is a pressing demand to leverage additional capacity across multiple radio access technologies.

Whereas the trend toward the use of WLAN in conjunction with cellular networks has emerged from the near-term need of operators to relieve congestion on cellular networks, the use of WiFi is expected to remain an integral part of operators' long-term strategy to address future capacity needs.

mechanisms that deliver significant gains in overall system performance and user QoS. We address both network- and user-centric approaches, wherein the control of how different radio technologies are utilized rests with the network or the user, respectively.

Enabling Architectures and Algorithms for Converged Heterogeneous Networks

As argued previously, the capacity and connectivity limitations faced by future 5G networks will continue to drive the need for closer integration across different RATs. Along these lines, Fig. 1a illustrates our vision of an operator's multi-RAT heterogeneous network (HetNet). It features a hierarchical deployment of wide-area macrocells for ubiquitous coverage, connectivity, and seamless mobility augmented with an overlay tier of inexpensive low-power smaller cells (picos, femtos, WiFi access points, integrated WiFi-LTE modules, etc.) to enhance capacity by moving infrastructure closer to the users in areas with higher traffic demand.

Whereas the trend toward the use of WLAN in conjunction with cellular networks has emerged from the near-term need of operators to relieve congestion in cellular networks, the use of WiFi is expected to remain an integral part of operators' long-term strategy to address future capacity needs. In the simplest case, no cooperation between WiFi and cellular RANs is available, and users are left to determine how the two RATs are utilized. However, when WiFi

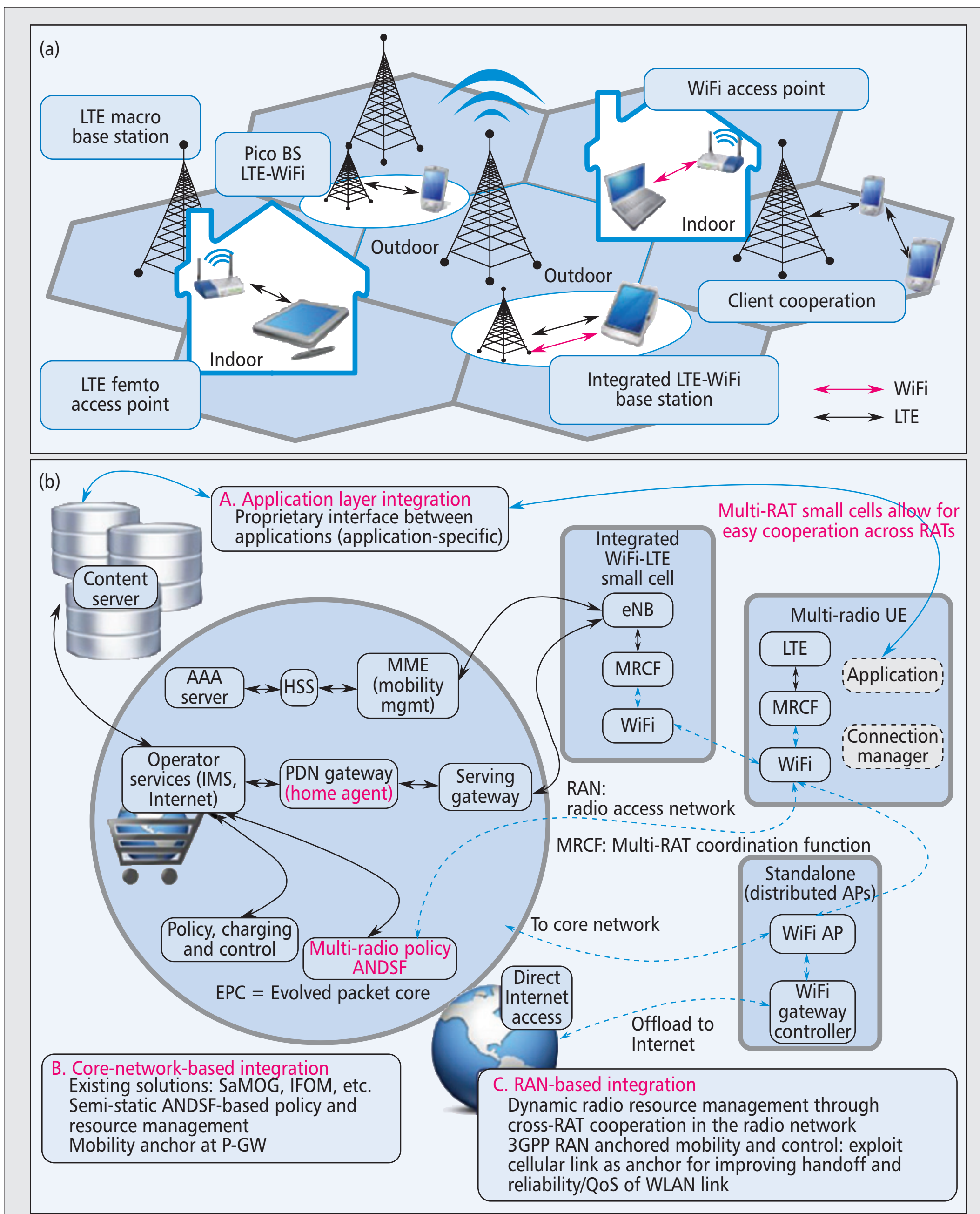


Figure 1. Topology and architecture of a converged heterogeneous network.

is managed as part of an operator's RAN, an increased level of cooperation between WLAN and 3GPP infrastructures may become feasible.

For instance, one may envisage an architecture where integrated WiFi-LTE small cells enable full cooperation between the two RATs, allowing for WiFi to simply become a "virtual carrier" anchored on the 3GPP radio network. We note that multi-RAT small cells with co-located WiFi and 3GPP interfaces are an emerging industry trend for lowering deployment costs by leveraging common infrastructure across multiple RATs. However, given that such deployments are presently not common, current standardization efforts aim to improve UE-centric interworking architectures while assuming only limited cooperation or assistance across a multi-RAT network.

Options for Integrating WiFi with 3GPP LTE

We continue by illustrating various architectural choices for integrating WiFi and LTE networks in Fig. 1b. These generally offer different mechanisms to implement important operations required for multi-RAT integration, including RAT discovery, RAT selection or assignment, control of multi-RAT radio resource management (RRM), protocols for inter-RAT mobility or session transfers, and so on.

Application Layer Integration — In Fig. 1b, *case A* corresponds to the application or higher-layer integration architecture. Accordingly, there is a proprietary or higher-layer interface allowing the UE and the content server to communicate directly by exchanging information over multiple RATs. As no coordination at the network layer is involved, such solutions are typically simple and have already been explored in the context of improving over-the-top applications. This choice of architecture is beneficial for boosting user quality of experience (QoE), but it remains largely application-dependent and may not fully account for underlying network conditions, especially when such conditions vary dynamically.

Core-Network-Based Integration — Furthermore, *case B* summarizes recent solutions proposed by 3GPP for cellular/WLAN integration based on interworking within the core network. Accordingly, ANDSF assists in discovery of WiFi access points and may also specify policies for network selection, but the overall network selection decision remains in control of the UE. Therefore, it can combine the local radio link state information, operator policies, and user preferences to make a decision that improves user QoE.

There are a number of benefits with this integration option, as it can more adequately account for both operator policies and user preferences. However, the performance of corresponding control procedures may still be rather limited. This is due to the fact that the UE may only have local knowledge of the network conditions and is thus likely to make greedy decisions, ultimately hurting overall system performance. Whereas the UE can be made to report its perceived radio link state to the core network, such information exchange cannot be updated dynamically due to prohibitive levels of associated signaling overhead. Hence, when wireless channel conditions change dynamically, local RRM directly on the RAN layer may deliver higher QoS. Therefore, advanced architectures allowing multi-RAT integration within the RAN are of increasing interest today, as they employ network-wide knowledge of radio link conditions.

RAN-Based Integration — Finally, *case C* details the emerging RAN-based 3GPP/WLAN integration architecture. Here, UE assistance may facilitate information exchange between cellular and WLAN infrastructures; for that matter, a dedicated interface may be introduced. The available levels of cooperation within the RAN are constrained by the capacity of the intercell/inter-RAT backhaul links. When high-capacity backhaul or integrated multi-RAT small cells are available, full cooperation across multiple RATs may become available, thus enabling more dynamic RRM for improved system and user performance.

In addition, the cellular RAT may be employed as a mobility and control anchor: a user thus utilizes 3GPP protocols for transferring sessions to multi-radio small cells and then uses local switching to steer sessions to/from WLAN with low latency. The benefits of this solution are obvious, as adaptations to dynamic variations in interference conditions can easily be performed without undesired session interruptions and packet drops. Furthermore, user and operator preferences may be accounted for through appropriate feedback by the UE or via a suitable configuration of the RAN by the operators.

In summary, the degrees of cooperation within the RAN can range from exploiting simple assistance information (e.g., network loading) by the radio network to tight coupling and joint/centralized RAN-based RRM. In what follows, we describe the various levels of cross-RAT cooperation options across a multi-RAT HetNet and then characterize the associated performance benefits. We pay particular attention to the more practical case when only limited assistance across multi-RAT network is available to users, in contrast to significantly more complex network-controlled approaches requiring higher signaling and computation overheads.

Algorithms for Radio Resource Management

In what follows, we detail various options for utilizing and managing multi-RAT radio resources available in the network. Both user- and network-controlled (or assisted) RRM may be considered for the range of architectural options described above. For application- or core-network-based integration (*options A and B*), only UE-based RRM schemes may be feasible. A richer set of choices is available for RAN-based multi-RAT integration (*option C*), which depends on the degree of inter-RAT cooperation achieved with different RAN topologies.

Generally, RAN can play a major role in multi-RAT resource management across the HetNet. Even if RAN does not directly control the RRM decisions, it may provide optimized

We pay particular attention to the more practical case when only limited assistance across a multi-RAT network is available to users, in contrast to significantly more complex network-controlled approaches requiring higher signaling and computation overheads.

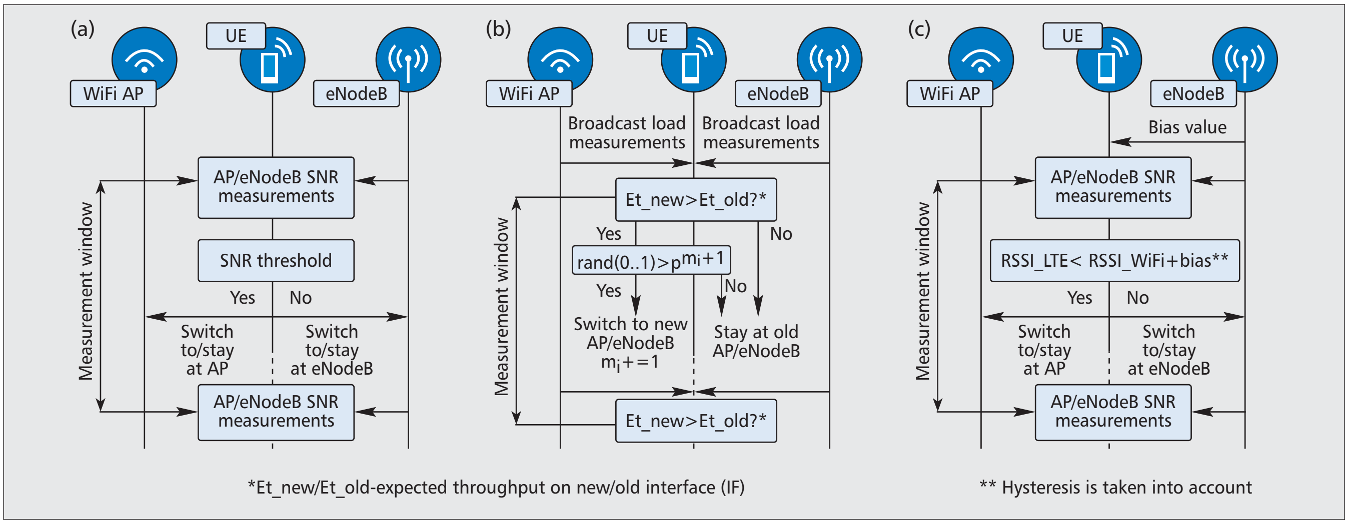


Figure 2. Alternative network selection algorithms for HetNets.

network assistance to enable better decisions by the UE. In *virtual* RAN architectures, where the mobility and control anchor is moved from the core network to the RAN, more dynamic RRM with fast session transfers between RATs (dynamic switching) may become feasible. For integrated multi-RAT small cells or where the delay between the interfaces is negligible, tighter cooperation involving joint RAT scheduling may also be enabled.

We continue by introducing specific RRM schemes that are investigated in our research. They range from typical implementations used by UEs today, where the UE always prefers to connect to the less expensive WiFi network if it is available (WiFi-preferred), to more intelligent cross-RAT access network selection for converged HetNets.

User-Centric Approaches — The simplest threshold-based algorithm serves as our baseline user-centric network selection scheme. With this solution, a UE is continuously monitoring the signaling messages from the neighboring WiFi access points (APs) to obtain timely signal-to-noise ratio (SNR) information. When a particular SNR value exceeds a predefined threshold (which we set equal to 40 dB as discussed in 3GPP), the user starts steering its traffic to the respective WiFi AP. Otherwise, it keeps transmitting on the LTE network (Fig. 2a).

Naturally, such behavior is an automatized version of what a human user would do: whenever a hotspot with a reliable signal is available, UEs switch to WiFi to enjoy higher data rates and reduce expenses associated with paid cellular traffic. Alternative user-centric algorithms include schemes based on preferring WiFi if certain minimum performance (coverage, QoS, etc.) is available, as well as solutions where the UE is able to transmit on both RATs without any intelligent coordination across them.

RAN-Assisted Approaches — Due to its simplicity, the baseline WiFi-preferred (SNR-threshold) scheme may experience limitations in dense interference-limited scenarios typical of modern urban deployments. For instance, a hotspot AP may experience overload conditions when a significant number of users try to steer their traffic through it. Moreover, nomadic WiFi users, such as those with laptops, could consume most of the WLAN capacity. To make matters worse, the WiFi medium access is contention-based, which results in nonlinear degradation of the throughput performance with increasing number of users.

Therefore, the load-agnostic SNR-threshold scheme is not expected to remain effective in environments with varying load. In such situations, UEs may attempt to combine SNR knowledge with additional knowledge of the loading information from the network infrastructure (cellular/WLAN). While accounting for WiFi load would certainly improve performance beyond the SNR-threshold scheme, it is easy to envision scenarios where accounting for WiFi load only will be insufficient. Hence, we focus our further investigation on schemes that account for both LTE and WiFi loading, and compare them with existing network-based schemes that have been standardized in 3GPP for small cell offload. Our proposed load-aware scheme works as follows (see simplified time diagram in Fig. 2b).

Throughput estimation: A user attempts to listen on both interfaces in order to monitor the SNR information in its neighborhood and estimate its expected throughput. For WiFi, such estimation is conducted based on predicted network capacity divided by number of UEs connected to a particular AP (as advertised by the AP through the load indicators in the beacon frames) as well as accounting for several weighting factors (SNR, contention, etc.). The motivation behind the SNR weighting is to exclude APs with low signal quality. Another coefficient may account for the contention-based nature of WiFi channel access and include signaling overheads as well as collision losses. For LTE, throughput prediction may simply be built on the scheduler advertisements by a base station (BS or eNodeB) and the used power control.

Randomization: A user may select the network with the highest expected throughput value probabilistically, $rand(0..1) < p^{m_i+1}$, where m_i is the number of recent connections to this AP/BS and p is the number in (0, 1) that represents the reconnection probability. The proper use of p reduces the number of concurrent reconnections to the same AP/BS, which will prevent uncontrollable hopping from one interface to another. If a network reselection occurs, m_i is incremented for AP/BS i. Other users take this information into account by dividing their expected throughput value for this AP/BS by $m_i + 1$. This allows for dynamic control of reselections on both networks.

Hysteresis: To additionally decrease the number of cell border switchings, an appropriate hysteresis value should be added to the current expected throughput value.

Filtering throughput estimations: Further improvement in throughput estimates is obtained through averaging. After each measurement window, the actual throughput obtained over this period may be filtered with a moving average filter. The resultant value, which combines the measured and predicted throughput, is then used as the expected throughput value for this AP/BS. This averaging is made to achieve more reliability, which could suffer due to contention-based channel access.

In summary, RAN-assisted approaches employ network assistance from the RAN to improve UE-based RAT selection decisions. Network assistance can be very simple in that the RAN may transmit certain assistance parameters (e.g., network load, utilization, expected resource allocation), but with increased cross-RAT cooperation, RAN assistance may also be improved.

RAN-Controlled Approaches — The above two network selection schemes are user-centric in nature. Hence, they may still result in suboptimal system-wide performance, which may otherwise be improved through network-based centralized mechanisms. Consequently, RAN-controlled approaches place the control of the RRM in the radio network so that the BS can assign the UEs to use certain RATs. Such network control may be distributed across base stations, or may utilize a central RRM entity that manages radio resources across several cells/RATs.

Below we consider the conventional cell-range extension schemes applied in cellular networks to steer users to small cells employing a network-optimized received signal strength indication (RSSI) bias value. We use the RSSI bias to increase/decrease the effective WiFi AP coverage area depending on the network capacity expectations. One limitation of this method is that the optimal bias value needs to be adapted based on network-wide knowledge of user distribution. For example, our results show that the optimal bias depends on the user deployment model as well as the interference levels in the network, which may not always be available as typically WiFi cells may not have a direct interface to a cellular BS. In what follows, we evaluate RSSI-based cell range extension with bias values optimized for the target scenario. We also use hysteresis for the RSSI-based algorithm. The time diagram of this method is shown in Fig. 2c.

More generally, network-controlled schemes may utilize proprietary or standardized interfaces between cells/RATs. Distributed network-controlled schemes have recently been discussed as part of the 3GPP study on WLAN/3GPP RAN interworking. Here, the network establishes certain triggers for UE to report measurement on their local radio environment. The final RAT selection decisions are then made by the 3GPP BS based on UE measurement reports. Other examples of centralized RAN-controlled architecture is the emerging dual connectivity, or "anchor/booster" architecture, where the UE always maintains a control link to the macrocell tier, and the macrocell centrally manages the user offload to smaller cells. Hence, the macrocell can centrally determine the optimal offload mechanisms.

In summary, RAN-assisted approaches employ network assistance from the RAN to improve UE-based RAT selection decisions. Network assistance can be very simple in that RAN may transmit certain assistance parameters, but with increased cross-RAT cooperation RAN assistance may also be improved.

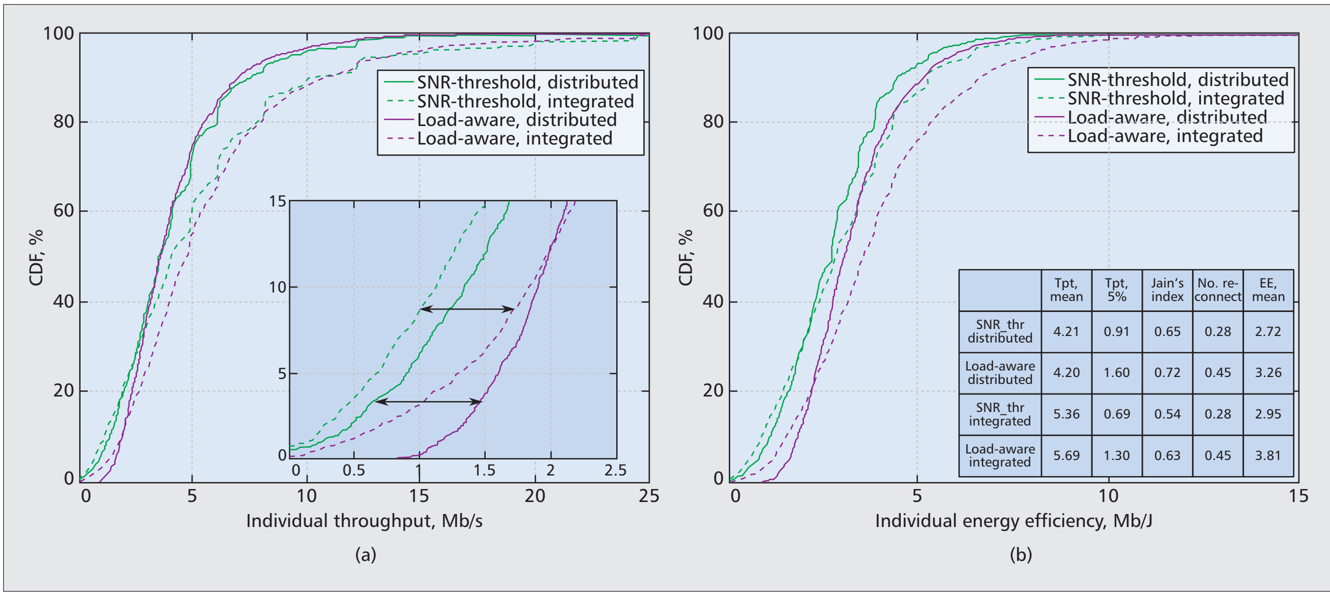
Analyzing Intelligent Access Network Selection

In what follows, we concentrate on the important problem of network selection between LTE and WiFi RATs [11], assuming that WLAN is a part of an operator deployed and managed multi-RAT HetNet. We target feasible practical extensions to improve the performance of UE-centric network selection schemes. To be consistent with current network deployments, we consider distributed small cell overlay with standalone WiFi APs, assuming that there is no interface between the WiFi and 3GPP radio networks [10]. Additionally, we discuss benefits of deploying integrated WiFi-LTE small cells and quantify the respective performance gains.

In particular, we investigate distributed RAT selection schemes that account for network loading information across the LTE and WiFi technologies, and compare them with solutions that only rely on signal strength measurements. We also benchmark the performance of UE-centric RAT selection with optimized network-based load balancing mechanisms. Intuitively, network-centric solutions may seem to offer better performance than UE-based approaches as network-wide radio link information across users can be employed to develop optimum RAT assignment algorithms. However, with distributed architectures assuming no direct cooperation between LTE and WiFi RATs, such solutions may only be implemented through extensive UE feedback, which could result in significant overheads. UE-centric RAT selection may also be preferred as the UE can better account for user preferences and application QoE.

System-Level Evaluation Scenario and Results

In the course of this study, we have developed an advanced system-level simulator (SLS) that mimics a complete LTE-WiFi system deployment compatible with the 3GPP LTE Release 10 and IEEE 802.11-2012 specifications. Presently, neither free nor commercially available simulation tools are readily applicable for evaluating heterogeneous multi-RAT systems, as they lack the necessary features, as well as the scalability to



	Tpt, mean	Tpt, 5%	Jain's index	No. re-connect	EE, mean
SNR_thr distributed	4.21	0.91	0.65	0.28	2.72
Load-aware distributed	4.20	1.60	0.72	0.45	3.26
SNR_thr integrated	5.36	0.69	0.54	0.28	2.95
Load-aware integrated	5.69	1.30	0.63	0.45	3.81

Figure 3. Comparing SNR-threshold (WiFi-preferred) and our load-aware (RAN-assisted) network selection schemes

adequately capture the dependences between the studied variables. In contrast, our SLS is a flexible tool designed to support diverse deployment strategies, traffic models, channel characteristics, and wireless protocols. It comprises several software modules modeling the deployment of wireless infrastructure and user devices, control events related to transmission of signals between several distinct types of transmitters and receivers, abstractions for wireless channels, mechanisms for collecting measurements, statistics for quantifying system performance, and so on.

Below we construct a multi-RAT simulation model representative of an urban deployment, where WiFi small cells are overlaid on top of the 3GPP cellular network. Outdoor deployments are considered based on recommendations in [12]. A brief summary of the parameters is provided in Table 1. Specifically, we consider a loaded (full-buffer) WiFi network with WLAN APs uniformly distributed across the cellular coverage area. Most UEs cluster around the APs, which recreates a hotspot area (airport, restaurant, shopping mall, or university campus) with many bandwidth-hungry users loading the WiFi network. Moreover, around one third of UEs are still deployed uniformly across the cellular network mimicking regular mobile users. While this scenario may not be characteristic of all practical urban conditions, it represents a harmonized 3GPP vision of a characteristic HetNet deployment.

The major expected outcome of leveraging WiFi small cells is efficient offloading of cellular user traffic resulting in significant user benefits. For that reason, our primary metric of interest is the *uplink* UE throughput (in contrast to many existing studies concentrating on downlink performance), which, in turn, determines the overall system capacity. The cumulative distribution function (CDF) of individual user throughput comparing performance between SNR-threshold (WiFi-preferred) and our proposed load-aware (RAN-assisted) scheme is shown in Fig. 3a. The results indicate that the load-aware scheme gives visible benefits at the cell edge (e.g., over 75 percent of improvement is observed in the 5 percent quantile), as well as some improvement in the average throughput for integrated deployments (i.e., with co-located WiFi-LTE interfaces).

Energy efficiency (EE) is also becoming increasingly important for 5G wireless systems due to the limited battery resources of mobile clients [13], and we confirm significant gains in bits per Joule for both distributed (19 percent) and integrated (29 percent) scenarios in Fig. 3b. Furthermore, as QoS may be equally important, we also account for fairness between the users, which indicates how large the deviation between actual user throughput and the cell-average performance is. In terms of fairness, Jain's index (table, Fig. 3b) of the load-aware scheme (0.72/0.63) is also higher than that for the SNR-threshold scheme (0.65/0.54). The stability of UE-centric schemes is another very important aspect of UE-centric RAT selection, as excessive ping-ponging between RATs is undesirable due to the overhead and latency of mobility protocols as well as EE considerations. In Fig. 3b (see table), we additionally report the number of cellular/WLAN reconnections (in number of reconnections per second) and employ hysteresis mechanisms (an optimized 3 dB value has been used in our experiments) to improve performance.

We also account for the performance of an optimized cell range extension (RAN-controlled) scheme based on RSSI bias, where the network-wide optimization is expected to result in improved performance. The main feature of the considered cell range extension scheme is that it increases the effective WiFi/LTE small cell radius with respect to the bias level. This could work well in the scenario with uniformly deployed UEs, but in the clustered case the interference between WiFi users needs to be considered as well, which is what our load-aware scheme does explicitly. To this end, we perform optimization of small cell offloading bias based on network-wide knowledge of user distribution in Fig. 4a.

However, from Fig. 4b we learn that even with a network controlled bias value (the optimal value of 14 dB is chosen), the individual user throughput is very close to that in the load-aware case (and even smaller at the cell edge). In more detail, Fig. 4b (see bar chart) also highlights the average percentage of time spent by users on each interface. It may be seen from our results that the load-aware scheme is effective in utilizing the available WLAN capacity while efficiently balancing capacities across the LTE macro and pico tiers.

Analytical Space-Time Methodology for Converged HetNets

The above performance results addressed loaded multi-RAT HetNets, but such networks may also be substantially underutilized during off-peak hours. Hence, the load on a heterogeneous deployment can vary significantly, and it is crucial to capture network dynamics explicitly when modeling HetNet performance. However, given the associated complexity, dynamic systems have not been studied as broadly as their static counterparts with a fixed set of active users. Consequently, our proposed analytical methodology suggests assessing flow-level network performance enabling user, traffic, and environment dynamics.

Recently, we have made progress along these lines and have results demonstrating that the locations of the network users relative to each other highly impact the overall system performance [14]. Indeed, given that users are not regularly spaced, there may be a high degree of spatial randomness that needs to be captured explicitly. We thus adopt a range of random spatial models where user locations are drawn from a particular realization of a random process. Coupling such topological randomness with system dynamics requires a fundamental difference in characterizing user signal power and interference. Fortunately, the field of stochastic geometry provides us with a rich set of powerful results and analytical tools that can capture the network-wide performance of a random user deployment [15].

More specifically, every data flow in a dynamic network may generally represent a stream of packets corresponding to a new file transfer, web-page browsing, or real-time voice/video session. As an example, consider an isolated cell of a macro network with radius R encompassing a macro BS together with several distributed pico BSs and WLAN APs. All the BSs/APs are capable of serving uplink data from their wireless users concurrently. The considered traffic is characteristic of real-time sessions with some target bit rate. Based on the recent 3GPP specifications, we further assume non-overlapping frequency bands for all three tiers. However, all WLAN/pico links share the frequency bands of their respective tiers and thus interfere, whereas the macro tier may be considered interference-free (with appropriate intercell power control).

To explicitly model topological randomness in our network (Fig. 5a), we employ several stochastic processes and, to this end, adopt a number of simplifications based on a Poisson point process (PPP). The key novelty of this approach is that we consider a *space-time* PPP with the rate function $\Lambda(x, t)$, where $x \in R^2$ is the spatial component, and $t \in R^+$ is the time component. While random network topology is the primary focus of our model, we also couple it with flow-level system dynamics. This involves an appropriate queuing model, where the session arrives and leaves the system after being served (the service time is determined by the random session length). When a new session arrives or a served session leaves the system, the centralized assisting entity in the RAN performs admission and power control on all tiers by deciding whether the session would be admitted to a particular tier or not and/or advising on the users' transmit powers.

Our general system model is illustrated in

Parameter	Value
LTE/WiFi configuration	10 MHz FDD/20 MHz
Macrocell layout	7 cells, 3 sectors each
LTE signaling mode	2 out of 20 special subframes, short CP, 10 ms frame
Inter-site distance (ISD)	500 m
LTE macro antenna configuration	1 × 2 (diversity reception)
UE to eNodeB/pico/AP pathloss	ITU UMa/UMi
eNodeB antenna gain	14 dB
eNodeB/AP/UE maximum power	43/20/(23/20 LTE/WiFi) dBm
LTE power control	Fractional (α = 1.0) [12]
WiFi power/rate control	Max-power/ARQ
UE/eNodeB/AP antenna height	1.5/25/10 m
UE noise figure/feeder loss	9 dB/0 dB
Feedback/control channel errors	None
Traffic model	Full-buffer
Number of UEs/APs	30/4 per macrocell (3 sectors)
AP/UE deployment type	Uniform/clustered (4b in [12])
AP/UE-eNodeB, AP/UE-UE distance	> 75/35 m, 40/10 m
WiFi MPDU	1500 bytes
Modeling time	3 s
Number of trials per experiment	30

Table 1. Important simulation parameters. ITU: International Telecommunication Union.

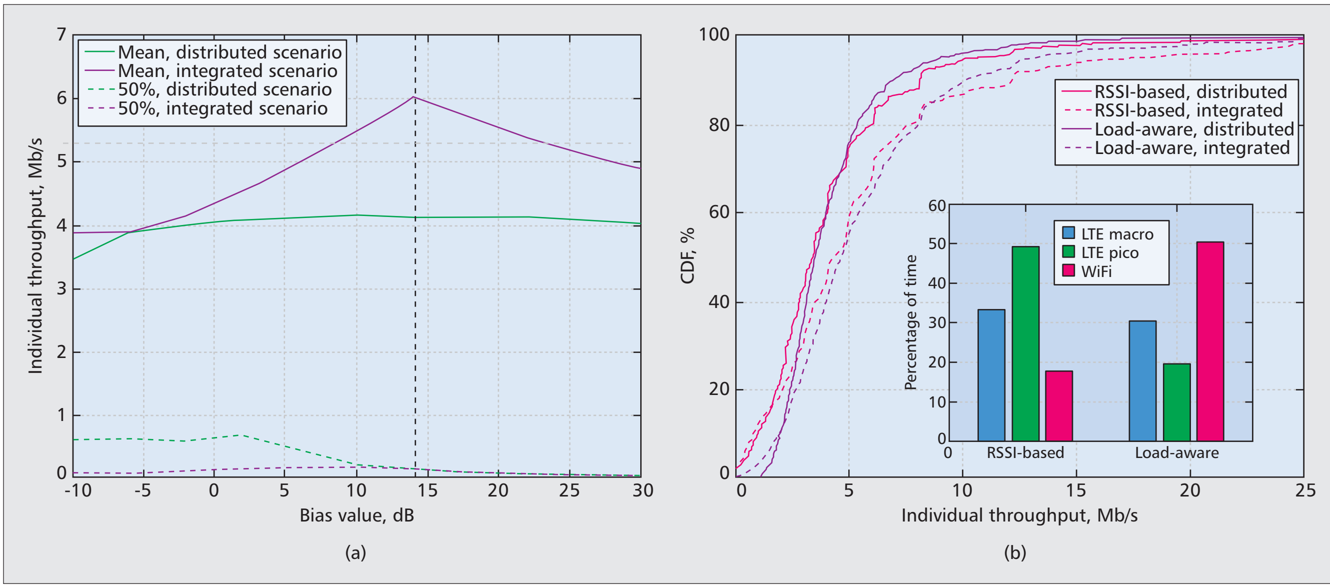


Figure 4. Comparing our load-aware (RAN-assisted) and RSSI-based (RAN-controlled) network selection schemes.

Fig. 5b representing areas of the macro, pico, and WLAN tiers together with the corresponding users and infrastructure nodes. We consider the following *cascade* network selection when a new session arrives into the system. First, the RAN-based network selection assistance entity attempts to offload the newly arrived session onto the nearest WLAN AP by performing centrally managed WLAN admission control. If the session is accepted on the WLAN tier, it is served there without interruption until it successfully leaves the system. Otherwise, if this session cannot be admitted onto the WLAN, the pico network admission control is executed, and either the session is accepted on the pico tier and served by the nearest pico BS, or the macro network itself attempts to serve this session. Eventually, if the session cannot be admitted onto the macro tier either, it is considered permanently blocked and leaves the system unserved.

In Fig. 5c, we detail the overall blocking probabilities for the converged HetNet as well as for the three tiers in, macro, pico, and WLAN, individually. Our main observation is that with two additional overlay tiers, HetNet performance improves significantly over what can be achieved in the macro-only networks (cellular baseline). Remarkably, we actually witness visible performance improvement even with only a few additional infrastructure nodes, such as two WLAN APs and two pico BSs in this example. Therefore, we believe that multiple RATs and the associated network selection intelligence for their efficient use will become a characteristic feature of future 5G HetNets.

Main Takeaways and the Way Forward

In summary, this article reviews major challenges in delivering uniform connectivity and service experience to future heterogeneous 5G networks. It discusses several architecture choices and associated algorithms for intelligent access network selection in multi-RAT HetNet deployments, both when the control of how radios are utilized rests with the network and when it rests with the user. In particular, it compares simulated performance of RAN-assisted load-aware network selection schemes with the conventional/existing UE- and network-based solutions employed in current systems. We primarily focus on uplink performance as it has not been fully addressed in past literature.

The main advantages of load-aware schemes stem from the fact that the SNR-threshold (WiFi-preferred) scheme, as well as the network-centric cell range extension scheme, do not explicitly account for the loading and interference on WiFi APs typically encountered in clustered UE deployments. Our results show that the load-aware user-centric scheme, which augments SNR measurements with additional information about network load, could improve the performance of a WiFi-preferred scheme based on minimum SNR threshold. We observed over 75 percent improvement in 5 percent cell edge throughput as well as significant gains in energy efficiency for both distributed and integrated deployment scenarios.

Comparison with more advanced network-controlled schemes has also been completed across various heterogeneous deployments to confirm attractive practical benefits of distributed user-centric solutions. Next steps include further investigation of UE-based algorithms while explicitly considering load variation in the network and accounting for application-layer statistics. System behavior in the presence of uncoordinated (rogue) WiFi interference must also be accounted for, and the hysteresis mechanism may further be improved to combat the uncertainty in estimating user throughput.

Building on the system-wide simulation data, we also propose a novel dynamic methodology for RAN-assisted network selection capturing the spatial randomness of HetNets together with

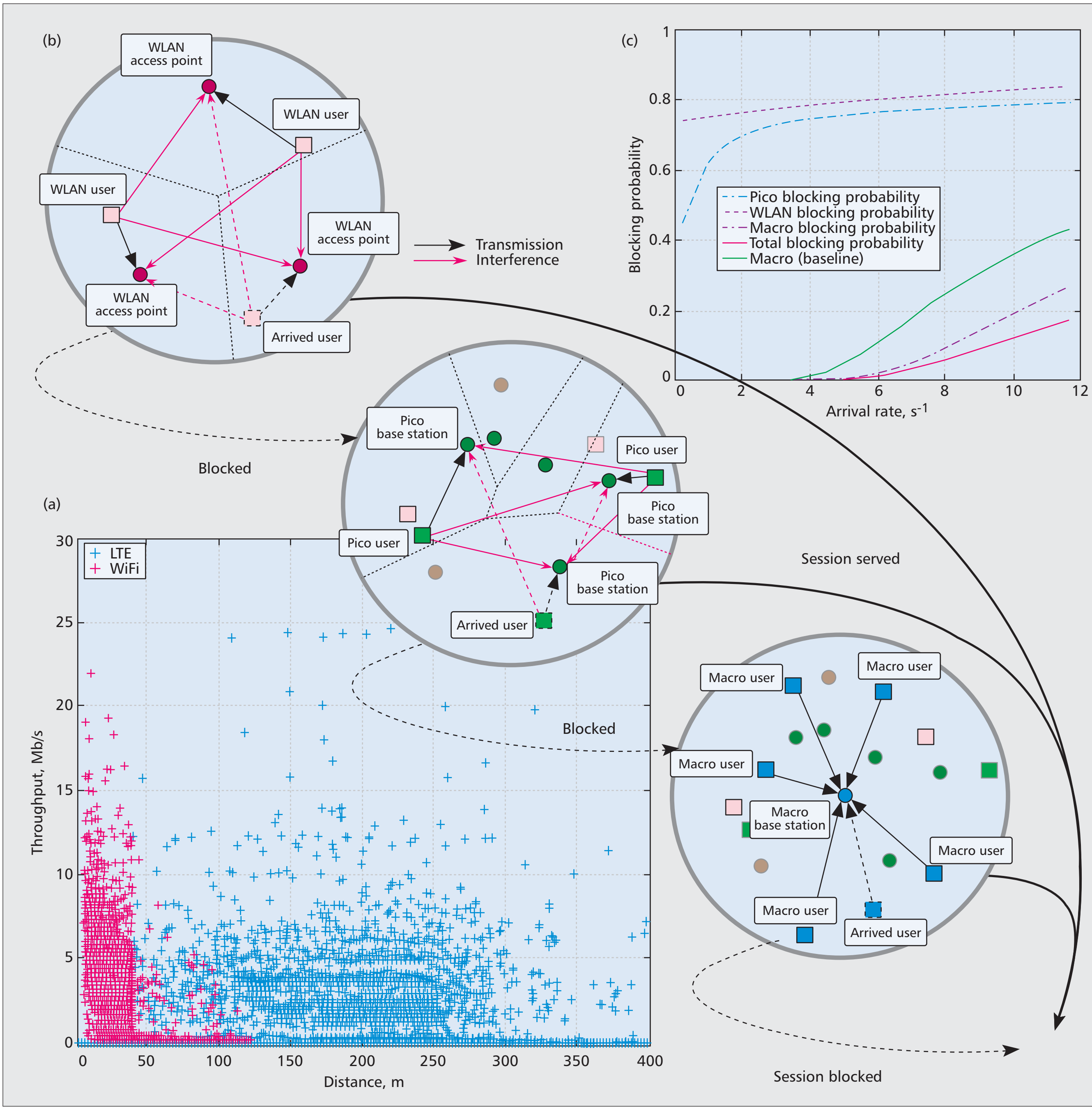


Figure 5. Illustration of the proposed space-time methodology for multi-RAT networks.

unsaturated uplink traffic from its users. Our stochastic-geometry-based analysis enables in-depth characterization of dynamic interactions between macro and pico cellular networks, as well as WLAN, mindful of user QoS and based on intelligent RAT selection/assignment. Going further, we expect our space-time methodology to be capable of encompassing other technologies beyond LTE and WiFi, as well as additional use cases beyond simple aggregation of capacity across unlicensed bands.

More generally, studying the ultimate capacity of multi-radio wireless networks remains an open problem in the field of information theory, and stochastic geometry has the potential to shed light on it given that it can explicitly capture new interference situations and hence the achievable data rates. This challenging objective may require novel advanced analytical tools to interconnect and apply techniques and methods coming from the area of point processes, probability theory, queuing theory, and percolation theory, as well as modern engineering insights.

Acknowledgment

This work is supported by Intel Corporation, GETA, and the IoT SRA program of Digile, funded by Tekes. The work of the first author is supported with a Postdoctoral Researcher grant by the Academy of Finland, as well as with a Jorma Ollila grant by Nokia Foundation.

Our stochastic-geometry-based analysis enables in-depth characterization of dynamic interactions between macro and pico cellular networks, as well as WLAN, mindful of user QoS and based on intelligent RAT selection/assignment.

References

[1] B. Bangerter *et al.*, "Networks and Devices for the 5G era," *IEEE Commun. Mag.*, vol. 52, 2014, pp. 90–96.

[2] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018, Feb. 2014.

[3] S.-P. Yeh et al., "Capacity and Coverage Enhancement in Heterogeneous Networks," *IEEE Wireless Commun.*, vol. 18, 2011, pp. 32–38.

[4] P. Marsch *et al.*, "Future Mobile Communication Networks: Challenges in the Design and Operation," *IEEE Vehic. Tech. Mag.*, vol. 7, 2012, pp. 16–23.

[5] J. Andrews *et al.*, "An Overview of Load Balancing in HetNets: Old Myths and Open Problems," *IEEE Wireless Commun.*, vol. 21, no. 2, 2014, pp. 18–25.

[6] S. Andreev *et al.*, "Cellular Traffic Offloading onto Network-Assisted Device-to-Device Connections," *IEEE Commun. Mag.*, vol. 52, no. 4, 2014, pp. 20–31.

[7] R. Baldemair *et al.*, "Evolving Wireless Communications: Addressing the Challenges and Expectations of the Future," *IEEE Vehic. Tech. Mag.*, vol. 8, 2013, pp. 24–30.

[8] J. Andrews, "Seven Ways that HetNets Are a Cellular Paradigm Shift," *IEEE Commun. Mag.*, vol. 51, 2013, pp. 136–44.

[9] M. Bennis *et al.*, "When Cellular Meets WiFi in Wireless Small Cell Networks," *IEEE Commun. Mag.*, vol. 51, 2013, pp. 44–50.

[10] 3GPP TR 37.842, "3GPP/WLAN RAN Interworking Study Item Report," 2013.

[11] L. Wang and G.-S. Kuo, "Mathematical Modeling for Network Selection in Heterogeneous Wireless Networks — A Tutorial," *IEEE Commun. Surveys & Tutorials*, vol. 15, 2013, pp. 271–92.

[12] 3GPP TR 36.814, "Further Advancements for E-UTRA Physical Layer Aspects," 2010.

[13] S. Navaratnarajah *et al.*, "Energy Efficiency in Heterogeneous Wireless Access Networks," *IEEE Wireless Commun.*, vol. 20, 2013, pp. 37–43.

[14] O. Galinina *et al.*, "Capturing Spatial Randomness of Heterogeneous Cellular/WLAN Deployments with Dynamic Traffic," *IEEE JSAC*, vol. 32, no. 6, 2014, pp. 1083–99.

[15] H. ElSawy *et al.*, "Stochastic Geometry for Modeling, Analysis, and Design of Multi-Tier and Cognitive Cellular Wireless Networks: A Survey," *IEEE Comm. Surveys & Tutorials*, vol. 5, 2013, pp. 996–1019.

Biographies

SERGEY ANDREEV is a senior research scientist in the Department of Electronics and Communications Engineering at Tampere University of Technology (TUT), Finland. He received his Specialist degree (2006) in information security and his Cand.Sc. degree (2009) in wireless communications from St. Petersburg State University of Aerospace Instrumentation, Russia, as well as his Ph.D. degree (2012) in technology from Tampere University of Technology, Finland. He has (co-)authored more than 80 published research works. His research interests include wireless communications, energy efficiency, heterogeneous networking, cooperative communications, and machine-to-machine applications.

MIKHAIL GERASIMENKO is a research assistant at TUT in the Department of Electronics and Communications Engineering. He received his Specialist degree from St. Petersburg University of Telecommunications in 2011. In 2013 he obtained his M.Sc. degree from Tampere University of Technology. He started his academic career in 2011 and in recent years he appeared as a (co)-author on several scientific journal and conference publications, as well as several patents. Moreover, he also acted as a reviewer and participated in education activities. His main subjects of interest are wireless communications, machine-type communications, and heterogeneous networks.

OLGA GALININA is a Ph.D. candidate in the Department of Electronics and Communications Engineering at TUT. She received her B.Sc. and M.Sc. degrees in applied mathematics from the Department of Applied Mathematics, Faculty of Mechanics and Physics, St. Petersburg State Polytechnical University. She has published work on mathematical problems in the novel telecommunication protocols in internationally recognized journals and high-level peer-reviewed conferences. Her research interests include applied mathematics and statistics; queueing theory and its applications; wireless networking and energy efficient systems, and machine-to-machine and device-to-device communication.

YEVGENI KOUCHERYAVY is a full orofessor and laboratory director at the Department of Electronics and Communications Engineering of TUT. He received his Ph.D. degree (2004) from TUT. He is the author of 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, as well as nanocommunications. He is an Associate Technical Editor of *IEEE Communications Magazine* and an Editor of *IEEE Communications Surveys and Tutorials*.

NAGEEN HIMAYAT is a senior research scientist with Intel Labs, where she performs research on broadband wireless systems, including heterogeneous networks, cross-layer radio resource management, MIMO-OFDM techniques, and optimizations for M2M communications. She has over 15 years of research and development experience in the telecom industry. She obtained her B.S.E.E. from Rice University and her Ph.D. in electrical engineering from the University of Pennsylvania in 1989 and 1994, respectively.

SHU-PING YEH is a research scientist in the Wireless Communications Laboratory at Intel. She received her M.S. and Ph.D. from Stanford University in 2005 and 2010, respectively, and her B.S. from National Taiwan University in 2003, all in electrical engineering. Her current research focus includes interference mitigation in multitier networks utilizing multi-antenna techniques, machine-to-machine communications, and interworking of multiple radio access technologies within a network.

SHILPA TALWAR is a principal engineer in the Wireless Communications Laboratory at Intel, where she is conducting research on mobile broadband technologies. She has over 15 years of experience in wireless. Prior to Intel, she held several senior technical positions in the wireless industry. She graduated from Stanford University in 1996 with a Ph.D. in applied mathematics and an M.S. in electrical engineering. She is the author of numerous technical publications and patents.