SHARING

SELF-ORGANIZED HETEROGENEOUS ADVANCED RADIO NETWORKS GENERATION

Deliverable D4.6

Performance assessment of the most promising techniques in Heterogeneous Wireless networks

Date of delivery	11/02/2016
Contractual date of delivery	28/02/2016
Project number	C2012/1-8
Editor(s)	Mehdi Bennis (OULU)
Author(s)	Grégory Gougeon (SIR), Florian Letourneux (SIR), Mathieu Brau (SIR), Yoan nCorre (SIR), Yves Lostanlen (SIR), Engin Zeydan (AVEA), A. Serdar Tan (AVEA), Mourad Khanfouci (MERCE), Mehdi Bennis (UOULU), Mohammed Saad El Bamby (UOULU), Edgar Ramos (ERICSSON), Antti Ratilainen (ERICSSON), Kimmo Hiltunen (ERICSSON), Mohamad Assaad (SUP)
Dissemination level	PU
Workpackage	4
Version	V1.0
Total number of pages	79

Abstract:

This document represents the final deliverable of the SHARING project as far as work package 4 is concerned. It describes a concise summary of the technical innovations carried out in the four research tasks. The deliverable is composed of two parts. The first part provides a comprehensive summary of the major technical contributions carried in WP4 with an outlook on the opportunities and challenges facing the deployment of multi-dimensional heterogeneous networks. The second part describes, for the sake of completeness, other innovations carried out within WP4 not included in the first part of this deliverable or in previous deliverables. Finally, conclusions and recommendations are drawn.

Keywords: Heterogeneous networks, spectrum, small cells, offloading, energy efficiency, carrier aggregation.

Document Revision History

Version	Date	Author	Summary of main changes
0.1	16.11	Mehdi Bennis	first draft
0.2	2.12	Mehdi Bennis	editorial work
0.3	30.1	Mourad Khanfouci and Yolanda Fernandez	Review
0.4	21.1	Mehdi Bennis	editorial work based on reviewer's comments + adding summary tables, conclusion
0.5	31.1	Arturo Ortega	Final version
0.6	11.2	Mehdi Bennis	Final version

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This document constitutes the final deliverable of SHARING work package 4, providing a performance assessment of the most promising techniques carried out within the four research tasks of work package 4 (Figure 1). In the first part of this deliverable (Section A), a summary of the most innovative solution concepts is presented, with their most notable assets and their impact on current standardization activities in 3GPP. Subsequently, for the sake of completeness, other WP4 technical contributions, not included in Section A or in other previous deliverables, are presented in the second part of this deliverable (Section B).

Resource optimization in heterogeneous networks is very challenging due to time-varying traffic demands, more pronounced interference conditions among different tiers, backhaul congestion and uplink and downlink data traffic asymmetry. To overcome these challenges, novel innovative solutions, summarized in the Section A, have been called upon:

- **Intra-system radio access offloading**: More small cells (i.e., network densification) are advocated in conjunction with the legacy macrocellular network where capacity is needed the most. This entails intercell interference mitigation techniques in time, frequency and space; with solution concepts like cell range expansion, dynamic cell switch ON/OFF, dynamic TDD and dual connectivity. Efficient use of advanced multiple antenna techniques are fundamental in improving the network spectral efficiency. This is achieved by techniques such as massive MIMO, decentralized coordinated multipoint transmission and reception, and other variants.

- **Inter-system radio access offloading**: A smooth coexistence between LTE and WiFi is essential to smartly combine the benefits of both radio access technologies. This is achieved via dynamic and intelligent load balancing techniques among both radio access technologies. The problem to be solved is when to switch from LTE to WiFi and vice-versa in a dynamic and flexible manner.

- **Enhanced energy efficiency or energy savings**: Simply adding small cells is not energy efficient per se. Therefore, dynamic and load-aware energy saving mechanisms for small cell base stations or clusters of base stations are needed taking into account interference levels, traffic load and user density. Which small cells need to be switched off as a function of density of users and traffic load is one of the main problems to be solved.

- **Advanced spectrum management**: One of the challenges of heterogeneous networks is optimizing the use of spectrum via aggregated component carriers coming from the same serving node or different serving nodes. This is commonly referred to as carrier aggregation. The fundamental problem is to know how many component carriers should be aggregated and from which serving nodes.

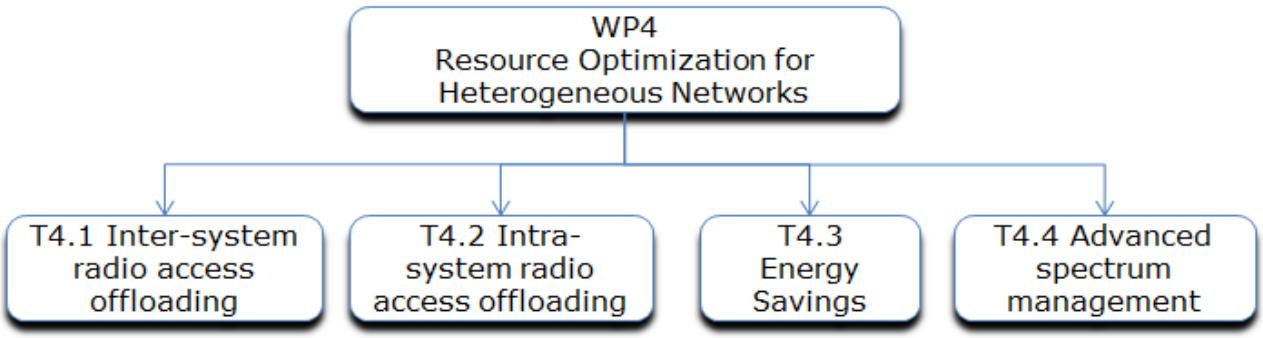


Figure 1: Structure of SHARING work package 4.

In the second part of the deliverable (Section B), we provide a technical description of other SHARING WP4 innovations, not included in Section A or in earlier deliverables. This includes backhaul offloading via caching and big data analytics, dynamic uplink-downlink TDD optimization and energy savings using consensus algorithms.

1 INTRODUCTION

According to recent estimates, data traffic demands are expected to grow rapidly in the coming years. This traffic surge is mainly caused by the evolution of mobile terminals (handsets, tablets and other handheld devices) and the advent of connected devices encompassing all aspects of our lives. Among the most promising ways to address these compelling challenges and increase network capacity is: 1) exploiting available spectrum by dynamically sharing spectrum among different tiers and radio access technologies; 2) improving spectral efficiency by exploiting recent advances in multiple antenna techniques (e.g., MIMO, beamforming and massive MIMO); 3) network densification by replacing existing macrocellular networks with low-power base stations or small cells when more capacity is needed. Figure 2 depicts the various degrees of freedom that can be harnessed when optimizing heterogeneous wireless networks.

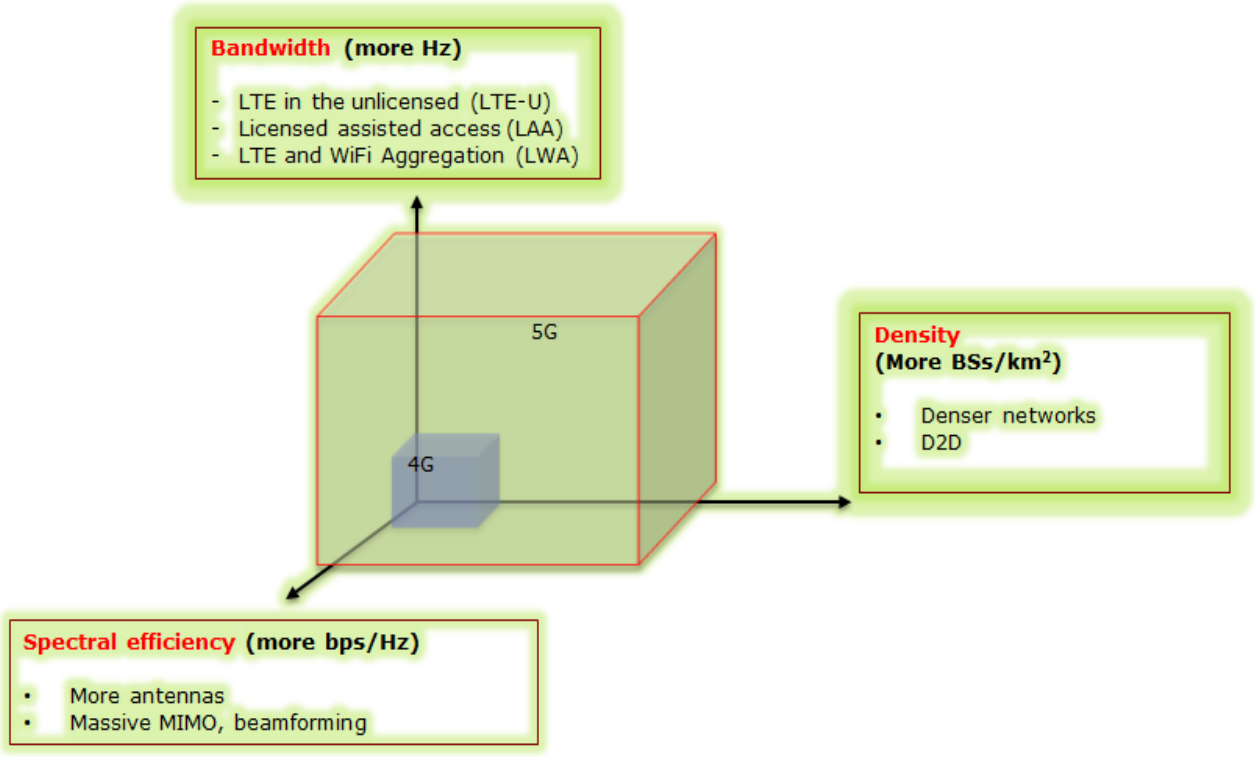


Figure 2: Degrees of freedom for resource optimization in heterogeneous networks.

The total cost of the network deployment consists of large number of different components, related to both the capital expenditure (CAPEX) and the operational expenditure (OPEX). CAPEX-related components include the base station equipment, site equipment and network roll-out. The OPEX-related components include site rental, energy, and operation and maintenance. The total cost of the network deployment can be reduced for example by improving the cell capacity and coverage, by introducing various types of SON functions, and by introducing different kinds of energy saving mechanisms.

This final deliverable addresses these challenges by proposing a number of innovative solutions grouped into four main categories:

- **Intra-system radio access offloading:** Mobile traffic offloading has been recognized as a promising solution for data traffic boost in mobile cellular networks. In LTE heterogeneous networks, offloading and load balancing strategies among eNodeBs and small cells are key techniques to improve network resource utilization.

- **Inter-system radio access offloading**: The aim is to improve the use of radio resources and unlicensed bands in order to offload traffic from LTE to WiFi. In fact, improving only licensed bands for offering additional capacity will lead to a strong increase of deployment cost for operators. A cost efficient solution is simply to offer to broadband mobile networks the possibility to exploit both licensed and unlicensed bands. This solution is particularly attractive for cellular networks with small cells (micro, pico, femto), which is the current trend.

- **Energy saving mechanisms:** Currently deployed base stations need to permanently signal their presence and availability, and continuously listen to the radio environment to detect incoming users. At the same time, the massive large scale deployment of small cells can be ecologically worthwhile only if armed with smart energy efficiency and power saving mechanisms. Motivated by this, energy-efficient sleep mode strategies are required so as to reduce the power consumption of cellular networks. These strategies allow the hardware components in the base station to be smartly switched off in idle conditions, such that the energy consumption is modulated over the variations in traffic load and space and is carried out across all available tiers.

- **Spectrum resource allocation**: One of the important aspects in future network deployment will be the increasing bandwidth demand. Therefore, multi-flow carrier aggregation will play an important role in providing the operators with the maximum flexibility in using their available spectrum. With carrier aggregation, LTE-A will be able to deliver much higher throughputs, by combining spectrum blocks up to 100MHz either from the same base station or sets of non-collocated base stations.

2 SECTION A

This section is divided into four distinct subsections spanning the four key innovation areas of work package 4, namely intra-system radio access offloading, inter-system radio access offloading, energy saving mechanisms, and advanced spectrum management. In what follows, the most promising innovations are summarized.

2.1 Intra-System Radio Access Offloading

The surge of new network services and shifting of networks from a circuit-switch focused revenue to data driven services have increased the search for improving the efficiency of the use of radio network resources and cost efficient deployments. A better utilization of available resources and increase of efficiency (spectrum, load, scheduling, etc.) without requiring additional access technologies is one of the ultimate goals of network operators. Therefore, the intra-system radio access offloading has become one of the most valued and targeted improvements in the search for a cost efficient solution to increase network performance and capacity.

The deployment of offloading solutions also comes with an operation risk that might affect the same aspect tried to be improved, the experienced network performance of the end users. The solutions are challenged by several aspects that affect their outcome, effectivity and performance. Fast changing environments where user density in several locations might vary according to time or events (as it happens in train stations for example), terminal capabilities not being uniform, meaning that legacy systems have to be always in place, and equipment limitations that might be reached after sudden peaks of traffic (backhaul capacity, processor load, buffer overflow, etc.) are some of the aspects. In some cases, the introduction of additional elements in a deployment might cause inadverted consequences to the initial network planning that were not considered when establishing the initial macro sites, for example unintended strong intercell interference from small cells, increase in the signaling overhead or loss of coverage in some areas. Another aspect to the increased efficiency is the search of flexibility, meaning that some deployments might be used "when required" to assist the network to provide the capacity or performance expected. The mentioned flexibility requires monitoring of the environment from particular areas and the overall network to provide the decision bases to controllers in the network to activate or deactivate features and deployment options. This might be the case of enabling additional carriers, activating and deactivating small cells, switching between transmission/reception modes, etc. The collection of the required KPIs, measurements and activity in the network is an additional challenge together with the actual organization and processing of the data that in many cases is distributed and might accumulate large volumes of information and the final execution of the configuration decision, especially challenged in distributed environments. It is important to keep in mind, that the best strategies correspond to striking a balance between the three key factors, performance, capacity (including resource utilization) and energy saving.

In order to address some of the mentioned issues, SHARING has proposed several innovations. Following there is a selection of the key innovations that have significant impact on solving the problems mentioned above. The presented innovations aim to make the usage of radio resources and the installed equipment (base stations, backhaul connections) more efficient. Hence, the target is to find innovative ways to reduce the total cost of network deployment for the operator, while still maintaining the desired network performance. The innovation scopes are different and highly situational, meaning that they apply to specific concepts and in some cases deployments. This makes them difficult to compare because their final objectives target different KPIs. In any case their main advantages and disadvantages are being mentioned and described to provide some comparison framework.

First, in Section 2.6 of [3], the concept of virtual small cells (VSC) is presented, where large antenna arrays at macrocell sites are used to focus the energy towards a traffic hotspot, see Figure 3. As in traditional heterogeneous networks with picocells deployed at hotspots, this creates areas with enhanced signal to noise and interference ratio (SINR) and increases the

resource reuse of the system (cell splitting gain). VSCs are managed at the macrocell level, giving place to a completely centralized system free of coordination and backhaul latency constraints. VSCs can reuse the same carrier frequency of the macrocell allowing an efficient reuse of spectrum, bringing a high spectral efficiency and considerable aggregated capacity gains. Furthermore, the use of dense array antenna systems could result in energy savings since they allow the energy to be focalized where and when it is needed. Also, the introduction of the new equipment, which use more efficient electronics make an important difference in the network power consumption. Some of the challenges of the deployment of VSC are to recognize the hotspot areas and the resource density optimization for such changing scenarios. Therefore, this solution is feasible for very predictable and "stable" scenarios.

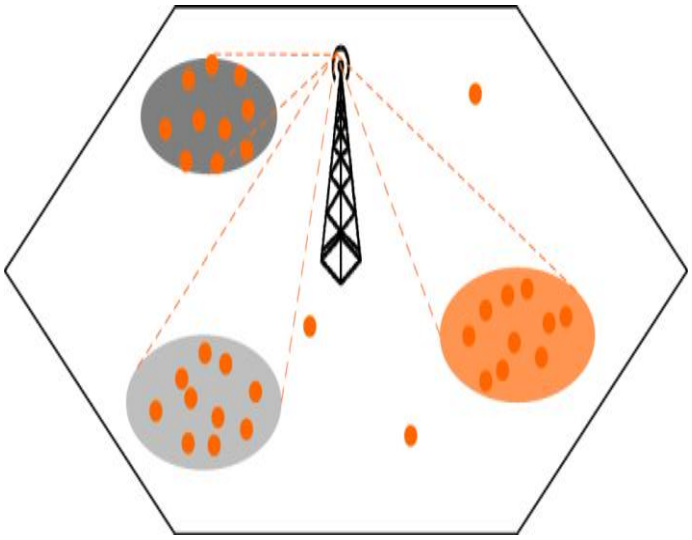


Figure 3: Antenna array virtual small cells (VSCs)

The results show that the introduction of virtual small cells can improve the mobile network throughput when compared with a situation where only macrocells are deployed. For a dense urban scenario with full buffer model, system-level simulations show that the average user throughput is increased by 50%. Conventional picocells outperform VSC in terms of user throughput in dense urban scenarios if we consider limited number of antennas (limited directivity for the VSC beam), but this comes at the price of increased cost and reduced flexibility. It should also be noted that these results do not take into account the changes in the traffic location that occur along the day. Such simplification artificially benefits the performance results of traditional heterogeneous networks versus VSCs, which can actually adapt their location to the moving traffic. When it comes to energy consumption, the results indicate that the introduction of the VSCs allow consumed power saving of about 27% and 43% in the dense urban and in the rural scenarios with respect to the only-macrocell and of 36% and 73% with respect to the macrocell-picocell case. The introduction of VSCs has very little standardization impact if none, which makes it very attractive to be deployed by operators in outlined scenarios where the benefits are visible.

SHARING project has also proposed and evaluated the performance of several innovations aiming to enhance the traffic offloading gains, and to control the interference between neighboring cells. Three of those innovations are highlighted here.

- A centralized recursive self-optimization algorithm is proposed for heterogeneous LTE-A networks in Section 2.4 of [3]. The algorithm aims to find the optimum values for both, the Cell Range Extension (CRE) for low power nodes and the eICIC muting ratio, which determines sub-frames during which the macro layer uses Almost Blank Subframes (ABS) to reduce the interference caused to the low power nodes. The algorithm is based on the use of a surrogate of the network model and a search and

poll algorithm for optimization. The proposed approach is advantageous as the algorithm approaches to the global optimum in the first update thanks to the surrogate, and the network continues to operate at this near optimal parametric setting as the algorithm locally searches the global optimum. This makes the algorithm well suited for self-optimization on an operational network. Furthermore, no major impact on standardization is expected in order to be able to implement the proposed algorithm. From the presented evaluation results it can be clearly seen that the user link quality (in terms of SINR and Average Bit Rate) is improved. However, this is at the cost of the increase in blocked call rate which is a consequence of muting sub-frames at the macro layer. It can be seen that the optimized HetNet setting improves the performance despite a decrease in the available network capacity with respect to the heterogeneous network without CRE+eICIC.

- A distributed algorithm which dynamically optimizes the uplink/downlink TDD switching point configuration is presented and evaluated in Section 2.8 of [3]. The algorithm relies only on the local information available at each small cell, thus taking into account the cell traffic load and the interference experienced from neighboring cells. Therefore each small cell creates an estimation of the optimal switching point according to the information available for them and updates it dynamically. The purpose of the innovation is to increase the spectrum efficiency by reducing the possibilities of interference and maximizing the opportunities of transmission. Some challenges facing this solution are the different requirements of quality of service for the transmissions and the coordination between several base stations deploying slow or limited backhaul. The optimization is made based on SON schemes that do not necessarily require standardization, but mainly an implementation of O&M that can make use of services that are to be found in traditional TDD deployments and requiring measurements and input that are collected also for other means. The performance results show that depending on the traffic distribution, gains up to 100% are achieved compared with fixed uplink/downlink TDD switching point. The fixed scheme achieves the same performance in case the UL and DL traffic loads in the system are assumed to be equal, but as the difference between the UL and DL traffic loads increases; the proposed scheme performs better compared to the fixed scheme showing clearly improved spectral efficiency. The proposed scheme seems to achieve considerable gains also in high cross-link interference environment, where half of the cells have opposite uplink-to-downlink ratios compared to the other half.

  Description of the latest research results on this innovation can be found in the second part of this deliverable (Section 3.1.1).

- In heterogeneous networks, the optimal cell borders might be different for the downlink and the uplink due to the transmission power difference between the macro and the small cell. The downlink cell border is pushed relatively close to the small cell, where both cells are perceived equally strong. In uplink on the other hand, the maximum UE transmission power is constant regardless of serving cell and therefore the uplink cell border should be in between the two nodes where the path loss is the same to both nodes. To improve the efficient usage of radio resources in such situations, a method called uplink/downlink split is proposed. In case of uplink/downlink split the UE would connect both links separately to the best serving cells, in practice the downlink to the best macrocell and uplink to the closest node, whether it is the same macrocell or a small cell. This concept is presented in Section 3.2 [3], and an illustration of the innovation can be seen in Figure 4. Uplink/downlink split has very little standardization impact, as Dual Connectivity was introduced already in 3GPP LTE Release 12 and specification TS 36.323[30] contains variable *ul-DataSplitDRB-ViaSCG* which is used to determine whether the uplink is transmitted via secondary or master node. Transmitting the downlink traffic via e.g. master node only can be done by scheduling decisions, i.e. scheduling the UE only in master node and not in serving node. This might require a new implementation for scheduling function to eNodeB.

The simulations show, that using uplink/downlink separation results in gains in uplink transmissions. The gains are measured from the mean FTP rates of the users in scenarios both with and without uplink/downlink separation. The highest gains, up to 50% for all users on average and 218% for cell-edge users, seem to result from a highly loaded system and relatively high Cell Uplink Range Expansion (CURE). The load of the system does have an effect on the gains, since with low load the average gains were much lower, for all users the gain reaches up to 5% gain and for cell-edge users up to 65%. Also, the users with already good throughput seemed to have decreased performance in the low load scenario. With downlink traffic, the simulations show that using uplink/downlink separation results in decreased performance. This is explained by downlink transmissions not benefitting from the increased capacity in uplink, since the uplink transmissions contain only small size TCP acknowledgements. On the contrary, the increased round trip time caused by the backhaul delay between the secondary and the master node negatively affects the TCP slow start. The delay in the uplink signaling causes delay in the downlink transmissions and therefore decreased performance. Hence, from the results it can be seen that UL/DL split should be activated, when the traffic in the system is uplink oriented and the system load is medium to high. In low loaded system and when the traffic is downlink only, the performance is slightly decreased, so it might be preferable not to activate UL/DL separation in such conditions.

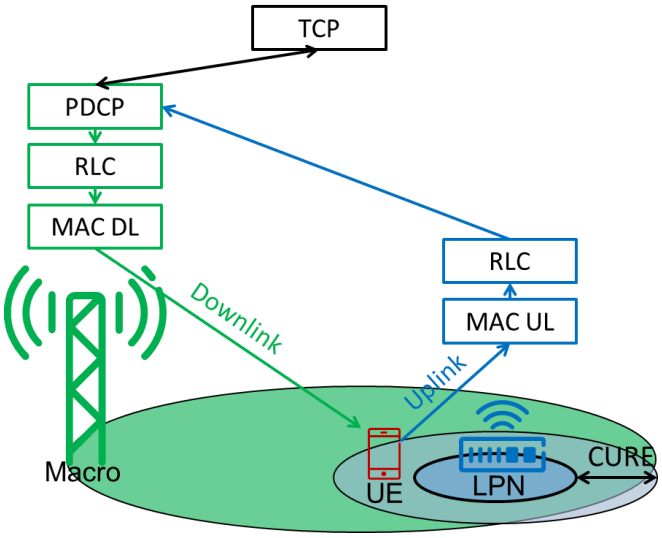


Figure 4: Uplink/Downlink traffic split on PDCP level between a macrocell and a low-power node (LPN)

Finally, the cost of providing a sufficient backhaul connection to each of the site locations is playing a major role when considering the total cost of network deployment. The situation is envisioned to become even more crucial for future networks, where the number of small cells is expected to be considerably larger compared to current network deployments. Related to that problem, dynamic content caching has recently attracted much attention for beyond 4G networks. By harnessing recent advances in storage and computing, dynamic caching can help alleviate backhaul congestion, reduce loads at peak times and minimize latency, by pre-caching contents at strategic network edge locations. If smartly coupled with meta-data analytics, network operators can further exploit the vast amount of users' context information (location, speed, etc.) for a better predictability of future demands, to proactively cache popular contents before users actually request them. Backhaul offloading by proactive caching is described in Section 4.1 of [3] and an illustration of the proposed concept can be seen in Figure 5.

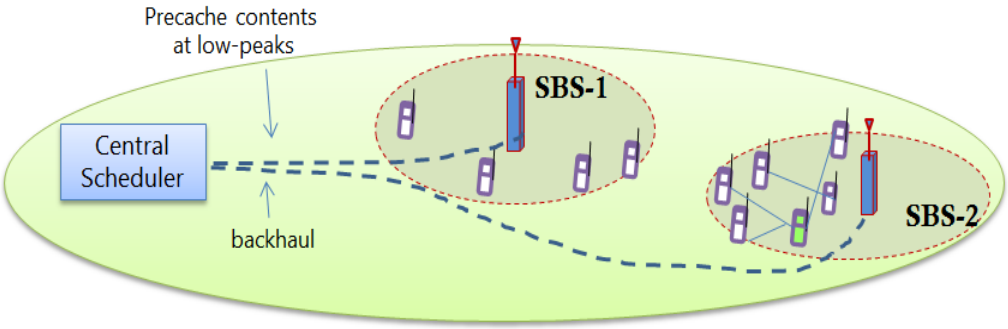


Figure 5: Proactive caching at the small cell nodes

The concept exploits the storage capabilities at the small cell base station in which contents are precached at strategic times to satisfy users' QoS requirements. The proposed caching procedure requires some known information about the popularity matrix, which tells the popularity of the files among the users, in order to perform its first training and prediction step. It should be noted that the solution does not elaborate in the actual architectural impact in order to provide the connection between the bearer connection and the cached data, which at this level would suit better a proprietary implementation of the caching infrastructure.

The caching strategy has been evaluated in a variety of scenarios as a function of storage constraints, wireless backhaul links, and content popularity. The studied factors are satisfaction ratio, which translates to realized QoS, and backhaul load. The obtained results show that the proposed proactive caching method outperforms the baseline reactive procedure in almost all of the evaluated cases. With very low amount of users' requests the reactive approach generates fewer loads for the backhaul. But when the number of requests increases, the proactive method performs better. The highest gains of proactive caching are achieved when some files are requested more often than others, and this gain is further increased when the high load regime is studied.

Description of the latest research results on backhaul management can be found in the second part of this deliverable (Section 3.1.2).

2.2 Inter-System Radio Access Offloading

The exploitation of multiple radio access technologies (RAT) holds the promise of tacking the increasing traffic demands, by leveraging both licensed and unlicensed bands for a more efficient traffic offload. Inter-system radio access offloading includes LTE-WiFi interworking, LTE-WiFi aggregation (LWA) and other multi-RAT coexistence approaches, all of which depend on whether RATs are collocated or not. This topic has recently attracted significant interest from academia, industry, and standardization bodies alike, whereby the complementary benefits of both RATs can be leveraged in various network deployment scenarios. On the one hand, due to the uncontrolled, unlicensed nature of WiFi, the competition for resources among a potentially large number of hotspot users and other devices transmitting on the same unlicensed band can yield dramatically poor throughput. In such a scenario, offloading some of this traffic to a well-managed small cell network operating over the licensed spectrum can improve the performance. On the other hand, the inherent constraints of small cell networks, particularly due to cross-tier and co-tier interference, motivate offloading some of the traffic to the WiFi band both to reduce the interference and to ease the congestion. Clearly, a best of both world approaches is of utmost importance. Figure 6 depicts the variants of LTE-WiFi coexistence scenarios.

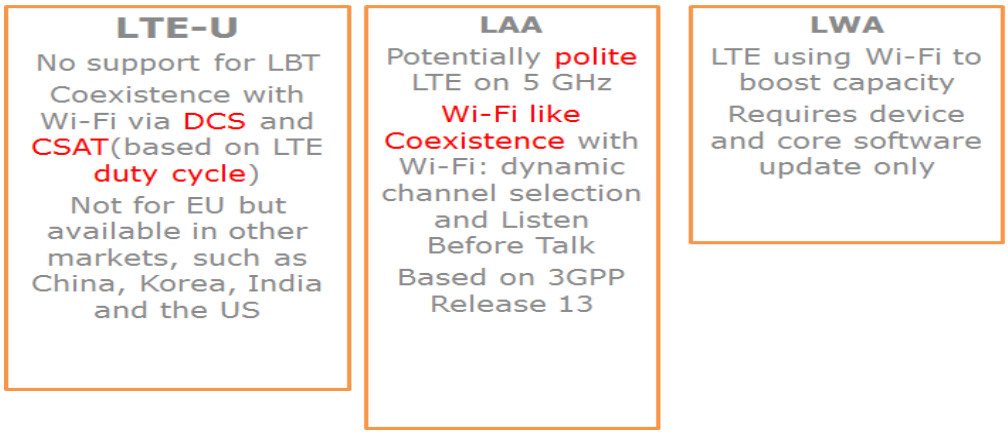


Figure 6: Variants of LTE-WiFi coexistence scenarios, namely LTE-U, LAA, and LWA

The fundamental LTE-WiFi coexistence problem boils down to investigating novel flexible and dynamic strategies to offload LTE traffic towards WiFi as well as load balancing between different RATs either in a collocated (dual mode) or non-collocated setting (e.g., LWA), as a function of interference levels, traffic load, congestion levels, etc. Needless to say, a solution that combines the benefits of both RATs is of utmost importance. More specifically, the first set of contribution in this key innovation examines offloading and load balancing strategies to improve resource utilization via the definition and implementation of a middleware as described in detail in Chapter 2 of [31]. The middleware is in charge of determining at each location point and for each user which is the best base station to be attached to from the list of potential candidates. A decision module is then shown to be able to interact with existing user/spectrum databases allocating optimum RATs based on diverse key performance indicators while maintaining quality of service guarantees.

Another solution concept is a method to seamlessly offload traffic from LTE to WiFi in which the total system performance is increased by making use of several network and user attributes for centralized offloading decision, as described in Chapter 4 of [31] and illustrated in Figure 7. Moreover, it is shown that by taking into account the base station protocol functions such as access, error and congestion control, network and channel capacity, performance metrics including data transmission latency and reliability parameters are improved. Results capturing users' quality of experience (QoE) in the context of video transmission during a handover process from an LTE to WiFi have shown that significant packet losses can be avoided. Description of the latest research results on this solution can be found in the second part of this deliverable (Section 3.2.1)

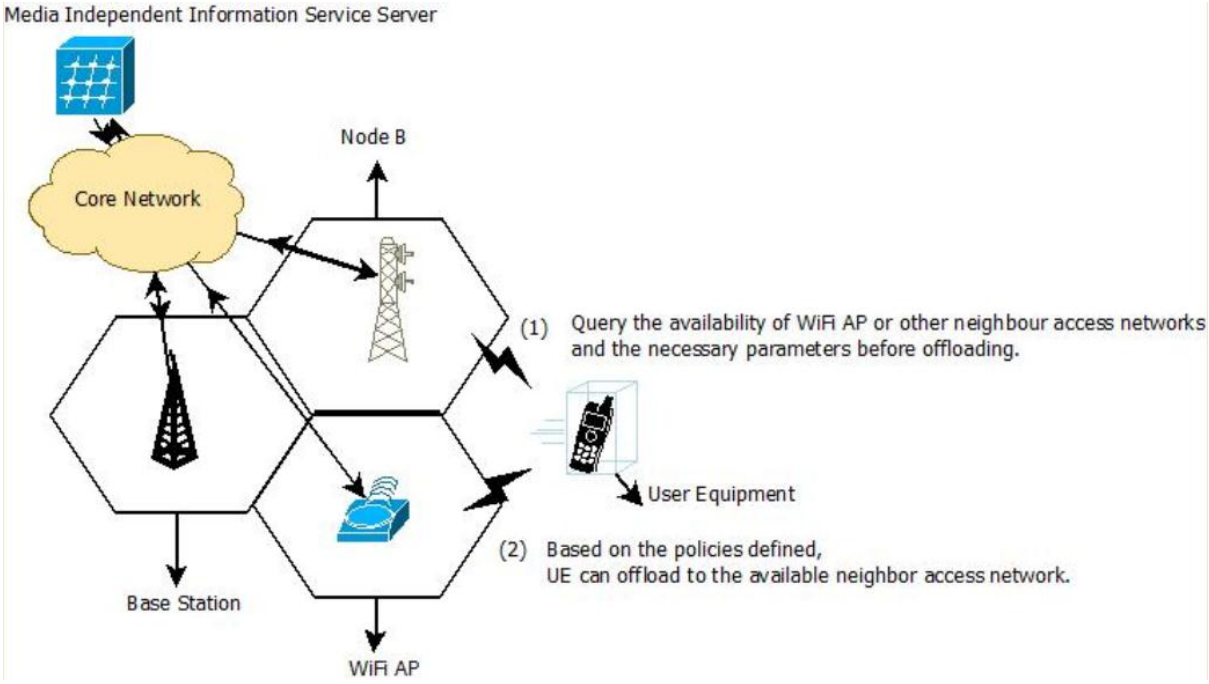


Figure 7: Variants IEEE 802.21 and user experience based offload scenario.

Yet another offloading strategy considers the case of dual mode small cells able to simultaneously transmit on LTE and Wi-Fi, see Figure 8. A cross-system learning framework is proposed in Chapter 5 of [31], which optimizes the usage of both licensed (LTE) and unlicensed (WiFi) spectrum resources leveraging tools from reinforcement learning. In particular, decentralized load balancing algorithms are proposed to learn the best subband for transmission, and a traffic aware scheduler taking into account users' channel, and delay. Numerical results demonstrate a cost-effective integration of both LTE and WiFi whose gains depend on type of deployments (indoor/outdoor), density of users as well as the traffic classes (best-effort). For instance, while the proportionally fair scheduler cannot cope with the increasing number of UEs, the traffic-aware scheduling approach judiciously steers users' traffic in an intelligent and dynamic manner over both licensed and unlicensed spectrum, with a 160-fold increase for 300 UEs.

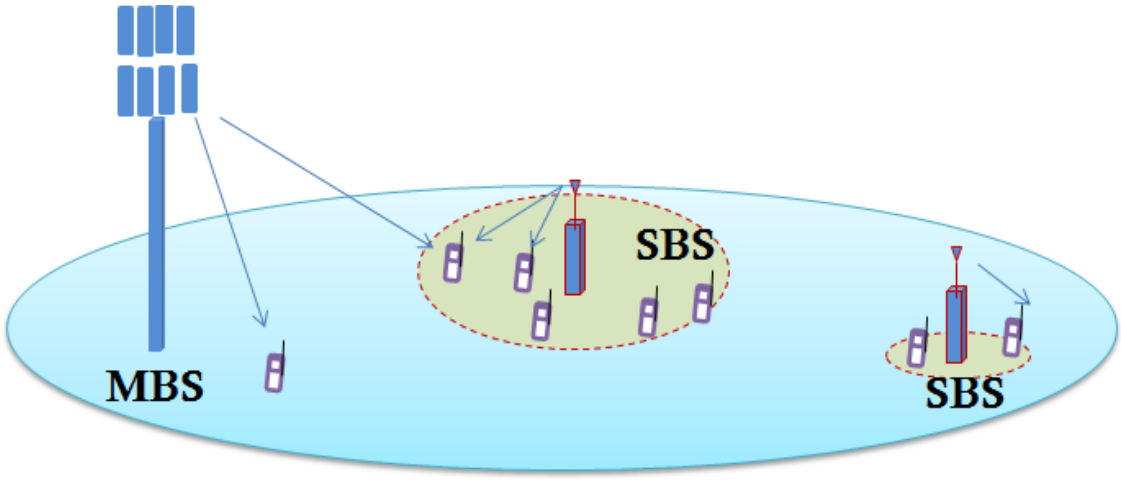


Figure 8: Multi-RAT coexistence assuming dual mode small cell base stations

In terms of impact on standardization, the proposed cross-system learning solution involves an integration of LTE and WiFi at the PDCP layer, which is similar to the LWA concept. This research topic will further gain importance as evidenced by the recent discussions in 5G research whereby the possibility of LTE operation in the unlicensed band (5GHz) referred to

LTE-U or LAA is being discussed. Already now, SHARING proposal namely the LWA variant solution will be part of the diverse set of solutions to be agreed upon between 3GPP and IEEE.

2.3 Energy Saving Mechanisms

Currently, 3% of the world-wide energy is consumed by the Information and Communication Technologies (ICT) infrastructure which generates about 2% of the world-wide CO_2 emissions, comparable to the world-wide CO_2 emissions by all commercial airplanes or one quarter of the world-wide CO_2 emissions by all vehicles. The ICT sector's carbon foot print is expected to quickly grow to 1.4 Giga ton CO_2 equivalents by 2020, nearly 2.7% of the overall carbon footprint from all human activities. Therefore, lowering energy consumption of future communication systems and networks meanwhile increasing its total energy-efficiency is demanding greater attention not only within government, industry and standardization bodies but also within international research communities.

The cost of energy is also one important component contributing to the total OPEX for the cellular network operators. Furthermore, the total cost of energy is expected to increase together with the network densification caused by the rapidly increasing traffic. Therefore, innovations aiming to reduce the total network energy consumption will be increasingly important for the future. Yet another challenge is that the state-of-the-art base stations are consuming a considerable amount of power even when they are idle, i.e., not serving any users. This is caused partly by the fact that the base station hardware is kept powered on during the time instants (symbols) when the base station has nothing to transmit. Another fact contributing to the idle state power consumption is the design of the downlink control signaling, which has to be "continuously" transmitted.

To target the objectives listed above, SHARING has provided innovative results on new opportunities, challenges and concepts for energy saving in heterogeneous network deployments both from the radio resource management and the base station hardware point of view. The innovations can be divided into three main groups:

- Fast, symbol-level base station micro sleep scheme.
- Longer term base station ON/OFF or sleep mode mechanisms.
- Improvements in the energy efficiency of the base station hardware, i.e. the power amplifier.

In case of the base station micro sleep scheme, see Section 2.1 in [4], the base station power amplifier is switched off during the periods (symbols) when the base station has nothing to transmit. Since all the necessary downlink control signals are transmitted, the cell stays accessible to the users and hence, the system performance will not be affected. This also means that micro sleep can be applied to all cells within the network. Furthermore, the micro sleep scheme can be implemented without any impact on the network architecture or standardization. As indicated by the evaluation results in Section 2.1 of [4], the proposed micro sleep can result in 16% lower average energy consumption.

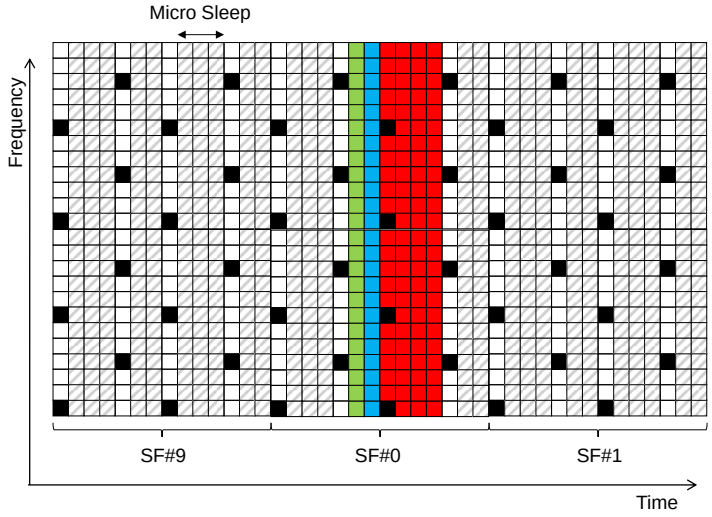


Figure 9: In case of micro sleep the base station power amplifier is switched off in between the downlink control signals.

In case of the base station ON/OFF or sleep mode scheme, underutilized base stations are switched off during low-traffic periods. Since all the downlink transmissions are deactivated, the cells are not accessible to the users, and hence, the maximum system performance will potentially be affected. Therefore, the main challenge becomes to decide which cells should be switched off or on.

SHARINGhas proposed four different innovations to solve the problems related to base station ON/OFF scheme: base station sleep mode utilizing measurements, compensation-based ON/OFF energy saving, dynamic cell ON/OFF and opportunistic sleep mode.

- A base station sleep mode scheme applicable for heterogeneous network deployments is proposed in Section 2.1 of [4]. In the proposed scheme, a small cell is switched to sleep mode as soon as it has been idle for a given time. Hence, the decision to enter the sleep mode is made by the small cell itself, without the control of the neighbouring cells. When it comes to the reactivation of a sleeping small cell, the decision can be based either on the monitored macrocell load or on the level of uplink interference measured by the sleeping small cell itself. Furthermore, a combination of these two schemes is possible. It is envisioned that the proposed sleep mode scheme can be implemented without any impact on architecture or standardization, utilizing the already existing X2 messages to wake up neighbouring cells. Finally, the obtained evaluation results indicate that the proposed scheme can offer a great energy saving potential, resulting up to 45% lower average network energy consumption.
- In the context of base station sleep mode, the compensation of coverage by nearby base stations is a key issue. Therefore, a concept for a compensation-based ON/OFF energy saving has been proposed; see Section 4.1.2 of [2] and Section 2.2 of [4]. When triggering the system state into energy saving state, the compensating base stations are transmitting with full power, while the non-compensating (energy saving) base stations are set to OFF. The compensating base stations are determined through dominant sets calculated from the neighborhood graph of the considered deployment scenario, i.e. campus of base stations as visualized in Figure 10. The presented evaluation results indicate that the proposed ON/OFF energy saving technique based on dominant sets is effective in both energy saving and the optimization of the overall system performance in terms of average throughput. Compensation-based ON/OFF energy saving technique based on maximum independent sets of the graph is showing the best performance. Furthermore, the achievable energy saving gain is shown to depend on the size of the campus: 15% gain for a large campus and 28% for small and medium sized campus. The compensation based ON/OFF energy saving is having an impact on 3GPP RAN WG3 standardization since it introduces new messages over the X2 interface signaling. The

impact of this innovation over the legacy LTE-A architecture is to introduce a new node, a coordination gateway, that proxies the X2 interfaces of the deployment. The coordination gateway receives the measurement reports from the user terminals of the network and/or the base stations, calculates the compensation base stations and sends commands for the energy saving base stations to be set OFF when the traffic intensity is low in the coverage.

Description of the latest research results on this innovation can be found in the second part of this deliverable (Section 3.3.3).

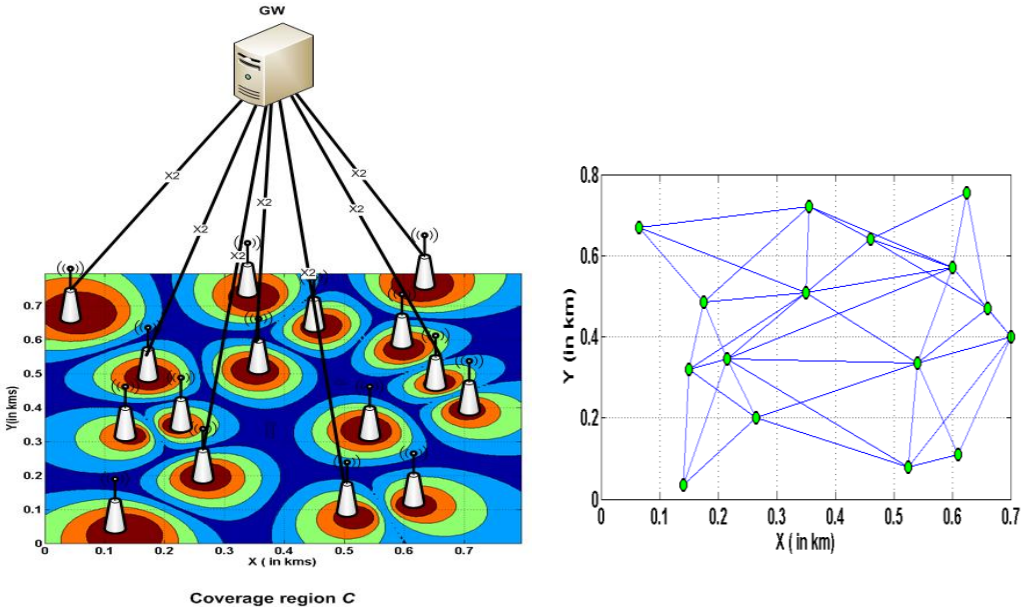
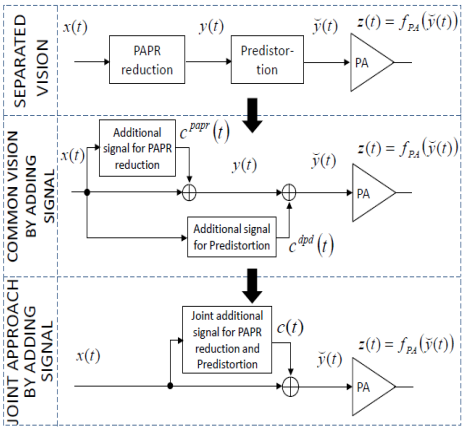


Figure 10: Campus deployment scenario and coverage map, and the corresponding graph representation.

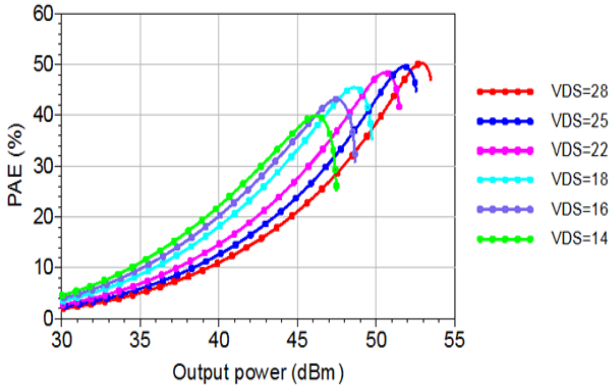
- Dynamic cell ON/OFF (see Section 2.3 of [4]) presents a study on the applicability of a cell ON/OFF scheme utilizing the estimated cell load as a measure to either switch cells off or on. Both a centralized and a distributed scheme has been proposed and evaluated. In case of the centralized scheme, the controlling node will transmit activation/deactivation commands to the cells it is controlling. In case of the distributed scheme, each cell can take the decision by itself, based on the known status of the neighboring cells. Hence, in both cases, the cells need to exchange either control or cell load/status information between each other, which may require some new X2 messages to be standardized. Finally, similar to the other studies, the obtained evaluation results clearly indicate the gain of being able to deactivate underutilized cells during low-traffic periods.
- A concept called opportunistic sleep mode has also been proposed and evaluated, see Section 2.4 of [4]. The proposed scheme enables the small cell base stations to optimize their downlink performance while balancing the load among each other and satisfying the users' QoS requirements. The problem is formulated as a non-cooperative game among the base stations that seek to minimize a cost function which captures the tradeoff between energy expenditure and load. To solve this game, a distributed learning algorithm is proposed, where the base stations autonomously choose their optimal transmission strategies. The results indicate again that the gain of switching off small cells can be considerable. Furthermore, the results demonstrate that the gains become larger if the cells are divided into clusters of coordinated cells. In terms of standardization impact, indeed there is a need for local coordination via X2/Xn interface among picoeNodeBs.

The performance of base station ON/OFF scheme has also been confirmed by system-level simulations based on realistic radio propagation models (see Section 2.5 of [4], as well as the description in the second part of this deliverable (Section 3.3.4)), which highlight the interest in switching off small cells outside busy hours to optimize the network energy consumption. A given scenario shows that a 29% gain can be achieved when switching off small cells having

less than 3% of traffic load, without compromising the user QoS. In addition, theoretical measurements of the power saving are derived to validate smart base station ON/OFF switching principle in periods where the total amount of traffic served by the network is such that can be satisfied with just a subset of active base stations.



Joint crest factor and predistortion approach for energy saving



Power amplifier efficiencies with reconfigurable operating points

Figure 11: Innovations improving the energy-efficiency of the base station hardware.

When it comes to improving the energy-efficiency of the base station hardware, two innovations have been proposed (see Figure): merging the peak factor reduction and linearization steps prior to power amplification, and adapting the operating point of the power amplifier. Both innovations focus on the power amplifier, which is responsible for a majority of the total power consumption of the transmitter. Furthermore, these innovations can be implemented within the base station without any impact on the network architecture or standardization.

- The first innovation is proposing to smartly merge the peak factor reduction (also known as PAPR reduction) and linearization steps prior to power amplification, as described in Section 3.2 of [4]. Indeed, the smaller the peak factor of transmitted waveforms, the larger the power amplifier efficiency. In the presented concept, the PAPR reduction and linearization are introduced by adding signal based on the Bussgang's theorem, enabling a common vision of PAPR and predistortion and finally a joint approach by adding signal. Two algorithms to generate the additional signal have been proposed: a) algorithm based on polynomials' roots finding with clipping and filtering and b) algorithm based on error compensation on the useful sub-carriers. The presented evaluation results have shown that both algorithms are able to improve the transmission performance by jointly tuning linearity and peak factor reduction depending on needs and load, which directly mitigate the power amplifier consumption. Furthermore, compared to the polynomials-based algorithm, the error compensation algorithm has better performance and lower complexity.
- The second innovation proposes a method to adapt the operating point of the power amplifier in order to reduce the power consumption, see Section 3.3 of [4]. Indeed, the power amplifier is usually designed for maximum load, performing the highest energy efficiency at maximum RF output power. Nevertheless, when the traffic load decreases, lower RF output power levels are required and the energy efficiency gets worse due to the power amplifier characteristics. Therefore, a reconfigurable power amplifier is proposed defining different operating points which can be optimized to different power levels, providing energy savings at medium and low traffic load in the base station. Performance of the proposed scheme has been evaluated for different transistor technologies and

different output power levels corresponding to different deployment (macrocell and microcell) scenarios. Based on the results, the power amplifier efficiency could be improved by 20% using the proposed dynamically reconfigurable solution compared to the conventional power amplifier. As power amplifier could represent 64% DC power consumption breakdown in macrocell scenarios, and 47% in microcell scenarios, the energy efficiency improvement in an urban area could be around 13% and 9% respectively.

2.4 Advanced Spectrum Management

Self-organization based spectrum and resource management is an important building block for the design of heterogeneous networks based on LTE-A access as well as a key technology for the development of future 5G resource allocation systems. In a heterogeneous network deployment, several nodes with different capabilities are expected to coexist in the network and the problem of spectrum resource management is to use this network heterogeneity for improving the performance, i.e. the QoS of the end user and the resource usage in the network.

SHARING has studied several techniques for spectrum management and resource allocation, which has led to the definition of advanced carrier aggregation techniques as one of the key innovations of Work Package 4. This key innovation extends the LTE-A carrier aggregation paradigm through:

- Choosing primary and secondary component carriers in order to improve the resource usage in the network and combining the carrier aggregation with coordinated multipoint transmission.
- Combining dual connectivity with carrier aggregation in order to achieve both performance improvement and flexibility of the system.

Coordinated carrier aggregation (CCA) described in Section 2.1.1 of [28] combines the advantages of carrier aggregation with the advantages of coordinated multipoint transmission and reception (CoMP). In coordinated carrier aggregation, primary and secondary component carriers are chosen for a cluster of randomly deployed home base stations such as to minimize the interference for coordinated multipoint transmission and the number of resources used for the cluster. The spectrum optimization is formulated as a graph coloring problem on an interference graph obtained from the aggregated measurements of the crowd of the user terminals deployed in the coverage area of the cluster. Figure 12 illustrates an example of such interference graph construction and a corresponding node coloring for a cluster of 25 home base stations deployed outdoor.

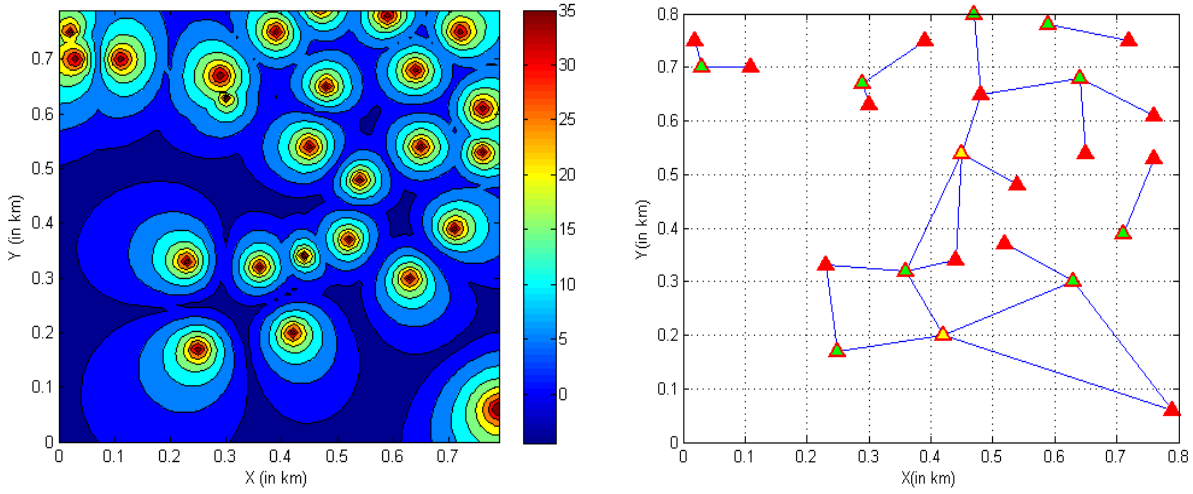


Figure 12: Coverage map and interference graph coloring of a cluster of 25 home base stations

Two popular graph coloring heuristics have been considered for the performance evaluation of the CCA that are based on the degrees of the nodes of the interference graph. In Figure 13, the cumulative distribution of the throughput of the user terminals in the coverage of the cluster where it is seen that the proposed CCA improves the throughput performance of 5%. The spectrum usage over the cluster, i.e. the number of component carriers used to achieve this throughput improvement is optimized of 60%. The proposed innovation introduces additional signaling over the X2 interface and may need additional node in the legacy architecture of LTE-A.

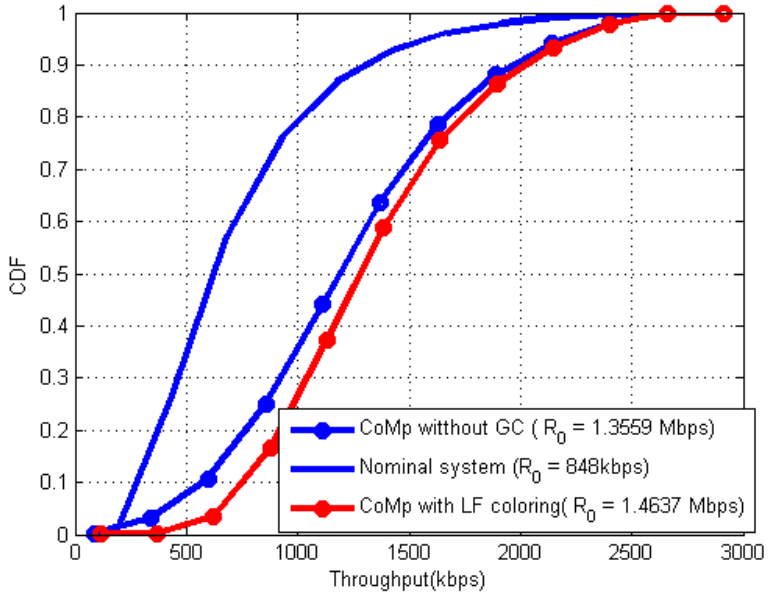


Figure 13: Throughput CDF of the cluster of 25 base stations

Description of the latest research results on coordinated carrier aggregation can be found in the second part of this deliverable (Section 3.4.1).

Multiflow carrier aggregation combines the advantages of carrier aggregation with the dual connectivity and has been presented in Section 2.1.2 of [28].In multiflow carrier aggregation (MF-CA), multiple base stations (from different tiers) simultaneously transmits data to a UE on different component carriers through dual connectivity transmission scheme. In the MF-CA, the control plane (C-plane) is always provided at low frequency band at the macro base stations to maintain good connectivity and mobility while the user plane (U-plane) is provided by both the macro base stations and the small cells (deployed at higher frequency bands) for data transfer. The small cells are not transmitting reference signals in this innovation since the C-plane is transmitted only through the macro base stations. The proposed MF-CA concept is illustrated in Figure.

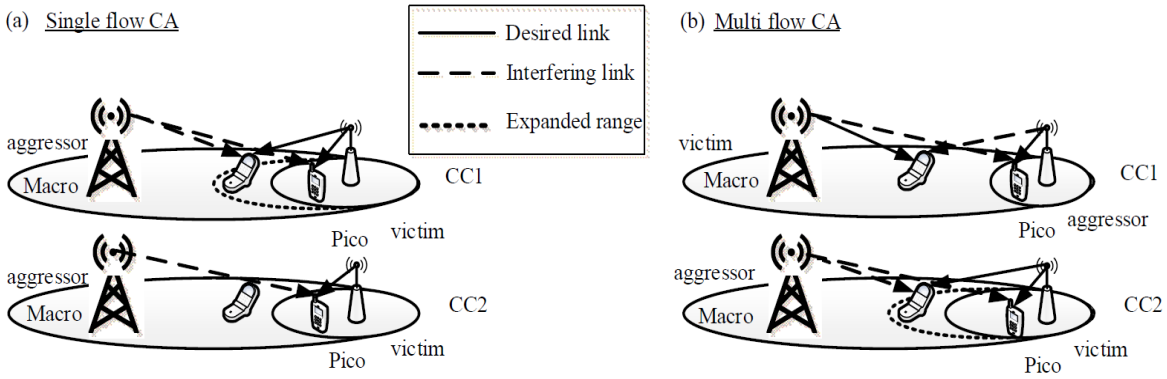


Figure 14: Single and multiflow carrier aggregation techniques

The optimization of the time domain inter-cell interference coordination (e-ICIC) is performed in multiflow carrier aggregation through a dynamic Q-learning algorithm that finds the network parameters, e.g., cell range extension, powers, component carriers, with the global objective of maximizing the sum rate of the system. Figure 15is showing the performance as a function of the user distribution in the macro coverage. The performance is evaluated for different learning strategies and for multiflow and single flow carrier aggregation. It is seen from the results that the multiflow CA with dynamic Q- learning is showing the best sum rate performance.

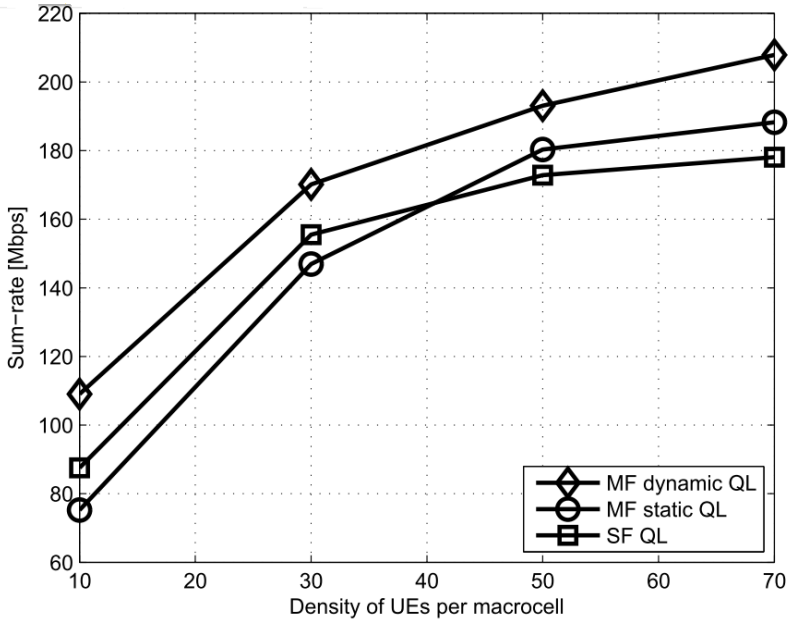


Figure 15: Sum rate performance of single and multiflow CA techniques

3 SECTION B

While the purpose of Section A is to showcase a summary of the most innovative solutions carried out in work package 4, section B presents the latest innovations carried out in work package 4.

3.1 Intra-system Radio Access Offloading

In order to address the fast data traffic boost experienced in today's mobile cellular networks, several techniques and deployment strategies have been identified, allowing the network to cope with localized peak of traffic load and optimizing the utilization of network resources. These techniques are known as traffic offloading, and in the context of SHARING Task 4.1 they are carried out within the same radio access technology. Among these technologies are dynamic TDD via decoupled uplink and downlink, traffic-aware feedback strategies in MIMO systems, and interference management in TDD small cells taking into account queue stability.

3.1.1 Dynamic TDD via decoupled uplink and downlink

Data traffic in cellular mobile networks can vary dynamically and rapidly due to the arrival and departure of users requesting different services. Since the next generation cellular systems will be based on a dense deployment of small cell base stations (SCBSs) and a massive number of connected devices, more spatio-temporal traffic load variations are expected. These variations are expected to result in asymmetric and dynamically changing uplink (UL) and downlink (DL) traffic loads. Generally, a time-division duplex (TDD) system has ability to deal with asymmetric UL/DL traffic. In particular, a flexible splitting of UL and DL resources can dynamically vary the percentage of subframes allocated to both UL and DL (duty cycle). The main challenge that limits the performance of dynamic TDD is the cross-link interference resulting from two cells transmitting in opposite directions. To avoid cross-link interference, strict synchronization within clusters of locally-coupled SCBSs (in terms of mutual interference) is typically considered to align their TDD configuration. This limits the flexibility of handling asymmetric load conditions in neighboring cells. The proposed dynamic TDD framework consists of the following:

1) A dynamic load-aware base station clustering algorithm: where clusters of mutually-coupled SCBSs are formed based on their distance and UL and DL load. Each group of SCBSs aligns its UL/DL configuration and controls its power allocation to mitigate cross-link interference.
2) An UL/DL decoupled user association to ensure flexible UL and DL load balancing within clusters. By decoupling the UL and DL user association, each user connects to different cells within the same cluster for both UL and DL, based on the traffic load in each direction.
3) A dynamic TDD optimization scheme level:where each cluster optimizes its TDD frame and UL/DL power levels in a distributed manner. Since the network configuration of a given cluster affects the performance of other clusters, we formulate the inter-cluster interaction as a noncooperative strategic game among clusters of SCBSs. Then, we propose an online learning algorithm with regret to reach equilibrium for this game.

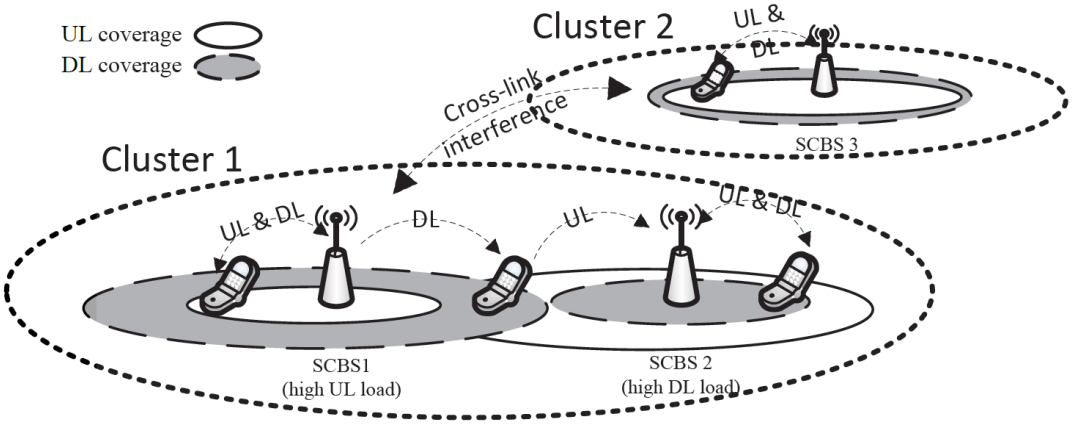


Figure 16: An illustration of the proposed clustering and decoupled user association scheme. Different UL and DL coverage areas are considered in the case of neighboring cells with asymmetric load.

The proposed framework proceeds as follows: A dynamic clustering of SCBSs is performed when the UL and DL load conditions significantly change. After performing clustering, each cluster learns its UL and DL loads autonomously relying on local observations. Using this information, each cluster finds the optimum TDD configuration and power allocation so that clusters are not severely interfering with each other. In this case, SCBSs within a cluster align their TDD configuration and limit their power allocation to the limit determined by the cluster. Furthermore, an intra-cluster power control to limit the interference between SCBSs belonging to the same cluster is used. With this scheme, each cluster can find the optimum TDD configuration that matches its current load conditions. And load is balanced within the cluster using the load-aware UL/DL decoupled user association. The proposed framework is explained in Figure 17. In the first step, clustering is performed to group SCBSs into clusters as explained. User association is performed in both link directions. Subsequently, each cluster starts to learn its optimum TDD strategy based on local observations and perform its inter-cluster power allocation. The user association process is performed only once in the beginning of each learning phase in order not to interrupt the convergence of the learning process.

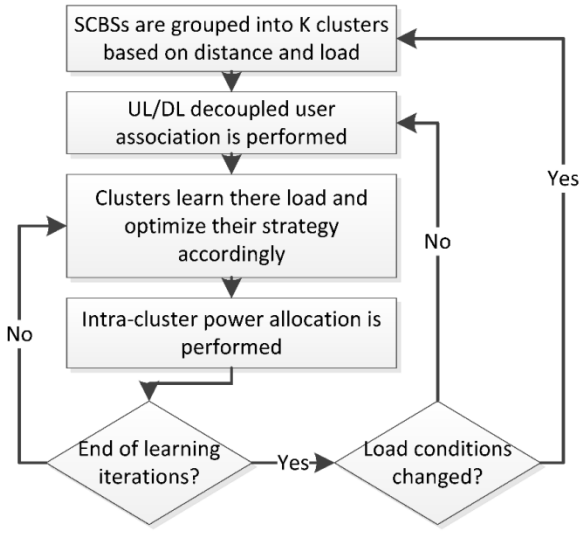


Figure 17: Illustration of the proposed dynamic TDD framework.

We evaluate the performance of the proposed dynamic UL/DL TDD scheme. To show the gains of the proposed scheme, we consider three baseline schemes; 1) fixed TDD scheme, in which we assume that SCBSs have the same synchronous frame of 3:3 UL/DL subframe ratio, 2) instantaneous traffic scheme, in which the ratio of UL and DL schemes are adapted following the instantaneous buffer size of UL and DL, and 3) no-clustering scheme, in which learning is performed in each SCBS separately without clustering or UL/DL decoupling. For the proposed scheme with clustering, we simulate it with and without the proposed decoupling and load-aware user association; we denote them as decoupled, and coupled, respectively. In the baseline schemes, user association is performed based on received signal strength.

Figure 18 illustrates the effect of traffic load on the performance of the proposed scheme by comparing it against the baselines schemes for different load conditions. We vary the load by changing the value of the packets arrival rate per user, for the same number of users per SCBS. The first plot in Figure 18 shows the packet throughput performance for the case of 40 users per SCBS and UL-to-DL ratio of 13 dB. It is shown that increasing the traffic load degrades the performance of the different schemes due to experiencing higher interference levels. However, the proposed scheme achieves up to 43% and 50% gains over the fixed and instantaneous schemes, respectively. The figure also shows how clustering and UL/DL decoupling behave under different traffic loads. At low loads, Decoupling slightly improves the

performance as compared to the coupled scheme. This is because the load has low impact on the user association in low traffic. Therefore, the performance is close to the coupled scheme. However, in the case high traffic loads, the gap increases between the coupled and decoupled cases. In the second plot in Figure 18, we investigate the average delay performance in light, medium and high load conditions, respectively. We can clearly see the gain of the proposed scheme in terms of lower service delays in different load conditions. While clustering and UL/DL decoupling gains are more observable in higher traffic loads, the proposed scheme also achieves lower service delays even in lightly loaded systems.

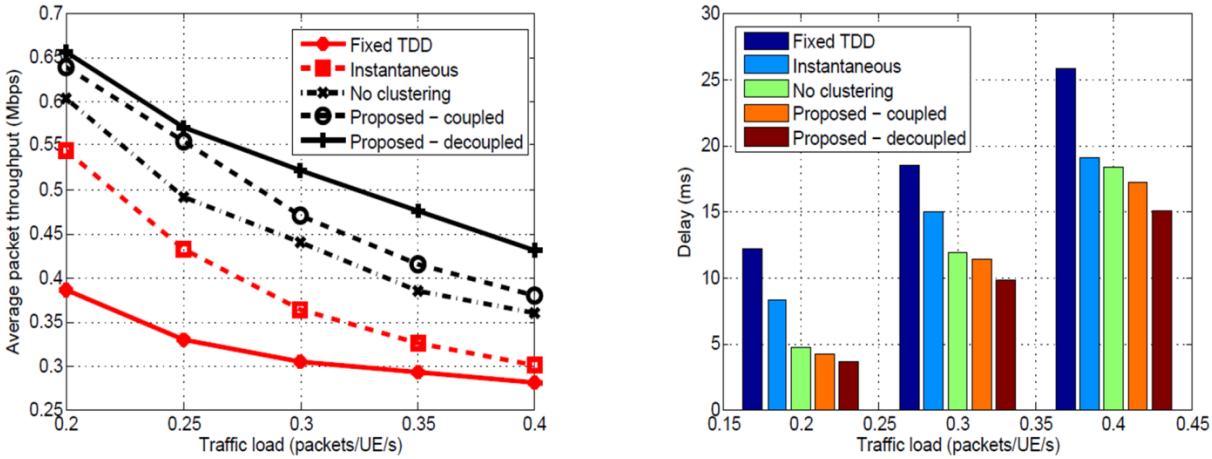


Figure 18: Packet throughput and average delay performance of the proposed scheme at different traffic load condition.

3.1.2 Backhaul offloading via caching: a real case study

Driven by the surge in mobile video content, current solutions based on the reactive paradigm in which contents are stored at the core network are obsolete and fail to cope with the increasing user demands. At the same time motivated by the highly predictable human behavior, the proposed solution concept exploits the contextual information (such as user's viewing history, location information, social ties, etc.) to predict users' patio-temporal demand to proactively cache judiciously selected contents at the network edge. Hence, users' demands are highly satisfied yielding low latency, backhaul offloading gains and higher users' QoE. In this work, in a joint collaboration with AVEA, OULU and SUPELEC, we have validated our proposed caching algorithms using real data traces from AVEA. In the numerical setup, we assume that *D* contents are requested from the processed data (namely final-traces table) over a time interval of 6 hours 47 minutes. Information on arrival times (FRAME-TIME), requested contents (HTTP-URI) and content sizes (SIZE) are taken from the final traces-table. Then, the requests are pseudo-randomly assigned to *M* base stations.

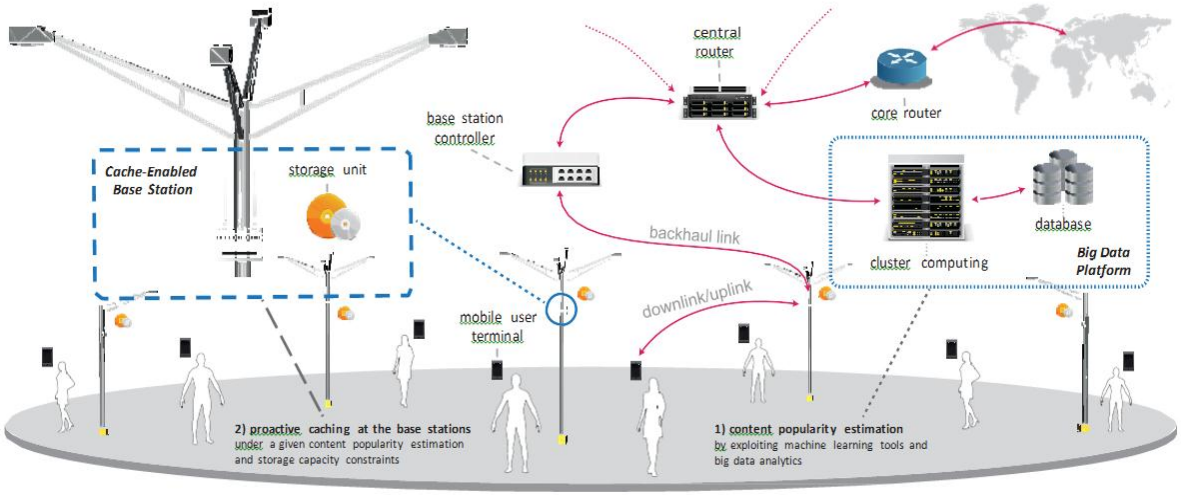


Figure 19: Illustration of the proposed architecture leveraging storage and cluster computing at the controller.

In order to solve the backhaul offloading problem redundant transmission from the core network of the same contents need to be avoided. This hinges on proactively predicting file popularity matrix based on spatio-temporal patterns. In particular let P, denote the content popularity matrix and caching strategy by X. In order to construct P, the following methods are examined in the numerical setup:

GroundTruth: The content popularity matrix **P** is constructed by considering all available information in the final traces-table. The matrix has 6.42% of rating density in total.

Collaborative Filtering: 10% ratings available in the final traces-table are picked uniformly at random for training of the content popularity matrix estimation. Then, based on the constructed model, the remaining missing entries/ratings in the traces are predicted via the regularized singular value decomposition (SVD) from the collaborative filtering (CF) method. After constructing **P** based on these above-methods, the cache decision (modelled by the matrix is made by storing the most-popular contents greedily at the SBS until no storage space remains. Then, the base stations serve their users starting from the first content request at t=0 to the delivery of the last request. Having these contents cached proactively at the SBSs at t=0, the requests are then served until all of the contents are delivered.

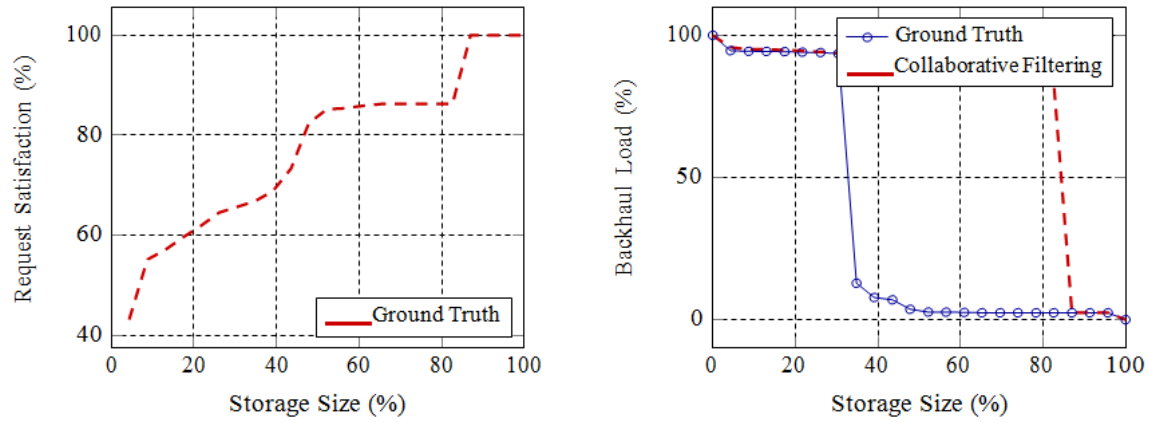
As regards the performance metrics: i) *request satisfaction* is defined as the percentage of contents delivered at a given rate (which should be at least equal to the bit rates of contents) over all content requests, and ii) *backhaul load* corresponds to the percentage of the traffic passing over the backhaul links over the total possible traffic volume induced by the content requests. The impact of storage size on the users' request satisfaction is plotted in Figure 20. Therein, 0% of storage size corresponds to no caching, whereas 100% of storage is equivalent to caching the entire library (17.7GByte). In the figure, we note that the users' request satisfaction has a monotonically increasing behavior, and somewhat intuitive, 100% of satisfaction is achieved in both methods when the complete library is stored. However, a performance gap between the ground truth and CF is observed until 87% of storage size which is mainly due to the estimation errors. For instance, when the base stations have 40% of storage size for caching, the ground truth yields 92% of satisfaction whereas the performance of CF stays at 69%.

Figure 20 also shows the impact of storage size on the backhaul load/usage. It can be seen that both methods yield less backhaul load (namely higher offloading gains). For instance, having 87% of storage size at the base stations, both methods offload 98% of backhaul. However, the ground truth out performs the CF method since it has complete information of the content ratings. On the other hand, after a certain storage size, a dramatical decrease of backhaul is observed in both approaches. Compared to previous works which mostly consider identical content sizes, we are dealing with real traces with non-identical content sizes. In

other words, popularity based caching used here fails to capture relatively less popular (but sufficiently big sized) contents, thus causing higher backhaul loads in case of cache misses. These points show the importance of taking into account contents sizes when making caching decision.

Table 1: Description of all parameters for validation purpose

Parameter	Description	Value
T	Time duration	6hours47minutes
D	Nr. of requests	422529
F	Nr. of contents	16419
M	Nr. of small cells	16
L_{min} L_{max}	Min.size of a content Max.size of a content	1Byte 6.024GByte
$B(f)$	Bit rate of content f	4Mbyte/s
$_{m}C_{m}$	Total backhaul link capacity	3.8Mbyte/s
$_{m}C_{m}$	Total wireless link capacity	120Mbyte/s



(a) Evolution of satisfaction with respect to the storage size.

(b) Evolution of backhaul usage with respect to the storage size.

Figure 20: Impact of proactive caching on both users satisfaction and backhaul load

In this study, it was demonstrated that caching is a viable strategy to offload backhaul traffic, which will be an important cornerstone for upcoming 5G networks. This is already obvious from current activities at ETSI on mobile edge computing.

3.1.3 Traffic-aware feedback strategies in MIMO systems

In this work, we develop a novel feedback strategy for MIMO systems taking into account the traffic patterns. The key performance metric considered in this work is the queuing stability region. We consider the downlink of a wireless cellular network where Zero-Forcing (ZF) precoding is used at the transmitter. Acquisition the channel state is done via uplink training in the Time Division Duplex (TDD) mode. Due to limited feedback time, only a subset of users can perform uplink training and the selection of this subset must be done such that the queuing stability region is as big as possible. In this work, we have developed a decentralized

feedback scheme where the users must compete to send their uplink training sequences. We show that, using infrequent signaling between the BS and the users, the decentralized policy outperforms the centralized one. This result implies that, as far as stability is concerned, the users must be involved in the active user selection and feedback decision.

3.1.3.1 System Model

We consider a single cell wireless system where a BS equiped with N antennas serves K users having each one antenna. The channel is assumed to be a Rayleigh block fading, i.e. the channels stay constant in a slot of T_s channel uses and change independently in the next slot. The channel of user k can be written as an N-dimensional complex vector $\mathbf{h}_k(t) = \sqrt{\bar{g}_k}\hat{\mathbf{h}}_k(t)$ where $\hat{\mathbf{h}}_k(t) : CN(0, \mathbf{I}_N)$ is a complex Gaussian vector that represents the small scale fading and $\bar{g}_k$ represents the channel gain due to large scale fading. The channel vector can then be written as $\mathbf{h}_k(t) = \sqrt{g_k(t)}\mathbf{u}_k(t)$, where $\mathbf{u}_k(t)$ is an isotropically distributed unitary vector and $g_k(t) = \|\mathbf{h}_k(t)\|^2$ is the channel magnitude. Multiple users can be served simultaneously by the BS thanks to MISO. This is allowed by the use of linear precoding at the BS.

We denote the linear precoder by $\mathbf{W}(t) = [\mathbf{w}_1(t),...,\mathbf{w}_K(t)]$. The signal received by user k at slot t is

$$y_k(t) = \mathbf{h}_k^H(t)\mathbf{w}_k(t)s_k(t) + \sum_{j\neq k}\mathbf{h}_k^H(t)\mathbf{w}_j(t)s_j(t) + n_k(t)$$

where $s_j(t)$ is the data symbol intended to user j, $n_k(t) : \mathrm{CN}(0, \sigma^2)$ is the white noise at the receiver or user k. The achievable rate of user k at slot t is $r_k(t)$ (bits/channel use). We assume that a finite number of modulation and coding schemes is used and therefore the achievable rates take values from the set $\mathsf{R} = \{R_1,...,R_l,..,R_L\}$, with $R_1 = 0$, $R_{l-1} < R_l$. Also we assume that rate R_l can be supported if the SINR at the receiver is above some threshold S_l.

We denote by $\mathsf{F}(t)$ the set of users that are scheduled at slot t, $F(t) = |\mathsf{F}(t)|$ and $k(1),..,k(i),...,k(F(t))$ the corresponding permutation of user indices. Also, define

$$\mathbf{H}_{k(i)}(t,\mathsf{F}) = [\mathbf{h}_{k(1)}(t),...,\mathbf{h}_{k(i-1)}(t),\mathbf{h}_{k(i+1)}(t),...,\mathbf{h}_{K(F(t))}(t)].$$

The precoding vector for user $k(i), \forall i \in \{1,..,F(t)\}$ is given as the projection of the channel of the user on the nullspace generated by the channels of the other users:

$$\mathbf{w}_{k(i)}(t) = \sqrt{\frac{P}{F}}\frac{\left(\mathbf{I}_N - \mathbf{H}_{k(i)}(t,\mathsf{F})(\mathbf{H}_{k(i)}^H(t,\mathsf{F})\mathbf{H}_{k(i)}(t,\mathsf{F}))^{-1}\mathbf{H}_{k(i)}^H(t,\mathsf{F})\right)}{\left\|\left(\mathbf{I}_N - \mathbf{H}_{k(i)}(t,\mathsf{F})(\mathbf{H}_{k(i)}^H(t,\mathsf{F})\mathbf{H}_{k(i)}(t,\mathsf{F}))^{-1}\mathbf{H}_{k(i)}^H(t,\mathsf{F})\right)\mathbf{h}_{k(i)}(t)\right\|}\mathbf{h}_{k(i)}(t).$$

The corresponding SINR for user k is then

$$\mathrm{SNR}_{k(i)}(t) = \frac{P\|\mathbf{h}_{k(i)}(t)\|^2}{\sigma^2 F(t)}\mathbf{u}_{k(i)}^H(t)(\mathbf{I}_N - \mathbf{H}_{k(i)}(t,\mathsf{F}(t))(\mathbf{H}_{k(i)}^H(t,\mathsf{F}(t))\mathbf{H}_{k(i)}(t,\mathsf{F}(t)))^{-1}\mathbf{H}_{k(i)}^H(t,\mathsf{F}(t)))\mathbf{u}_{k(i)}(t).$$

One can see that the transmission requires an accurate channel state information at the BS. This must be acquired by using feedback or training from the receivers. In TDD mode (which our case here), the channel acquisition is done by uplink training from the users.

One can see that when the CSI is acquired for too many users, the time resources consumed by the CSI acquisition is high and there is little time left to transmit the data in each timeslot.

The objective of this work is to develop opprtunistic CSI acquisition policies (called scheduling policies in the sequel) that determine the users that must send their training sequences. Our feedback policies take into account the fact that the traffic is dynamic. For that, we must first provide some description about the queuing model.

Each of the K users in the cell has an incoming traffic process $a_k(t)$ i.i.d. in time and independent across users with $a_k(t) < A_{max}$ almost surely and mean rate $\mathsf{E}\{a_k(t)\} = \lambda_k$. Let $q_k(t)$ denote the queue size in bits for user k at the beginning of slot t. Let $z_k(t)$ be the schedule in timeslot t, that is $z_k(t) = 1$ if user k is scheduled. We consider that $F(t)$ users are scheduled at each timeslot, with $F(t) \le N$. We denote by $\tau(t)$ the number of channel uses used for training and signalling in the slot t. The queues evolve then as follows, $\forall k \in \{1,..,K\}$:

$$q_k(t+1) = \left[q_k(t) - \lfloor (T_s - \tau(t)) r_k(\mathbf{W}(t), \mathbf{H}(t)) \rfloor z_k(t)\right]^+ + a_k(t), t \ge 0$$

Where $r_k(\mathbf{W}(t), \mathbf{H}(t))$ is the bit rate of the scheduled user. In this work, we are interested in the stability of the system. Formally, its definition is as follows:

Definition (Strong Stability) *A system is said to be strongly stable if*

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathsf{E}\{q_k(t)\} < \infty, \forall k \in \{1,..,K\}$$

Definition (Stability Region) *The stability region Λ of a resource allocation policy is defined as the set of vectors of mean arrival rates $\boldsymbol{\lambda} = [\lambda_1,..,\lambda_K]$ for which the system is stable under this policy.*

3.1.3.2 Proposed feedback policies

Before describing the scheduling and training policies, we provide the following defintions and assumptions. We define R_0 (bits per channel use) the rate of the control information transmitted by the BS. from the to the users can be broadcasted. We denote by βF the length (in channel uses) required to perform uplink training by F users where β is an integer. The downlink pilots are assumed to require a length of β_p channel uses.

Centralized policy

In the centralized policy, the BS decides which users must send the uplink training using the statistics of the channel conditions of all users and the knowledge of their current queue length. For this scheme, the BS sends also a downlink pilot of length β_p channel uses to allow the users to estimate their channel and decode the control messages. After the pilot, there is a control phase, where the BS broadcasts the set of scheduled users at this timeslot. This control phase takes $\beta_c F$ where

$$\beta_c = \frac{\log_2 K}{R_0}$$

The BS selects the set of scheduled users as follows:

$$\mathsf{F}(t) = \arg\max_{\mathsf{F}\in 2^K} \left\{ (T_s - (1 + \beta_p + \beta_c F + \beta F)) \sum_{k\in\mathsf{F}} q_k(t) \mathsf{E}\{r_k(t) \mid F\} \right\},$$

where the expectation is taken with respect to the channel statistics.

Decentralized policy

The decentralized policy is based on the observation that each user knowns instantaneously its own channel realization thanks to the downlink training sent by the BS. The decentralized policy is then based on the following: (i) the BS periodically broadcasts quantized values of the queue lengths of the users, (ii) the BS decides on the *number* of users to be scheduled in the next T slots, and iii) the users compete between each other to send their uplink training

In more details, the BS broadcasts quantized versions of the queue lengths of the users each T slots, i.e. $\mathbf{q}(mT), m = 0,1,...$

In addition, the BS broadcasts the number $F(mT)$ of users to report the channel each timeslot for the next $T-1$ consecutive timeslots (one can refer to [22] for more details). Let $\hat{\mathbf{q}}(t) := \tilde{\mathbf{q}}(T\left\lfloor \frac{t}{T} \right\rfloor)$ be the most recent information about the queue state of the users. At each timeslot the BS sends a downlink pilot with duration β_p channel uses in order to let the users estimate their channels. The users will then compete between each other to send their uplink training using the following contention algorithm. We assume that the contention period has a length of τ channel uses. Each user k waits until time

$$\tau_{k'} = \frac{\tau_{c'}}{\hat{q}_k(t)\mathsf{E}\{r_k(t) \mid g_k(t), F(t)\}}.$$

and sends a short contention access signal. Note that in the above equation $g_k(t) = \left\|\mathbf{h}_k(t)\right\|^2$. This implies that the F users with the biggest values of $\hat{q}_k(t)\mathsf{E}\{r_k(t) \mid g_k(t), F(t)\}$ will get scheduled. Once the contention period is over, the F first users to have a signal broadcasted will perform uplink training and then the BS transmits to them using Zero Forcing. The total time for transmission is then $\left(T_s - \beta_p - \tau - (\beta + \beta_c)F(m)\right)$ channel uses.

Mixed policy

The mixed policy is simply a combination of the centralized and decentralized policies. The BS at every slot $t = mT, m = 0,1,...$ decides, for the next $T-1$ slots, that either the decentralized policy will be used, with the optimal number of users to get scheduled as above, or select a set $\mathsf{F}(m)$ of users to schedule according to the centralized policy. This decision is made based on which of the two policies will maximize the quantity $\mathsf{E}\left\{\sum_{k=1}^{K} \hat{q}_k(mT)(T_s - \tau(t))r_k(t) \mid \mathbf{q}(mT)\right\}$, with the expectation taken over the channel distributions. The maximization of the above quantity is of type max weight policy that results in minimizing the Lyapunov drift of the queues. In other words, the above optimization will stabilizes the queues one can refer to [22] for more details).

3.1.3.3 Special Case: The 2-User MISO BC

In this section, we provide a complete characterization of the stability region for the 2-user MISO BC. The corner points of the region can be fully defined and the comparison between the regions achieved by the above three policies can then be performed. This is the main reason behind studying the 2-user case. The insights obtained by the 2-user case are very useful for the design of the system in the general case. In order to characterize the stability region, we first provide some definitions.

Let us define the following variables

$$\begin{aligned}\bar{p}(1) &= 1-\frac{\gamma\left(\frac{\hat{S}}{\bar{g}}\right);N}{\Gamma(N)}\\ \bar{p}(2) &= 1-\frac{\gamma\left(\frac{2\sigma^2}{\bar{g}P}\hat{S};N-F+1\right)}{\Gamma(N-F+1)}.\end{aligned}$$

$$\bar{\mu}_c(2) = (T_s-(\beta_p+2\beta_c+2\beta))\bar{p}(2)R$$

$$\bar{\mu}_c(1) = (T_s-(1+\beta_p+\beta_c+\beta))\bar{p}(1)R,$$

$$\begin{aligned}\bar{\mu}^{d,(1)}_{k(1)}(t) &= (T_s-(\beta_p+\tau_c+\beta))\bar{p}(1)R := \bar{\mu}_d(1)\\ \bar{\mu}^{d,(1)}_{k(2)}(t) &= (T_s-(\beta_p+\tau_c+\beta))\bar{p}(1)(1-\bar{p}(1))R.\end{aligned}$$

$$\bar{\mu}_d(2) = (T_s-(\beta_p+\tau_c+2\beta))\bar{p}(2)R$$

define further $\bar{\mu}_m(\{k\}) = (T_s-(\beta_p+\beta))\bar{p}(1)R$ and $\bar{\mu}_m(\{1,2\}) = (T_s-(\beta_p+2\beta))\bar{p}(2)R$.

Theorem *The stability regions of the centralized, decentralized and mixed policies in the 2 user case with one rate level R are given respectively as,*

$$\Lambda_c^{(2)} = \mathrm{Cv}\{(0,\bar{\mu}_c(1)),(\bar{\mu}_c(2),\bar{\mu}_c(2)),(\bar{\mu}_c(1),0)\}.$$

$$\Lambda_d^{(2)} = \left(1-\frac{1}{T}\right)\mathrm{Cv}\{(0,\bar{\mu}_d(1)),(\bar{\mu}_d(1)(1-\bar{p}(1)),\bar{\mu}_d(1)),(\bar{\mu}_d(2),\bar{\mu}_d(2)),(\bar{\mu}_d(1),\bar{\mu}_d(1)(1-\bar{p}(1))),(\bar{\mu}_d(1),0)\}$$

$$\begin{aligned}\Lambda_m^{(2)} = \left(1-\frac{1}{T}\right)\mathrm{Cv}\{\quad &(0,\bar{\mu}_m(\{k\})),(\bar{\mu}_d(1),(1-\bar{p}(1))\bar{\mu}_d(1)),(\bar{\mu}_m(\{1,2\}),\bar{\mu}_m(\{1,2\})),\\ &((1-\bar{p}(1))\bar{\mu}_d(1),\bar{\mu}_d(1)),(\bar{\mu}_m(\{k\}),0)\}.\end{aligned}$$

where Cv stands for the convex hull.*Proof.* The proof is provided in [22].

Comparison between the feedback policies

Proposition *A sufficient condition for the mixed policy to achieves a bigger stability region than the centralized policy is*

$$T > \max\left[\frac{T_c-\beta_p-\beta}{1+\beta_c},\frac{T_c-\beta_p-2\beta}{2\beta_c}\right].$$

In addition, this increase is with a factor at least equal to

$$\rho(T) = \frac{T}{T-1}\min\left[\frac{T_c-\beta_p-\beta}{T_c-(\beta_p+\beta_c+1+\beta)},\frac{T_c-\beta_p-2\beta}{T_c-(\beta_p+2\beta_c+2\beta)}\right]$$

that is $\Lambda_m^{(2)} \supseteq \rho(T)\Lambda_c^{(2)}$.

An illustration of both regions and the comparison is given in the following figure.

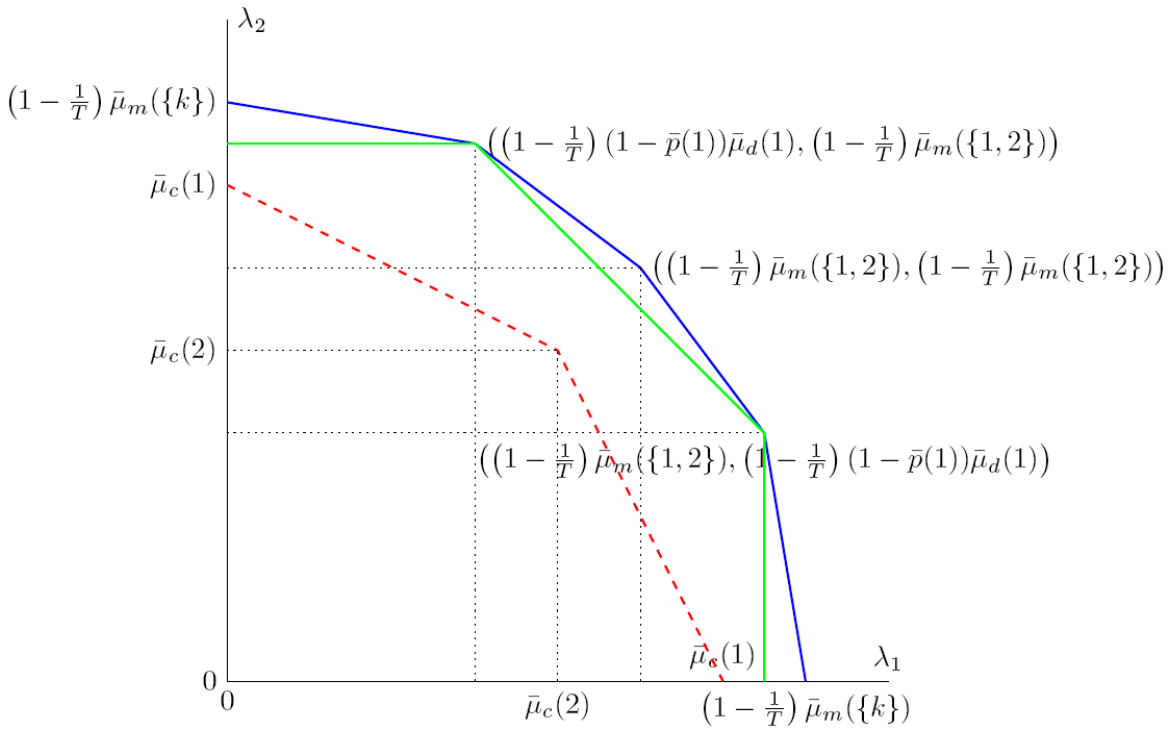


Figure 21: Comparing the stability regions of the centralized (dashed line) and mixed (continuous line) policies in the 2 user system with single rate level

3.1.3.4 General Case

In this Section we consider the general case with K users and L possible transmission rates. Although the general form of the stability region can be shown, the exact derivation of the corner points in this case is hard to obtain as one can see in the following results. However, a comparison between the regions obtained by the three feedback policies can still be performed and the obtained result is in line with the previous 2-user case results. Under some conditions, the decentralized and mixed policies outperform the centralized one.

We begin by considering the centralized policy and characterizing its stability region. We denote by $\mu_k^{(c)}(t)$ the service in bits given to user k at slot t under the centralized policy.

Theorem *The stability region Λ_c of the centralized policy is the convex hull that consists in all rate vectors $\boldsymbol{\lambda}=[\lambda_1,..,\lambda_K]$ for which there exist $0\le p(\mathsf{F})\le 1, \mathsf{F}\in 2^{\mathsf{K}}$ with $\sum_{\mathsf{F}\in 2^{\mathsf{K}}} p(\mathsf{F})=1$ such that*

$$\lambda_k < \sum_{\mathsf{F}\in 2^{\mathsf{K}}} p(\mathsf{F}) I_{\{k\in\mathsf{F}\}}(T_s-(1+\beta_p+(\beta+\beta_c)\,|\,\mathsf{F}\,|))\mathsf{E}\{r_k\,\|\,\mathsf{F}\,|\},\forall k\in\mathsf{K}.$$

Let us Denote by $\Lambda_d(F)$ the stability region of the decentralized policy for a fixed number of users to feed back every timeslot, i.e. setting $F(m)=F,\forall m\ge 0$. The following result is usefull to characterize the stability region.

Lemma *The region $\Lambda_d(F)$ consists in all mean arrival rate vectors $\boldsymbol{\lambda}\in\mathsf{R}_+^K$ for which there exist $\phi_{\mathsf{F}}(\mathbf{g})\ge 0$ such that*

$$T\lambda_k < (T-1)\int_0^\infty p_1(g_1)dg_1 \ldots \int_0^\infty p_K(g_K)dg_K \sum_{\mathsf{F}:|\mathsf{F}|=F} \phi_{\mathsf{F}}(\mathbf{g})\mathsf{E}\{\mu_k(t) \mid g_k, \mathsf{F}\}$$
$$:= (T-1)\mu_k^*(F) \quad .$$
$$\sum_{\mathsf{F}:|\mathsf{F}|=F} \phi_{\mathsf{F}}(\mathbf{g}) \le 1, \forall \mathbf{g} \in \mathsf{R}_+^K$$

The expectation is with respect to the directions of the channel vectors and $\mathbf{g}$ is the vector containing a realization of the channel magnitudes. The proof is given in [22].

Theorem *The stability region of the decentralized policy is the convex combination of the regions given by the previous lemma for every F, i.e.*

$$\Lambda_d = \mathsf{Cv}\{\Lambda_d(1), \ldots, \Lambda_d(F), \ldots, \Lambda_d(F_{max})\}.$$

We then characterize the stability region of the mixed policy. For that, denote

$$\Lambda_{c'} = \left(1-\frac{1}{T}\right)\mathsf{Cv}\left\{\frac{T_s-(\beta_p+\beta F)}{T_s-(1+\beta_p+\beta F+\beta_c F)}\Lambda_c(\mathsf{F})\right\},$$

where

$$\Lambda_c(\mathsf{F}) = \left\{\boldsymbol{\lambda} \in \mathsf{R}_K^+ : \lambda_k < \mathsf{E}\left\{(T_s-(\beta_p+\beta F+\beta_c F))r_k(t) \mid F\right\} I_{\{k\in\mathsf{F}\}}\right\},$$

Then we have the following result,

Theorem *The stability region of the mixed policy is*

$$\Lambda_m = \mathsf{Cv}\{\Lambda_{c'}, \Lambda_d\}.$$

Comparison between the feedback policies

By selecting a high enough value of T, we can guarantee that the mixed scheme increases the stability region of the system. Denote

$$\hat{m} = \min_{1\le F\le F_{max}} \left\{\frac{T_s-(\beta_p+\beta F)}{T_s-(\beta_p+\beta F+\beta_c F+1)}\right\}.$$

Then the following result holds:

Proposition *A sufficient condition for the mixed scheme to have greater stability region than the centralized scheme is*

$$T > \frac{1}{1-\hat{m}^{-1}}.$$

In this case, $\Lambda_m \supseteq \rho(T)\Lambda_c$ with $\rho(T) = \left(1-\frac{1}{T}\right)\hat{m}$.

Proof. The proof of all the above results is provided in [22].

The above theorem shows that by taking the signaling period very high (but finite), the mixed policy outperforms the centralized one.

3.1.4 Queueing Stability of interference management techniques in TDD systems

Interference is one of the key issues in wireless communication systems. Interference Alignment (IA) is introduced in [23] as one of the most efficient interference management

techniques. It is based on the concept of aligning the interferences in a reduced dimensional subspace, so that the desired signal can be transmitted with less interference (or no interference) in a larger subspace. The main disadvantage of IA is that it requires global channel state information (CSI) at each of the transmitting nodes. CSI acquisition is therefore an important issue in wireless networks using IA technique. Although the performance of IA has been studied in the literature, the impact of dynamic traffic has been widely ignored. In this work, we analyze the performance of IA taking into account the traffic pattern and the CSI acquisition cost. We consider a network where multiple transmitter-receiver pairs operate in TDD mode and apply the IA technique under backhaul links of limited capacity. The CSIs are then obtained by decoding the pilots sent by the receivers and then the transmitters exchange their estimated CSIs over the backhaul. A major contribution of this work is the precise characterization of the queuing stability region of the system. Furthermore, we provide a scheduling algorithm that selects the users that must send their pilots in each time slot. Another main contribution is the comparison between IA and TDMA-ZF (zero forcing) techniques in terms of stability regions. We identify the regime in which a simple TDMA-ZF scheme outperforms the IA scheme.

3.1.4.1 System Model

We consider the MIMO interference channel with N transmitter-receiver pairs. All transmitters and all receivers are equipped with N_t and N_r antennas, respectively. Each transmitter communicates with its intended user via $d \le \min(N_t, N_r)$ independent data streams, and interferes with all other unintended users. We denote by $\mathbf{H}_{ki}$ is the $N_r \times N_t$ channel matrix between transmitter i and receiver k with independent and identically distributed (i.i.d.) zero mean and unit variance complex Gaussian entries. Let $\mathbf{v}_i^j$ is the $N_t \times 1$ precoding vector of unit norm and P be the total power at each transmitting node, which is assumed to be equally partitioned over the data streams i.e. $\alpha = \frac{P}{d}$. It is worth mentioning that in order to *ensure the feasibility of the interference alignment problem, the condition* $N_t + N_r \ge d(L+1)$ *must be satisfied where L is the number of active pairs (* $L \le N$ *).*

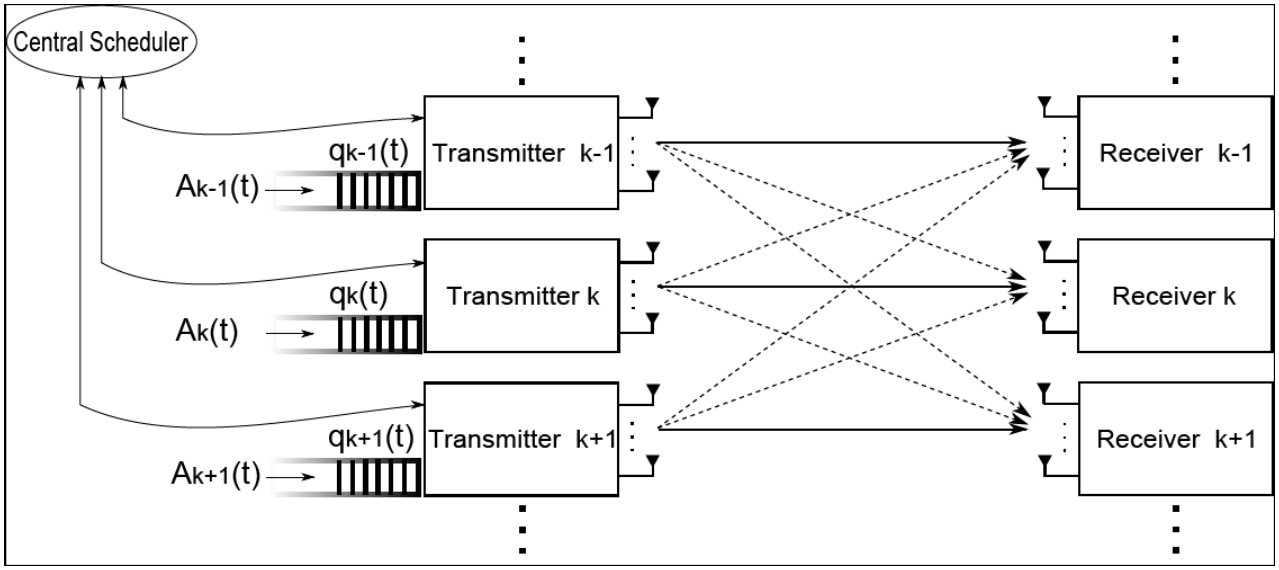


Figure 22 MIMO interference network with limited backhaul

At the receiver, a per-stream zero-forcing $\mathbf{u}_k^m$ is applied. The interference is mitigated by the design of a set of combiner and precoder satisfying the following conditions

$$\left(\mathbf{u}_k^m\right)^* \mathbf{H}_{ki} \mathbf{v}_i^j = 0, \forall (k,m) \ne (i,j).$$

Clearly the above conditions requires perfect CSI knowledge. In this work, we consider a TDD system where the transmitters can estimate their local channels toward different receivers by decoding the pilot sequences sent by the receivers. However, this knwoledge is not sufficient since the whole channel matrix of all cross links must be known at the transmitters. The transmitters must exchange between each other their estimated CSIs via a backhaul links of limited capacity. This can be done by codebook-based quantization. The transmitters have then a quantized version of the global CSIs in the network which has a direct impact on the perofrmance of IA technique since the cross links interference cannot be perfectly cancelled. In this case, the SINR for stream m at receiver k can be written as

$$\gamma_k^m = \frac{\alpha\left|\left(\hat{\mathbf{u}}_k^m\right)^* \mathbf{H}_{kk} \hat{\mathbf{v}}_k^m\right|^2}{\sigma^2 + \alpha \sum\limits_{\substack{i=1 \\ i\neq k}}^{L} \left\|\mathbf{h}_{ki}\right\|^2 e_{ki} \sum\limits_{j=1}^{d} \left|(\mathbf{w}_{ki})^* \mathbf{T}_{k,i}^{m,j}\right|^2},$$

where $\mathbf{w}_{ki}$ is a unit norm vector isotropically distributed in the null space of $\hat{\mathbf{h}}_{ki}$, $\mathbf{T}_{k,i}^{m,j} = \hat{\mathbf{v}}_i^j \otimes (\hat{\mathbf{u}}_k^m)^*$, $\hat{\mathbf{v}}_k^m$ and $\hat{\mathbf{u}}_k^m$ are the combining and precoding vectors, respectively, designed based on the available quantized global CSI at transmitter k. e_{ki} is the error resulting from the quantization process.

For each user, we assume that the incoming data is stored in a respective queue (buffer) until transmission and we denote by $\mathbf{q}(t) = \left[q_1(t),...,q_N(t)\right]$ the queue length vector. Let $\mathbf{A}(t) = \left[A_1(t),...,A_N(t)\right]$ denote the vector of number of bits arriving in the buffers in time slot t, which is an i.i.d. time process, independent across users and with $A_k(t) < a_{max}$. The mean arrival rate for receiver k is denoted by $a_k = \mathsf{E}[A_k(t)]$. We denote by $\mathbf{B}(t) = \left[B_1(t),...,B_N(t)\right]$ the vector of number of bits served at time slot t with $B_k(t) < b_{max}$.

In each slot, the central scheduler selects a subset L (recall that $|\mathsf{L}| = L$) of users that must send their pilots so that their corresponding transmitters can estimate the CSIs. Note that acquiring the CSI of one user takes fraction θ of the slot. Then, the scheduling decision must depend on the probing cost $L\theta$. If a large number of pairs L is scheduled for transmission, many pairs can communicate but a high CSI acquisition cost ($L\theta$) is needed which leaves a small time for data transmission. The optimal number of pairs to schedule at a given time is therefore a challenging task.

The scheduling decision can be represented by an indicator vector $\mathbf{s} \in \mathsf{N}^N$ where the kth component is equal to 1 if the kth queue (pair) is scheduled and equal to 0 otherwise. We denote S as the set of all possible vectors $\mathbf{s}$, thus the cardinality of this set is equal to $|\mathsf{S}| = 2^N$.

3.1.4.2 Queuing Stability Analysis

Definition *The stability region can be defined as the set of arrival rate vectors for which all the queues of all users are strongly stable. The condition for strong stability can be expressed as*

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathsf{E}[q_k(t)] < \infty, \forall k \in \{1,...,N\}.$$

Note that a scheduling policy that stabilizes the system for all this set of arrivals is called throughput optimal. Under a policy Δ, the queue length dynamics can be given by

$$\mathbf{q}^{(\Delta)}(t+1) = \max\left\{\mathbf{q}^{(\Delta)}(t) + \mathbf{A}(t) - \mathbf{B}^{(\Delta)}(t), \mathbf{0}\right\},$$

We now provide a precise characterization of the stability region. Let S_L be a subset defined as $S_L = \{\mathbf{s} : \mathrm{PsP}_1 = L\}$, where $\mathbf{s} \in \mathrm{Z}^N$ is the vector whose coordinates take values 0 or 1. Let G_L be a subset defined as $G_L = \{r_L \mathbf{s}, \forall \mathbf{s} \in S_L\}$ where $r_L = Re^{-\frac{\sigma^2 \tau}{\alpha}} L(1-L\theta)F^{L-1}$, $F = \kappa_1^{-Q}\, {}_2F_1(\beta_2, Q; \beta_1+\beta_2; \kappa_2^{-1})$, ${}_2F_1$ is the hypergeometric function with paramters $\kappa_1 = (1+\frac{d\tau}{2^{\frac{B}{Q}}})$, $\kappa_2 = (1+\frac{2^{\frac{B}{Q}}}{d\tau})$, $\beta_1 = \frac{(Q+1)d}{Q} - \frac{1}{Q}$ and $\beta_2 = (Q-1)\beta_1$. Let $L_1 = \frac{\frac{1}{\theta} - \frac{2}{\log F} - \sqrt{\left(\frac{2}{\log F} - \frac{1}{\theta}\right)^2 + \frac{4}{\theta \log F}}}{2}$ and L_m be the nearest integer to L_1 that maximizes the total average rate function.

We define the set R and its complementary set $\overline{\mathrm{R}}$ as $\mathrm{R} = \{G_1, G_2, ..., G_{L_m}\}$ and $\overline{\mathrm{R}} = \{G_{L_m+1}, ..., G_N\}$. Notice that $|\mathrm{R}| + |\overline{\mathrm{R}}| = |\mathrm{S}|$.

Lemma *Each point in* $\overline{\mathrm{R}}$ *is inside the convex hull of* R*.*

Proof. The proof can be found in[24].

Theorem *The stability region of the adopted system can be characterized as*

$$\Lambda_c = \mathrm{Cv}\{G_1, G_2, ..., G_{L_m}\} = \mathrm{Cv}\{\mathrm{R}\},$$

where Cv stands for the convex hull.

Proof. The proof is provided in [24].

This theorem provides an exact specification of the corner points (vertices) of the stability region.

In order to choose the users who will send their pilots, we use the following scheduling policy

$$\Delta^* : \mathrm{L}(t) = argmax_{\mathbf{s} \in \mathrm{S}} \{r(\|\mathbf{s}\|_1)\mathbf{s} \cdot \mathbf{q}(t)\},$$

where $\|\mathbf{s}\|_1$ gives the number of active pairs. Unlike the standard max-weight, our policy depends on the average rate instead of the instantaneous one.

Proposition *The scheduling policy* Δ^* *is throughput optimal, i.e.* Δ^* *stabilizes the system for every arrival rate vector* $\mathbf{a} \in \Lambda_c$*.*

Proof. The proof is provided in [24].

3.1.4.3 Comparison between IA and TDMA-ZF

In this subsection, we compare between the stability regions of IA and TDMA-ZF. The TDMA-ZF refers to the case where only one terminal is active at aech time and ZF pre-coding is used at the transmitter.

Proposition *If we apply TDMA-ZF technique, the stability region of the corresponding system can be given by*

$$\Lambda_{\mathrm{zf}} = \mathrm{Cv}\{J_1\},$$

where $J_1 = \{r_{\text{zf}}\,\mathbf{s}, \forall \mathbf{s} \in S_1\}$, $r_{\text{zf}} = R(1-\theta)\left(1 - F_{\chi^2_{2(N_r-N_t+1)}}\left(\frac{\tau}{snr}\right)\right)$, $F_{\chi^2_{2(N_r-N_t+1)}}$ is the cumulative density function of $\chi^2_{2(N_r-N_t+1)}$ and $snr = \frac{\alpha}{\sigma^2}$.

Theorem *In terms of stability, interference alignment can outperform TDMA zero forcing if there exists a number* L *(with* $1 \le L \le L_m$*) such that* $Lr_L > r_{\text{zf}}$*. If this condition is not satisfied, then it is better to use TDMA zero forcing technique.*

Where r_L *and* r_{zf} *are defined above.*

Proof. Please refer to [24] for the proof.

3.1.4.4 Numerical Results

We consider a system where the number of antennas $N_t = N_r = 10$, $\text{SNR} = \frac{P}{\sigma^2} = 10$ dB, $d = 2$, $\theta = 0.01$, $\tau = 1$. We take $N = 9$ and we assume that all users have Poisson incoming traffic with the same average arrival rates as $a_k = a$. We set the slot duration to $T_s = 1$ ms and we consider a bandwidth $BW = 10$ MHz. Therefore, the assigned transmission rate per active user can be given by $R = d\,BW \log_2(1+\tau)$ bits/s = 20 Mbits/s (20 Kbits/slot). In order to have insightful results, we consider that all the direct links have a path loss coefficient of 1 and all the cross links have a path loss of ζ (with $\zeta \le 1$). To show the stability performance of the system, we plot the total average queue length for different values of a in the following two figures.

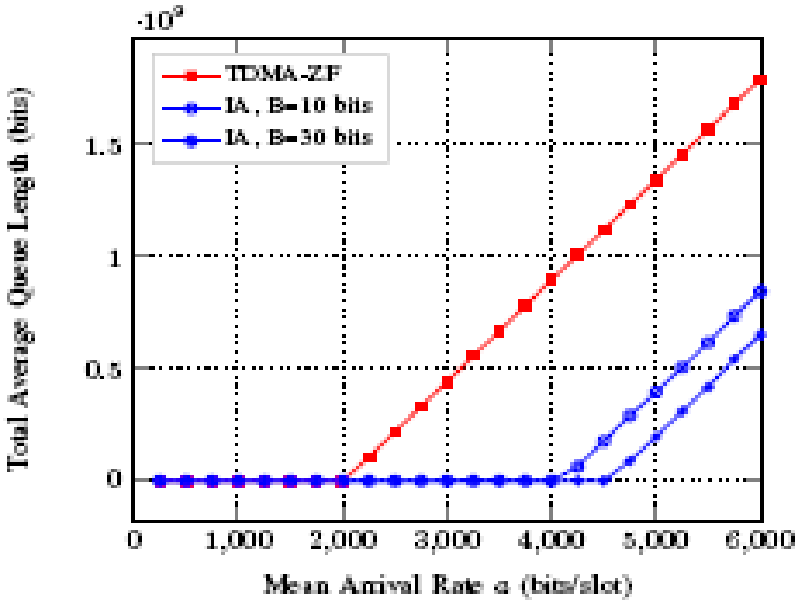


Figure 23: Total average queue length vs. mean arrival rate a. $\zeta = 0.1$.

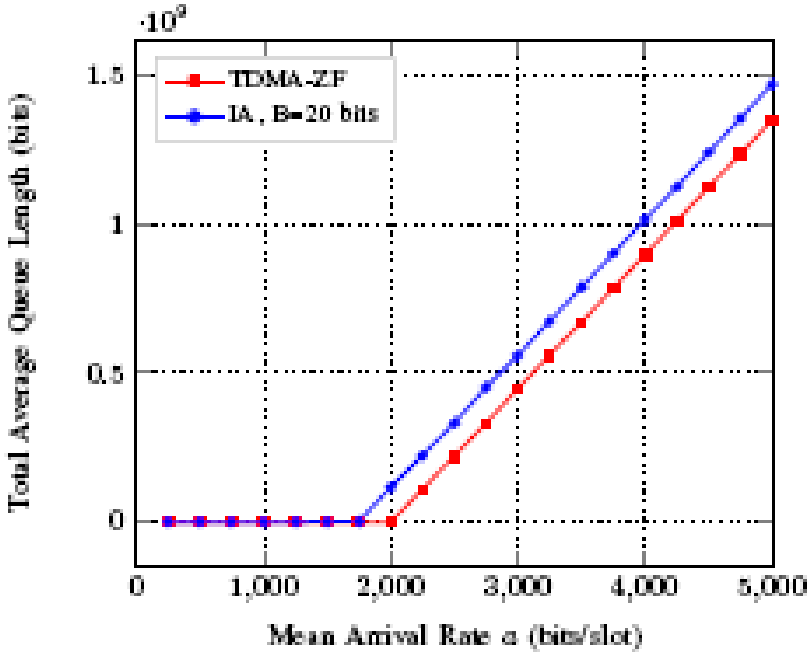


Figure 24: Total average queue length vs. mean arrival rate a. $\zeta = 0.3$.

From these figures, we can see that TDMA-ZF outperforms IA when the interference impact is high (for instance $\zeta = 0.3$), whereas we obtain the converse for less interfering system ($\zeta = 0.1$). This can be explained by the fact that when ζ increases the performance of IA is more sensitive to the number of quantization bits B.

3.2 Inter-system access offloading

An efficient and cost-effective integration of cellular and WiFi technologies, referred to as inter-RAT (or inter-system) offloading, has recently attracted significant interest from academia, industry, and standardization bodies alike, whereby the complementary benefits of both RATs can be leveraged. On one hand, due to the uncontrolled, unlicensed nature of WiFi, the competition for resources among a potentially large number of hotspot users and other devices transmitting on the same unlicensed band can yield dramatically poor throughput. In such a scenario, offloading some of this traffic to a well-managed small cell network operating over the licensed spectrum can improve the performance. On the other hand, the inherent constraints of small cell networks, particularly due to cross-tier and co-tier interference, motivate offloading some of the traffic to the WiFi band both to reduce the interference and to ease the congestion. This section includes the latest findings in capacity-aware multi-user offloading for heterogeneous networks using real traces from a mobile operator (AVEA).

3.2.1 Capacity Aware Multi-User Offloading For Heterogeneous Networks

The main purpose of this work is to balance the load between 3GPP (LTE) and WLAN networks deployed within an integrated architecture. A novel capacity-aware multi-user multi-attribute decision making algorithm is presented and evaluated in terms of mobile user distribution and total channel utilization in the heterogeneous network. The proposed algorithm is shown to enhance total channel utilization of heterogeneous networks compared to standard single-user decision making algorithms.

3.2.1.1 Multi-user Offloading Algorithms for Heterogeneous Networks

The objective of this work is to investigate handover decision making algorithms in heterogeneous wireless networks and point out the metrics and factors influencing data offloading and related open research issues to the research community. To this extent, a

capacity aware multi-user multiple attribute decision making (MADM) algorithms based on Quality of Experience (QoE) metrics has been developed and evaluated. The proposed capacity aware multi-user load balancing algorithm optimizes total benefit of the system that is balanced according to total channel utilization among different heterogeneous wireless networks. The proposed algorithm is shown to enhance total channel utilization of heterogeneous networks compared to standard single-user decision making algorithms.

In this study, a multiple user multiple attribute decision making (MADM) problem, which is targeted for heterogeneous network access within integrated mobile network architecture, is considered. For the system model, the following conditions are assumed to represent the MADM problem.

- The total users set in the system is denoted as $U = \{u_1, u_2, u_3, ..., u_k\}$ where k (k>=2) denotes number of users.
- The multiple users' set involved in the decision making process are denoted as $V = \{v_1, v_2, v_3, ..., v_{k'}\}$ where k' (k'<=k) denotes number of users under multiple coverage.
- The multiple attribute set is denoted as $S = \{s_1, s_2, s_3, ..., s_m\}$ where m (m>=2) denotes number of possible attributes.
- The multiple decision point set is denoted as $E = \{e_1, e_2, e_3, ..., e_P\}$ where there are p ($p \geq 2$) possible decision points.

The weight set is denoted as $w = \{w_1, w_2, w_3, ..., w_m\}$, where each weight w_i is the weight assigned to attribute $s_i i \in \{1,2,...,m\}$. In this study, TOPSIS is used as the core algorithm [32], due to its easy implementation, as a way of selecting the best target network for a set of given users. The decision to use this algorithm was made based on the other multiple attribute decision making (MADM) algorithms' performance comparison results. In [33], four different MADM algorithms (MEW, SAW, GRA, TOPSIS) were evaluated and it was concluded that they all performed very similar.

3.2.1.1.1Set of Attributes

The attributes that will be used in this work are listed below. Note also that the weights of the attributes can be dynamically altered by the mobile operators based on their quality of experience requirements.

- RSSI, s_1, weight: w_1

- Average Latency, s_2, weight: w_2

- Battery Level, s_3, weight: w_3

- Number of Connected Users, s_4, weight: w_4

- Backhaul Capacity, s_5, weight: w_5

- Remaining Capacity, s_6, weight: w_6

- Roaming Status, s_7, weight: w_7

3.2.1.1.2TOPSIS

TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) [32], due to its easy implementation, is a suitable candidate to select the optimal target network for a given a set of given observed attributes for a user.

In the first step of TOPSIS algorithm a decision matrix **A** is created:

$$\mathbf{A} = \left[a_{ij} \right] = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ . & . & \ldots & . \\ . & . & \ldots & . \\ . & . & \ldots & . \\ a_{p1} & a_{p2} & \ldots & a_{pm} \end{bmatrix} (i = 1, \ldots., p; \; j = 1, \ldots., m)$$

In matrix **A**, m refers to size of the multiple attribute set such as link quality, MOS of the target network for the given application, user preference (cost security), etc. and *p* refers to size of the multiple decision point set decision points target networks which can be LTE, WLAN or D2D (device-to-device). Note that that all the attributes are transformed to have positive impact if necessary.

In second step, a normalized decision matrix is formed by using the following equation:

$$r_{ij} = \frac{a_{ij}}{\sqrt{\sum_{k=1}^{p} a_{kj}^2}}$$

Then the normalize matrix R is obtained as:

$$\mathbf{R} = [r_{ij}] = \begin{bmatrix} r_{11} & r_{12} & \ldots & r_{1m} \\ r_{21} & r_{22} & \ldots & r_{2m} \\ . & & & . \\ . & & & . \\ . & & & . \\ r_{p1} & r_{p2} & \ldots & r_{pm} \end{bmatrix}$$

In third step, a weighted normalized decision matrix is created by multiplying each column of the matrix **R** by corresponding weight w_i where $\sum_{i=1}^{m} w_i = 1$ by using the following equation:

$$\mathbf{v_i} = w_i * \mathbf{r_i}, \qquad \mathbf{r_i} = [r_{1i}, \ldots., r_{pi}]^T, \qquad i = \{1, 2, \ldots, m\}$$

In fourth step, the positive (A^*) and negative (A^-) solutions are formed by using the following formulas:

$$A^* = \left\{ (\max_i v_{ij} \,|\, j \in \{1, 2, ..m\}) \right\}$$

$$A^- = \left\{ (\min_i v_{ij} \,|\, j \in \{1, 2, ...m\}) \right\}$$

At the end of fourth step, sets are formed as $A^* = \left\{ v_1^*, v_2^*, \ldots, v_n^* \right\}$ and $A^- = \left\{ v_1^-, v_2^-, \ldots, v_n^- \right\}$.

By calculating the Euclidean distance S_i^* of each multiple decision point from the positive point A^* and S_i^- of each multiple decision point from the negative point A^-.

$$S_i^* = \sqrt{\sum_{j=1}^{p} (v_{ij} - v_j^*)^2}, \quad i = \{1, \ldots, p\}$$

$$S_i^- = \sqrt{\sum_{j=1}^{p} (v_{ij} - v_j^-)^2}, \quad i = \{1, ..., p\}$$

In the final step, the relative similarity of the alternatives from the positive and negative points is calculated as:

$$C_i = \frac{S_i^-}{S_i^- + S_i^*}, \quad i = \{1, ..., p\}$$

Where $0 \le C_i \le 1$ the final solution is selected by:

$$e^* = e_{i^*} \quad \text{where} \quad i^* = \arg\max_i C_i, \quad i = \{1, ..., p\}$$

In the next two sub-sections, two algorithms are defined. First algorithm developed is a Multi-user TOPSIS with capacity-aware characteristic where channel utilization parameter is of the utmost importance for the 3GPP network to balance the channel allocations. With this type of multi user algorithm, the total system benefit is considered as important. Second algorithm is Standard TOPSIS (ST) algorithm. With this method each user's individual benefits are considered individually as they arrive.

3.2.1.1.3 STANDARD MULTI-USER TOPSIS (ST) ALGORITHM

With standard TOPSIS method, user's individual's benefits are considered. The method details are explained in the following steps.

Input: Set of networks (multiple decision set)E, and the TOPSIS matrix of user i' denoted by $\mathbf{A}^{i'}$.
Output: Standard TOPSIS channel utilization vector **CU^e**= [CU_1^e, CU_2^e, …. ,$CU_{k'}^e$], e ∈ E.

Step 1: Run TOPSIS algorithm for all $\mathbf{A}^{i'}$ simultaneously and select the optimal decision points e^* for all users.
Step 2: Update the channel utilization vector **CU^e** by summing the channel utilization demands of each user on the selected decision point.

3.2.1.1.4 CAPACITY AWARE MULTI-USER TOPSIS (CAT) ALGORITHM

In order to obtain certain benefits for access channel selection and resource allocation problem between multiple users, Capacity aware iterative multi-user TOPSIS algorithm is proposed. In the capacity aware approach, a new network level attribute vector $\hat{\mathbf{u}} = [\hat{u}_1, \hat{u}_2, \dots, \hat{u}_p]$ is inserted to the TOPSIS matrix of all users and the decision is calculated sequentially in a centralized platform as users arrive, thus optimizing the total system benefit as well. The network level attribute $\hat{u}_i$ denotes the remaining available capacity in the corresponding decision point e_i and calculated by,

$$\hat{u}_i = CU_{TH}^{e_i} - \sum_{\forall i'} CU_{i'}^{e_i}$$

where $CU_{TH}^{e_i}$ represents the capacity threshold of e_i.

Input: Set of Set of networks (multiple decision set) E, and the extended TOPSIS matrix of i'th arriving user, $\hat{\mathbf{A}}^{i'} = \left[\mathbf{A}^{i'} \quad \hat{\mathbf{u}}_i^{\mathrm{T}}\right], \forall i'$.

Output: Capacity-aware channel utilization vector **CU^e**= [CU_1^e, CU_2^e, …. ,$CU_{k'}^e$], e ∈ E.

Step 1: Set **CU^e**=[0] and i'=0 (i' ≤ k' is the user number)

Step 2: When a new user arrives set i' = i' +1.

Step 3**:** Run TOPSIS algorithm using Ai' and select the optimal decision point e* =e_n∈ E

Step 4**:** Update the channel utilization vector **CU^e** by $CU_{i'}^{e*}$ **= channel demand of user i'.**

Step 5: Recalculate network level attribute vector $\hat{\mathbf{u}}$.

Step 6: If i' = k', stop, else goto to Step 2.

3.2.1.2 Performance Evaluations

Offloading platforms are designed to overcome capacity problems in dense heterogeneous networks. In this context, in order to evaluate the performances of the proposed algorithms as well as the proposed SHARING platform, we have performed both simulations and experimental studies.

3.2.1.2.1 Simulation Results

For simulations, we created a 10K capacity stadium scenario with several Wi-Fi and eNB access points deployed. Out of the 10K people in the stadium 100 terminals are selected randomly to represent the target users. A sample user distribution for 100 users in a 10K capacity stadium is provided in Figure. The assumed locations of 8 Wi-Fi APs and 2 eNB APs are also depicted in Figure. The locations of APs are chosen heuristically equally spaced to the mid of the rows and can be modified based on requirements.

In this simulation scenario, the SHARING platform users demand to download streaming video from 360p to 1080p with bit rates adapted from those of the worldwide video streaming service, YouTube. The platform users can download the content either from the closest WiFi AP or the closest LTE eNodeB.

The attributes for the TOPSIS algorithm are chosen as the ones that are most related with video transmission. First attribute is received signal strength which affects the maximum bit rate that the user can download the content. For simulations, the received signal strength is calculated by free space propagation principles based the distance between the transmitter and receiver. Second attribute is the video bit rate (i.e., throughput) calculated based on a mapping table using the received signal strength. Another attribute is the latency which is assumed to be constant for a given access technology. The last attribute is the remaining capacity in the access node that affects only the CAT algorithm and is ignored in ST. For all simulations, the attribute weights given in Table 2 are used in order to obtain a balanced offloading result. The latency attribute is assigned the lowest weight due to low sensitivity of delay in the target scenario.

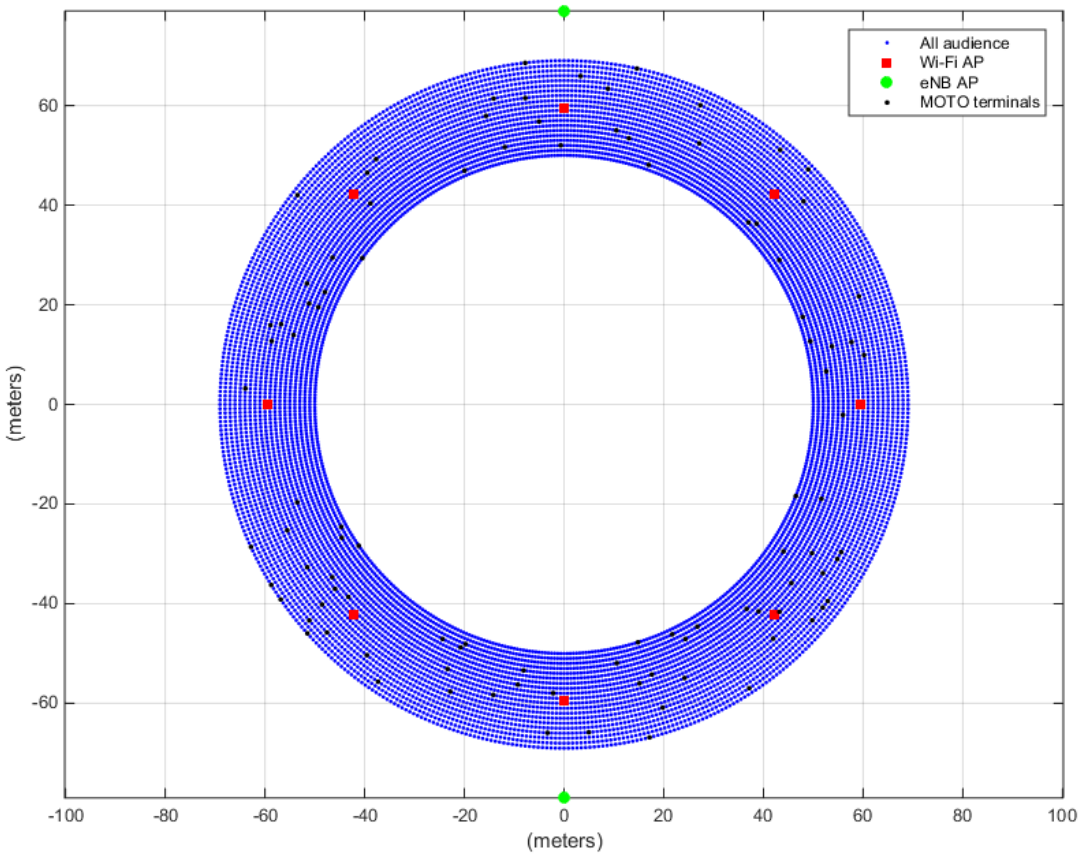


Figure 25: A sample of users' distribution and locations of APs

Table 2: Attributes for simulations

Attribute Set	**Weight**
Received Signal Strength	0.3
Throughput	0.3
Latency	0.1
Remaining Capacity	0.3

The simulation results are presented mainly based on two different performance metrics. First one is the total capacity demand (utilization demand) of the SHARING platform users from the existing access nodes based on their individual video bit rates. If the total capacity demand is above 100% in the results, this directly implies that the network is congested and some users cannot receive the video properly. The other performance metric is the subjective user satisfaction metric. The SHARING platform users are assumed to be satisfied if they can download the content with the bit rate determined based on the received signal strength. Due to higher demand compared to available capacity in the access nodes, some users will become unsatisfied. The simulation results are given in Table 3 which depicts the user distribution and resulting total capacity demands when ST and CAT algorithms are utilized for 100 SHARING platform users. The simulation results demonstrate the average of 500 simulations. It is assumed that capacity of 75Mbps per each eNB and 36 Mbps per each WiFi AP is reserved (available) for SHARING platform users. The rest of the capacity is assumed to

be utilized by other mobile users in the stadium. In the results of Table 3, all users are assigned either to a WiFi AP or an LTE eNB.

Table 3 : User distribution, capacity demands and user satisfaction for CAT and ST

Type of MADM	**USERS' DISTRIBUTION (avg.%)**		**TOTAL CAPACITY DEMAND (%)**		**MOTO User Satisfaction (%)**
	3GPP	**WLAN**	**3GPP**	**WLAN**	
CAT	37.34%	62.66%	76.13%	53.25%	99.80%
ST	80.82%	19.18%	158.40%	21.01%	67.50%

As observed from the results, the CAT method, compared to ST, yields improved balance between 3GPP and WLAN utilizations and significantly better user satisfaction (above 30%) owing to the multi-user sequential TOPSIS algorithm extension. The ST algorithm over-utilizes the 3GPP network due to higher RSSI, which in turn deviates the user satisfaction ratio. Therefore, the results clearly indicate the superiority of CAT algorithm owing to the utilization of network level information based on the remaining available capacity.

Experimental Results

In this section, network selection experiment and sensitivity analysis for a mobile SHARING Client is performed. The experimental set-up is provided in Figure 26. In the experimental set-up there is one WLAN AP A, one eNodeB, one mobile phone running a SHARING client application and one PC to configure the SHARING server deployed at Amazon Web Services.

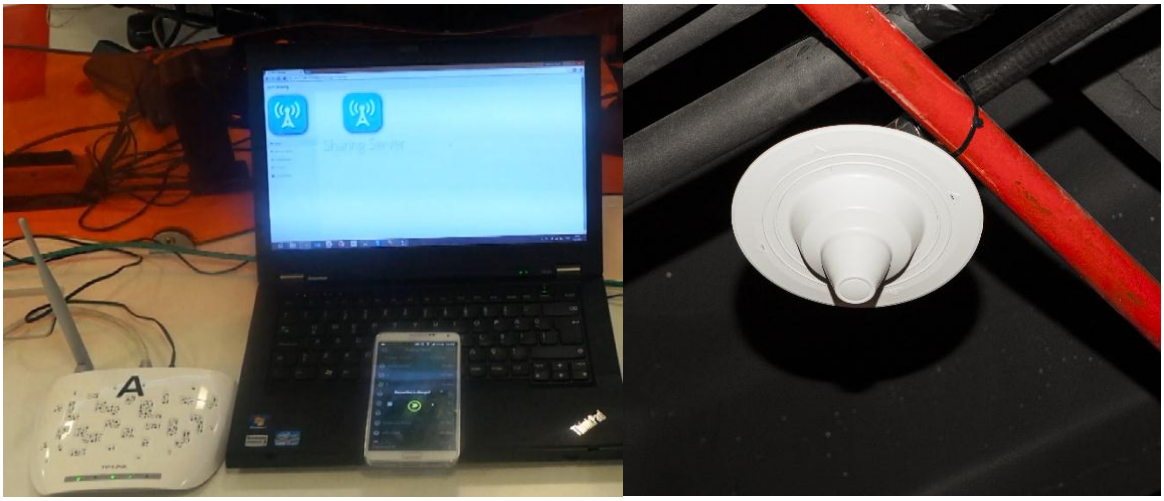


Figure 26: Experimental demo set-up

The SHARING platform ensures that the mobile terminal is connected to the best available network based on the observed attributes. In order to provide a simplified demonstration of the performance, two attributes, RSSI and number of connected users, are set as the most important attributes. The candidate network set consists of two wireless networks. Both candidate networks are previously registered to the SHARING server. SHARING client is moved back and forth between different distances to WLAN AP A and eNodeB and the total number of connected users is varied to simulate a dynamic heterogeneous network environment.

The values of observed attributes, RSSI and number of connected users and output of the experiment (i.e., the selected network) are presented in Figure 27. The experiment covers around 450 seconds of real-time observations and ST algorithm is used in the SHARING server. Initially, there are 6 users connected to eNB, 3 users connected to WLAN AP and the

RSSI values for the LTE network are higher. Based on this initial setting, the SHARING server decides for SHARING client to connect to WLAN network at the beginning due to lower number of connected users although LTE network has higher RSSI values. During the experiment, around time 470, number of users connected to eNodeB drops to 4, and the SHARING server selects LTE as the best available network, thus the mobile terminal switches to LTE. Later on, around time 545, RSSI value of WLAN AP suddenly surpasses that of LTE, and SHARING client is requested to switch to WLAN. Then, around time 665, the number of users connected to WLAN AP experiences a sudden increase, and SHARING client switches to LTE. Then, until the end of the experiment, the mobile terminal stays connected to LTE despite larger RSSI values of WLAN, since number of connected users becomes the dominant factor in this region.

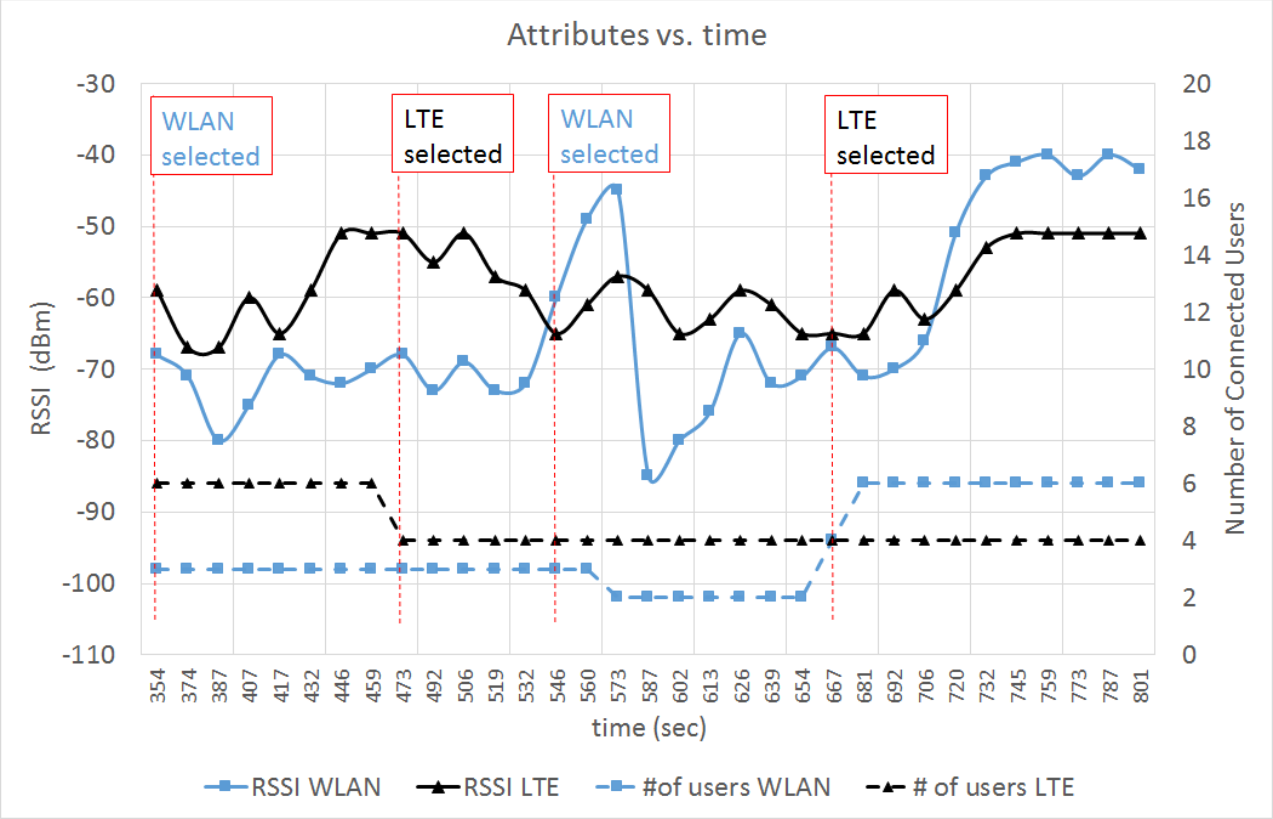


Figure 27: Network Selection Experiment Results

For the sensitivity analysis experiment, on the contrary to the network selection experiment, the SHARING client is now static during realization of the experiment in order to test the sensitivity of the attribute weight representing the number of connected users. In the experiment, the SHARING client stands static close to WLAN AP A and far away from the eNodeB and initially SHARING client is attached to WLAN AP A.

In Figure 28, the experiment result is given, where the average RSSI values of WLAN AP A and eNodeB are represented with dotted blue and red lines, respectively. In this set-up, since initially SHARING client is attached to WLAN AP A, we want to identify the sensitivity of increasing connected users to WLAN AP A before it switches to eNodeB for different values of weight values, where $w_4$represents the weight for the number of connected users. In this type of experiment, number of connected users to eNodeB is 5 and number of connected users to WLAN AP A is varied from 1 to 100 inside SHARING server. For our experiments the value of w_4 is varied from 0.1 to 1.0 with 0.1 increments, while other attributes are assigned equal values after subtracting w_4 from 1.

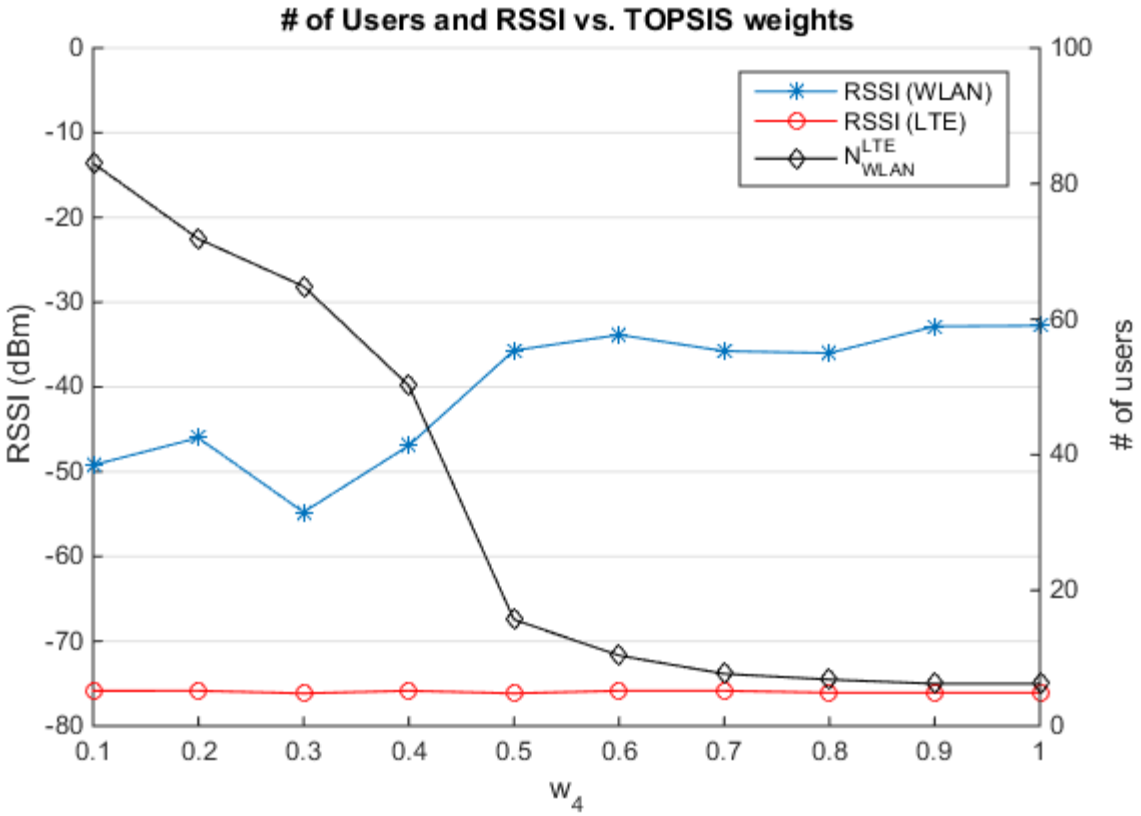


Figure 28: Sensitivity analysis experiment

Figure 28 shows average number of users at WLAN AP A before SHARING client switches to eNodeB as the attribute weight of number of users increases (the solid black line, represented by N_{WLAN}^{LTE}). As observed from the results, the increase inw_4 results in lower N_{WLAN}^{LTE}. For example, when$w_4 = 0.1$, N_{WLAN}^{LTE} becomes 83 and when $w_4 = 0.9$, $N_{WLAN}^{LTE} = 6$. The results also illustrate that when w_4 is greater than 0.6, as the number users at WLAN AP A becomes marginally higher than eNodeB (which is fixed to 5), SHARING client hands-off from WLAN AP A to eNodeB. This means that for $w_4 < 0.5$, number of users connected to WLAN AP A has to be significantly larger than the number of users connected to eNodeB for SHARING client to perform hand-off. In addition, when $w_4 < 0.5$, in some cases of the experiment, SHARING client did not switch to eNodeB even if the number of connected users in WLAN AP A became very large, since the experiment results present only the cases when there is a hand-off. Note also that, this is the main reason behind the fall in the value of the average RSSI of WLAN AP A for $w_4 < 0.5$ since the number of handovers only occurs in cases when WLAN AP A RSSI level becomes lower. This result also indicates that when RSSI difference between the access technologies is large the balancing value for w_4 should be held between 0.5 and 0.6.

It is important to note that the reference values for w_4 obtained via these experiments can be used by operators to adapt the offloading platform into their specific operational requirements. In fact, there is an observed trade-off where as the w_4 value increases, the capacity of users for eNodeB and WLAN AP A decreases, whereas when w_4 value is too low, the performance optimization does not count this attribute which may yield sub-optimal results.

3.3 Energy Saving Mechanisms

The scope of SHARING Task 4.3 is to investigate the performance of various types of energy saving mechanisms both at the network and the base station hardware level. At the network level, the proposed solutions will be based on load balancing and on RRM reconfiguration. More specifically, the common factor is to propose and evaluate the performance of (self-optimizing) energy saving mechanisms based on the use of fast cell DTX (e.g. micro sleep),

eNodeB sleep mode and/or dynamical cell switch-off. The investigations will consider both the QoS of the users and the overall network performance (e.g. capacity). The latest findings are showcased in this section.

3.3.1 Centralized/decentralized techniques for coverage control in HetNet campus scenario

Coverage control is an important aspect of future resource management of 5G communication systems and networks. The basic problem of coverage control is to maximize the minimum cell edge signal to interference and noise ratio $SINR_i$ .of the deployment scenario, described by Figure 29.

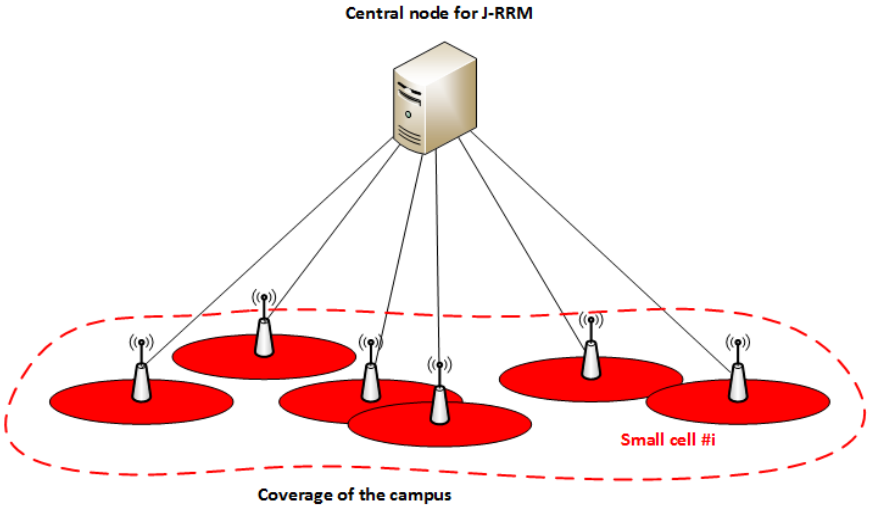


Figure 29: deployment scenario for the coverage control techniques

The maximization is performed through the adjustment of the DL transmit powers of the base stations of the scenario and the corresponding optimization problem can be formulated as follows:

$$\underline{P} = \arg\max\left(\min_{i=1,\ldots N}\{SINR_i\}\right)$$
$$P_{\min} \le P_i \le P_{\max} \quad (1)$$

The cell edge signal to interference and noise ratio is defined as the following with respect to the path gain of the serving and interfering base stations as

$$SINR_i = \frac{\alpha_{ii} P_i}{\sum_{j \neq i} \alpha_{ij} P_j + N_0} \quad (2)$$

Where: P_i is the DL transmitted power of the base station *i*, α_{ij} is the path gain of the base station *j* to the position of the minimum SINR of the coverage of the base station I and N_0 is the noise power.

The optimization problem described in equation (1) can be solved either through centralized technique based that is based on bisection algorithm and the successive application of the ellipsoid technique for the feasibility testing of the constraints. The key idea of the centralized technique is to transform the problem (1) into the following optimization problem

$$\begin{aligned} &\arg\max(\gamma) \\ &s.t.\ SINR_i \geq \gamma; i \in \{1,..,N\} \\ &P_{\min} \leq P_i \leq P_{\max} \end{aligned} \tag{3}$$

The ellipsoid technique is applied to the power and SINR constraints while the bisection search is used to find the optimum value of the minimum of the cell edge SINR γ .[21].In this contribution we develop distributed techniques for solving the global optimization problem described by the equation (1). . These techniques solves a relaxed optimization problem instead of the original coverage control problem and use consensus averaging techniques [20] for computing the gradient of the relaxed objective function. Then, each base station independently updates its power in the direction of the gradient in order to increase the value of the objective function. The first step towards the development of the distributed techniques is to develop relaxation of the original optimization problem. To this end we will use the following exponential minimum approximation of the original minimum objective function as:

$$\min\{SINR_i\}_{i=1}^N \approx f_\alpha\left(\{SINR_i\}_{i=1}^N\right) = \frac{\sum_{i=1}^N SINR_i \exp(\alpha SINR_i)}{\sum_{i=1}^N \exp(\alpha SINR_i)} \tag{4}$$

The optimization problem is then reformulated as

$$\begin{aligned} &\underline{P} = \arg\max\left(f_\alpha\left(\{SINR_i\}_{i=1}^N\right)\right) \\ &P_{\min} \leq P_i \leq P_{\max} \end{aligned} \tag{5}$$

The second step is to reformulate the optimization problem as a geometric program. To this end let us consider the following variable changes: $y_i = \log(SINR_i)$; $x_i = \log(P_i)$; $a_{ij} = \log(\alpha_{i,j})$. The SINR equation (2) can be written as

$$y_i = a_{ii} + x_i - \log\left(\sum_{j \neq i} \exp(a_{ij} + x_j) + N_0\right) \tag{6}$$

The geometric program can be expressed as the following:

$$\underline{x} = \arg\max\left(f_\alpha\left(\{y_i\}_{i=1}^N\right)\right)$$
$$x_{\min} \le x_i \le x_{\max} \quad (7)$$

The base stations are updating their transmit powers in the direction of the gradient of the objective function in order to maximize the relaxed minimum of logarithm of the SINRs $f_\alpha\left(\{y_i\}_{i=1}^N\right)$. The update equation for each base station is given by the following:

$$x_p(k+1) = \left(x_p(k) + \delta_i \left(\frac{\partial f_\alpha}{\partial x_p} \right)_k \right)_{I(x_p)} \quad (8)$$

For performing this update, the base stations need to calculate the component of the gradient. In the next section we propose a distributed technique for calculating this gradient and evaluate the performance of the scheme.

3.3.2 Consensus averaging for coverage control

Assuming the deployment is in interference limited scenario and using the log map approximation, it is possible to rewrite the equation (6) as the following:

$$y_i \approx a_{ii} + x_i - a_{ij_0(i)} - x_{j_0(i)} \quad (9)$$

Where $j_0(i)$ is the index of the most interfering base station for the cell edge user terminals of the base station i. The equation (9) allow us to calculate the gradient component at the base station *p* of the objective function as

$$\frac{\partial f_\alpha}{\partial x_p} = \left(\phi_p - \phi_{j_0(p)}\right)\left(1 - \alpha f_\alpha\right) + \alpha\left(\phi_p y_p - \phi_{j_0(p)} y_{j_0(p)}\right) \quad (10)$$

Where the function ϕ_p is defined as the following selection function $\phi_p = \frac{\exp(\alpha y_p)}{\sum_{i=1}^{N} \exp(\alpha y_i)}$.

For the calculation of the gradient at the base station *p,* the selection function for the base station and the most interfering base station should be obtained as well as the value of the global minimum given by the function $f_\alpha\left(\{y_i\}_{i=1}^N\right)$. The calculation of the gradient is shown at the Figure 30.

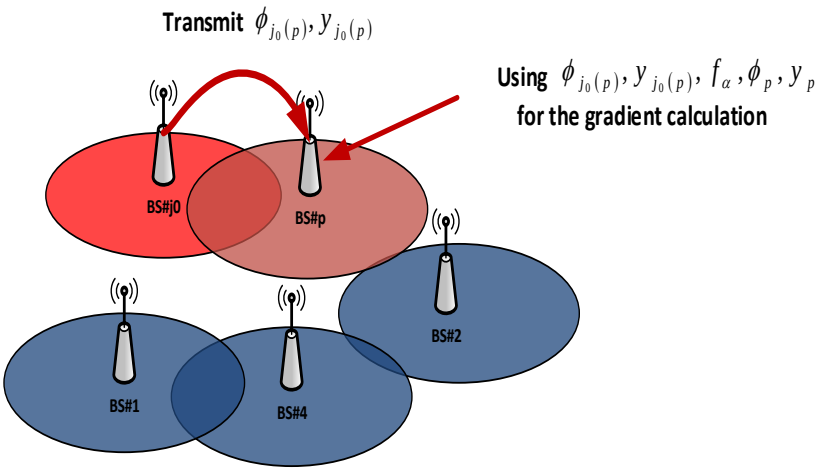


Figure 30: information needed for the calculation of the gradient for distributed CC

The selection function ϕ_p and the relaxed objective function f_α are expressed as functions of averages/weighted averages of the logarithm of the SINRs. In this contribution, it is proposed to use consensus averaging techniques [20] for the calculation of these network wide averages.

Consensus averaging techniques are iterative techniques for reaching an agreement in a network on a function that is dependent on the states of the base stations. The basic principle of the consensus averaging technique can be summarized through the following:

- each base station at the iteration *n* receives the states from the neighboring nodes
- each node updates its state on the iteration *n+1* with the average of the difference between its state and the neighboring states

Assuming that the state of the base station *p* at the iteration *n* is denoted as $s_p(n)$, the state update equation of the consensus averaging is given as following:

$$s_p(n+1) = s_p(n) + \varepsilon \sum_{j \in N(p)} \left(s_j(n) - s_p(n) \right) \tag{11}$$

Where ε is the consensus averaging parameter and $N(p)$ is the set of neighbors of the base station p. The output of the consensus averaging algorithm is the average of the states of the base stations of the network as given in the equation (12) below

$$\lim_{n \to \infty} s_p(n) = \frac{1}{N} \sum_{i=1}^{N} s_i(0) \tag{12}$$

Consensus averaging technique can be applied to the calculation of the gradient parameters as follows:

1. Each base station *p* initializes its states as: $s_{p,1}(0) = y_p \exp(\alpha y_p)$ and $s_{p,2}(0) = \exp(\alpha y_p)$.
2. The base stations execute the state update step given by the equation (11) until convergence, i.e. $s_{p,1}(n) = \frac{1}{N} \sum_{i=1}^{N} y_i \exp(\alpha y_i)$ and $s_{p,2}(n) = \frac{1}{N} \sum_{i=1}^{N} \exp(\alpha y_i)$.

3. The parameters of the gradient are obtained for each base station as $\phi_p = \frac{s_{p,2}(0)}{Ns_{p,2}(n)}$; $f_\alpha = \frac{s_{p,1}(n)}{s_{p,2}(n)}$.
4. Each base station receives from the most interfering neighboring base station its parameters $\phi_{p_0}; y_{p_0}$ and calculates the gradient through the equation ((10))
5. The base station update its log transmit power through the equation ((8)), calculate the logarithm of its cell edge SINR value and go to step (1).

The power adaptation will continue until the relaxed objective function, i.e. f_α is above a given coverage threshold that is given by the network operator or when a maximum power adaptation steps is reached.

3.3.2.1 Performance evaluation

The deployment scenario used for the evaluation of the distributed coverage control technique is a randomly deployed cluster of home base stations with density of (5, 15 and 25 home base stations per square km). The radio parameters of the base stations are given by Table:

Table 4 : radio parameters of the cluster of base station

Parameter	**Value**
Maximum Tx power	20dbm
Minimum Tx power	-10dBm
Bandwidth of the transmission	10Mhz
Antenna gain	5dbi
Pathloss parameters	a=127
	b=30
Noise density	-174dbm/Hz
Noise factor	5dB
Shadowing	NA

The simulations considered different values for the relaxation parameter of the objective function α as: $\alpha = -1, -10, -25$. The Figure 31is showing the effect of the relaxation parameter variation on the approximation of the objective function f_α. The blue curve stands for the relaxed exponential minimum f_α and the red curves stands for the exact minimum function. It is seen that the approximation is good, starting from a value of $\alpha = -10$

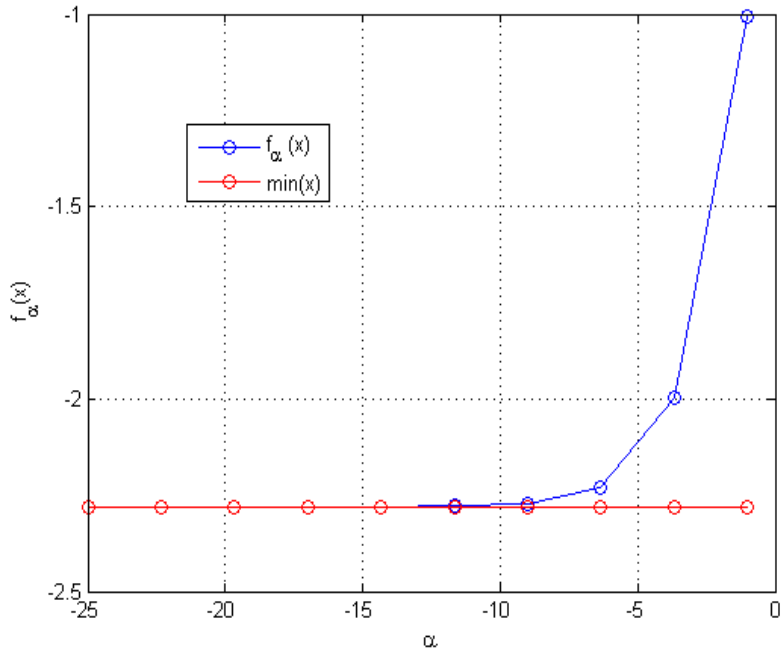


Figure 31: effect of the relaxation parameter on the approximation of the objective function

For Figure 32, 1000 independent deployment configurations of a cluster of 15 home base stations where simulated. Each deployment consideration consists of independent deployment of the home base stations and user terminals in coverage region of 1 km^2. For each deployment configuration, the minimum cell edge signal to interference ratio was evaluated for a baseline system where all the base stations are transmitting at maximum power and a system using the proposed distributed coverage control technique. The parameters of the gradient and the consensus, i.e. ε, δ, are set to be common to all the base stations and the relaxation parameter is set as $\alpha = -10$. The Figure 32 shows the histogram of the minimum cell edge SINR for the baseline and optimized system

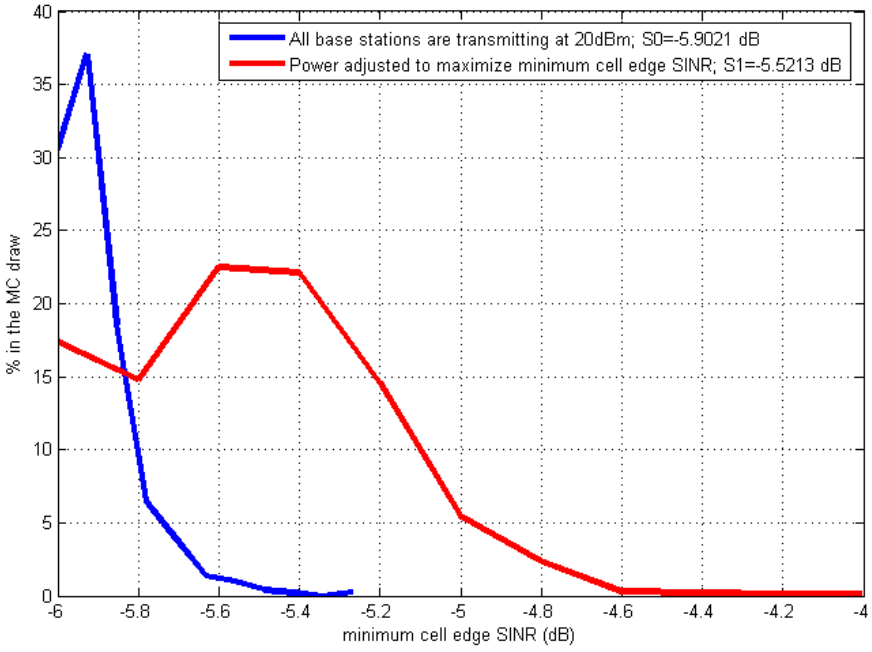


Figure 32 : Histogram of the minimum cell edge SINR for coverage control

The results are showing that the proposed distributed coverage control technique improves the minimum cell edge SINR of 0.4dB on the average (around 8% improvement with respect to the baseline system). At best, the improvement is around 1.2 dB (around 24%). In the Figure 33, we plotted the average cell edge SINR wrt to the density of the cluster for $\alpha = -10$ and the same gradient parameters as in the Figure 32.

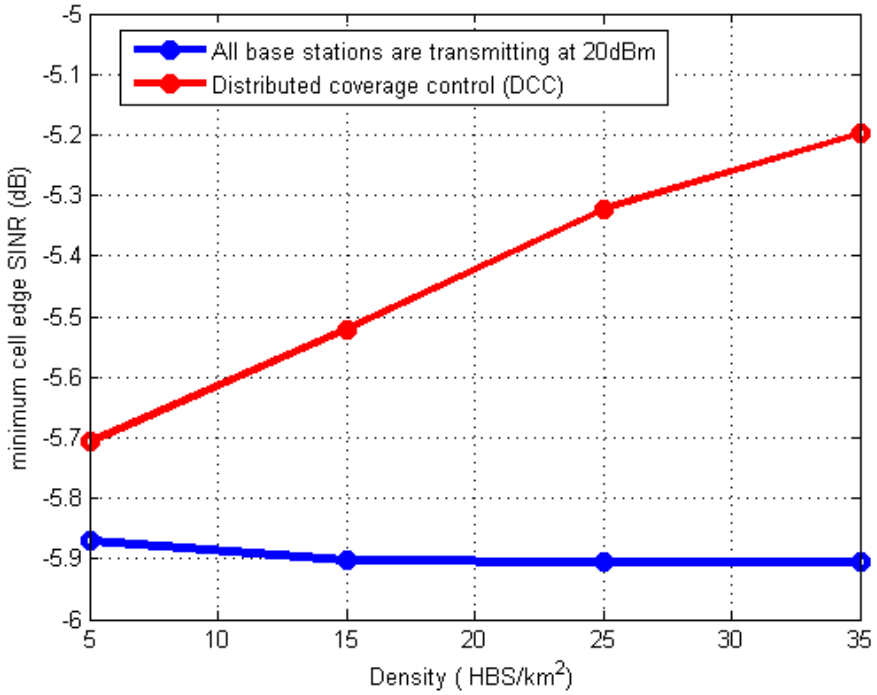


Figure 33 : average cell edge SINR wrt the density of the cluster

The Figure 33 is showing that the proposed distributed coverage control technique is improving the minimum cell edge SINR on the average for a wide range of cluster densities. The improvement is low for low density clusters (5 home base stations/square km). It is around 4% of the baseline minimum cell edge SINR. For higher cluster densities, i.e. 35 home base stations/square km, the improvement is around 12% of the nominal minimum cell edge SINR. For the most favorable configurations, the minimum cell edge SINR improvement is around 24 % for low density clusters and 28% for clusters with high density.

In this contribution we developed coverage control techniques that distributively adjusts the transmit powers for maximizing the minimum cell edge SINR of a cluster of randomly deployed home base stations. The coverage control techniques are distributed in the sense that each base station autonomously adjusts its power in order to maximize a relaxed function of the minimum of the cell edge SINR. Consensus averaging techniques are used to obtain network wide parameters that are used by the base stations for the calculation of the gradient that is used for the power adjustment. The simulation results are showing that the proposed coverage control technique improves the average minimum cell edge SINR of around 4% for clusters of low density (5 base stations/square km) and up to 12% for clusters of higher density (35 base stations/square km). The best case performance improvement is around 24% for low density clusters and 28% for high density clusters.

3.3.3 Compensation based ON/OFF energy saving for cluster of HetNets

In [21], we proposed and evaluated the performance of ON/OFF energy saving techniques based on the calculation of dominant sets over a compensation graph obtained from the measurements of the active user terminals in the coverage area. The performance evaluation proposed in [21] evaluates the energy saved by the cluster and the average throughput offered to the active user terminals as function of different deployment configurations of the cluster and user spatial distributions.

In this contribution, it is proposed to evaluate the performance of the compensation based ON/OFF energy saving techniques from energy efficiency perspective. Energy saving performance metric is shown to be increasingly important for the development of future 5G communication networks [26], [27] where 5G networks are expected to increase the energy efficiency of the legacy 4G networks of 1000 times. In this contribution, it is proposed to evaluate the performance of the compensation based ON/OFF in terms of energy efficiency in order to evaluate how the proposed techniques are approaching the design goals of 5G.

3.3.3.1 HetNet cluster deployment scenario and compensation graph construction

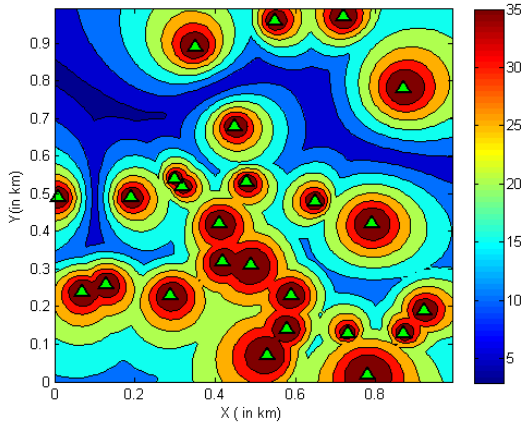
The HetNet cluster deployment scenario is shown in Figure 34 (a) for 25 base stations. In this deployment scenario, two tiers of base stations are randomly deployed in the coverage region of 1km^2.

- The first tier is a group of N_1 pico base stations
- The second tier is a group of N_2 femto base stations

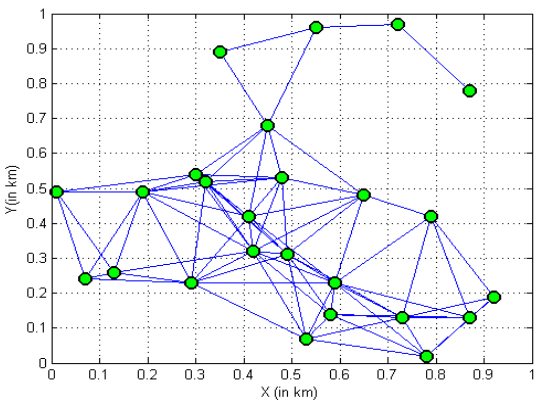
The total number of nodes in the cluster is $N = N_1 + N_2$ and the normalized heterogeneity of the cluster is defined as $d = N_1 / N$. The graph representation of the cluster deployment is defined as follows: the nodes of the graph represent the nodes of the cluster (picos, HeNBsetc). The edges of the graph are present if the neighboring base station is reported with sufficiently high signal to interference and noise ratio (above a predetermined threshold) to the current, attachment base station by a given active user terminal.

For example two nodes i and$j$ of the graph representation of the cluster are connected by an edge if the signal to interference and noise ratio (SINR) of at least one UE attached to the base station i, of base station j is above -10dB and the SINR reported by at least one UE attached to the base station j of the base station i is above -10dB. Another possible method for the graph construction is to link the two nodes i and j by an edge if the downlink interference level seen in the two base stations is above a threshold fixed by the network operator.

The Figure 34 (b) is showing the graph representation of the cluster of 25 base stations of pico and home base stations with normalized heterogeneity of $d = 10\%$ corresponding to 2 pico base stations and 23 home base stations. The corresponding coverage of the cluster, i.e. SINR levels are shown in the Figure 34 (a). The compensation graph is built from the processing of the measurements from a crowd of 65 user terminals in the coverage area with SINR threshold of -10dB.



(a)



(b)

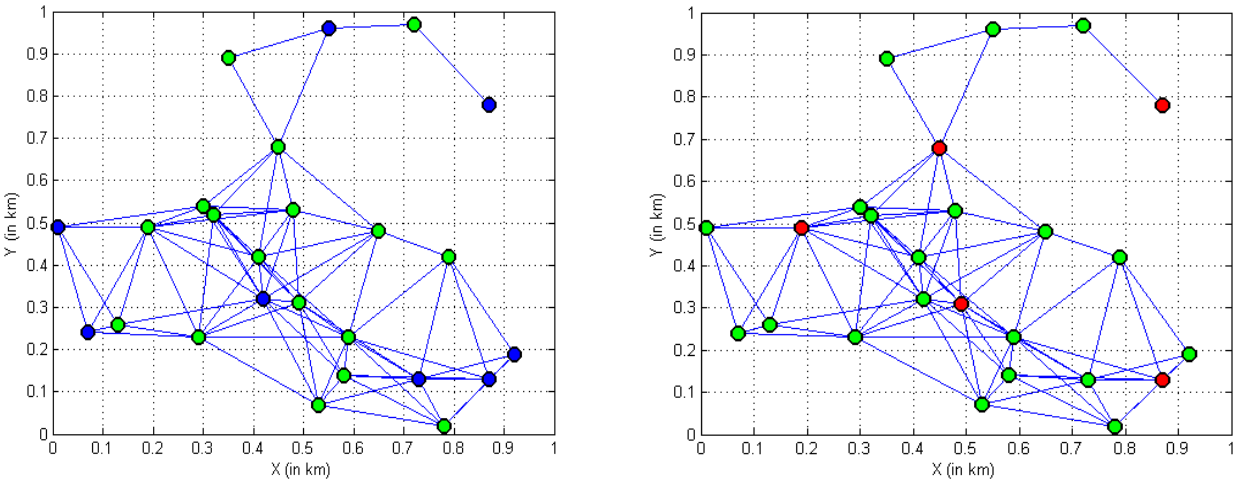
Figure 34 : Campus deployment scenario and coverage map (a) and corresponding graph representation (b)

The compensation base stations are determined as dominating set over the compensation graph. The dominating set is the set of nodes that is a neighbor to every node in the compensation graph. The SINR of the UEs in the coverage region of the cluster of base stations are optimized even when the energy saving base stations are OFF. The dominating set may be either a set of connected nodes in the graph or a set of independent, non-connected nodes in the compensation graph. This leads to the definition of the following two compensation nodes classes:

1. Minimum connected dominant set (CDS): where the nodes of the dominant set are allowed to be connected by an edge in the compensation graph.
2. Maximum independent set (MIS): where the nodes of the dominant set are not connected in the compensation graph and the coverage of the dominant set is maximized.

The Figure 35illustrates an example of the construction of CDS and MIS sets from the compensation graph of Figure 34 **Figure 34** : Campus deployment scenario and coverage map (a) and corresponding graph representation (b)

(b): the blue nodes are CDS nodes and red nodes are MIS compensating nodes



(a) **(b)**

Figure 35 : Dominant sets based compensating base stations (a) connected dominating set compensation (b) maximum independent sets compensation.

Finding minimum dominant sets for general graph topologies is a non deterministic polynomial time (NP) hard problem. We have used in this contribution, greedy heuristic algorithm for the determination of MIS and CDS dominant sets from the compensation graph. The greedy heuristic algorithm, shown in Figure 36 , sort the nodes with respect to their degrees in the compensation graph, then the first node of the ordering is added to the compensation graph set, the neighbors of the added node are market and the overall process is reiterated over the unmarked nodes.

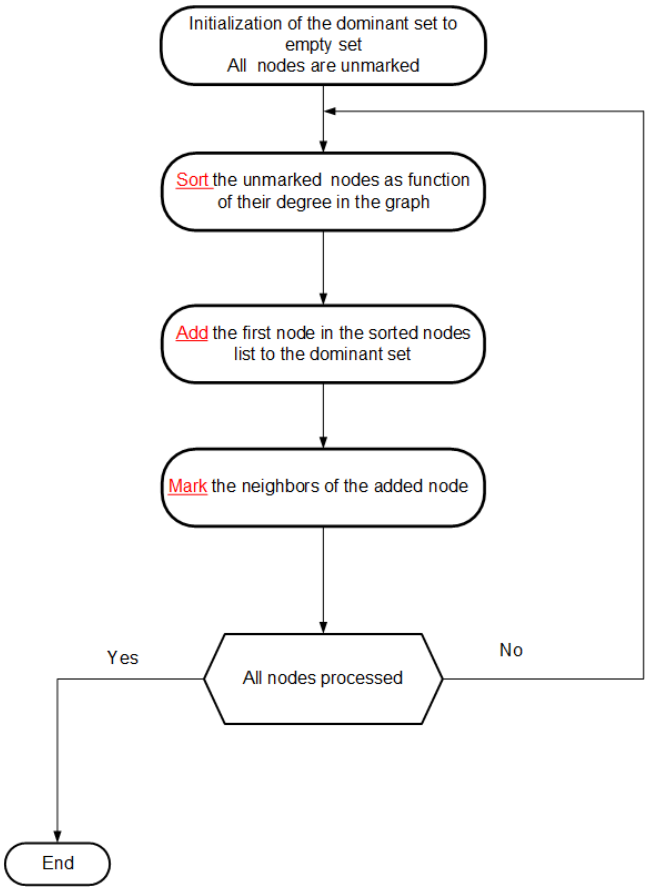


Figure 36: Greedy algorithm for the calculation of CDS and MIS dominating sets

We have presented in the [25], the overall performance evaluation for the compensation based energy saving, including a comparison between CDS and MIS based energy saving. The performance metric used in the evaluation was essentially the throughput and the power consumption of the cluster. Next, we propose performance evaluation with respect to energy efficiency, defined as the average throughput offered to the user terminal divided by the total power consumption of the cluster.

3.3.3.2 Simulation scenario assumptions

The simulation scenario considered is a square region of 1000x1000 meters where a HetNet cluster is deployed. The HetNet cluster is formed of two tiers of nodes, pico base stations and femto base stations and is characterized by its cluster density N and its normalized heterogeneity d.

The radio parameters of the pico and femto base stations are shown in the Table 5 **Table 5** : Radio parameters of the pico base stations campus

and both nodes are deployed in outdoor.

Table 5 : Radio parameters of the pico base stations campus

Parameter	**Value**
Maximum Tx power (pico)	20dBm
Maximum Tx power (femto)	16dBm
Bandwidth of the transmission	10Mhz

Antenna gain (pico)	5dBi
Antenna gain (femto)	0dBi
Pathloss parameter	a=140.7
	b=36.7
Noise density	-174dbm/Hz
Noise factor	5dB

The scheduling used in the simulations is round robin with 50 physical resource elements are assumed per base station with elementary resource bandwidth of 180 kHz. The scheduling and the elementary resource elements bandwidth will be used for the calculations of the throughput of the system and the load of the pico/femto base stations of the cluster. The pico/femto base stations are assumed to be deployed outdoor and are open base stations and the signal to interference ratio for the user terminal, *j* at base station *i* is calculated by taking into account the neighboring base stations loads. The power consumption model of the pico and the femto base stations is given by the following:

$$P_c = \begin{cases} P_0 + \Delta_p P_r & \text{if } 0 < P_r \le P_{\max} \\ P_{Sleep} & \text{if } P_r = 0 \end{cases}$$

Where P_c is the consumed power by the base station, P_r is the radiated power and $P_{\max}$ is the maximum transmit power of the base stations. The rest of the parameters of the model are detailed in the following table for the different types of base stations.

Table 6 : Power consumption model parameters

BS type	$P_{\max}$ [W]	P_0 [W]	Δ_p	P_{sleep} [W]
Pico	0.13	6.8	4	4.3
Femto	0.05	4.8	8	3.9

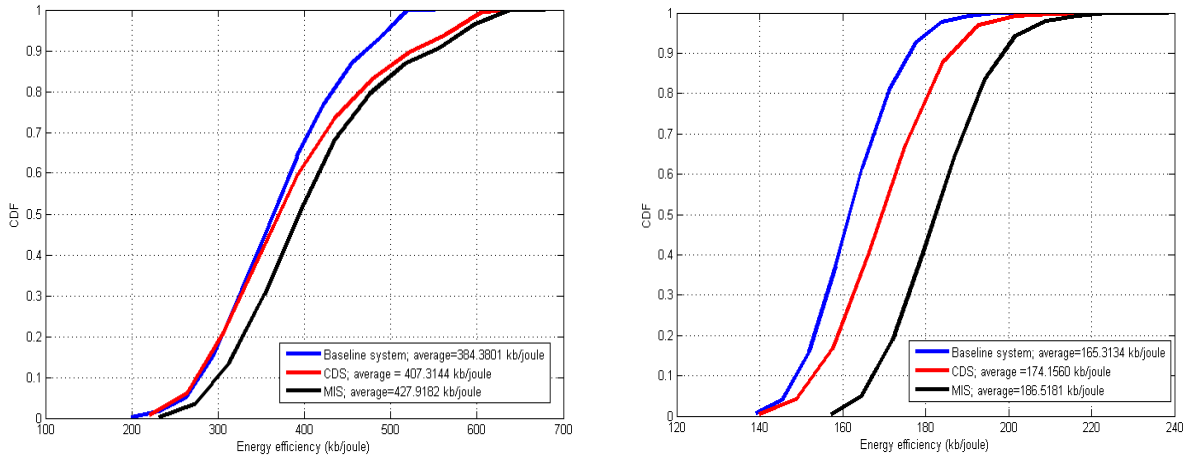
The spatial distribution of the user terminals in the coverage area of the cluster is random and uniform. The users are attaching to the base stations on the basis of maximum received SINR, i.e. no cell range extension is considered and two load states are considered:

- Low load situation where the average number of user terminals in the campus coverage is fixed to $\lceil 0.1N \rceil$ users per square km.
- High load situation where the average number of user terminals in the campus coverage is fixed to $\lceil 2N \rceil$ users per square km.

Simulation Results

Figure 37 is showing the energy efficiency cumulative distribution function CDF with respect to 1000 independent and random configurations of the cluster deployment, i.e. relative base

stations positions and the user terminals positions in the coverage area of the cluster. The Figure 37 (a) is showing the CDF of the energy efficiency of a cluster of 25 home base stations, i.e. d=0% for the low load situation of the cluster. The Figure 37 (b) is showing the CDF of the energy efficiency of the same cluster of base stations for the case of high load situation.



(a) **(b)**

Figure 37 : Energy efficiency CDF low load situation (d=0%) (a) Energy efficiency CDF for high load situation (d=0%) (b)

In Figure 37, the energy efficiency CDF is shown for a cluster of base stations transmitting at maximum power, i.e. 20dBm for the pico base stations and 16 dBm for the femto base stations. This first cluster performance is shown in blue and is considered as the baseline system. The CDF of the energy efficiency of CDS based ON/OFF energy saving scheme is shown in plain red and the performance of MIS is shown in plain black curve. The energy efficiency results are showing that both CDS and MIS ON/OFF energy saving systems outperform the baseline system of 6% for CDS and 11 % for MIS in low load situation. In high load situation, the CDS and MIS based system outperforms the baseline system of 5% for CDS system and 13% for the MIS system. The MIS based energy saving outperforms the CDS energy saving systems of 6 %. .

We have shown in Figure 38 the CDF of the evolution of the energy efficiency of the HetNet cluster with d=0% with respect to the node densities for low and high load situations. The Figure 38(a) shows the performance for low load situation and the Figure 38 (b) is showing the performance for high load.

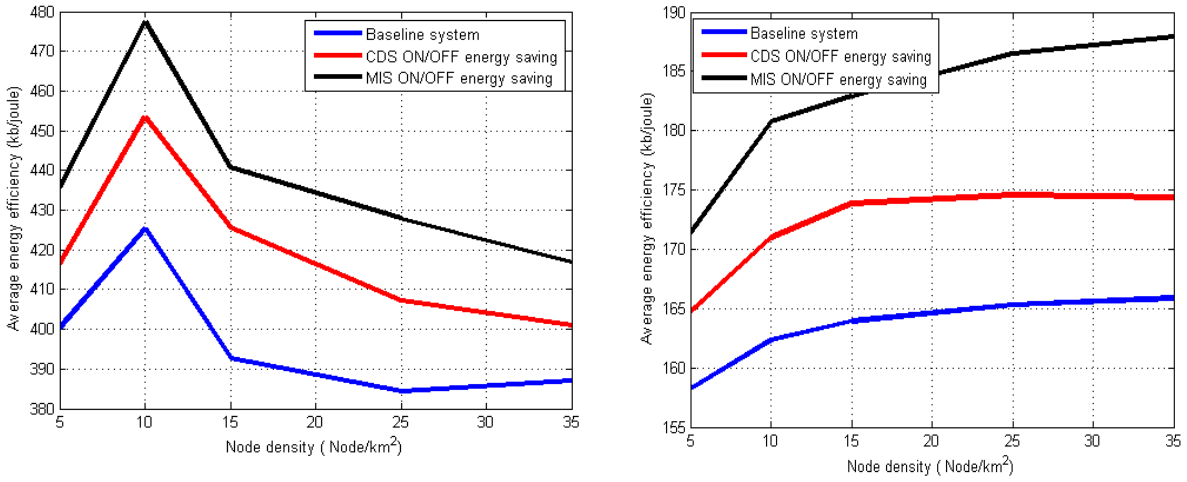


Figure 38 : Average energy saving wrt node density (low load, d=0%) (a) Average energy saving wrt node density (high load, d=0%) (b)

For low load situation, the results of the Figure 38(a) are showing that CDS and MIS based ON/OFF energy saving systems are improving the energy efficiency of the baseline system. The maximum improvement of CDS over the baseline system is around 8% and the improvement for MIS is around 12% for a density of 15 nodes/km^2. For high load situation, the results of the Figure 38(b) are showing similar improvement in energy efficiency wrt to the baseline system. The maximum improvement for the high load situation for the CDS is around 6% for a density of 15 nodes/km^2 and the maximum improvement for the MIS is around 13% for a density of 35 nodes/km^2 . It is interesting to note that the overall energy efficiency for low load scenario is higher than the energy efficiency for high load scenario.

In the Figure 39, the CDF of the energy efficiency is shown for a HetNet cluster of 25 nodes with normalized heterogeneity of d=10%.

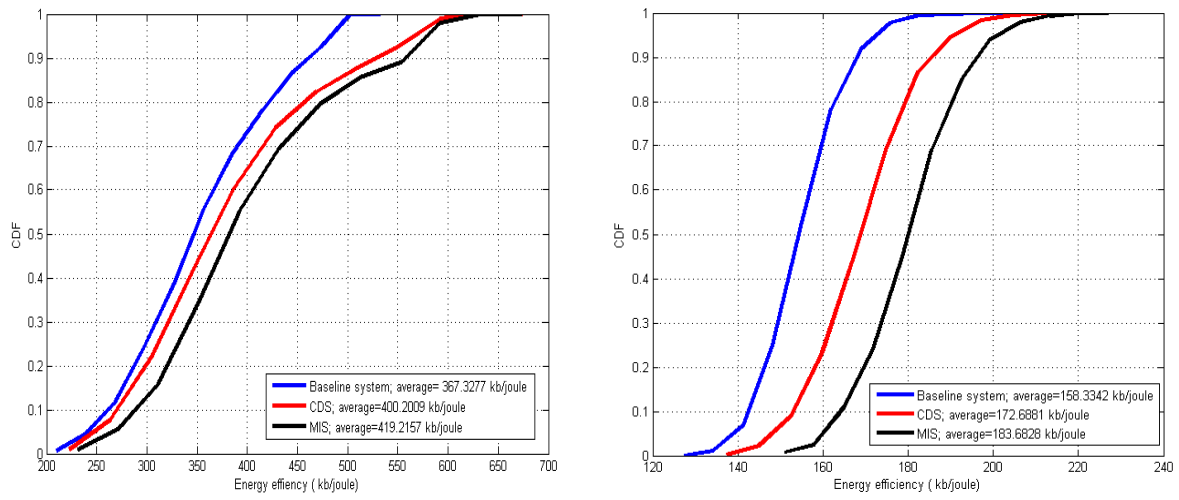


Figure 39 : Energy efficiency CDF low load situation (d=10%)(a) Energy efficiency CDF high load situation (d=10%) (b)

The Figure 39 (a) is showing the CDF of the energy efficiency of the HetNet campus for low load situation while the Figure 39 (b) is showing the results for high load scenario. The results are showing that the CDS based ON/OFF energy saving is improving the performance of the baseline system of 9 % for low load scenario. In the same scenario, the MIS based ON/OFF energy saving is improving the performance of the baseline system of 14%. For high load scenario, the CDS is improving the performance of the baseline system of 9%, the MIS ON/OFF energy saving is improving the energy efficiency of the baseline of the system of 16%. When comparing the results of Figure 37 to the results of the Figure 39, it is seen that increasing the heterogeneity of the cluster decreases the energy efficiency both for low and high load scenarios.

In Figure 40(a) and Figure 40(b) we show the evolution of the average energy efficiency with respect to the node densities in the HetNet cluster deployment with normalized heterogeneity of d=10%.

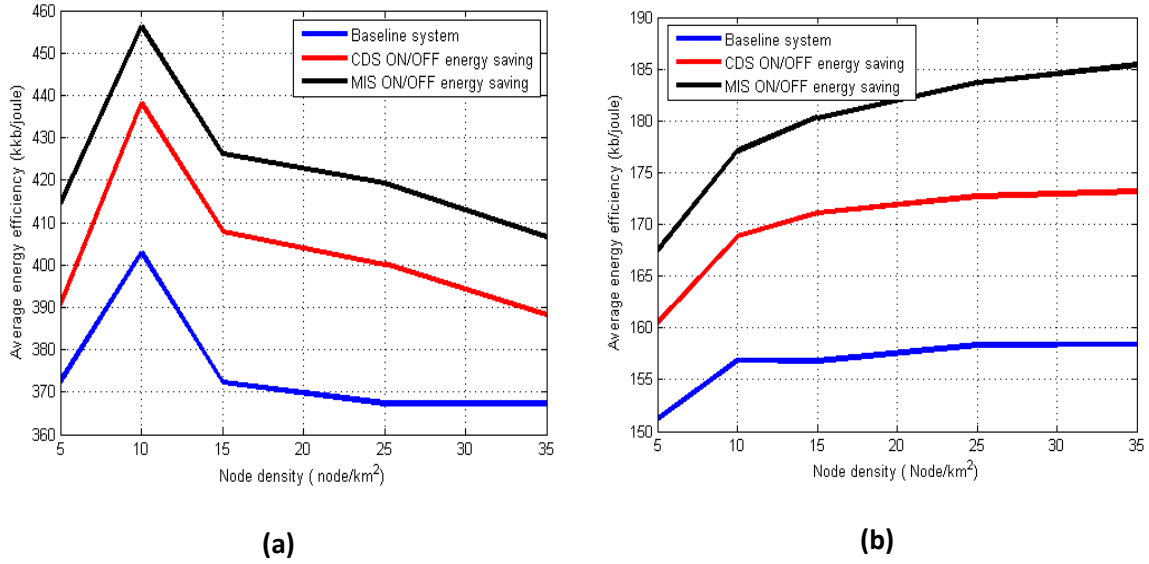


Figure 40 : Average energy efficiency wrt node density (low load, d=10%)(a) Average energy efficiency wrt node density (high load, d=10%) (b)

In Figure 40(a), the CDS based ON/OFF energy efficiency is improving the performance of the baseline system of 9% at maximum and the MIS is improving the baseline of the system of 14 % for low load situation and low heterogeneity factor d=10% .In Figure 40(b), it is shown that CDS based ON/OFF energy saving system improves the performance of the baseline system of 9% at maximum. The improvement of the MIS w.r.t to the energy efficiency of the baseline system is 17% at maximum for highly loaded HetNet cluster. When comparing the results of the Figure 38 with the results of the Figure 40, it is seen that on the average the energy efficiency is reduced when increasing the degree of the normalized heterogeneity index *d*. Another comment is that the maximum energy efficiency improvement of the MIS based ON/OFF energy saving is increasing when the normalized heterogeneity index *d*. is increased.

The Figure 41 is showing the average energy efficiency evolution with respect to node density for low and high load situation when the normalized cluster heterogeneity is increased to d=50%.

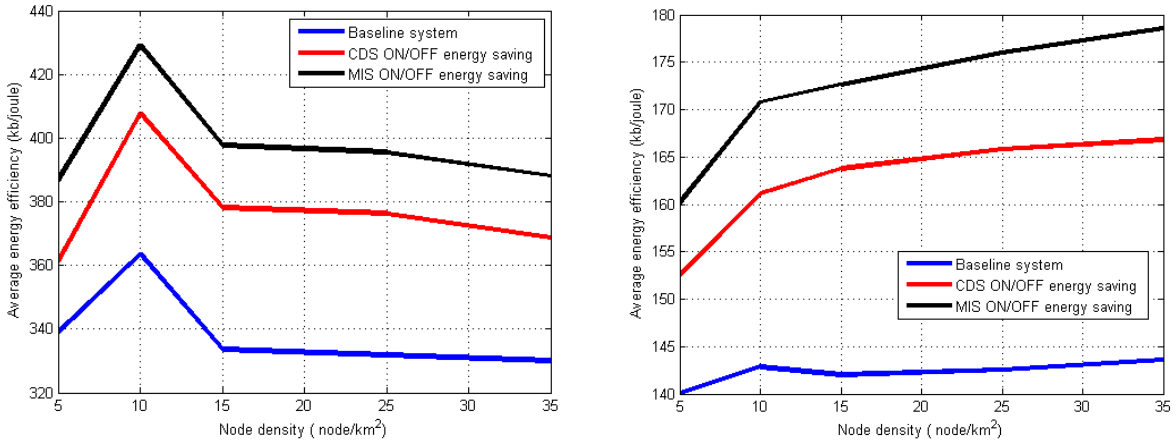


Figure 41 : Average energy efficiency wrt node density (low load, d=50%)(a) Average energy efficiency wrt node density (high load, d=50%) (b)

The results are showing that for the considered HetNet cluster deployment the MIS is the best ON/OFF energy saving technique. It is improving the baseline system of 19% at maximum for low load situation and 24% for high load situation. The CDS based ON/OFF energy saving technique is improving the performance of the baseline system of 13% at maximum for low load situation and 16% for high load situation. These results are showing that the performance improvement of MIS is increasing with the normalized heterogeneity of the cluster, i.e. 7% increase when the heterogeneity parameter d is increased from 10 to 50%. It is shown also that the average energy efficiency of the cluster deployment is lower when the heterogeneity is increased.

The Figure 42 is showing the average energy efficiency evolution with respect to node density for low and high load situation when the normalized cluster heterogeneity is further increased to d=100%, meaning that the cluster is formed of pico nodes. .

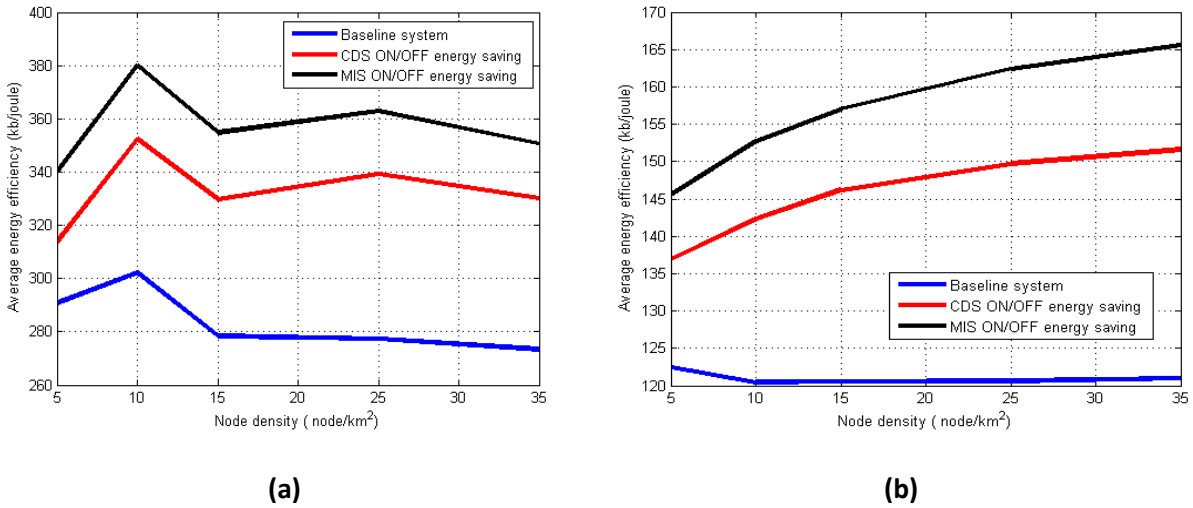


Figure 42 : Average energy saving wrt node density (low load, d=100%)(a) Average energy saving wrt node density (high load, d=100%) (b)

The results are showing that for the considered HetNet cluster deployment the MIS is the best ON/OFF energy saving technique. It is improving the baseline system of 28% at maximum for low load situation and 37% for high load situation. The CDS based ON/OFF energy saving technique is improving the performance of the baseline system of 21% at maximum for low load situation and 28% for high load situation. These results are showing that the performance improvement of MIS is increasing with the normalized heterogeneity of the cluster, i.e. 7% increase when the heterogeneity parameter d is increased to 100%. It is shown also that the average energy efficiency of the cluster deployment is lower when the heterogeneity is increased.

In Figure 43 shows the evolution of the *best* average energy efficiency gain with respect to normalized cluster heterogeneity for CDS and MIS based ON/OFF energy saving.

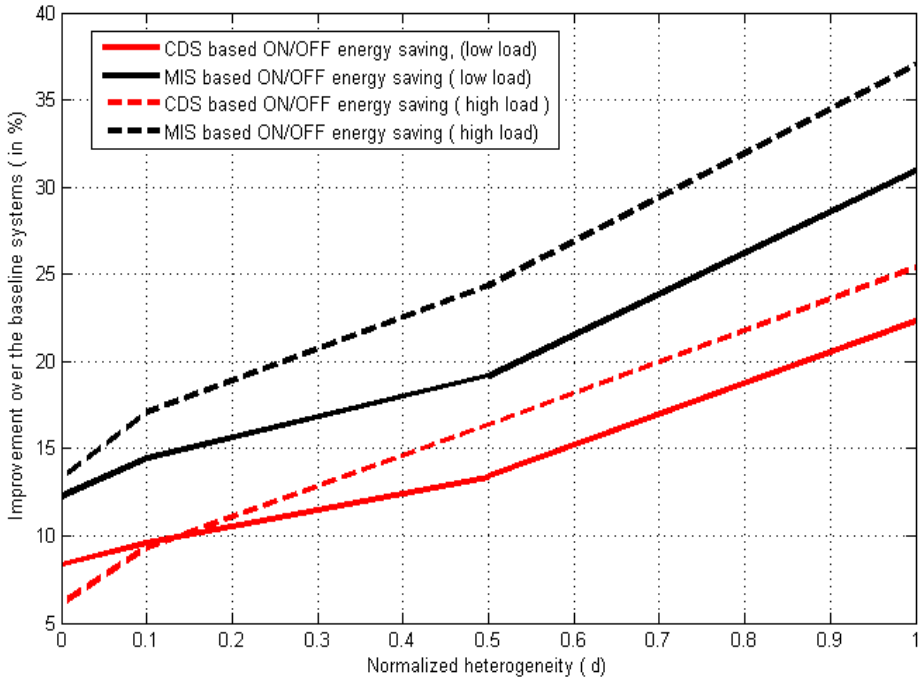


Figure 43 : Maximum average energy saving gain with respect to normalized heterogeneity

The results are showing that the energy efficiency gain of CDS and MIS based ON/OFF energy saving are increasing with the normalized heterogeneity of the HetNet cluster. For low values of the normalized heterogeneity, the cluster is dominated by femto base stations and the average energy efficiency gain obtained from using CDS and MIS energy are lower than a cluster dominated by pico base stations and MIS based ON/OFF energy saving technique is showing the best performance in terms of energy efficiency gain over the baseline system.

3.3.3.3 Discussion and final comments

In this contribution we have updated and finalized our initial results reported in [25] regarding the design of compensation based ON/OFF energy saving systems and provided an extensive simulation study of the energy efficiency of these techniques for the deployment of isolated HetNet cluster.

In [25], we studied the performance of the homogeneous cluster of pico base stations showing that the MIS based ON/OFF energy is the best ON/OFF strategy and is providing a gain of around 15% in terms of power consumed by the cluster with respect to a baseline system defined as a cluster of base stations transmitting at maximum power.

In this contribution we investigated the validity of CDS and MIS based ON/OFF energy saving strategy when considering heterogeneous clusters where the normalized average number of d% picos are deployed within a cluster of femto base stations. Varying the normalized

heterogeneity factor allowed us to simulate various levels of heterogeneity in the HetNet cluster. The final results, provided in Figure 43, are showing that CDS and MIS based ON/OFF energy saving are improving the energy efficiency of the HetNet cluster for different heterogeneity values.

So, the proposed ON/OFF energy saving techniques are both effective in the optimization of the energy efficiency of the baseline system and MIS based ON/OFF is showing the best performance.

From energy efficiency perspective, the maximum gain is obtained for pico only cluster and is around 37% for MIS. The minimum gain is around 13% for femto only cluster and the expected gain is around 25 % for a fairly heterogeneous cluster (50% femto and 50% picos). Another result is that when the normalized heterogeneity. The achieved energy efficiency is reduced when the normalized heterogeneity is reduced.

3.3.4 Energy efficiency of heterogeneous network using ON/OFF small cells in real large scale environments

The global mobile data traffic grew 69% in 2014 and is expected to increase nearly 10-fold between 2014 and 2019 [5]. Heterogeneous networks (HetNet), which consist of legacy macro-cells and additional small-cells to significantly increase the spectrum reuse, are seen as a key solution for handling this data traffic increase problem. It brings new research and design challenges on both radio access (RAN) and backhaul networks [6].

First, in the RAN, the deployment of a large number of small-cells has been shown to be a viable solution to largely increase network capacity. It brings many new challenges, including interference management, performance modeling and network design [6]. Many published studies [7], [8], and [9] address them using innovative 3D simulation approaches and realistic urban scenarios to provide some meaningful insight into the performance of 4G indoor and outdoor small-cell networks. 3D ray-based propagation models are used along with high-resolution geographical map data to accurately predict link budgets. 3D traffic distributions are generated using network measurement data and/or social network data. However, they typically focus on the busy hour, i.e. time of the day where the total traffic demand is at its maximum; and on one single aspect of the network (for example, network capacity or energy consumption).

Second, another key challenge with dense outdoor small-cell deployments is the backhaul network that relays the user data between the Base Station (BS) and the core network. Wired backhaul is usually the preferred option if available, but in many cases the small-cell backhaul will rely on wireless or a mix of technologies [10]. The location of the small-cells, typically on lampposts and below surrounding buildings, makes them hard to reach with traditional Line-Of-Sight (LoS) wireless backhaul. Therefore, new backhaul technologies and topologies are required to ensure high-throughput backhaul links even in Non-LoS (NLoS) [11]. Some recent works leverage advanced simulation tools similarly to RAN studies to address propagation modeling, performance assessment, network design and optimization [11], [12], [13], [14], and [15]. However, they are limited to relatively small areas and not very dense networks.

This contribution resumes the works initiated during the SHARING project including traffic modelling [2] intra-system offloading [3], wireless backhaul [3] and energy efficiency [4]. It aims to give a holistic view of a dense outdoor small-cell network in an urban environment considering a unified simulation scenario. Our contribution is twofold. First, the RAN is simulated and analyzed along the day in terms of network performance and power consumption, and user experience. Then, to support this RAN topology, the backhaul network performance and design to wirelessly connect 124 small-cells within 0.6 km² is studied. Towards this goal, the tools and case studies presented in [9] and [14] are extended and applied to a realistic 4G urban scenario.

3.3.4.1 Methodology and Modeling Tools

This section presents a methodology and associated modeling tools for holistic analysis of radio access and backhaul in 4G HetNet, including metrics for network performance and

power consumption, user experience, etc. This framework can be employed for technology assessment, deployment strategy definition, and radio planning and optimization.

Site-specific 4G HetNet simulator

The LTE-A simulator introduced in [9] is an extended version of the coverage simulation tool presented in [8]. Simulations rely on real environments (represented by high-resolution geo map data) and a calibrated ray-based propagation [16] developed by SIRADEL. The user throughput demand is modeled by a 3D time-variant traffic map generated using the methodology presented in [9]. It leverages different types of data, including social network data and geo map data.

The interference calculation relies on a simple abstraction of the MAC layer protocols assuming the allocation of the resources is uniformly distributed in time and frequency domains, and performed independently in each cell. The average received power from an interfering cell is then given by its maximum transmit power times the channel path-loss and its DL Traffic Load (TL) ranging from 0 to 100%. The available spectral efficiency at a user location is given by a Signal to Interference plus Noise mapping table [17], then the amount of resources required to serve all the cell users is compared to the amount of available resources and gives the cell TL. Finally, the centralized sleep mode mechanism presented in [9] is applied here. Small-cells are switched off when their TL is below a user-defined threshold (4% in this study).

Small-Cell RAN and Backhaul design

Network densification with small-cells needs to rely on user experience assessment in the legacy network and on the space and time distribution of the user traffic demand [9]. Combining site-specific ray-based propagation prediction with optimization algorithms is now commonly used for technology assessment, deployment strategy definition, and radio planning of small-cell access points and backhaul links. Several studies show the benefit of this approach for the selected target metrics for radio access studies [8], [9] and backhaul planning [11], [12], [13], [14], and[15]. Depending on operator organization and constraints, radio-planning can be done either in two separate stages (without considering the backhaul constraints in the small-cell design) or jointly. The latter one can provide an additional optimization gain [14].

The study presented here relies on a two-step design process. First, the small-cells are automatically placed in the street (typically on lampposts) based on traffic demand and existing network outage to reach the network capacity target. Then, wireless backhaul links are designed to connect the small-cells to hubs (typically macro-cell sites) using a novel backhaul design tool [14] developed by Siradel. The tool includes automated hub selection, small-cell attachment and antenna orientation for any frequencies, technologies and topologies.

Holistic HetNet analysis

A holistic view on small-cell impact for operators and users is proposed through the analysis of various types of metrics:

- Network performance and user experience
  - Macro-cell and small-cell TLs, linked to the space- and time-varying user demand;
  - Service outage, which results from overloaded cells;
  - Cell-edge user peak throughput defined as the 5% quantile of peak throughput.
- Network power consumption: using the model presented in [18] that depends on TL and small-cell mode (active or sleeping).

3.3.4.2 Heterogeneous radio access network study

The scenario is based on a typical macro-cell network layout in a dense urban environment (7th district of Paris), which is then upgraded with a co-channel outdoor small-cell network. Macro-cells are deployed over a larger area, on two rings around a central three-sector site in

order to simulate realistic interference patterns. The four scenarios described in **Erreur ! Référence non valide pour un signet.** are analyzed and compared. The main simulation parameters are given in Table 8. Busy hour defines the 60-min period with the maximum total traffic demand, which is from 10pm to 11pm here. The traffic-driven small-cell deployment is given in Figure 44 along with the spatial distribution of outdoor users.

Table 7 : Scenarios description.

ID	Network layers	Description	At busy hour	
			Capacity density (Mbps/km²)	***Service outage (%)***
1	Macro-cells	Max. capacity	55	10
2	Macro-cells + small-cells	Deployment	55	0
3	Macro-cells + small-cells	Upgraded max. capacity	332	3
4	Macro-cells + small-cells	Sleep mode	332	3

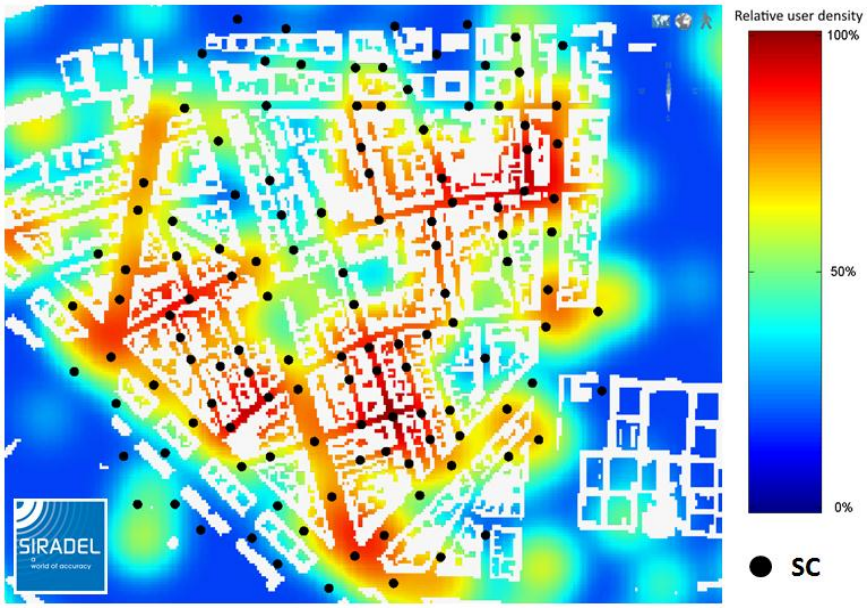


Figure 44: Small-cell deployment and outdoor user distribution at busy hour.

Table 8 RAN simulation parameters.

System	• LTE FDD 2x10 MHz • Central frequency: 2.6 GHz • MIMO configuration: 2x2, diversity
Macro-cell	• Hexagonal site deployment: two rings around the central site, i.e. 19 sites corresponding to 57 cells • Inter-site distance (ISD): 450 m • Maximum total transmit power: 20 W per antenna • Antenna: directional, 14 dBi, 6° electrical down-tilt, 32 m above ground on avg. • ICIC FFR scheme: 5% of total radio resources being allocated to each sub-band, re-use factor of 3 • Power consumption model: a_{ma}= 3.77; b_{ma}= 68.73 W
Small-cell	• Spectrum usage: co-channel. • Traffic-driven deployment of 124 small-cells with average ISD of 60 m

	• Maximum total transmit power: 2.5 W • Antenna: omnidirectional, 5 dBi, 6 m above ground • eICIC parameters: ABS duty cycle: 25%; CRE: 9 dB • Power cons. model: a_{mi}=1.11, b_{mi}=26.59W, c_{mi} = 1W
UE	• Antenna: Omni-directional, 0 dBi, 1.5 m above floor. • DL noise figure: 9 dB.

Results and analysis

First, the impact of the dense small-cell deployment is analyzed by comparing the results from macro-only and two-tier networks (scenarios #1 and #2 in dashed blue and pink lines, respectively) for the same traffic demand. Densifying the network leads to a more homogeneous load distribution, illustrated by very small traffic loads for both macro- and small-cells. This results in a very large increase in peak user throughput (+ 234% and +188% on average for DL and UL, respectively). The drawback of this infrastructure upgrade is a very large increase in network power consumption (+148%) even with very low loads. Indeed, 39% and 85% of power consumption are independent of traffic load for macro- and small-cells, respectively (considering a maximum traffic load of 70%).

Nevertheless, this upgrade and associated consumption increase is unavoidable as the macro-only network cannot cope with the high data traffic demand considered in scenario #3. The deployment of small-cells increases the network capacity by 504% (to 332 Mbps/km²) compared to the macro-only network (scenario #1). **Erreur ! Source du renvoi introuvable.** shows that the macro layer is used as an "umbrella", mainly for coverage (very low TL) and that the small-cells provide the capacity boost. Cell-edge user peak throughputs increase in busy hour by 211% in DL and 125% in UL, as shown in Figure 38.

Finally, in scenario #4, a sleep mode allows a central unit (e.g. in the core network) to switch off small-cells when their TL is below 4% [9]. Figure 39 shows an increase in TL for the macro- and active small-cells, which corresponds to the part of the traffic that cannot be handled by the small-cells in sleep mode. The same user experience is guaranteed as peak throughputs are very similar (except in quiet hours where only a few users are active) and service outage remains the same. Furthermore, the power consumption is reduced by 42% between 1am and 9am and by 16% overall, as shown in Figure 47.

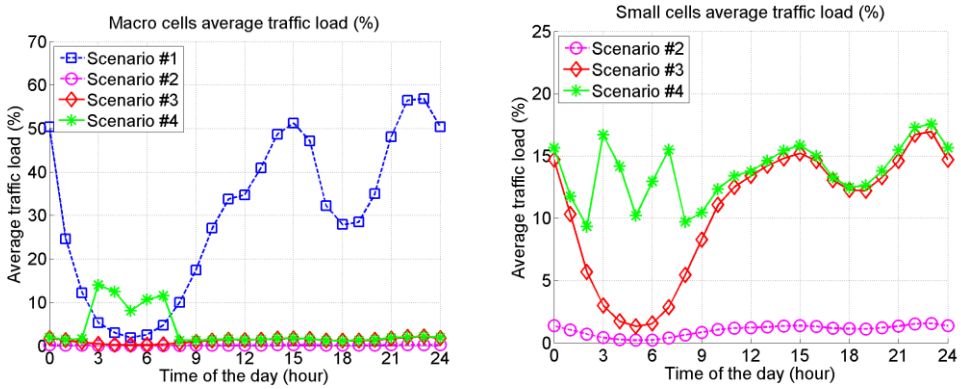


Figure 45 Cell-edge user peak throughputs along the day.

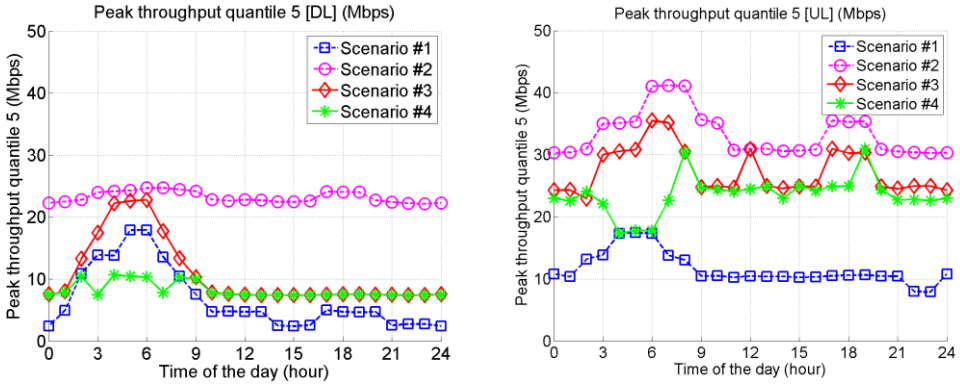


Figure 46 Network power consumption along the day.

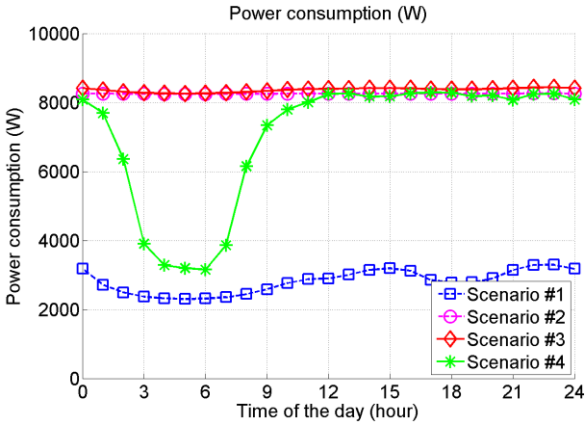


Figure 47 Network power consumption along the day.

3.3.4.3 Small-cell wireless backhaul network study

To support the radio network topology analyzed above, the small-cells (collocated with the transport remotes) need to be backhauled to the transport hubs (collocated with the macro BS). This section presents an efficient backhaul design that meets the small-cell network traffic demand presented in the previous section.

The scenario is based on the HetNet studied in the previous section. Fiber technology is not considered in this study (e.g. operator with no fixed network infrastructure). Two complementary wireless technologies are leveraged:

- P2P (Point-to-Point) backhaul in V band (60 GHz) to connect small-cells in LoS of site candidates. Multi-hop is not allowed here.
- P2MP (Point-to-MultiPoint) backhaul in 3.5 GHz band to connect small-cells in NLoS.
- The design tool [14] performs automated hub selection, small-cell remote attachment and antenna orientation leading to the backhaul network presented in Figure 48. The backhaul network properties are given in Backhaul planning aims to provide each small-cell with the required network capacity. In this study, it is computed by aggregating the user throughputs in each small-cell in busy hour. .

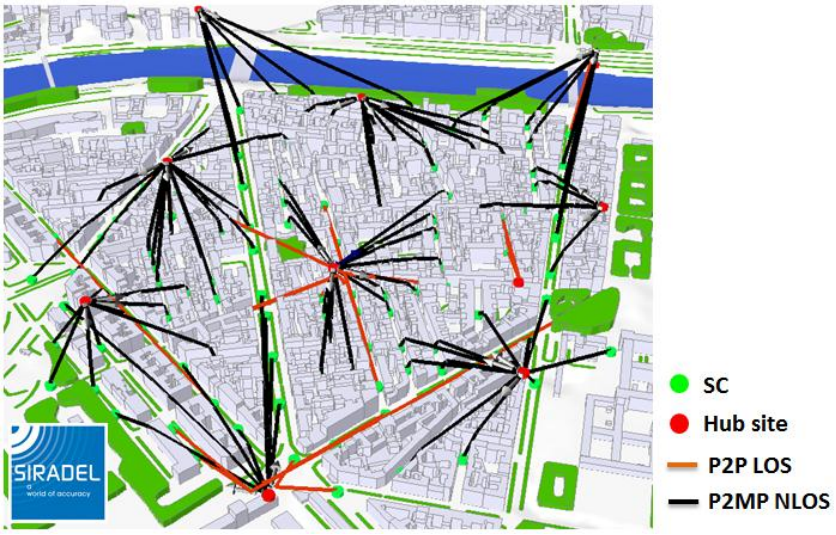


Figure 48: Small-cell wireless backhaul network.

Backhaul planning aims to provide each small-cell with the required network capacity. In this study, it is computed by aggregating the user throughputs in each small-cell in busy hour. These values, shown in Figure 49 (in red), are the target mean backhaul link throughput. Note that this metric is computed for P2MP assuming a fair allocation of the hub resources to all its served small-cell remotes (the backhaul throughput allowed per small-cell remote is proportional to the demand of each small-cell). Furthermore, the backhaul performance is also characterized by the peak link throughput, which defines the maximum peak throughput that a small-cell can deliver to a single-user who benefits from all its resources (especially at quiet time when interference is low and link quality is at its maximum) [19].

Table 9 Backhaul network parameters.

System	**V-band P2P:** FDD 2x500 MHz. Central frequency: 60 GHz MIMO configuration: no (single antenna) **Sub-6GHz P2MP:** FDD 2x20 MHz. Central frequency: 3.5 GHz MIMO configuration: Dual-polar. Spatial multi-plexing.
Hub	Collocated with macro sites (see Table 8) **V-band P2P:** 30 TRx located in 5 sites connecting 30 small-cells Maximum total transmit power: 6 dBm per antenna Antenna: directional, 2° HPBW, 40 dBi UL noise figure: 4.5 dB **Sub-6GHz P2MP:** 25 hubs located in 9 sites connecting 94 small-cells. Maximum total transmit power: 27 dBm per antenna Antenna: directional, 65° horizontal HPBW, 18 dBi UL noise figure: 4.5 dB
Remote	Collocated with small-cell sites (see Table 8) **V-band P2P:** same as hub node. **Sub-6GHz P2MP:** Maximum total transmit power: 27 dBm per antenna Antenna: directional, 24° horizontal HPBW, 17 dBi UL noise figure: 4.5 dB

First, all small-cells located in LoS of site candidates are backhauled with a P2P link at 60 GHz. Thereby, 30 small-cells out of the 124 get a dedicated short-distance link, thus

providing a very high throughput, as shown in Table 10. Mean and peak throughputs are the same as the bandwidth is not shared with other small-cells.

Then, the 94 remaining small-cells are backhauled with a P2MP system using a FDD 2x20 MHz band at 3.5 GHz. The resulting network is composed of 25 hubs, located in 9 sites, which connect from 1 to 10 small-cells each. The mean and peak throughputs are much lower than the ones from the P2P backhaul links, but still outperform the mean targets (see blue and red bars in Figure 49). The main differences with the P2P backhaul links are: the larger propagation loss due to NLoS reception that is not fully compensated by the expected better wave propagation at a lower frequency, and the P2MP topology, where the resources are shared among small-cell remotes connected to the same hub. It results in heterogeneous link performance, as shown in Figure 49 and in Table 10.

In summary, the studied hybrid wireless backhaul approach is suitable for dense 4G outdoor small-cell networks. Small-cells in LoS of hub sites (typically macro-cell sites) benefit from great mean and peak throughputs. Furthermore, the designed sub-6GHz P2MP backhaul network outperforms the mean throughput targets, allowing future traffic demand growth and possibly small-cell BS upgrade with more bands. It also allows the small-cells to provide a good user experience with min and average peak throughputs of 40.0 and 98.72 Mbps respectively.

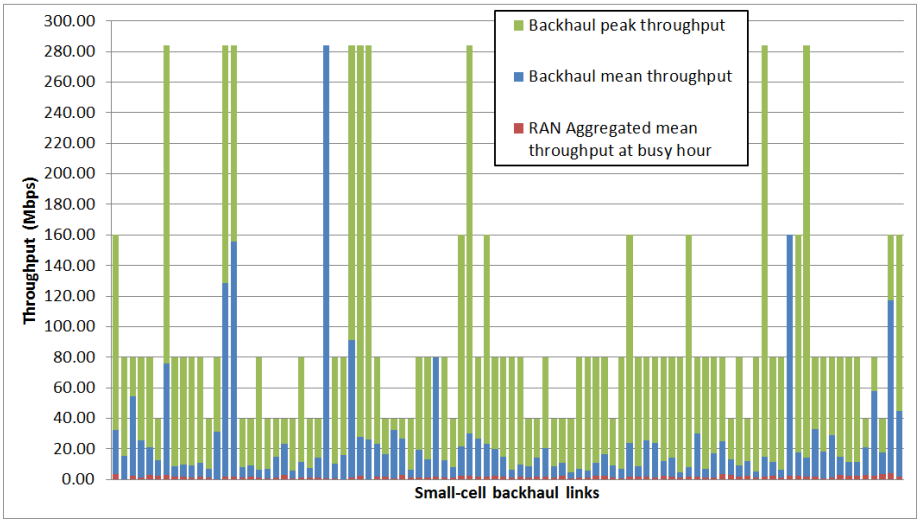


Figure 49; 3.5 GHz P2MP backhaul network DL performance.

Table 10: Backhaul DL performance overview.

Backhaul layer		**LoS 60GHz P2P**	**NLoS 3.5GHz P2MP**
Backhaul mean throughput (Mbps)	Minimum **Average**	700 **700**	5.0 **26.8**
Backhaul peak throughput (Mbps)	Minimum **Average**	700 **700**	40.0 **98.72**
Hub load	Maximum **Average**	3.8% **3.8%**	64.2% **25.4%**

3.3.4.4 Conclusion

This contribution presents a detailed analysis of a dense outdoor small-cell network in a realistic urban environment. First, the RAN is simulated and analyzed along the day. The deployment of a dense small-cell layer leads to a 504% capacity increase (to 332 Mbps/km²) compared to the macro-only network but also to a 109% power consumption rise. User experience is largely improved (cell-edge user peak throughputs increase in both DL and UL). Finally, the impact of traffic demand variations in space and along the day is analyzed and leveraged to optimize the power consumption with a sleep mode (-42% between 1am and

9am). Then, to support this RAN topology, the backhaul network performance and design to wirelessly connect the 124 small-cells is investigated. A hybrid approach, combining LoS V-band P2P and NLoS sub-6GHz P2MP technologies, is shown to be suitable and scalable. Small-cells in LoS of hub sites (typically macro-cell sites) benefit from great mean and peak throughputs. Furthermore, the designed P2MP backhaul network outperforms the mean throughput targets, allowing future traffic demand growth and possibly small-cell BS upgrade with more bands.

3.4 Spectrum management

One of the important aspects in future network deployment will be the increasing bandwidth demand, which is in confrontation with spectrum availability. In case of LTE-A networks, carrier aggregation (CA) will play an increasingly important role in providing operators the maximum flexibility for using their available spectrum. With CA, LTE-A will be able to deliver much higher throughputs than otherwise possible, by combining spectrum blocks up to 100MHz.

The scope of SHARING Task 4.4 is to study various self-organizing techniques for resource allocation, taking into account carrier aggregation features and capabilities of the future networks in the context of heterogeneous networks deployment. For example, the task will develop distributed and centralized radio resource management (RRM) strategies that ensure the stability of the network in the context of bursty traffic. Furthermore, the task will develop a learning framework in order to improve the performance of RRM strategies. A fundamental question to answer is how users can learn and improve their network utilities (rate, QoS, QoE, probability of success, etc.) in heterogeneous networks based on local information and noisy measurements?

3.4.1 Enhanced carrier aggregation mechanisms in campus of home base stations

In this work we update and finalize our contribution initially presented in [25] on coordinated carrier aggregation (CCA) by providing simulation results and performance evaluation for the deployment of single HetNet cluster with pico and femto base stations. The basic assumptions in this study are similar to the assumptions of [25], i.e. the problem we are addressing is the minimization of the number of component carriers (CCs) needed to improve the throughput performance of coordinated multipoint transmission (CoMP) systems that are using carrier aggregation. This optimization problem was formulated as graph coloring over the interference graph of the cluster and sequential heuristic coloring algorithms were tested for its solution.

The results of [25] were that the proposed coordinated carrier aggregation scheme outperforms the CoMP without CCA of 8% in terms of average throughput and that the best performance is achieved using the smallest last greedy graph coloring heuristic.

The main objective of this contribution is to study the performance of the coordinated carrier aggregation for a cluster with both home base stations and pico base stations.

The considered HetNet cluster deployment scenario is a group of randomly deployed nodes with N_1 pico base stations nodes and N_2 femto base stations nodes. The total number of nodes in the cluster is $N = N_1 + N_2$ and the normalized heterogeneity of the cluster is defined as $d = N_1 / N$.

The graph representation used in this contribution is the interference graph representation that we have developed in [25] and the graph coloring techniques are essentially, greedy coloring techniques where the nodes of the graph are sorted by the means of their degrees. Two heuristics were considered in the contribution: largest first ordering where the nodes with

high degree are colored first and smallest last ordering where the nodes with the smallest degrees and colored last in the ordering.

For sake of brevity we will not recall the graph construction or the problem formulation as graph coloring.

We will present in the next section, simulation scenario and results that are showing the performance of the coordinated carrier aggregation (CCA) system in HetNet campus with different normalized heterogeneities *d*.

3.4.1.1 Simulation scenario assumptions

The simulation scenario considered in this section is a square region of 1000x1000 meters where a HetNet cluster is deployed. The HetNet cluster is formed of two tiers of nodes, pico base stations and femto base stations and is characterized by its cluster density N and its normalized heterogeneity *d*.

The radio parameters of the pico and femto base stations are shown in the Table 5 : Radio parameters of the pico base stations campus

and both nodes are deployed in outdoor.

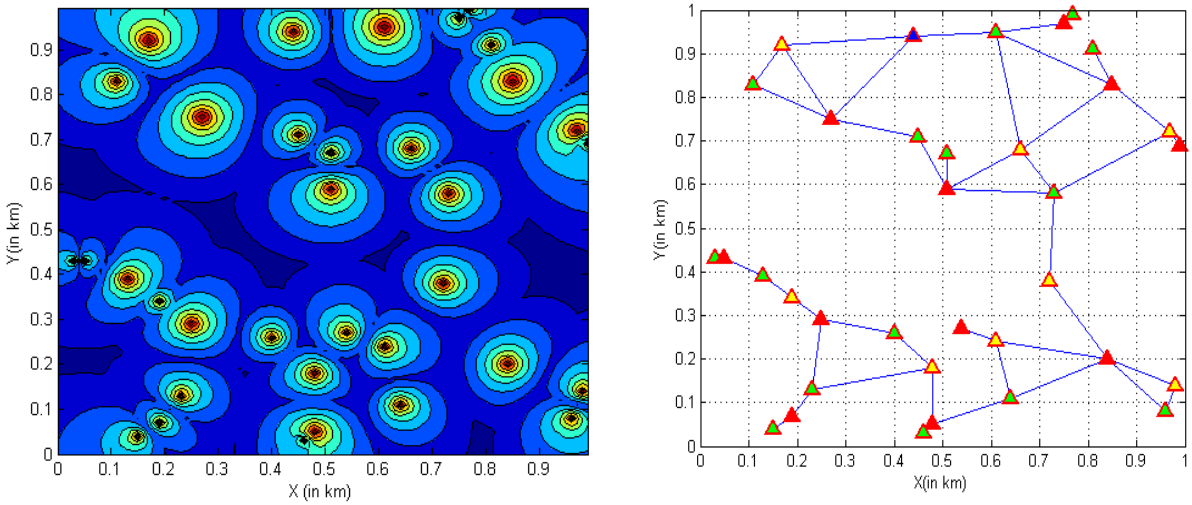
Table 11 : Radio parameters of the pico base stations campus

Parameter	**Value**
Maximum Tx power (pico)	20dBm
Maximum Tx power (femto)	16dBm
Bandwidth of the transmission	10Mhz
Antenna gain (pico)	5dBi
Antenna gain (femto)	0dBi
Pathloss parameter	a=140.7
	b=36.7
Noise density	-174dbm/Hz
Noise factor	5dB
Shadowing	NA

The scheduling used in the simulations is round robin secluding where 50 physical resource elements are assumed per base station with elementary resource bandwidth of 180 kHz. The scheduling and the elementary resource elements bandwidth will be used for the calculations of the throughput of the system and the load of the pico/femto base stations of the cluster. The performance where considered for an average load per base station of 50%.

3.4.1.2 Simulation Results

The Figure **37** is showing the interference graph and a coloring example for a cluster of density *N*=35 with normalized heterogeneity of *d*= 50%.



(a) **(b)**

Figure 50 : Coverage of a HetNet cluster (d=50%) (a) Interference graph and coloring example (d=50%) (b)

The figure is showing that the algorithm is effectively working for the HetNet cluster deployment configuration. The number of component carriers needed for optimized for improving the carrier aggregation systems is 4 component carriers

The Figure 51 shows a plot of average throughput of the cluster with respect to the nodes density, for *d*=100%.

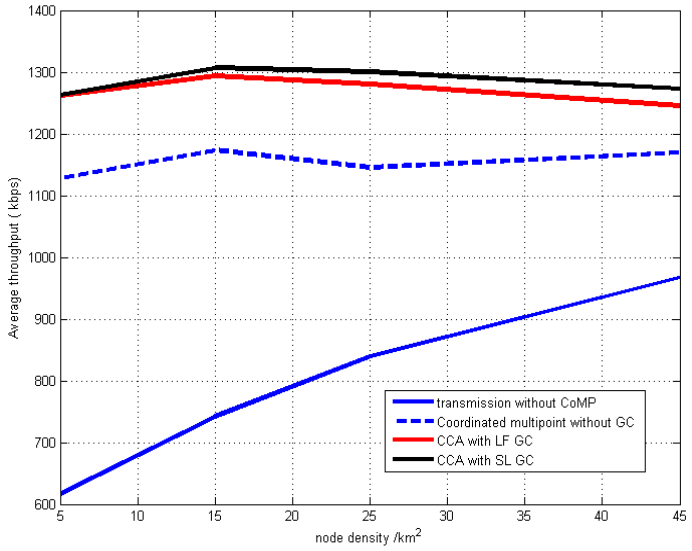


Figure 51 : Average throughput with respect to the node density

From the figure it is seen that the smallest last and largest first based CCA improves the performance of the system without coordinated multipoint transmission of 10%. The performances of largest first and smallest last CCA systems are very similar and constant with respect to the node density.

The Figure 51 shows the performance for a node density of 35 nodes/km^2 of coordinated multipoint transmission without graph coloring and CCA with largest first and smallest last graph coloring heuristics when the normalized heterogeneity of the cluster, *d* is changed from 0 to 100%.

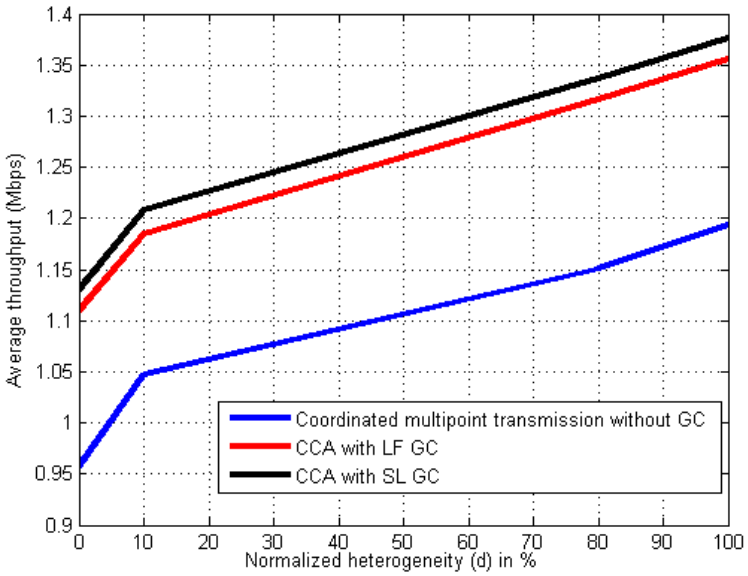


Figure 52 : Average throughput with respect to normalized heterogeneity *d*

From the figure it is seen that CCA based on graph coloring are improving the performance of the coordinated multipoint system of 15% at maximum for largest first based CCA and 17% for smallest last based CCA. The improvement of the smallest last based CCA with respect to the largest first based CCA is around 2%.

3.4.1.3 Conclusion

In this work we have presented an update on graph coloring based carrier aggregation for coordinated multipoint transmission systems. In [25], we presented the results of CCA based on two graph coloring techniques and for a uniform cluster (*d*=100%) showing that these techniques are improving the performance of the baseline of around 10% for a density of 25 home base stations per km^2.

Here we extended the study to the deployment of heterogeneous cluster with pico base stations and home base stations.

It is shown that the performance improvement of CCA with respect the baseline coordinated multipoint system is similar for low cluster densities, i.e. around 10% for largest first and smallest last CCA.

For HetNet cluster of density N=35 nodes, it is seen that the performance of CCA is improving of 15 % the performance of the baseline coordinated multipoint system

4 CONCLUSION

This final deliverable summarizes the main technical research activities carried out in SHARING WP4 for Intra-system radio access offloading, Inter-system radio access offloading, energy saving mechanisms and advanced spectrum management. The main takeaways are summarized as follows:

- **Intra-system radio access offloading**: More small cells (i.e., network densification) are advocated in conjunction with the legacy macrocellular network where a number of small cells are deployed in locations where capacity is needed the most. This entails inter cell interference mitigation techniques in time, frequency and space, with solution concepts like cell range expansion, dynamic cell switch ON/OFF, dynamic TDD and dual connectivity. For this purpose, the concept of virtual small cell has been defined alongside self-organizing capability to optimize the cell range expansion bias per small cell base station as well as the duty cycle. Moreover, the novel concept of decoupled uplink and downlink has been shown to provide significant gains as opposed to the case where the UE is served on both UL and DL by the same base station. Last but not least, it was demonstrated that caching will play an important role in 5G regarding backhaul offloading.
- **Inter-system radio access offloading**: In order to combine the benefits of both LTE and WiFi, different strategies have been proposed. First, a middleware-based technique can help users flexibly and dynamically attach to the right base station or access point depending on the interference levels, traffic load and other KPIs. In the case of collocated base stations (i.e. dual model small cells), the idea is that every small cell base station carries out load balancing by serving distinct sets of users on different bands. In the non collocated case, a handover from LTE to WiFi and vice-versa is carried out as a function of interference, load and user's location.
- **Enhanced energy efficiency or energy savings**: It is advocated to implement sleep mode for isolated and clustered deployment of small cells either by leveraging measurements, or compensation based technique. The gains depend mainly on the user density and rate requirements. The techniques were evaluated in terms of energy efficiency for the considered deployments and are can be viewed as paving the way towards the optimization of energy efficiency of future 5G deployments, viewed as a very dense network of small cells.
- **Advanced spectrum management**: Two strategies have been identified for the optimized use of spectrum via the concept of carrier aggregation. In the first one spectrum optimization is formulated as a graph coloring problem on an interference graph obtained from the aggregated measurements of the crowd of the user terminals deployed in the coverage area of the cluster. In the second one, multi-flow carrier aggregation is leveraged where instead of obtaining all component carriers from one serving nodes, the UE aggregates different sets of component carriers coming from distinct base stations. This provides more gains than the single flow carrier aggregation.

REFERENCES

[1] 3GPP TSG-RAN1#48 (Orange; China Mobile; KPN; NTT Docomo; Sprint; T-Mobile; Vodaphone& Telecom Italia) "LTE physical layer framework for performance verification" 2007

[2] SHARING deliverable D4.1 – "New opportunities, challenges and innovative concept candidates for SON/Heterogeneous Networks".

[3] SHARING deliverable D4.2 – "Intra-system offloading: innovative concepts and performance evaluation".

[4] SHARING deliverable D4.4 – "Energy saving: innovative concepts and performance evaluation".

[5] Cisco Visual Networking Index: Global mobile data traffic forecast update, 2014–2019. White paper, February 2015.

[6] J.G. Andrews, "Seven ways that HetNets are a cellular paradigm shift,"IEEE Communications Magazine, vol. 51, no. 3, pp. 136 – 144, Mar. 2013.

[7] C. Coletti, P.E. Mogensen, R. Irmer, "Deployment of LTE In-Band Relay and Micro Base Stations in a Realistic Metropolitan Scenario", Proc. IEEE 74thVeh. Tech. Conf., September 2011.

[8] F. Letourneux, et al. "3D coverage analysis of LTE urban heterogeneous networks with dense femtocell deployments." EURASIP Journal on Wireless Communications and Networking 2012.1 (2012): 1-14.

[9] G. Gougeon, et al. "Energy Efficiency of Heterogeneous Network Using On/Off Small Cells in Real Large Scale Environment." Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st. IEEE, 2015.

[10] Small-cell forum, *Backhaul Technologies for Small-cells*, February 2013.

[11] F. Letourneux, S. Guivarch, and Y. Lostanlen, "3D Propagation and Environment Modeling for NLoS Wireless Small-cell Backhaul", EuCAP 2014, The Hague, April 2014.

[12] Islam, Md Nurul, et al. "Wireless backhaul node placement for small cell networks." Information Sciences and Systems (CISS), 2014 48th Annual Conference on. IEEE, 2014.

[13] E. Karamad, et al. "Optimizing placements of backhaul hubs and orientations of antennas in small cell networks." Communication Workshop (ICCW), 2015 IEEE International Conference on. IEEE, 2015.

[14] G. Gougeon, et al. "Impact of wireless NLoS backhaul design on small-cells deployment and end-user experience." Antennas and Propagation (EuCAP), 2015 9th European Conference on. IEEE, 2015.

[15] Jafari, Amir H., et al. "Small cell backhaul: challenges and prospective solutions." EURASIP Journal on Wireless Communications and Networking 2015.1 (2015): 1-18.

[16] Y. Corre and Y. Lostanlen, "Three-dimensional urban EM wave propagation model for radio network planning and optimization over large areas", IEEE Transactions on Vehicular Technology, Vol. 58, No. 7, pp. 3112-3123, Sept. 2009.

[17] C. Mehlführer, M. Wrulich, J. C. Ikuno, D. Bosanska and M. Rupp, "Simulating the Long Term Evolution Physical Layer", in Proc. of the 17th European Signal Processing Conference (EUSIPCO 2009), Aug. 2009, Glasgow, Scotland.

[18] H. Klessig, A. J. Fehske and G. P. Fettweis, "Energy efficiency gains in interference-limited heterogeneous cellular mobile radio networks with random micro site deployment", in Proc. 34th IEEE Sarnoff Symposium, Princeton, New Jersey, USA, 2011.

[19] NGMN Alliance, *Guidelines for LTE Backhaul Traffic Estimation*, white paper, Jul. 2011.

[20] R.O.Saber, J.A.Fax and R.Murray, "Consensus and cooperation in networked multi-agent systems", *Proceedings of IEEE*, vol. 95, N°.1, pp215-233, January 2007.

[21] Celtic-Plus SHARING, Deliverable D4.1, "New opportunities, challenges and innovative concept candidates for SON/Heterogeneous Networks", 2014.

[22] A. Destounis, M. Assaad, M. Debbah and B. Sayadi, "Traffic-Aware Training and Scheduling in MISO Downlink Systems," IEEE Transactions on Information theory, 61(5), pp. 2574-2599, 2015.

[23] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the k-user interference channel," IEEE Trans. Inform.Theory, vol. 54, no. 8, pp. 3425–3441, August 2008.

[24] M. Deghel, M. Assaad, M. Debbah, "Queuing Stability and CSI Probing of a TDD Wireless Network with Interference Alignment," in proc. of IEEE ISIT 2015.

[25] Celtic-Plus SHARING, Deliverable D4.1, "New opportunities, challenges and innovative concept candidates for SON/Heterogeneous Networks", 2014.

[26] R.L.G Cavalcante, et al, " Towards energy efficient 5G wireless communication technologies", *IEEE signal processing magazine*, vol 31, issue 6, pp24-34, Nov 2014.

[27] A.Zappone, L.Sanguinetti, G.Bacci, E.Jorswieck and M.Debbah, "Energy efficient power control: A look at 5G wireless technologies", available at http://arxiv.org/pdf/1503.04609.pdf, Nov 2015.

[28] Celtic-Plus SHARING, Deliverable D4.5, "Spectrum allocation: innovative concepts and performance evaluation", 2015.

[29] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming", Optimization and engineering, 8(1): 67-127, 2007.

[30] 3GPP TS 36.323, "Evolved Universal Terrestial Radio Access (E-UTRA); Packet Data Convergence Protocol (PDCP) specification (Release 12)", V12.5.0, December 2015.

[31] SHARING deliverable D4.3 – "Inter-system offloading: innovative concepts and performance evaluation", 2015.

[32] Z. Markovic "Modification of TOPSIS method for solving of multi-criteria tasks". In: YugoslavJournal of Operations Research. 20, p.117-143, 2010

[33] E. Stevens-Navarro and V. W. S. Wong "Comparison between vertical handoff decision algorithms for heterogeneous wireless networks", Vehicular Technology Conference, VTC 2006-Spring. IEEE 63rd, vol. 2, Melbourne, Vic., May 2006, pp. 947-951

GLOSSARY

ACRONYM	DEFINITION
3G	Third Generation cellular system
3GPP	Third Generation Partnership Project
AAS	Active Antenna Systems
ABR	Average Bit Rate
ABS	Almost Blank Subframes
ACPR	Adjacent Channel Power Ratio
ANDSF	Access Network Discovery and Selection Function
AP	Access Point
BB	Baseband
BCR	Block Call Rate
BS	Base Station
CA	Carrier Aggregation
CAPEX	Capital Expenditure
CCO	Capacity and Coverage Optimization
COMP	Coordinated multi-point
CN	Core Network
CRE	Cell Range Expansion
CRS	Cell-specific Reference Signal
DCR	Drop Call rate
DTX	Discontinuous Transmission
EARTH	Energy Aware Radio and Network Technologies
eICIC	Enhanced InterCell Interference Coordination
EVM	Error Vector Magnitude
eNodeB	Evolved NodeB
FDD	Frequency Division Multiplexing
FTT	File Transfer Time
GaN	Gallium Nitrate

HetNets	Heterogenous Networks
HM	Handover Margin
IBO	Input Back Off
ICIC	Inter-Cell Interference Coordination
ISD	Inter-Site Distance
JPM	Joint Performance Metric
KPIs	Key Perforamance Indicators
LDMOS	Laterrally Diffused MOS
LHS	Latin Hypercube Sampling
LTE	3GPP Long Term Evolution
LB	LoadBalancing
LTE-A	LTE-Advanced
LTE-U	LTE in the unlicensed
LWA	LTE WiFIAggregation
LOS	Line Of Sight
MIMO	Multiple-Input Multiple-Output
MLB	MobilityLoadBalancing
MLE	Maximum Likelihood Estimation
MSC	Message Sequence Chart
NLOS	Non-Line Of Sight
NP	Network Parameter
OOB	Out Of Band
OMC	Operations and Maintenance Center
OPEX	Operational Expenditure
OFDM	Orthogonal Frequency Division Multiplexing
PA	Power Amplifier
PAE	Power Added Efficiency
PAPR	Peak to Average Power Ratio
PBCH	Physical Broadcast Channel

PTS	Partial Transmit Sequence
PSS	Primary Synchronization Signal
PU	Public
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RE	Range Extension
RETs	Remote Electrical Tilts
RF	Radio Frequency
RRM	Radio Resource Management
RSRP	Reference Signal Received Power
SAP	Service Access Point
SCBS	small cell base station
SHARING	Self-Organized Heterogeneous Advanced Radio Networks Generation
SINR	Signal to Interference Noise Ratio
SLM	Slective Mapping
SON	Self Optimizing/Organizing Network
SSPA	Solid State Power Amplifier
SSS	Secondary Synchronization Signal
TCP	Transmisionn control protocol
TL	Traffic Load
UE	User Equipment
Ul	uplink
VBR	Variable Bit Rate
WCDMA	Wideband Code Division Multiple Access
WiFi	Wireless Fidelity
WP	Work Package

X2	Interface between eNodeBs