

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Zolotukhin, Mikhail; Sayenko, Oleksandr; Hämäläinen, Timo

Title: On Optimal Deployment of Low Power Nodes for High Frequency Next Generation Wireless Systems

Year: 2018

Version: Accepted version (Final draft)

Copyright: © 2018 Elsevier B.V.

Rights: CC BY-NC-ND 4.0

Rights url: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Please cite the original version:

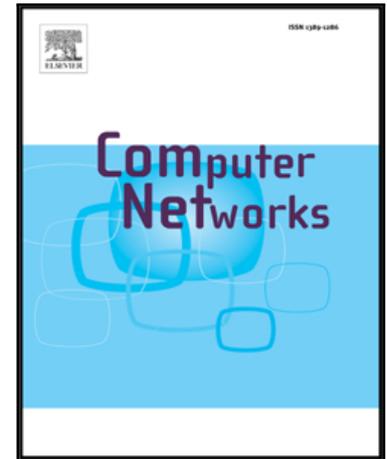
Zolotukhin, M., Sayenko, O., & Hämäläinen, T. (2018). On Optimal Deployment of Low Power Nodes for High Frequency Next Generation Wireless Systems. *Computer Networks*, 144, 120-140. <https://doi.org/10.1016/j.comnet.2018.07.029>

Accepted Manuscript

On Optimal Deployment of Low Power Nodes for High Frequency Next Generation Wireless Systems

Mikhail Zolotukhin, Alexander Sayenko, Timo Hämäläinen

PII: S1389-1286(18)30637-6
DOI: <https://doi.org/10.1016/j.comnet.2018.07.029>
Reference: COMPNW 6559



To appear in: *Computer Networks*

Received date: 29 November 2017
Revised date: 5 July 2018
Accepted date: 31 July 2018

Please cite this article as: Mikhail Zolotukhin, Alexander Sayenko, Timo Hämäläinen, On Optimal Deployment of Low Power Nodes for High Frequency Next Generation Wireless Systems, *Computer Networks* (2018), doi: <https://doi.org/10.1016/j.comnet.2018.07.029>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

On Optimal Deployment of Low Power Nodes for High Frequency Next Generation Wireless Systems

Mikhail Zolotukhin, Alexander Sayenko and Timo Hämäläinen
Faculty of Information Technology, University of Jyväskylä, Finland
{mikhail.m.zolotukhin, sayenko, timo.t.hamalainen}@jyu.fi

ACCEPTED MANUSCRIPT

On Optimal Deployment of Low Power Nodes for High Frequency Next Generation Wireless Systems

Mikhail Zolotukhin, Alexander Sayenko and Timo Hämäläinen
 Faculty of Information Technology, University of Jyväskylä, Finland
 {mikhail.m.zolotukhin, sayenko, timo.t.hamalainen}@jyu.fi

Abstract

Recent development of wireless communication systems and standards is characterized by constant increase of allocated spectrum resources. Since lower frequency ranges cannot provide sufficient amount of bandwidth, new bands are allocated at higher frequencies, for which operators resort to deploy more base stations to ensure the same coverage and to utilize more efficiently higher frequencies spectrum. Striving for deployment flexibility, mobile operators can consider deploying low power nodes that could be either small cells connected via the wired backhaul or relays that utilize the same spectrum and the wireless access technology. However, even though low power nodes provide a greater flexibility in terms of where they can be deployed, they also create new challenges. In particular, it is often the case that operators need to balance carefully between how many additional low power nodes it is necessary to install versus potential gains of the whole system. Thus, in this study we aim to develop a model that can find optimal network configuration for low power nodes assisting operators network deployment process. The outcome of the analytical model is complemented by extensive dynamic system level simulations, by means of which we analyze overall system performance for the obtained solutions. We also show that deviations from optimal configurations can lead to significantly worse system performance.

Keywords: Multi-hop network, relay, small cell, optimization, genetic algorithm

1. Introduction

One of the recent development trends in the wireless communication is characterized by adopting higher frequency ranges. While most of the existing large scale network deployments use frequencies up to 2.5 GHz, there is also a commercial interest in exploring higher frequency band ranges. In particular, the wireless industry has already been considering 3.5 GHz and 4.9 GHz frequency bands for the next generation communication systems. Furthermore, the latest WRC-15 agreed to study new bands in range of 24-40 GHz and 66-86 GHz, and some countries have already allocated spectrum in 24-28 GHz bands. One of the immediate challenges of higher frequency bands is the shrunk communication range due to the radio signal propagation properties. For the urban and dense urban deployment cases, this will be compensated by techniques such as analog beamforming and/or increased number of base stations with smaller inter-site distance (ISD). Referring to TR 38.913 [1], inter-site distance as small as 200m has been assumed for certain urban deployment cases. Even though connecting increased number of base stations to the wired backhaul network will require noticeable efforts, it is still feasible in dense urban environments because the next generation wireless system deployments will most likely leverage existing 3G HSPA and 4G LTE macro sites, for which quite dense backhaul infrastructure is already built. At the same time, 5G TR 38.913 puts requirements on supporting next generation wireless system also in the sub-urban and rural areas, whereupon the latter has the

most typical assumption for the inter-site distance of around 1500..1700m (as a more extreme case, there is an option for 5000m). It is difficult to anticipate presence of sophisticated wired infrastructure in the sub-urban and rural areas because it would require quite noticeable investments, which might not be compensated by somewhat sparse population density.

One of the brute force deployment approaches is to place next generation base stations only to the existing sub-urban/rural macro base station sites without introducing new ones. However, as the frequency range gets higher, it will not be possible to assume that a mobile terminal residing at the cell edge will be able even to decode successfully reference signals from the base station. As an example, one of the most typical macro base station transmission power densities is -30 dBm/Hz. For the 10GHz center carrier frequency, the path-loss at the distance of 1500m will be around 150dB (the exact value depends on the exact path-loss model and its parameters), resulting in the received signal being below even the thermal noise power of -174 dBm/Hz. Furthermore, exact link budget calculations must also account for other aspects, such as receiver noise figure, shadowing, fast fading margin, outdoor to indoor penetration loss, etc. As a result, there will be regions out of the high frequency transmission coverage, and thus a lot of spectrum resources will be wasted. For exactly the same reason regulatory requirements in certain countries do mandate mobile operators to ensure sufficient coverage once they acquire license for a certain spectrum.

Another approach to utilize more efficiently higher frequencies spectrum for the sub-urban and rural areas is to consider deploying additional low power nodes. However, one of the immediate challenges that an operator faces is how to connect all these base stations to the backhaul network. Wired connection is the most reliable and straightforward solution, but it limits immediately an area where such a base station can be deployed as fiber connection could be simply absent in the desired area. An operator could rely upon the microwave wireless connection, which is a common solution for macro sites, but its cost and installation expenses become quite noticeable when compared to the cost of the low power base station. Furthermore, the microwave beam is typically just a few degrees and due to propagation properties of a high frequency range where it operates, it becomes not attractive as a wireless connection for the low power base stations or small cells (especially if it is deployed below roof-top). At this point it bears mentioning that all the recent wireless technologies have an option for so-called relay networking. Practically speaking, the relay node works in exactly the same way as a low power small cell, but it uses the same wireless access technology to communicate with the super-ordinate base station. It provides quite an appealing option for an operator to deploy low power nodes and to connect them to the backhaul through the relaying mechanisms. As already presented in a number of studies, even though relay performance is obviously worse when compared to small cells, the overall system performance nevertheless becomes better [2] and coverage holes could be eliminated as with small cells. Nevertheless, in case of higher frequencies and large sub-urban/rural inter-site distance values, two-hop relays could be insufficient: if a relay is placed too far from the macro base station catering for better coverage, then macro-to-relay link performance might be a bottleneck. If a relay is placed closer to the macro base station to ensure good macro-to-relay link performance, then coverage could be compromised. Thus, as a natural solution to this trade-off question we can consider multi-hop relay systems [3].

Both small cells and relays share the same challenge that an operator should take a special care of deciding where to place (or not to place) a low power node. Of course, if an operator can identify clearly a hotspot with concentration of mobile stations and/or coverage holes, then a choice for placing a low power node becomes almost trivial. Otherwise, if concentration patterns cannot be identified and/or mobile stations tend to follow the uniform distribution, then this task becomes less straightforward. Furthermore, despite relatively low transmission power still the nodes should be placed so that it does not result in high mutual interference. In case of relays, another dimension of complexity concerns the fact that a relay uses a portion of the wireless access network resources to communicate with the super-ordinate base station. As a result, that link might become a bottleneck or it might take too much resources thus leaving less space for exchanging data with mobile terminals.

The contribution of this study is two-fold. Firstly, we provide a model that allows evaluating network system throughput in the presence of multi-hop relays or small cells, formulate

the process of searching for proper low power node positions, their connection scheme and frame partition as an optimization problem, and apply an evolutionary algorithm to solve this optimization problem. Secondly, we use the dynamic system level simulator to show that the proposed algorithm can indeed find optimal solutions that provide significant increase in network throughput when compared to other non-optimal solutions.

The rest of this paper is organized as follows. Related work is considered in Section 2. Section 3 describes the general principles of the multi-hop relays and small cells. The analytical system throughput model and a method that can be used to solve the resulting optimization problem are highlighted in Section 4. Section 5 presents system level simulation results for several deployment scenarios. Finally, Section 6 concludes the paper.

2. Related Work

As mentioned in Section 1, one of the biggest challenges with the small cells and multi-hop relays is where to deploy them provided that macro sites already exist. In case of relays, another dimension is also how to connect base stations between each other and to the backhaul network, and how to partition frame resources to avoid bottlenecks. In this section, we briefly discuss previous studies devoted to these issues.

2.1. Station Positions

One popular criteria of optimizing relay station positions is to minimize the outage probability. Study [4] evaluates outage probability based on path-loss for uniformly distributed users in a circular cell. The authors propose a simple algorithm that iteratively increases the distance between the base station and relays until a specific outage probability is reached. In case when relay coordinates reach the cell edge and there is no solution, the number of relays is increased by one. Thus, the algorithm allows finding the minimal number of relays and their placement over the studied area in order to reach the desirable outage probability. However, the algorithm completely ignores the link between the base station and relays.

A more complicated placement scheme is proposed in [5]. The paper considers a region which contains only one base station and proposes a method to deploy a fixed number of relay stations in such a way that the bandwidth requirement of mobile stations are satisfied. The proposed scheme takes into consideration the frame resource and the bandwidth requirement constraint, but it accounts for neither the interference caused by base and relay stations nor numbers of users connected to these stations.

Study [6] proposes an interesting approach that employs a distributed Neural/Genetic algorithm to compute the final mobile relay nodes positions and the more performing modulation schemes for each transmitter/receiver pair in order to guarantee an agreed QoS level. The proposed scheme allows obtaining the most suitable solution in terms of both coverage and connectivity in a dynamic environment. However, the study does not take into account the increase in interference potentially created by relays. Furthermore, the approach does not aim to increase

the throughput, but rather searches for the relay positions that would improve the coverage under QoS constraints. In particular, if a node is connected to a sink through a multi-hop path, there is probability that the link between the sink and the first relay in this path becomes a bottleneck for subsequent hops in terms of throughput.

There are also studies that look for relay positions based on the costs of deploying relay in each particular place. For example, study [7] formulates the problem of determining optimal positions of macro base stations and relays in the 802.16j network as an integer programming problem and solves it by applying branch and bound algorithm. The proposed solution selects coordinates from a given set of candidate sites based only on users demand and cost information without accounting for the additional interference caused by relay stations deployed.

As for the small cells, studies [8] and [9] propose to place small cells near hotspots with concentration of mobile stations due to the constant high demand in these geographical regions. However, it remains unclear where small cells are supposed to be installed when users are distributed uniformly inside of a macro-cell and as a result hotspots cannot be identified clearly.

Study [10] partially answers this issue by comparing three small cell deployment scenarios to improve the system performance: random deployment, deployment for the worst-performance users, and the cell-edge deployment. It concludes that the cell-edge deployment when small cells are located precisely on the boundaries between neighboring macro base stations shows the best performance.

In our previous works [2, 3], we considered optimal coordinates for the low power nodes aiming at improving the overall system performance. We used the analytical model that allowed us to find both relays and small cells coordinates, which improve performance of low throughput mobile stations by maximizing the minimal throughput value. However, placing relays and small cells in such a way that maximizes only the minimum throughput may lead into poor system capacity since in this case a lot of resources would be allocated for mobile stations with extremely low SINR. In this study, we try to find a compromise between improving performance of low throughput mobile stations and increasing the total system capacity.

2.2. Connection Scheme

There are not so many papers that study how multi-hop relays should be connected to each other and to the backhaul network. Most papers devoted to relays consider only two-hop deployment scenarios, in which all the relays deployed are simply connected to the macro base station. Papers that consider more transmission hops concentrate mostly on the impact of relay deployments on the network coverage extension. For example, study [11] demonstrates that the coverage of three-hop scenarios with six and nine relays can be five and seven times larger than that of a cell with no relays. Paper [12] compares four-hop relay network topologies that use either omni-directional or directional antennas in terms of total system costs. However, these studies do not provide any insights how the relays should be connected between each other and to the base stations in order to maximize the network coverage and/or throughput.

In [3], we present an analytical model that can provide us optimal relay coordinates for a particular multi-hop deployment configuration, however it cannot answer which particular two-hop or three-hop configuration would be the best one for the given environment. In this study, we extend our model in such a way that in addition to the coordinates it also searches for an optimal connection scheme between relays.

2.3. Frame Partition

A simplified resource allocation algorithm which maximizes the minimum users throughput in relay enhanced LTE-advanced networks is proposed in [13]. The algorithm calculates the resource allocation among transmission from Node B to its sub-ordinate relay nodes and from the relay nodes to mobile stations associated with them, as well as transmissions from Node B to mobiles in backhaul sub-frames and access sub-frames. An optimal frame partition is found based on solving a system of linear equations composed of the minimum users' throughput values.

Similar equations are used in an adaptive resource allocation scheme for dual-hop multi-relay LTE-advanced systems in study [14]. An optimal duration of time slots for the first hop and the second hop that maximizes the achievable system throughput is obtained by equalization of the achievable throughput in the first and in the second hop.

In our previous studies [3, 15], we partition the frame into access and relay zones in such a way that maximizes the worst user throughput values by simply testing for a range of all possible values, since that range is quite limited and is governed by the frame structure of a particular wireless technology. In this study, we look for such frame partition that would improve both the performance for low throughput mobiles and the total system throughput. The optimal frame partition is derived from equations similar to those utilized in [13, 14].

3. Relay and Small Cell Solutions

As briefly mentioned in Section 1, almost every wireless technology has an option for small cells because the latter is eventually just the same macro base station functionality in a different form factor providing somewhat reduced performance characteristics. Of course, almost every technology has a number of additional features for small cells and/or heterogeneous environment, but they are relatively minor enhancements rather than fundamental functional blocks.

As for the relay networking, one can find this functionality in wireless technologies such as IEEE 802.16 WiMAX and 3GPP LTE (3GPP HSPA does not have relays). For instance, IEEE 802.16 allows for a plethora of different relay options, technical description of which is given in [16] and a good overview is presented in [17]. From the viewpoint of the spectrum usage, relays can be either in-band or out-band; and from the viewpoint of the downlink (DL) management signaling, they can be either transparent or non-transparent. There is a strong motivation to consider the non-transparent in-band relays working in the distributed scheduling mode. Firstly, an in-band relay

reuses the existent spectrum instead of requiring a new frequency band. Secondly, a non-transparent relay with the distributed scheduling mode works in exactly the same way as a base station or a small cell from the viewpoint of all the key internal radio resource management algorithms. Regardless of the hop at which a particular relay is, it has to perform same basic radio resource management functions as the normal base station does, such as scheduling, link adaptation, power control etc. It allows for re-using same functional and software components at the macro base station and relay nodes. Of course, a relay node has to implement some additional functions of the mobile station to register itself to the network, ask for resources, and follow scheduling decisions from the super-ordinate node.

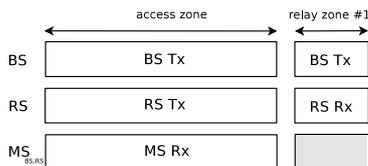


Figure 1: Non-transparent in-band two-hop relay frame structure (DL).

For the sake of further clarity, Figure 1 presents the frame structure of the non-transparent in-band relay working in the time division duplexing (TDD) mode. The relay zone is the one where the base station (BS) and relay station (RS) exchange data. It must be noted that both BS and RS transmit simultaneously in the DL access zone to the associated mobile stations (MS) thus mutually interfering. A similar situation occurs in the uplink (UL) access link when MSs associated with RS and BS start to interfere with each other. Figure 1 also shows quite eloquently that introduction of relays takes away some capacity as some portion of the DL and UL sub-frames is used for BS-to-RS communication, whereupon data cannot be exchanged with MSs. The frequency division duplexing (FDD) frame structure is identical to the one described in Figure 1 with the only difference is that DL and UL sub-frames are not following each other on the same carrier in the time domain, but rather reside on separate carriers with parallel and independent activity. Even though Figure 1 illustrates the frame structure of the IEEE 802.16 system, LTE relays works logically in exactly the same way [18, 19]. The only difference is that the LTE system allocates certain sub-frames for communication either between BS and RS (which corresponds to the relay zone in the IEEE 802.16 terminology), whereupon other sub-frames are for the communication with sub-ordinate mobile terminals (which corresponds to the access zone in the IEEE 802.16 terminology). The same principle is taken in the 5G/NR technology.

The multi-hop frame structure becomes more complex due to the fact that an intermediate relay station (RS) has to receive data from the super-ordinate base station (BS) and send data to the sub-ordinate relay(s). For the case when all the links are in-band, the frame must be partitioned in the time domain to allow for data transmission to the next relay node(s). As illustrated in Figure 2, there are three logical parts: access zone, in which data is scheduled to mobile stations (MS), and two relay zones. From the base station perspective, the first relay zone is used to

send data to the sub-ordinate relays, whereupon the second relay zone can be a "mute" zone, in which no transmission takes place. The intermediate relay utilizes all the zones where the first relay zone is used to receive data from the macro base station, and the second relay zone is used to send data to the next level relay node. It should be noted that even though Figure 2 present the DL sub-frame structure, exactly the same considerations apply also to the UL sub-frame.

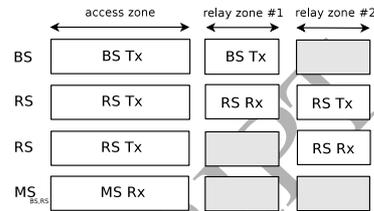


Figure 2: Non-transparent in-band multi-hop relay frame structure (DL).

Referring back to our considerations in Section 1, it can be seen that relay nodes additional dimension of complexity as an operator needs to decide how to partition properly resources between the access and relay links. In the multi-hop relay networks it becomes even more challenging as we need to decide how decide to allocate resources for the access link and several relay zones.

4. Analytical Model

In this section, we present the theoretical model that we rely on in order to find optimal serving station coordinates, connection scheme and frame partitioning. First, we evaluate the throughput in the case of small cells and multi-hop relays. After that we formulate an optimization problem that aims to find optimal serving station coordinates. Finally we propose a method that can be used to solve this optimization problem.

4.1. System Throughput

We consider a wireless network that includes set of mobile stations X and several serving stations Y . Each mobile station is connected to one of those stations such that received signal power from that station is maximal among signals from others. The power of the signal P^{rv} received at the point x from a source located at the point y is calculated as follows:

$$P^{rv}(x, y) = \frac{P^{tx} A^{tx}(x, y) A^{rv}(x, y)}{L(x, y)}, \quad (1)$$

where P^{tx} is the transmitted signal power, L is path-loss on the distance between the transmitting station located at y and receiver at x , and A^{tx} and A^{rv} are the antenna gains of the transmitting and receiving station respectively. A serving node can have an omni-directional antenna with fixed gain or directional antenna whose gain is calculated based on the angle between the antenna direction and the direction of the sender or receiver. Moreover, we assume that a mobile station cannot connect to any of the transmitting stations if the received signal power is less than thermal noise I_0 . Thus, the set of mobile stations

$m(y, Y)$ associated with a particular serving station located at point y can be defined as follows:

$$m(y, Y) = x \in X : \begin{cases} P^{rv}(x, y) \geq P^{rv}(x, \bar{y}), \forall \bar{y} \in Y \\ P^{rv}(x, y) \geq I_0. \end{cases} \quad (2)$$

In the worst case scenario, when all the serving stations transmit at full power, signal-to-interference-plus-noise ratio (SINR) $\zeta(x, Y)$ for the mobile station located at position x and associated with transmitting station at y can be found as follows:

$$\zeta(x, Y) = \frac{\kappa P^{rv}(x, y)}{I_0 + \sum_{\bar{y}: x \notin m(\bar{y}, Y)} P^{rv}(x, \bar{y})}, \quad (3)$$

where κ is the additional correction factor that allows for matching calculated SINR to the actual effective SINR.¹

In this study, we concentrate on the downlink traffic sent from the serving stations because its volumes are usually larger when compared to the uplink direction. The maximum achievable theoretical downlink throughput of the user at x without taking into account the part of wireless resources available for transmission on the link considered can be approximated based on the equation of Shannon channel capacity [20] as follows:

$$t(x, Y) = B \log_2(1 + \zeta(x, Y)), \quad (4)$$

where B is the overall transmission bandwidth. Since there can be many mobile stations associated with one serving station, each of them will receive a lower throughput. In this research, we assume that the scheduler performs fair resource allocation, i.e. the throughput of each mobile station associated with station y is the same:

$$t(m(y, Y), Y) = \alpha(x, Y)t(x, Y), \quad (5)$$

where $\alpha(x, Y)$ is the portion of the maximum achievable throughput available for station $x \in m(y, Y)$. This portion is inversely proportional to the maximum achievable throughput value at user's location x and also depends on the maximum achievable throughput values of all the other users associated with the same serving station:

$$\alpha(x, Y) = \frac{1}{\log_2(1 + \zeta(x, Y)) \sum_{\bar{x} \in m(y, Y)} (\log_2(1 + \zeta(\bar{x}, Y)))^{-1}}. \quad (6)$$

¹Firstly, this expression calculates SINR for a single sub-carrier, whereas a transport block comprises a number of sub-carriers making effective SINR value different as determined by a particular link-to-system mapping interface. In addition, average effective SINR is influenced negatively by a particular channel model that governs how the serving and interfering signals vary over the time. On the other hand, more Rx ports at the receiver side impact positively the combined effective SINR. All these dynamic factors in combination cannot be accounted for easily by a static model, and thus we rely upon a simpler approach when we just apply a certain correction factor determined based on comparing SINR CDF from this static model and effective SINR CDF from the dynamic simulator. Referring to the simulation parameters presented in section 5, we apply the correction factor of -1.5dB for the NLOS Rayleigh channel of mobile stations equipped with 2 Rx ports. At the same time, we apply 7.5dB correction factor for the LOS Rician channel of relay nodes with 4 Rx ports.

Therefore, the downlink throughput for each user $x \in m(y, Y)$ can be approximated as presented below:

$$t(m(y, Y), Y) = \frac{B}{\sum_{x \in m(y, Y)} (\log_2(1 + \zeta(x, Y)))^{-1}}. \quad (7)$$

In other words, we assume that the serving station scheduler strives for allocating same throughput for every mobile station $x \in m(y, Y)$ associated with it, and this value depends on SINR of all the mobile stations associated with this serving station.

Let us now consider the scenario when data between a macro base station and a mobile station associated with it can be transmitted in multiple hops by using N relay nodes distributed among n layers or one hop in case of small cells. Assuming that base stations occupy the 0-th layer, we denote position of the j -th serving station on the i -th layer as y_{ij} , where $i \in \{0, 1, \dots, n\}$ and $j \in \{1, 2, \dots, N_i\}$. We assume that each relay on the i -th layer is connected to the macro BS via $i - 1$ intermediate relays. Since small cells are connected directly to the backhaul network, they are also considered as located on the 0-th layer, i.e. $N = N_0 - N_{bs}$.

We introduce family of functions $S = \{s_l(y_{ij})\}$ for $i \in \{1, \dots, n\}$, $j \in \{1, 2, \dots, N_i\}$ and $l \in \{0, 1, \dots, i - 1\}$ to denote the super-ordinate station for serving station y_{ij} on the l -th layer, where $s_0(y_{ij})$ is one of the macro base stations. With the help of these functions, we can define set $M(y_{ij}, Y, S)$ of mobiles that are connected to station y_{ij} indirectly through relays on subsequent layers as follows:

$$M(y_{ij}, Y, S) = \{m(y, Y) : s_l(y) = y_{ij}\}. \quad (8)$$

Furthermore, set of relay nodes $R(y_{ij}, Y, S)$ associated with base or relay station y_{ij} can be defined as

$$R(y_{ij}, Y, S) = \{y_{i+1, k} : s_i(y_{i+1, k}) = y_{ij}\}. \quad (9)$$

In order to evaluate the throughput of mobile station x associated with node $y(x) = y_{ij}$ we have to take into account the throughput achievable for its super-ordinate relay stations on all the previous layers:

$$T(x, Y, S, Z) = \begin{cases} z_0 t(m(y(x), Y), Y) \\ z_1 t(R(s_0(y(x)), Y, S), Y_0) \frac{|R(s_0(y(x)), Y, S)|}{|M(s_0(y(x)), Y, S)|}, \\ \dots \\ z_i t(R(s_{i-1}(y(x)), Y, S), Y_{i-1}) \frac{|R(s_{i-1}(y(x)), Y, S)|}{|M(s_{i-1}(y(x)), Y, S)|}, \end{cases} \quad (10)$$

where $Y_k = \{y_{k1}, \dots, y_{kN_k}\}$ are positions of the nodes on the k -th layer and $Z = (z_0, z_1, \dots, z_n)$ is the distribution of resources in the downlink frame between access z_0 and n relay links z_1, \dots, z_n . It is worth noticing that the sum $\sum_{i=0}^n z_i = 1$.

Equation (10) does not take into account the fact that if mobile stations connected to some relay station do not fully use the capacity allocated for them by this relay's super-ordinate station, the remaining capacity can be distributed between mobile stations connected to other relays associated with the same

super-ordinate station. However, this formula allows us to estimate the maximum achievable throughput in the worst case scenario, when one of the links becomes a bottleneck.

In the case of small cells, the number of station layers is reduced to only the 0-th layer that consists of N_{bs} macro base stations and $N = N_0 - N_{bs}$ small cells. Since both base stations and small cells are connected directly to backhaul network and transmit at the same part of the frame, the throughput of mobile station x associated with node $y(x) = y_{0j}$ simply equals to

$$T(x, Y) = z_0 t(m(y(x), Y), Y). \quad (11)$$

4.2. Optimization Problem

To find optimal serving station coordinates, connection scheme and bandwidth resource distribution, we have to choose a certain performance metric that we are supposed to maximize or minimize. In principle, coverage area can play role of such metric. However, placing relays and small cells in such a way that maximizes only the coverage area may lead into poor system capacity since in this case a lot of resources would be allocated for mobile stations with extremely low SINR. Even though selecting the minimum throughput as the objective will guarantee that the throughput for every user will not fall below some value, it can cause similar issues as maximizing the coverage area. In addition to that, if there is no such set of serving station positions which would guarantee that the entire area is covered by the network, there is no way to compare different sets between each other, since they all result in the same objective function value, that is zero. Another natural choice could be to maximize the mean throughput. However, searching for maximum average throughput can lead to such solution that mobile stations that have high throughput will have even higher throughput at the expense of users whose throughput is low.

In order to account for both mobile stations with low throughput as well as the ones whose throughput is high, we may focus on maximizing a weighted sum of users' throughput values. The weights are supposed to be selected in such a way that lower throughput values have higher impact on the objective function's value. One example of the weight function $w(x, Y, S, Z)$ for throughput value of mobile x in the multi-hop relay network could be an isoquant-like function that relies on the number of mobile stations having throughput lower than $T(x, Y, S, Z)$:

$$w(x, Y, S, Z) = \left(1 + \frac{|\mathcal{X} \in X : T(\mathcal{X}, Y, S, Z) < T(x, Y, S, Z)|}{|X|} \right)^{-\alpha}, \quad (12)$$

where $\alpha \geq 0$ is the parameter that controls the slope of the weight function. Similarly, in the case of small cells, weights are calculated as follows:

$$w(x, Y) = \left(1 + \frac{|\mathcal{X} \in X : T(\mathcal{X}, Y) < T(x, Y)|}{|X|} \right)^{-\alpha}. \quad (13)$$

In case of such weight function, α value is governed by the goal that a provider has for deploying relay stations or small cells.²

In order to define constraints that all potential solutions are supposed to satisfy, we assume that the provider already has N_{bs} macro BSs and N_{lpm} low power nodes in the area considered, positions of which are constant and equal to $b = \{b_1, \dots, b_{N_{bs}}\}$ and $l = \{l_1, \dots, l_{N_{lpm}}\}$ respectively. Positions of the rest of the serving stations are defined by set of all possible positions \bar{Y} which depends on the natural landscape and artificial constructions in the area considered. In the multi-hop relay scenario, super-ordinate station for any relay on the 0-th layer is one of the macro base stations. We must also account for range Ξ of possible values of elements of Z . In fact, this range is quite limited and is governed by the frame structure of a particular wireless technology. Moreover, the sum of parts assigned for access and relay zones cannot exceed the portion \bar{Z} of the frame available in the downlink.

Thus, the optimal relay positions Y , the distribution of the relays between layers S and the distribution of the bandwidth resource between access and relay zones Z can be found by solving the following optimization problem:

$$\begin{aligned} \max_{Y, S, Z} \sum_{x \in X} w(x, Y, S, Z) T(x, Y, S, Z) \text{ subject to} \\ \left\{ \begin{array}{l} N_i > 0, \forall i \in \{1, \dots, n\}, \\ \sum_{i=1}^n N_i = N, \\ y_{0j} = b_j, \forall j \in \{1, \dots, N_{bs}\}, \\ y_{ij} \in \bar{Y}, \forall i \in \{1, \dots, n\}, j \in \{1, \dots, N_i\}, \\ \exists(i, j) : y_{i,j} = l_k, \forall k \in \{1, \dots, N_{lpm}\}, \\ s_0(y_{ij}) \in \{b_1, \dots, b_{N_{bs}}\}, \forall i \in \{1, \dots, n\}, j \in \{1, \dots, N_i\}, \\ z_i \in \Xi, \forall i \in \{0, \dots, n\}, \\ \sum_{i=0}^n z_i = \bar{Z}, \end{array} \right. \end{aligned} \quad (14)$$

where $T(x, Y, S, Z)$ and $w(x, Y, S, Z)$ can be obtained with (10) and (12) respectively.

In the case of small cells, the optimization problem is reduced to searching for optimal stations' positions:

$$\begin{aligned} \max_Y \sum_{x \in X} w(x, Y) T(x, Y) \text{ subject to} \\ \left\{ \begin{array}{l} y_{0j} = b_j, \forall j \in \{1, \dots, N_{bs}\}, \\ y_{0j} = l_j, \forall j \in \{N_{bs} + 1, \dots, N_{bs} + N_{lpm}\}, \\ y_{0j} \in \bar{Y}, \forall j \in \{N_{bs} + N_{lpm} + 1, \dots, N_0\}, \end{array} \right. \end{aligned} \quad (15)$$

where $T(x, Y)$ and $w(x, Y)$ are defined respectively by (11) and (13).

²For example, when parameter α is zero, each weight is equal to one, i.e. the objective function is the total throughput. If α is 8 then 10% of the mobiles with the worst throughput accumulate about half of the sum of the weights, i.e. the objective is to optimize low throughput stations performance. When α approaches infinity, all weights become zero except for the one that corresponds to the minimal throughput value and the objective function becomes to equal to the lowest throughput.

Since the objective function is neither convex nor concave, we cannot rely on standard optimization methods. Brute-force search would require to check $\binom{|Y|}{N-N_{ipm}} \times \binom{N-1}{n-1} \prod_{i=1}^n N_{i-1}^{N_i} \times \binom{|\Xi|-1}{n-1}$ and $\binom{|Y|}{N-N_{ipm}}$ candidate solutions for optimization problem (14) and (15) respectively. As one can notice, the space to be searched becomes incredibly large when the number of low power nodes (and, in case of relays, the number of hops) grow.

It is also worth noting that we search for stationary deployment solutions selected by the cellular network provider in the network planning phase to extend the cellular coverage and enhance the capacity of the mobile network within its coverage area. When the provider decides to introduce new low power stations in addition to the ones already deployed, a new solution should be computed with the positions of deployed nodes being considered as constraints for the new optimization problem. However, the node connection scheme and relay zone sizes are supposed to be recalculated.

Thus, the problem to be solved consists of the complex objective function defined on huge solution space in static environment. One well-known approach to solve this class of problems relies on employing an evolutionary algorithm.

4.3. Evolutionary Algorithm

Evolutionary algorithms represent a class of stochastic optimization algorithms in which the principles of organic evolution are used as rules in optimization. They are often applied to optimization problems when specialized techniques are not available or standard methods fail to give satisfactory answers. An evolutionary algorithm allows finding the global optimum of a problem even for the case of a complicated objective function. Another advantage of such algorithms is that they are well suited for parallel implementation.

There are different evolutionary optimization approaches, but as a rule they all include three main steps: initialization, breeding and selection. During the initialization phase the initial population of solutions is generated. After that, new solutions are introduced by means of some specific mutation operations. Finally, each individual's fitness is evaluated and least-fit solutions are discarded. The algorithm steps are repeated until some termination criteria is met.

In this study, we apply a genetic algorithm which is the most popular approach that relies on these evolutionary principles. A standard genetic algorithm seeks for the solution of the optimization problem in the form of a vector by applying operators such as recombination or mutation, or both of them. In our case, each solution contains station coordinates; and in case of relays, it also includes the actual connection scheme and frame partitioning.

Thus, in order to find the optimal combination of all the solution components (coordinates, connection scheme and zone sizes), we adjust the genetic algorithm's operators accordingly. While finding optimal serving station coordinates, the main aim of the algorithm is to reduce the outage probability of the last hop by moving serving stations "closer" to the mobile stations. At the same time, to achieve better system performance we try to reduce the probability of a bottleneck in the transmission

hops, for which we optimize the relay zone sizes and how relays are connected between each other.

4.3.1. Initialization

In order to generate a solution for the initial population, serving station coordinates are generated one by one. Let us assume that we have already generated set of coordinates Y . In order to generate coordinates for a new serving station y^* , C uniformly distributed candidate points $\{p_1, \dots, p_C\}$ are initialized in the area \bar{Y} . After that, one of these candidate points is selected randomly as the new serving station position using a weighted probability distribution, where a particular point is chosen with probability that is lower the higher the power of the received signal at this point:

$$\text{prob}(y^* = p_k) = \sigma \left(- \max_{y \in Y} P^{rv}(p_k, y) \right), \quad (16)$$

where $\sigma(x)$ is normalized exponential function that can be expressed as C -dimensional vector, the k -th element of which is defined as

$$\sigma(x)_k = \frac{\exp(x_k)}{\sum_{i=1}^C \exp(x_i)}. \quad (17)$$

Such an approach allows decreasing the probability of placing a relay or a small cell too close to another serving station, i.e relay or a small cell, thus avoiding excess cross-interference. Moreover, placing serving nodes into areas with lower signal strength reduces the outage probability [4] and, therefore, guarantees that more users are covered by the network.

An example of serving station positions initialization routine is demonstrated on Figure 3. In this example, we initialize four stations in a cell according to distribution (16). The gradient background corresponds to the received signal power values. As it can be seen from the figure, every time a station is generated in an area with low received signal power.

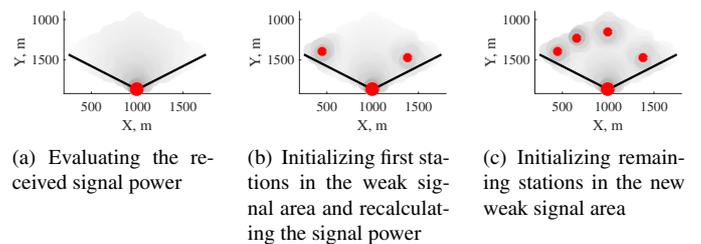


Figure 3: Example of serving station positions initialization.

In the case of small cells, the initialization of a solution is completed after positions of all the stations are generated. However, in the case of multi-hop relays, the distribution of the stations among the layers is supposed to be initialized. For this purpose, the number of relays on each layer is picked randomly from all the possible options by following a uniform distribution so that $\sum_{i=1}^n = N$. After that, the stations are supposed to be connected to the backhaul network and to each other. Since the throughput of mobiles associated with a relay depends directly on the throughput achievable by its super-ordinate relay stations on all the previous layers, we connect relays starting

from the first layer, and, therefore, reducing probability of the bottleneck on lower transmission hops. Let us assume that we have already connected $(i-1)$ layers of the stations. We connect the j -th station that occupies the i -th layer to the k -th station on the previous layer using the following weighted probability distribution that relies on the maximum throughput achievable at the super-ordinate station

$$\text{prob}(s(y_{ij}) = y_{i-1,k}) = \sigma \left(t(R(y_{i-1,k}, Y, S), Y_{i-1}) \frac{|R(y_{i-1,k}, Y, S)|}{|M(y_{i-1,k}, Y, S)|} \right) \quad (18)$$

in order to generate a connection scheme that decreases the probability of the bottleneck on the i -th transmission hop.

An example of relay connection scheme initialization procedure is demonstrated on Figure 4. In this example, we connect four relay stations by selecting connections on each layer according to distribution (18). The gradient color of each link corresponds to the value of the maximal achievable throughput of the link. As one can notice, relays are connected in an intuitively understandable way so that relays, which are closer to the base station, occupy the first layer, and the remaining relays are moved to the second layer.

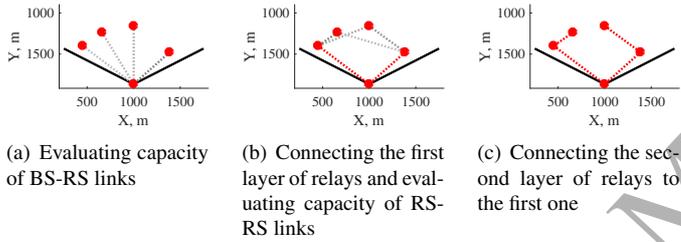


Figure 4: Example of multi-hop relay connection scheme initialization.

The same principle of reducing probability of the bottleneck is followed to select sizes of the access and relay zones. For this reason, we calculate the value of the objective function as if all relays were connected directly to the backbone network and entire frame in the downlink is assigned to the access link:

$$t_0 = \sum_{x \in X} \bar{w}(x) \bar{t}(x), \quad (19)$$

where

$$\bar{t}(x) = t(m(y(x), Y), Y) \quad (20)$$

and

$$\bar{w}(x) = \left(1 + \frac{|\chi \in X : \bar{t}(\chi) < \bar{t}(x)|}{|X|} \right)^{-\alpha}. \quad (21)$$

The objective function reaches $z_0^* t_0$ if there is no bottleneck in any of the relay links. The maximum achievable throughput of the worst relay link on the i -th layer that likely becomes the bottleneck can be found as:

$$t_i = \min_{j \in \{1, \dots, N_i\}} t(R(s_i(y_{ij}), Y, S), Y_i) \frac{|R(s_i(y_{ij}), Y, S)|}{|M(s_i(y_{ij}), Y, S)|}. \quad (22)$$

Sizes of the access and relay zones should be selected in such a way that at least guarantees

$$z_0^* t_0 = z_1^* t_1 = \dots = z_n^* t_n. \quad (23)$$

Taking into account that $\sum_{i=0}^n z_i^* = \bar{Z}$ the bandwidth resource distribution can be found as follows:

$$z_i^* = \frac{\bar{Z}}{t_i \sum_{i=0}^n t_i^{-1}}, \quad i \in \{0, \dots, n\}. \quad (24)$$

It is worth mentioning that the resulting distribution $Z^* = (z_0^*, z_1^*, \dots, z_n^*)$ does not necessarily belong to the set of distributions that satisfy the frame structure of the particular wireless technology used. For this reason, we select one of the distributions that closely resembles Z^* :

$$Z = \arg \min_{\lambda_0, \lambda_1, \dots, \lambda_n} \|Z^* - (\lambda_0, \lambda_1, \dots, \lambda_n)\|, \quad (25)$$

where $\lambda_i \in \Xi$ for $i \in \{0, \dots, n\}$ and $\sum_{i=0}^n \lambda_i = \bar{Z}$.

4.3.2. Mutation and Crossover

Once an initial set of solutions is defined, new solutions are generated. To obtain new serving station coordinates, we move from one serving station to another and select a station whose coordinates should be changed with some probability predefined (our current implementation uses mutation probability 0.25). The selected serving stations are removed and after that, for each station (i, j) position of which is chosen to be modified, we generate C candidate points $\{p_1, \dots, p_C\}$ that are distributed in the area \bar{Y} either uniformly or normally with mean value equal to the current station's position and the covariance matrix each diagonal element of which is equal to the square root of the distance to the closest transmitting station.

As previously, one of these candidates can be selected with probability that relies on the power of received signal at this point (16). However, in this case, we have to calculate power of the signal received from $\{(k, l) \neq (i, j)\}$ stations for C new points, which would increase the computational time of the algorithm's iteration. For this reason, we propose to approximate the received signal power at each candidate point with the signal received by the mobile station that is the closest to this candidate. Assuming that there are many mobile stations distributed uniformly in the area considered, for each candidate point, we can find the user that is close enough to this point to properly estimate the received signal power. The immediate benefit of such approach is that we do not need calculate the power values for mobile stations as they have been already calculated on the previous iteration.

An example of mutating three serving stations' positions is shown in Figure 5. In this example, we randomly remove two out of four stations. After that, we generate candidate points and select the new position with probability (16) from these candidate points.

We use standard multi-point crossover to add more diversity to the space of serving stations' positions. First, two parents are selected randomly from the current population. After that, $\nu - 1$ crossover points $1 \leq \eta_1 < \dots < \eta_{\nu-1} < N$ are selected randomly. A new solution is constructed by taking the coordinates of first η_1 stations from the first parent, next $\eta_2 - \eta_1$ stations from the second parent and so on. Stations from each parent solution are again selected according to (16). The stations of

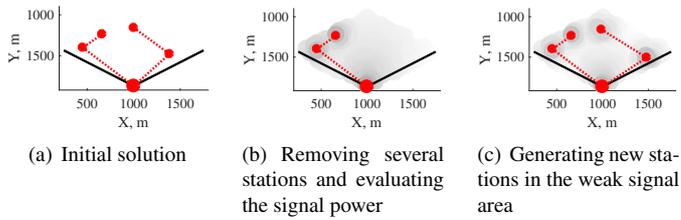


Figure 5: Example of serving station positions mutation.

the resulting set are connected between each other according to (18). We performed the crossover for every solution in the current generation, i.e. the crossover probability is equal to 1.

An example of two-point crossover is shown in Figure 6. In this example, we randomly select two parent solutions from the current population and generate a new solution by inheriting two different stations from each parent. After that, serving stations of the resulting solution are connected according to (18).

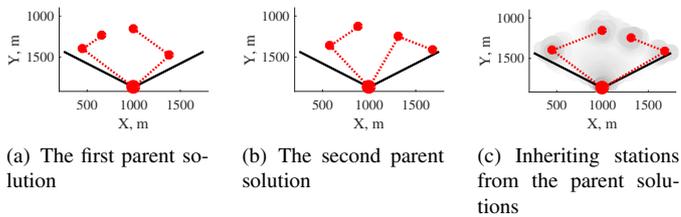


Figure 6: Example of station positions crossover.

For small cells, the process of generating new solutions is completed after positions of all the stations are generated. In case of multi-hop relays, we proceed to connect relays according to (18). For both resulting solutions, we move from one serving station to another and disconnect the station from the scheme with some probability predefined. We then proceed to again connect stations that have been disconnected using a weighted probability distribution where a connection between a relay and its super-ordinate station is chosen among other candidates with probability that relies on the throughput achievable at this relay (18). For all the schemes obtained, we select size of the access and relay zones according to (25) and add the resulting solutions to the population.

An example of scheme mutation is shown in Figure 7. In this example, we first randomly disconnect two out of four serving stations. After that, these two stations are connected to the network according to probability distribution (18).

4.3.3. Selection and Stopping Criteria

To generate a new population, a standard elitist selection method [21] can be applied that picks solutions that maximize the objective function. In order to combat crowding, i.e. a phenomenon in which some solution that is more highly fit than others in the population quickly reproduces, so that copies of this individual and very similar solutions take over a large fraction of the population, we apply the following strategy. In case of multi-hop relays, we find all unique distributions of relays

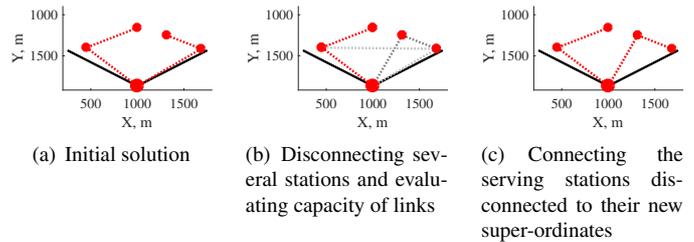


Figure 7: Example of scheme mutation.

among layers. We sort all solutions that follow each such distribution in descending order of the objective function, and proceed to select solutions one by one starting from the first one. A solution is selected if low power node coordinates of this solution are different enough from the solutions that have been already selected, i.e. average Euclidean distance between their node positions exceeds some predefined threshold. Otherwise, the solution is skipped. As a result, each solution competes only with the solutions that have the same connection scheme and similar node positions. If the number of solutions selected is not enough to form a new population, we generate more solutions using operators of mutation and crossover applied to the solutions of the current population until this number is achieved.

In addition to that, we conduct coarse grain approach to parallelization proposed in [22]. This approach relies on dividing the population into several distinct groups of individuals each of which is assigned to a different computational node, and algorithm search is performed at each node. Every given number of generations, we perform a global sort and selection of all of the fitness values obtained in these runs and the solutions selected are randomly distributed between the simulations. In case of multi-hop relays, the selection is carried out separately for each unique distribution of nodes among layers. This procedure can help to reduce the crowding problem often encountered in non-parallel implementations of the genetic algorithm.

For the solutions selected for the next generation, processes of mutation, crossover and selection are applied again. The process of forming new generations continues until all of the three following stopping criteria are met:

- A certain number of unique solutions is evaluated.
- A certain number of generations is reached.
- The highest value of the objective function remains unchanged during a certain number of the recent generations.

First two criteria guarantee that the current solution is the best among other candidates generated during the algorithm's iterations. The third criterion allows to state that there is less likely a solution that provides better result than the current one.

4.3.4. Complexity and Convergence

The fitness function requires calculating the power of the signal received by mobile stations from N serving nodes and finding maximum of the resulting N values. In case of multi-hop relays, the calculation of the throughput of links between relays and between relays and base stations requires evaluation

of SINR for $\sum_{i=1}^n N_i N_{i-1}$ links, rough upper bound of which is N^2 . Calculating capacity values requires summing elements for each layer which in total becomes N . Thus, the complexity of calculating fitness function is $O(N)$ in case of small cells and does not exceed polynomial $O(N^3)$ in case of multi-hop relays.

Complexity of mutation and crossover operators for low power node positions can be estimated as $O(N)$. In case of multi-hop relays, the number of operations required to mutate the connection scheme is proportional to $\sum_{i=1}^n (N - \sum_{j=1}^{i-1} N_j) N_i N_{i-1}$ and therefore can be bounded by $O(N^3)$. The complexity of the zone size calculation is $O(N)$ since each of N relays has only one super-ordinate station and therefore search for the worst link on each hop should not exceed N comparisons in total. Thus, time complexity of GA operators in the worst case scenario is $O(N^5)$. It is worth noting that this complexity values has to be multiplied by the number of individuals in the population.

The complexity of each algorithm iteration can be found as complexity of the fitness function multiplied by the complexity of genetic algorithm operators, which becomes $O(N^8)$ in the worst case. The total complexity of the solution can be estimated by multiplying the resulting value by number of generations. Despite the solution of the algorithm can be obtained in polynomial time, the degree of the polynomial is high. For this reason, the solution is recommended to be applied in case when the number of low power nodes to be deployed in the cell is relatively small.

Similar analysis can be carried out from the number of hops point of view. Throughput values on each of n hops are compared to each other in order to find which link is the bottleneck. Furthermore, the complexity of the zone size calculation is $O(n)$ since each relay on each of n hops has only one super-ordinate station. Thus, for any given number of hops, the complexity of the solution is $O(n^2)$.

Even though finding optimal low power node coordinates and connection scheme is a computationally expensive task, it is worth noting that it has to be performed only once when an operator decides to deploy low power nodes or considers to add a new node into environment where low power nodes already exist. In fact, a decision to deploy a new macro or a low power base station takes time and requires a lot of preparatory work, such as initial radio planning, site cost analysis, site acquisition request, etc. Thus, optimal coordinates and scheme calculations can be viewed as part of the initial radio planning step in the overall site deployment process.

Despite we cannot prove that our algorithm converges to the global optimum for the given deployment scenario, we can run several completely independent optimization task in order to have statistically meaningful results with high confidence. In case of relays, if the resulting solutions have different connection scheme between relay nodes, we select the scheme that is the most popular among the solutions found. The serving node coordinates are then calculated as average of the positions from the resulting solutions with the same connection scheme. In turn, the access and relay zone sizes are calculated according to (25).

We demonstrate the convergence of the algorithm for two

Table 1: Algorithm execution times.

Deployment scheme	Execution time
4 relay nodes, 2 hops	1287.58 seconds
4 relay nodes, 3 hops	3538.44 seconds
8 relay nodes, 2 hops	4379.43 seconds
8 relay nodes, 3 hops	8708.30 seconds

scenarios: 4 relays per a macro cell with ISD 1500m at 3.5GHz and 8 relays per a macro cell with ISD 2000m at 5 GHz. Figure 8 demonstrates performance of the algorithm with the proposed genetic operators and performance of the standard operators. Here, by "standard" we mean that both node positions and connections to other nodes are generated randomly with uniform distribution. All the results have been averaged over 10 different runs, each of which has included 100 generations with the size of population equal to 25. As one can see, the proposed operators result in faster convergence of the algorithm that can be well noticed in the case of three hops and 8 relays. In this case, the algorithm benefits from the proposed genetic operators applied to both serving node positions and connection scheme. The algorithm execution times for corresponding scenarios are provided in Table 1.

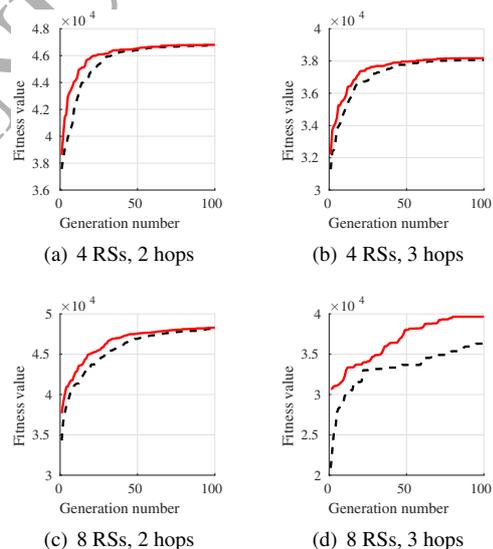


Figure 8: Evolution of the fitness function over generation for the algorithm with proposed operators (red) compared to standard ones (black dashed).

5. Simulations

In this section, we present simulation results for several scenarios aiming to analyze system performance when low power nodes are deployed according to the output of the proposed algorithm. As the general methodology for our simulations, we take the optimization problem considered in the previous section and input macro-level parameters of the environment, for which we need to find optimal coordinates of the low power nodes, e.g. relays or small cells. The output of the optimization problem is a set of coordinates for small cells, or a set of coordinates with frame partitioning in case of relays. Then, we

use the calculated coordinates in the full dynamic system level simulator to assess the final performance of the system.

As the common baseline environment we consider the sub-urban/rural area with the following parameters:

1. ISD 1500m at 3.5GHz.
2. ISD 2000m at 5GHz.

The premise idea behind choosing ISD of 1500m and 2000m is that they are the most common values used for the macro site deployments in the sub-urban and rural area (sometimes ISD of 1750m is also considered). As mentioned in Section 1, it is anticipated that an operator can and will re-use existing sites for further enhancements, and thus we take existing macro-site ISD value as a starting point. The reason behind considering 3.5GHz and 5GHz is that they are one of the most anticipated candidate bands for the next generation wireless systems that can provide more spectrum resources for operators.

We use a complex dynamic system level simulator based on NS-2 with IEEE 802.16 as a reference wireless broadband module [23]. Table 2 presents network and system level parameters for our simulation environment comprising three cells, which however belong to three different sites. In frequency reuse 1/3/1 networks these three macro base stations will have directional antennas facing each other and thus this simulation environment will capture dominating interfering nodes.

Table 2: Network and system parameters.

Parameter	Value
Inter-site distance	1500...2000 m
Center frequency	3.5...5 GHz
Bandwidth	20 MHz
Frequency reuse factor	1/3/1
BS / RS / MS Tx power	20 / 10 / 0.25 W
BS / RS / MS Rx antennas	4 / 4 / 2
BS / RS / MS antenna pattern	3GPP / omni / omni
BS / RS / MS antenna gain	17 / 5 / 0 dBi
BS / RS / MS antenna height	32 / 20 / 1.5 m
BS / RS / MS cable loss + noise figure	7 / 7 / 7 dB
BS-to-MS, RS-to-MS path loss	802.16m SMa
BS-to-MS, RS-to-MS shadowing+penetration	18 dB
BS-to-MS, RS-to-MS K-factor	0
BS-to-RS path loss	802.16j TypeD
BS-to-RS path shadowing+penetration	3.4 dB
BS-to-RS K-factor	10
Thermal noise	-174 dBm/Hz
Duplexing mode	TDD
Frame duration	5 ms
CP length	1/8 symbol
TTG+RTG	296+168 PS
OFDM symbols	47
DL/UL symbols	30/14
HARQ processes	16
HARQ retransmissions	4
HARQ advance timer	100 ms
ARQ block size	32 B
ARQ window	1024
ARQ block rearrangement	ON
ARQ deliver in order	ON
ARQ timers	
feedback interval	60 ms
retry	200 ms
block lifetime/Rx purge	500 ms

It is worth noting that both relays and small cells use an omni-directional antenna and have a lower transmission power. The motivation for this is that a lower power requires a simpler and a less expensive amplifier chain while the omni-directional

antenna simplifies the hardware design and installation efforts. Antenna height is set to 20m, which is a "compromised" value based on a range of values found from different studies and assumptions. Setting low power node antenna height to the same value as the macro base station, i.e. 32m, could be too optimistic, whereas antenna height of 5-12m is typically chosen for so-called uncoordinated deployments. For the access link the 802.16m SMa path loss model is used (which is effectively the modified Hata model [24]), and the relay link follows the 802.16j TypeD model from [25]. We do not make any special assumption with regards to the path loss model between a mobile station and a lower power node, i.e. we assume that a mobile station experiences the same path loss regardless of whether it is connected to the macro base station, relay node or a small cell.³

We run both HARQ and ARQ mechanism in the so-called hop-by-hop mode, i.e. both macro base station and all the relay nodes run HARQ and ARQ state machines. Even though ARQ can run in the so-called end-to-end mode, it has been shown in our previous study that hop-by-hop mode is more efficient, especially when the final access link is poor and might become a bottleneck [26].

To gather statistically reliable results, we run 20 different simulations, where each of them contains 16 mobile stations placed in random locations within each sector. In other words, 48 mobile stations are injected into the simulation environment upon each run. We place mobile stations at fully random locations within a cell, so a macro base station and relay nodes may have different number of associated mobile stations. In other words we do not strive for allocating same number of stations for macro base stations and low power nodes, neither do we try to distribute them evenly among them. The traffic model is full buffer, which is however sent over TCP as the application transport protocol with corresponding transport layer feedback messages. It means that effective DL throughput will depend on the TCP transmission window, which is in turn governed by how fast and reliable UL feedback messages are sent. It allows for better and more accurate estimation of the final system performance as UL failures will impact immediately the DL performance.

5.1. Simulation results (ISD 1500m, frequency 3.5GHz)

In this subsection, we present and analyze simulation results for deploying 2, 3, 4 or 6 low power nodes into the simulation environment with ISD of 1500m at 3.5GHz. First, we present our simulation results for relays showing performance with optimal and non-optimal configurations. Then, we present similar results for small cells.

³Since the macro base station has a greater antenna height, it will automatically ensure that a mobile station experiences stronger received signal strength when compared for the lower antenna height at a low power node. At the same time, adding more low power nodes should theoretically increase LOS probability as perceived by a mobile station associated them. However, we do not account for it in our simulations always assuming NLOS conditions. In other words we present somewhat more pessimistic performance numbers.

5.1.1. Relays

Table 3 presents additional configuration parameters for relay nodes, in particular, their coordinates and partitioning of frame resources. Since the relay coordinates are the same in each cell with respect to the macro base station antenna direction, we express them in the polar system through the angle and the distance from the base station (the same approach is followed in other simulation methodologies such as [25, 27]). As mentioned earlier, the output of the genetic algorithm is not only set of coordinates, but also how the frame resources should be partitioned between the access and relay zones and how relays should be connected between each other. Since there are 30 symbols in the DL sub-frame (refer to Table 2), they are further partitioned between the access and relay zone (X-Y) for two-hop relays, or between the access and two relay zones (X-Y-Z) for the multi-hop case. For the UL direction, we always use the fixed 12-2 or 10-2-2 partitioning of resources.

Table 3: Relay configurations for ISD 1500m at 5GHz.

Network configuration	Relay coordinates	Frame structure	
		DL	UL
1	-	30	14
1-2	(+/-47° 541m)	26-4	12-2
1-3	(0° 675m) (+/-39° 565m)	20-10	
1-1-2	(0° 655m) (+/-36° 602m)	16-8-6	10-2-2
1-4	(+/-13° 634m) (+/-43° 590m)	18-12	12-2
1-1-3	(0° 448m) (0° 673m) (+/-41° 579m)	18-6-6	10-2-2
1-2-2	(+/-13° 635) (+/-43° 592m)	16-10-4	
1-6	(+/-8° 608m) (+/-28° 522m) (+/-57° 631m)	18-12	12-2
1-1-5	(0° 391m) (0° 731m) (+/-25° 605m) (+/-57° 631m)	16-6-8	10-2-2
1-2-4	(+/-30° 367m) (+/-12° 671m) (+/-41° 641m)	16-8-6	
1-3-3	(0° 475) (+/-24° 567m) (0° 680m) (+/-52° 576m)	16-10-4	

For the sake of further clarity, Figure 9 shows macro base stations relays locations within the simulation environment. The macro base stations are denoted by large red dots in the hexagon vertices with their directional antennas pointing towards the center. In turn, small red dots show location of relay nodes and dotted red lines show how relays are further connected to the macro base station and to each other. As can be seen, relays are placed closer to the cell edge areas where SINR is worse. At the same time, relays are not placed right at the cell edge to maintain a good link between the macro base station and relay nodes. Referring back to Table 3, it is worth noting how the DL frame partitioning changes depending on the number of deployed relay nodes. When there are just 2 relay nodes, only 4 symbols are allocated for the relay zone because there are many mobile stations associated with every relay node, and not all of them have a good SINR. As the relay scheduler will cater for fair resource allocation among all the associated mobile stations, the effective throughput will be lower which in turn does not require a large capacity on the BS-RS link. On the contrary to it, once we deploy more relay nodes, there are fewer mobile stations with poor SINR, which motivates to allocate more resources to the BS-RS link. As can be seen, 10 and 12 symbols are allocated for the BS-RS link in case of 1-3 and 1-4 configurations, respectively.

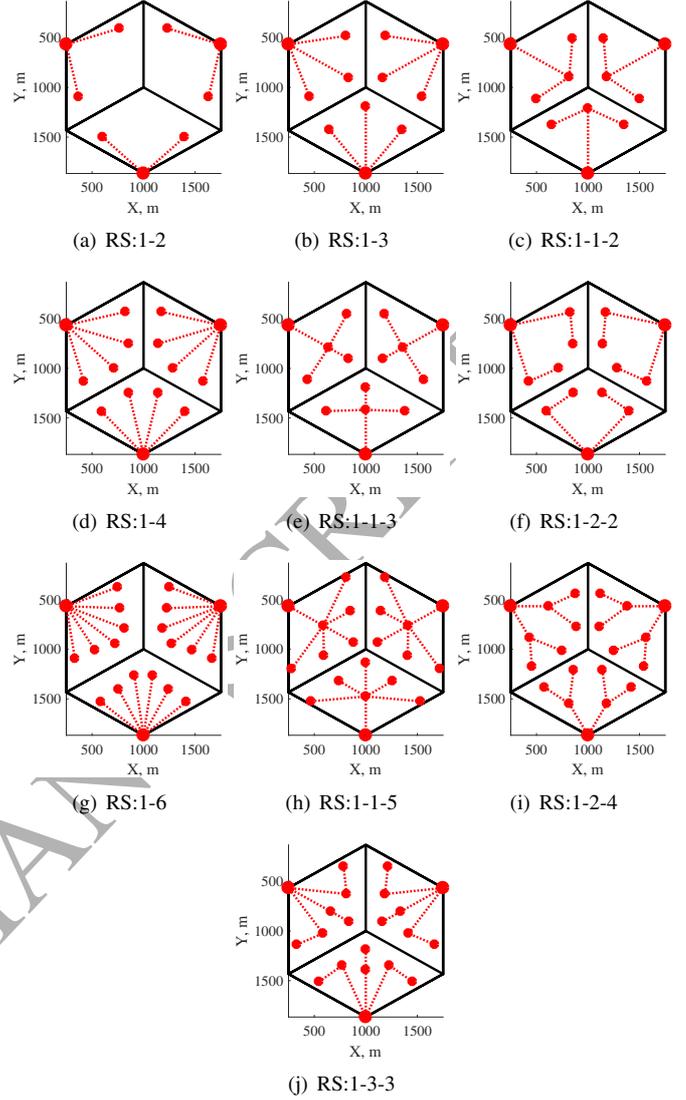


Figure 9: Optimal locations for relays (ISD 1500m, frequency 3.5GHz).

1-4 configurations, respectively.

Figure 11 presents the mean connection and total system throughput for different relay configurations for 2, 3, 4 and 6 relays per a macro base station. One of the first observations is that if there are no relays, then approximately 8% of mobile stations cannot receive any service. The matter is that ISD of 1500m has been considered predominantly for 2GHz spectrum, and switching to 3.5GHz might result in the shrunk service area (the exact number of mobile stations in outage depends of course on the exact path loss model and other factors, such as shadowing and the penetration loss). As shown in Figure 10, when no relays are deployed, then mobile stations located in cell vertices simply cannot get associated with the macro base station because they either cannot decode reliably DL reference signals and/or the UL transmission cannot be received by the macro base station. By deploying at least 2 relay nodes to the cell edge as illustrated in Figure 10, we can eliminate coverage holes. In fact, it also improves the overall performance as

the total system throughput increases by 56%, which becomes even better as we add more relays to the system. Relay configurations 1-3, 1-4 and 1-6 can improve total system performance by 127%, 175% and 205% respectively.

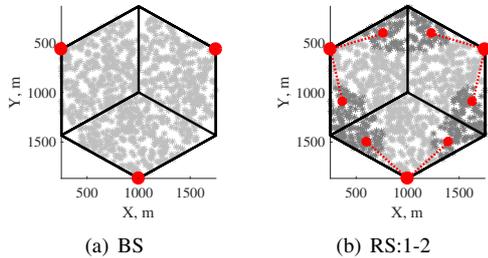


Figure 10: MS locations and associations (ISD 1500m, frequency 3.5GHz).

As can be seen from Figure 11, three-hop relays do not necessarily improve the overall performance compared to two-hop deployments, despite the fact that relays might have better channel properties when connected through another relay node. In fact, two-hop scheme 1-6 outperforms any of the three-hop 6-node configurations for the majority of connection throughput percentiles. One of the fundamental reasons why multi-hop relays cannot outperform two-hop configurations in this scenario is because two-hop relays already solve the coverage issue. At the same time, multi-hop relays require additional resources to forward data from one relay to another, which takes further resources from the access zone. As an example, configuration 1-6 has frame structure of 18-12 OFDMA symbols, while relay configuration 1-2-4 partitions frame as 16-8-6, i.e. additional resources are reserved for the relay zones making the access zone smaller. Thus, introduction of multi-hop relays does not always improve coverage, but it creates an additional transmission hop that requires allocating additional symbols.

Figure 11(c) presents the total system throughput for all the considered relay configurations (for the sake of clarity we use same colors as the throughput CDF curves in Figure 11). Both two-hop and three-hop schemes do not only eliminate the majority of blind spots, but also provide significant increase of the system throughput. Thus, an operator still can consider deploying relays in three-hop configurations, e.g. in case when there are limitations regarding preferred relay node locations and how they can be connected.

In order to show that relay coordinates are calculated optimally by the proposed genetic algorithm, we conduct a series of simulation runs in which we intentionally put relays nodes in different locations. During every simulation run, each relay node coordinates are generated randomly in a vicinity of the corresponding optimal location: we generate random angle and radius values, whereupon radius of the vicinity varies between 50 and 200 meters with 50 meters step. This approach is tested on the relay configuration 1-6 that provides best results in terms of both connection and total system throughput. As can be seen from Figure 12, relatively small deviations from the optimal coordinates up to 50m do not cause any noticeable degradation in performance, except somewhat worse performance in the low throughput range. However, as the maximum deviation

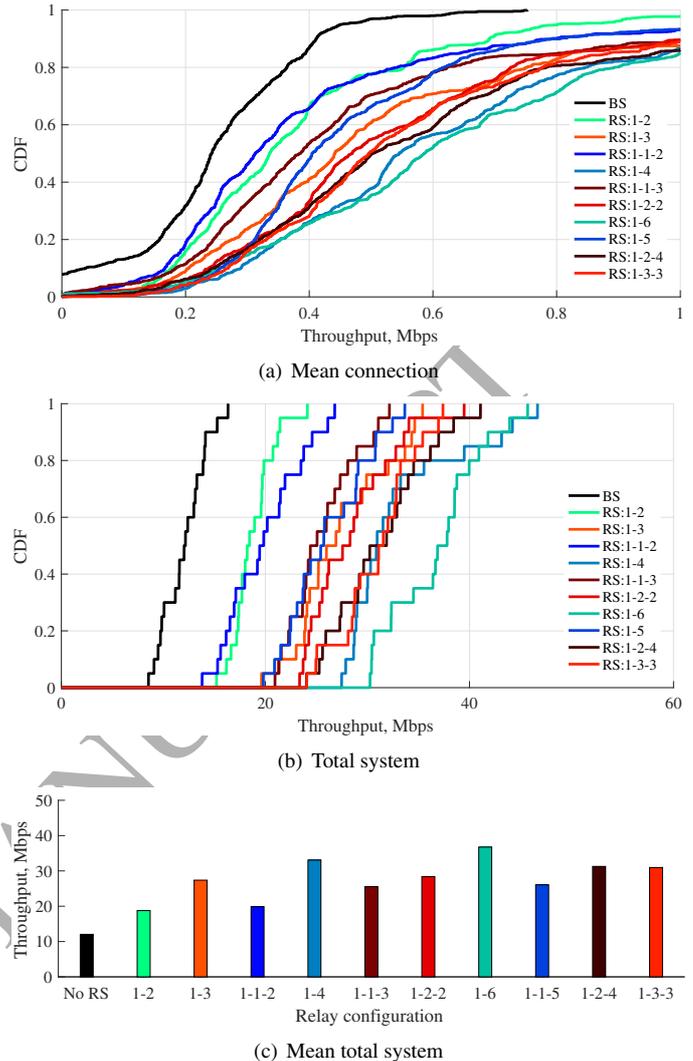


Figure 11: Throughput for 2, 3, 4 and 6 relays (ISD 1500m, frequency 3.5GHz).

distance becomes 50..100m, a clear degradation in the whole throughput range can be observed, which becomes even worse as the deviation increases. Referring to Figure 12, it is interesting to note that sometimes the overall system performance can be a bit better when relays have pseudo-random coordinates deviated from the optimal ones. The explanation is that there are obviously cases when a relay node coordinates deviate from optimal ones making the relay node closer to a particular mobile station, coordinates of which are fully random.

One of the reasons why deviations from optimal coordinates can cause such a noticeable performance loss is due to the fact that the genetic algorithm tries to find the best balance between the BS-RS link, i.e. effective SINR on that link, and the access/relay zone ratio. As a result, a big deviation from the optimal coordinates breaks this balance, which in particular increases SINR variance for the links between relays and their super-ordinate base stations. Figure 13 demonstrates that the more the relay node positions deviate from optimum the more dispersed the SINR distribution becomes. Nevertheless, it is very important to emphasize that relatively small deviations

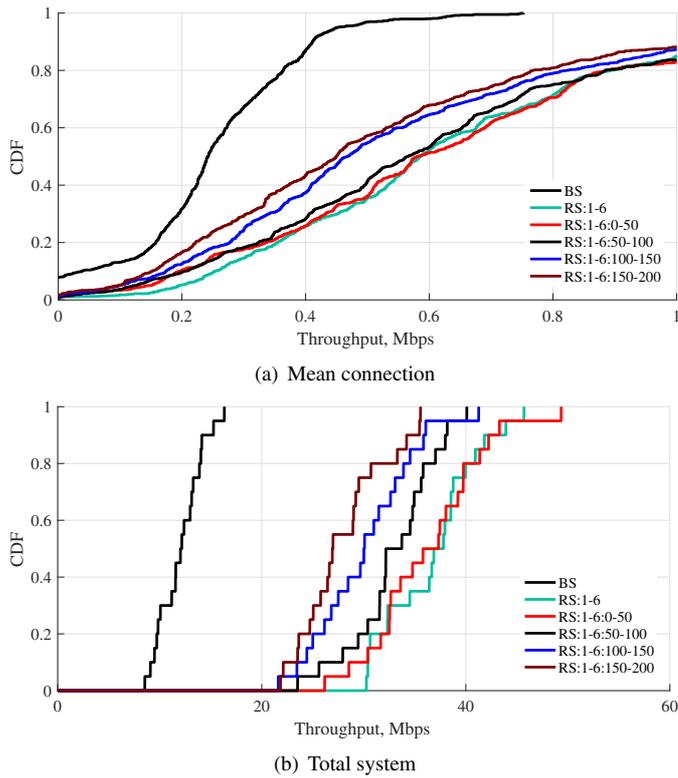


Figure 12: Throughput for 6 relays (ISD 1500m, frequency 3.5GHz), non-optimal coordinates.

from the optimal coordinates can be tolerated without a big loss in performance. This is an important observation because mobile operators do not always have a possibility to deploy relay nodes wherever they want, and thus some deviations from preferred locations could be simply inevitable.

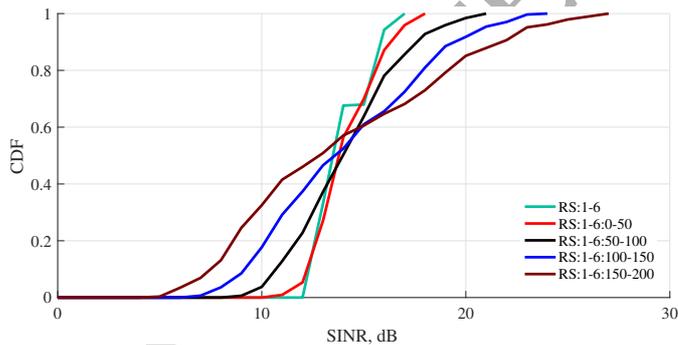


Figure 13: SINR of BS-RS link for 6 relays (ISD 1500m, frequency 3.5GHz), non-optimal coordinates.

Next, we demonstrate that proper partitioning of frame resources between the access and relay zones obtained with the genetic algorithm also plays a crucial role and impacts the final performance. For this purpose, we take relay configuration 1-6 with the optimal frame partitioning of 18-12 symbols and compare it to other configurations, in which we change the ratio between the access and relay zone size. As two extreme cases, we consider frame configurations of 22-8 and 14-16 symbols.

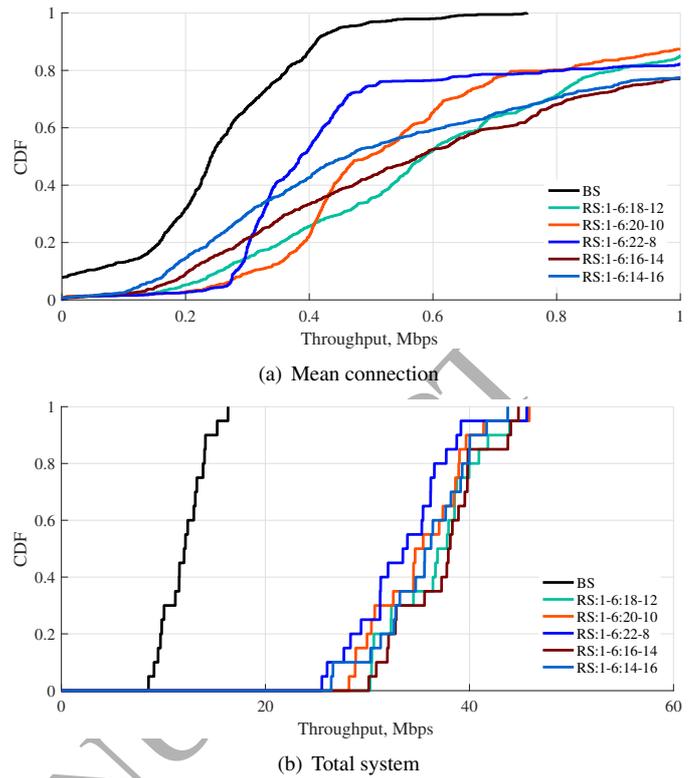


Figure 14: Throughput for 6 relays (ISD 1500m, frequency 3.5GHz), non-optimal relay zone sizes.

Figure 14 presents the mean connection and total system throughput for different frame configurations. Even though results look somewhat similar from the overall system throughput perspective, it is clear from the individual connection throughput that some frame configurations cause a disbalance between the low and high throughput stations. In some cases low throughput stations performance becomes better at the expense of sacrificing the overall performance (e.g. 22-8), while in some cases high throughput stations start to prevail. Figure 15 provides additional insight on the throughput distribution by showing separately throughput of mobile stations associated with the macro base stations and relays. As one can notice, decreasing the size of the relay zone increases the access links capacity thus allowing macro base stations to transmit more data to their sub-ordinate mobile stations. Naturally, the smaller the relay zone is, the better performance of mobile stations associated with the macro base stations becomes. At the same time, a very small relay zone starts to impact negatively stations associated with relay nodes, because the BS-RS links becomes a bottleneck not allowing to utilize the whole access zone resources at the relay node. As another extreme case, frame configuration 14-16 in Figure 14 shows that a larger relay zone size allows relays to receive more data from their super-ordinate base stations, but the overall system performance becomes worse due to a smaller access zone size.

It is important to note that even though the proposed algorithm calculates the optimal access/relay zone ratio in conjunction with the optimal coordinates, an operator still can consider

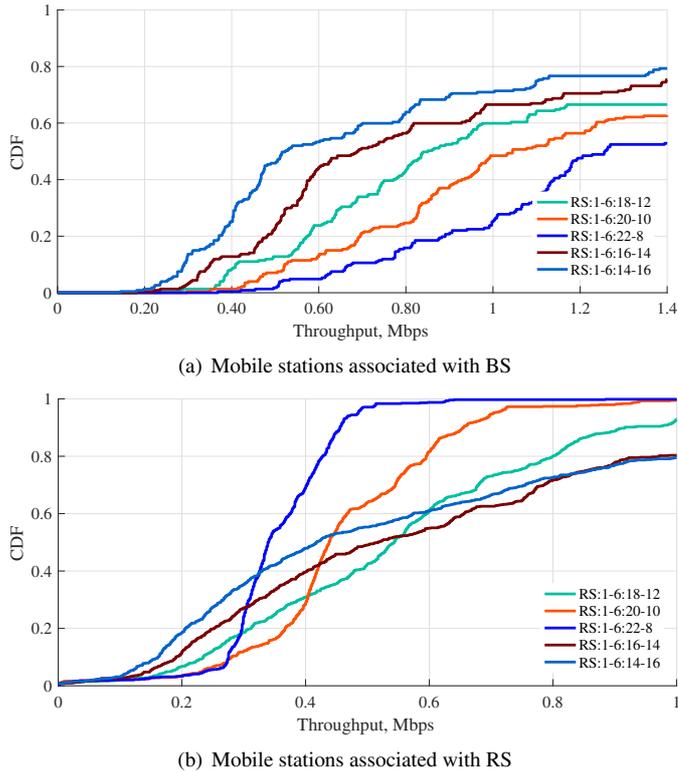


Figure 15: Mean connection throughput for 6 relays (ISD 1500m, frequency 3.5GHz), non-optimal relay zone sizes.

adjusting this ratio to favor more low or high throughput mobile stations. Referring to results in Figure 14, frame configuration 20-10 provides quite good results by improving performance of the low throughput stations. At the same time, frame configuration 16-14 is more suitable for scenarios when the maximum achievable throughput is preferred.

As a small summary, we have demonstrated that the proposed optimization problem allows us to calculate good coordinates for relay nodes and determine the optimal ratio between the access and relay zone sizes. Furthermore, if coverage is not an issue then there is no strong motivation to consider deploying multi-hop relays as they will just take more resources for the second relay zone. Thus, it is more beneficial to deploy two-hop relays for scenarios when either ISD is not large and/or when there is a strong link between the macro base station and relay nodes. At the same time, it is also worth noting that performance of certain multi-hop configurations was not considerably bad when compared to the corresponding two-hop cases, e.g. 1-2-4 and 1-6. In addition, one can notice from Table 3 that relay positions in case of 1-2-4 are somewhat different compared to 1-6. It could be the case that an operator cannot place two-hop relays where they ideally should be, but might resort for deploying relays only where it is possible, whereupon certain multi-hop configurations could be still beneficial.

5.1.2. Small cells

Similarly to relays, small cells are able to eliminate coverage holes and improve the system performance. Optimal small

cells coordinates are calculated by the same genetic algorithm; they are presented in Table 4 and are shown in Figure 16. Since small cells are connected directly to backhaul network there is no concern on the SINR between the macro base station and small cells, and thus they can be placed exactly on the cell edge in order to improve performance of low SINR users located in those areas. Furthermore, since there is no need to allocate resources for relay zones, all the OFDM symbols are allocated to the DL and UL access zones.

Table 4: Small cell positions for ISD 1500m at 3.5GHz.

Network configuration	Small cell coordinates	Frame structure	
		DL	UL
1	-	30	14
1-2	(0°605m) (38°758m)		
1-3	(17°767m) (+/-39°536m)		
1-4	(+/-11°624m) (+/-41°440m)		
1-6	(0°512m) (+/-36°496m) (12°789m)		
	(25°753m) (45°776m)		

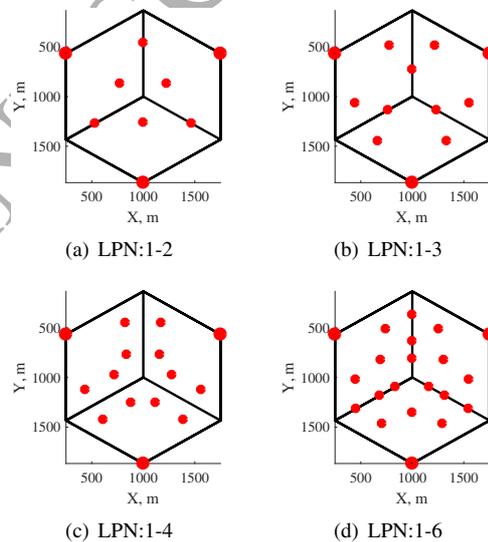


Figure 16: Optimal locations for small cells (ISD 1500m, frequency 3.5GHz).

By looking at Figure 17, one can see that the improvement in the system performance by almost 200% can be already achieved with only two additional small cells. In case of 6 small cells per a macro base station, the total system throughput can be improved by more than 475%. This is not surprising and predominantly due to the fact that each small cell is connected to a wired backhaul as opposed to relays that resort to exchanging data over the air interface.

For the configuration with 6 small cells, we test the positions obtained with the genetic algorithm against non-optimal ones generated in the same manner as it has been done for relays. As one can notice from Figure 16, the algorithm places 3 out of 6 nodes on the cell border in order to increase the received signal power for the users located in those areas. For this reason, even insignificant deviations of the node positions lead to much weaker signal power and, therefore, worse throughput values for the users located on the cell borders, as shown in Figure 18.

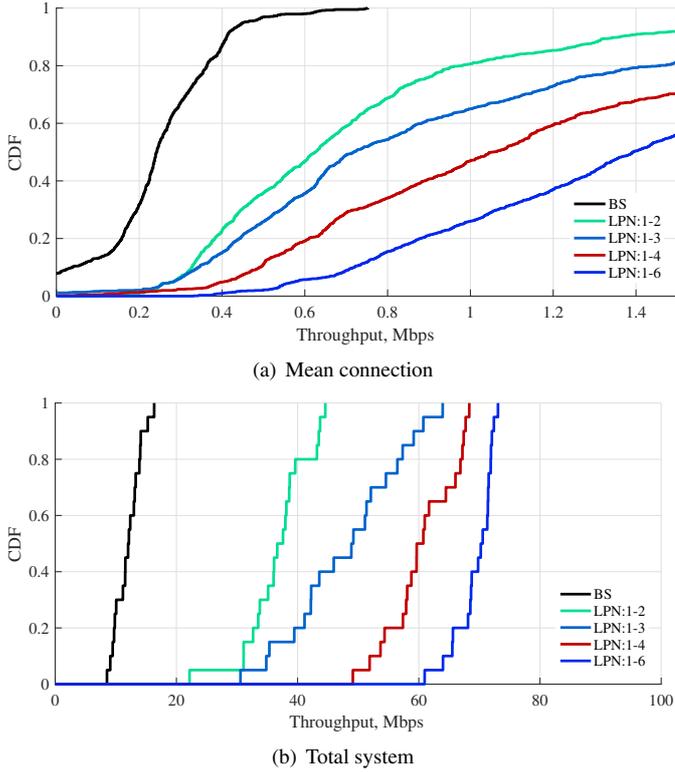


Figure 17: Throughput for 6 small cells (ISD 1500m, frequency 3.5GHz).

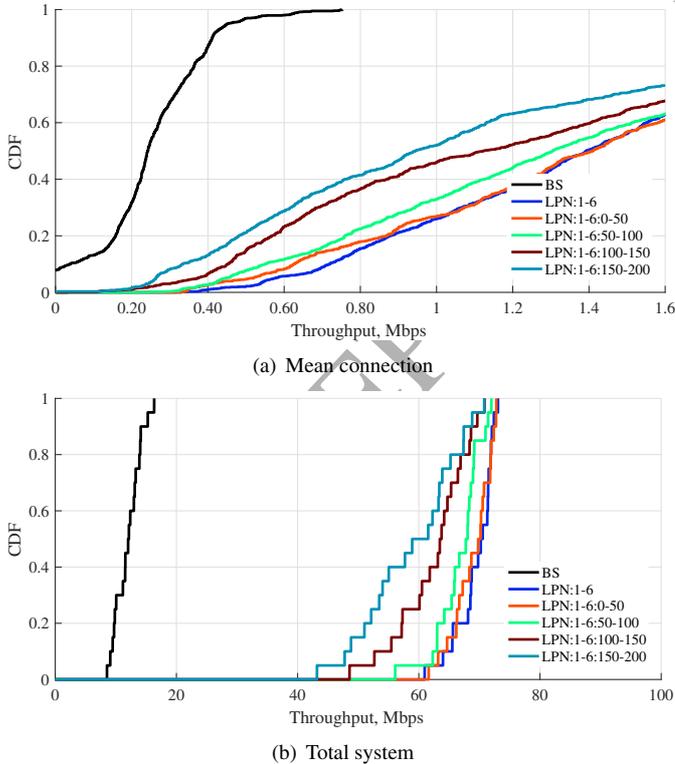


Figure 18: Throughput for 6 small cells (ISD 1500m, frequency 3.5GHz), non-optimal coordinates.

5.2. Simulation results (ISD 2000m, frequency 5GHz)

In this subsection, we present and analyze simulation results for a more challenging case when the inter-site distance is 2000m and the center frequency is 5GHz. Our preliminary simulations have indicated that 4 or 6 low power nodes are not enough to cover the whole area, and thus 8 and 10 nodes will be considered.

5.2.1. Relays

As in the scenario for 3.5GHz, Table 5 presents further parameters for relay nodes in different configurations, which includes exact coordinates of the relay nodes and access and relay zones size for each configuration scheme. As was already mentioned before, as we increase the number of relay nodes, so the relay zone size becomes larger to allow for sending more data to all the sub-ordinate nodes. And in case of 8 and 10 relay nodes the same trend can be observed: the relay zone size is 14 symbols for the 1-8 configuration and 16 symbols for the 1-10 configuration. Figure 19 shows where the relay nodes are deployed and how they are connected to each other and to the base station. As in the previous scenario for 3.5GHz, relays are placed closer to the cell edge to improve performance in areas with low SINR values eliminating coverage holes. At the same time they are not placed right at the cell edge to ensure a sufficient capacity of the BS-RS link.

Table 5: Relay configurations for ISD 2000m at 5GHz.

Network configuration	Relay coordinates	Frame structure	
		DL	UL
1	-	30	14
1-8	(0°895m) (0°570m) (+/-50°907m) (+/-40°619m) (+/-21°791m)	16-14	12-2
1-1-7	(0°563m) (0°860m) (+/-51°898m) (+/-37°648m) (+/-19°833m)	14-8-8	10-2-2
1-2-6	(+/-30°489m) (+/-60°864m) (+/-39°860m) (+/-10°902m)	16-8-6	
1-3-5	(0°714m) (+/-40°603m) (0°987m) (+/-54°904m) (+/-31°872m)	16-10-4	
1-4-4	(+/-41°569m) (+/-17°714m) (+/-47°961m) (+/-13°959m)	16-10-4	10-2-2
1-10	(+/-54°880m) (+/-44°496m) (+/-34°819m) (+/-17°642m) (+/-9°913m)	14-16	
1-1-9	(0°579m) (0°1032m) (+/-59°889m) (+/-41°584m) (+/-36°825m) (+/-17°894m)	12-8-10	10-2-2
1-2-8	(+/-36°529m) (+/-56°865m) (+/-33°820m) (+/-16°694m) (+/-9°907m)	14-10-6	
1-3-7	(0°583m) (+/-36°453m) (0°979m) (+/-15°833m) (+/-36°816m) (+/-57°882m)	14-10-6	
1-4-6	(+/-12°533m) (+/-37°609m) (+/-9°887m) (+/-35°823m) (+/-57°864m)	16-10-4	10-2-2
1-5-5	(0°599m) (+/-23°573m) (+/-42°571m) (0°970m) (+/-24°874m) (+/-49°933m)	16-10-4	

Figure 20 and 22 present the mean connection throughput collected from all the mobile stations over all the simulation runs, as well as the total system throughput calculated for each simulation separately for different configurations with 8 and 10 relays respectively. As can be seen, if there are no relay nodes, then more than 50% of the mobile stations are out of the coverage area. As further illustrated by Figure 21, high frequency and a large inter-site distance results in a situation when a noticeable area is out of service. On the contrary to it, 8 and 10 relay nodes can increase noticeably available service area eliminating coverage holes. In addition, as can be seen from

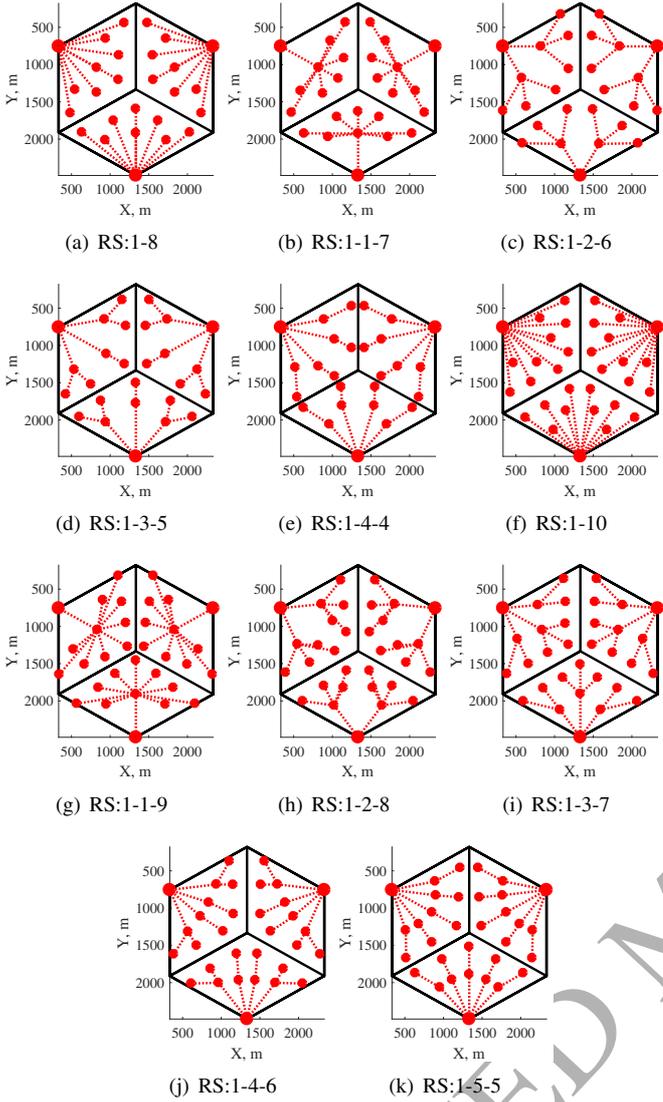


Figure 19: Optimal locations for relays (ISD 2000m, frequency 5GHz).

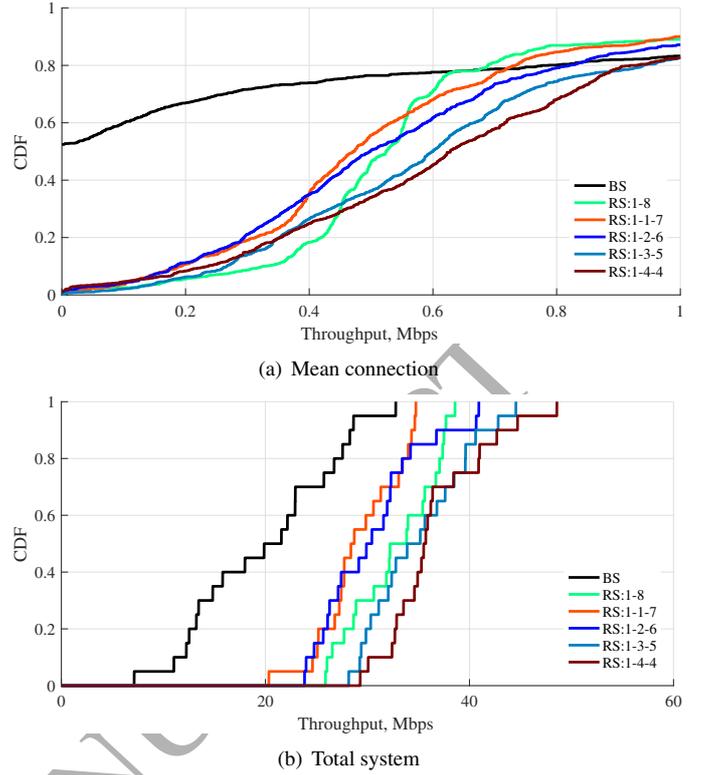


Figure 20: Throughput for 8 relays (ISD 2000m, frequency 5GHz).

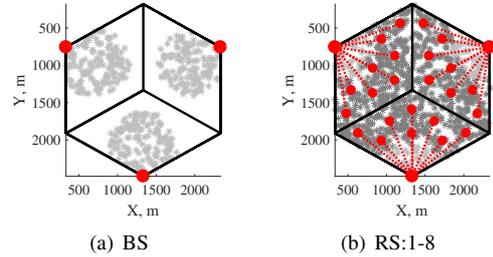


Figure 21: MS locations and associations (ISD 2000m, frequency 5GHz).

Figure 21, relay nodes also offload the macro base stations improving performance of mobile stations associated with them.

While comparing the mean connection and the total throughput of different configurations for 8 and 10 relay nodes, one can notice that two-hop configurations provide slightly better performance for low throughput users. However, some multi-hop configurations such as 1-4-4 and 1-5-5 outperform two-hop configurations from the viewpoint of the overall system results. As an example, 50-th percentile throughput for the three-hop configuration 1-4-4 is 19.6% greater than two-hop configuration 1-8. For 10 relay nodes, this difference is around 10% in favor of three-hop configuration 1-5-5. Reasons on why multi-hop relays start to provide better performance in this scenario are multi-fold. Firstly, if an operator caters for eliminating coverage holes with two-hop relays, then we will resort for placing relay nodes farther from the macro base station compromising the performance of the BS-RS link. In this case, multi-hop relays become more beneficial as the BS-RS and RS-RS links will

have better SINR. Secondly, as the performance of BS-RS and RS-RS link become better, fewer symbols are needed for the relay links allowing to allocate more resources for the access zone. As can be seen from Table 5, relay configuration 1-5-5 has 16 symbols allocated for the access zone, whereas relay configuration 1-10 has 14 symbols. All these observations in combination lead to a conclusion that by selecting proper coordinates and configurations for the multi-hop relays, an operator can achieve better performance for the given environment. This conclusion can be confirmed when comparing the total system throughput of different configurations for 8 and 10 relay nodes. As can be seen from Figure 23, total system throughput of three-hop configurations 1-4-4 and 1-5-5 exceeds the throughput of their two-hop counterparts by 11.8% and 9.6%, respectively.

Finally, it is also worth noting that some multi-hop relay configurations tend to provide better results, while some configurations usually have the worst performance. Referring to simu-

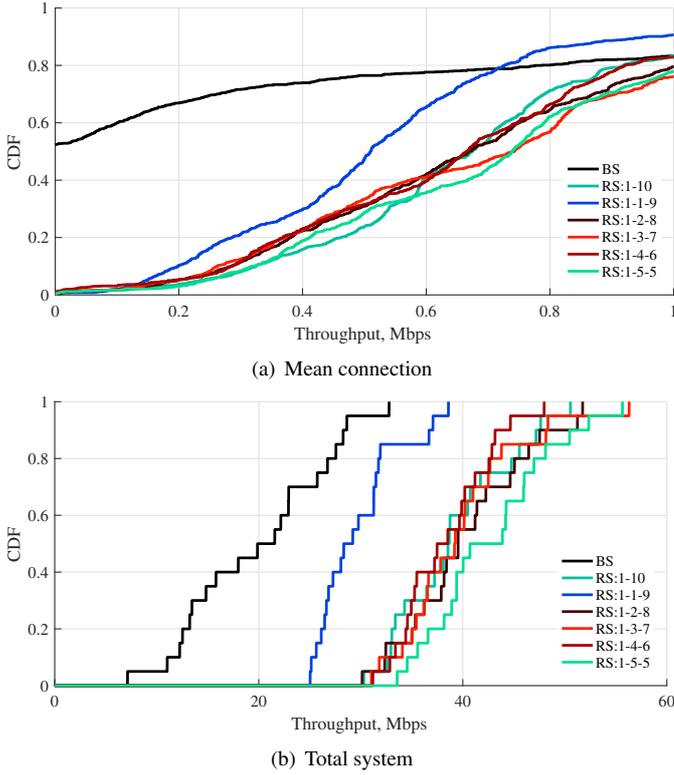


Figure 22: Throughput for 10 relays (ISD 2000m, frequency 5GHz).

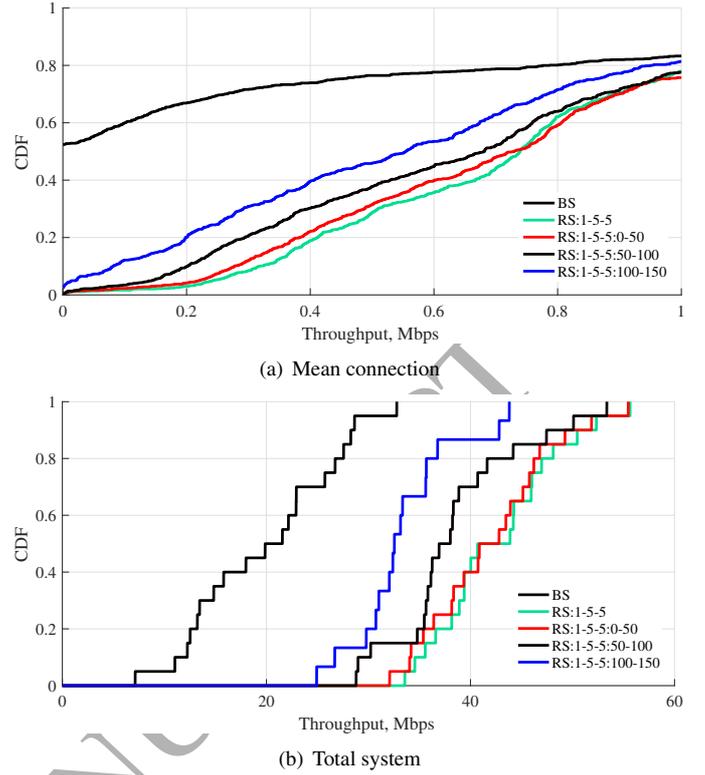


Figure 24: Throughput for 10 relays (ISD 2000m, frequency 5GHz), non-optimal coordinates.

lations results presented for both 3.5GHz and 5GHz, it can be seen that configurations, such as 1-1-2, 1-1-3, 1-1-5, 1-1-7, and 1-1-9 have the worst performance. In other words, it is not beneficial to put a single relay on the second hop as it will be a severe bottleneck in the system.

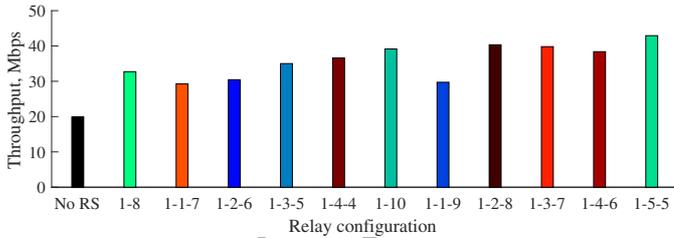


Figure 23: Mean total system throughput.

To test that calculated relay coordinates are indeed close to optimal, we use the same methodology as in the previous scenario to compare performance of the optimal coordinates and randomly generated coordinates in the vicinity of the optimal ones. We use relay configuration 1-5-5 as the baseline and run a series of additional simulations in which we generate randomly angle and distance from the optimal locations. The random distance changes from 50m to 150m at the step of 50m. In general, Figure 24 demonstrates the same outcome as we observed for 3.5GHz simulations. Relatively marginal deviations from the optimal coordinates, e.g. up to 50m, do not cause any noticeable change in the performance. However, once the deviation distance becomes larger, performance degradation be-

comes more severe.

Negative impact due to non-optimal coordinates is explained by same reasons as for 3.5GHz simulations: non-optimal relay coordinates impact the distribution of SINR for the BS-RS link even more than in the two-hop scenario. Furthermore, random relay locations can break the balance not only on the BS-RS link, but also on the RS-RS link as can be seen from Figure 25. It is also worth noticing that when the relay coordinates become too far away from optimal, there is another risk for a relay node to fail decoding control channels from its super-ordinate station. Referring to the 3.5GHz scenario simulation results, it is evident that selecting relay coordinates in the fully random or pseudo random way, as some simulation methodologies assume, may result in somewhat misleading conclusions regarding performance gains that relays could provide.

As for the optimal partitioning of frame resources, the genetic algorithm selects 16-10-4 symbols for the 1-5-5 relay configuration. Figure 26 and Figure 27 show that further changes in the ratio between the access zone and two relay zone result in shifting the balance between the mobile stations associated with the macro base station and the relay nodes, which in turn causes unfair resource allocation and even worse performance. As an example, by extending the access zone size at the expense of the first relay or the second relay zone, we can obviously improve performance of mobile stations connected to the macro base stations. However, it also increases the probability of the bottleneck on the first hop and, therefore, decreases the throughput for mobile stations connected directly to the first

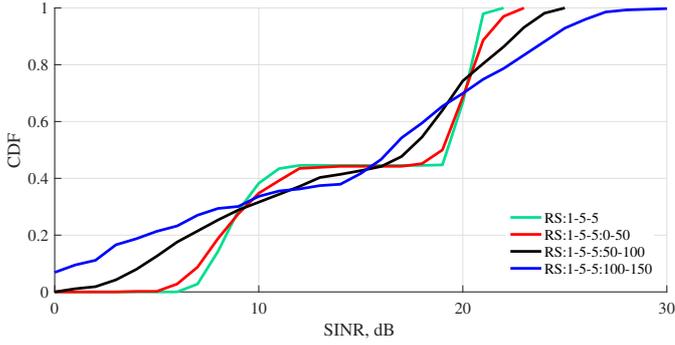


Figure 25: SINR of BS-RS and RS-RS links for 10 relays (ISD 2000m, frequency 5GHz), non-optimal coordinates.

layer relay nodes and stations connected to the second layer relays. As an example, frame configuration 20-6-4 provides the best performance for mobile stations associated with the macro base stations, but the stations associated with relays have much lower throughput when compared to other configurations. Similarly, frame configuration 18-10-2 has good performance for the macro base stations, but mobile stations associated with the second layer relays suffer from the small capacity of the RS-RS link. If we allocate much more resources for the relay zones, the outcome is even less predictable. On the one hand, relays could send more data, but the access link might become a bottleneck. As an example, frame configuration 12-14-4 has poor performance for both mobile stations associated with macro base stations and for low throughput stations associated with relays. All these results illustrate quite eloquently that any performance gain from optimal coordinates for relays can be diminished quite easily by inappropriate frame configuration and partitioning of resources between the access and relay zones.

5.2.2. Small cells

Similarly to the previous scenario with small cells at 3.5GHz, deploying small cells allows an operator to improve further the network capacity. Small cell coordinates are obtained with the genetic algorithm and are presented in Table 6 and Figure 28.

Table 6: Small cell positions for ISD 2000m at 5GHz.

Network configuration	Small cell coordinates	Frame structure	
		DL	UL
1	-		
1-8	(0°911m) (+/-16°625m) (+/-46°632m) (17°1027m) (35°1004m) (50°1063m)	30	14
1-10	(0°910m) (+/-17°628m) (+/-36°856m) (+/-44°559m) (+/-56°948m) (17°1026m)		

For the coverage extension scenario considered, small cells can provide much better performance compared to the baseline scenario improving the total system throughput by more than 250% on average, as can be seen from Figure 29. It is worth noticing that 10 nodes do not improve significantly the total throughput when compared to 8 small cells. However, for low throughput mobile stations, the impact of adding two more nodes is quite noticeable, e.g. the 20-th percentile of the connection throughput increases by 13%.

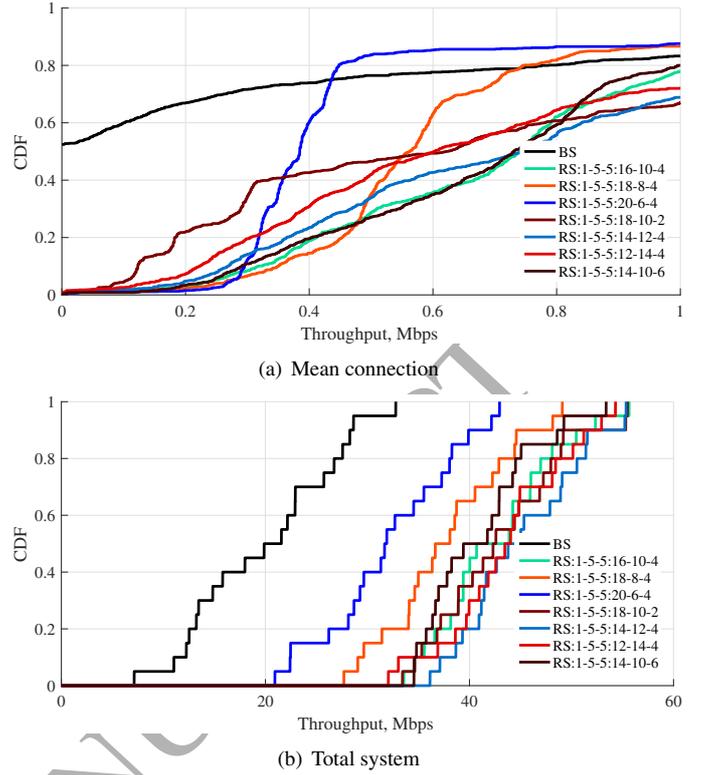


Figure 26: Throughput for 10 relays (ISD 2000m, frequency 5GHz), non-optimal relay zone sizes.

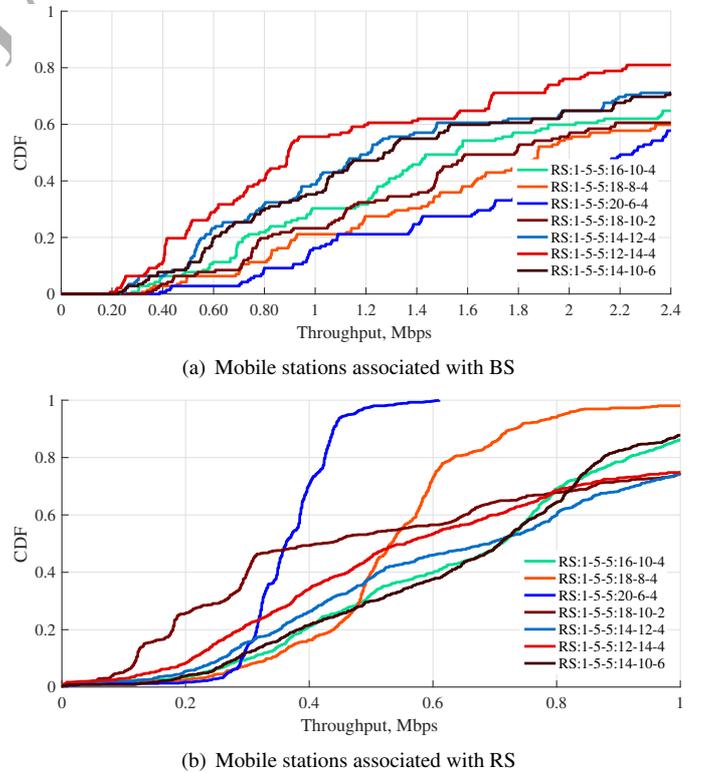


Figure 27: Mean connection throughput for 10 relays (ISD 2000m, frequency 5GHz), non-optimal relay zone sizes.

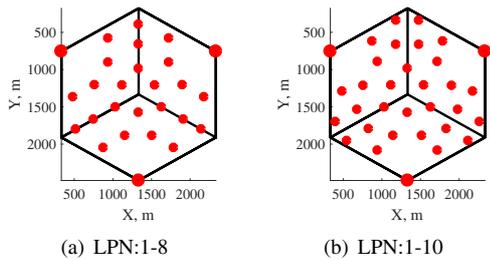


Figure 28: Optimal locations for small cells (ISD 2000m, frequency 3.5GHz).

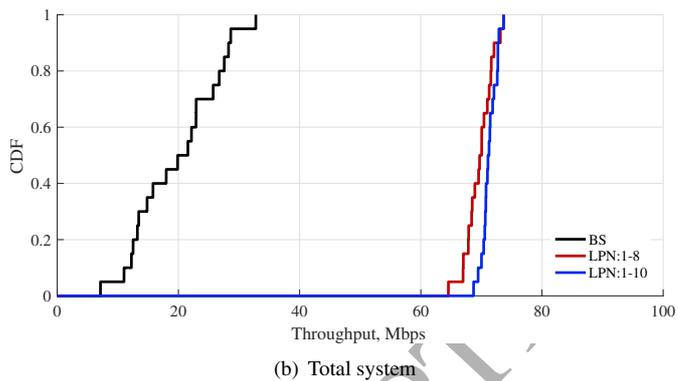
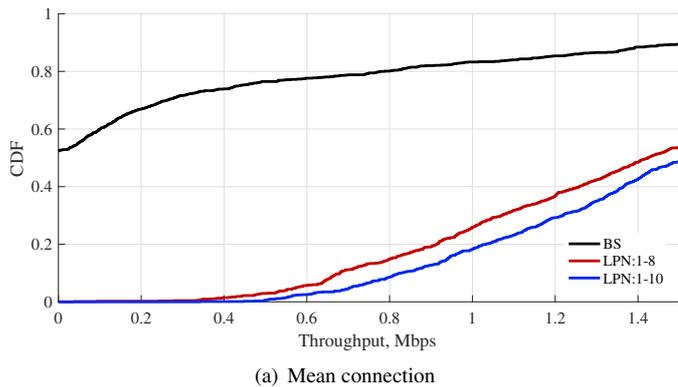


Figure 29: Throughput for 10 small cells (ISD 2000m, frequency 5GHz).

Speaking of the node locations, big numbers of small cells are not that sensitive to minor deviations from the optimal positions as can be seen from Figure 30. However, when positions are far from optimal, the performance gets worse noticeably, especially for low throughput mobile stations, e.g. value of the 20-th percentile of the connection throughput drops by 47%. The fact that minor deviations from the optimal coordinates do not have a great impact on final performance in case of small cells can be exploited by mobile operators to consider the following deployment strategy: first, relay nodes are deployed at optimal coordinates, and later they are switched to small cells once the wired fiber connections become available. Small cells coordinates will not be optimal in this case, but they will still provide quite noticeable performance gains.

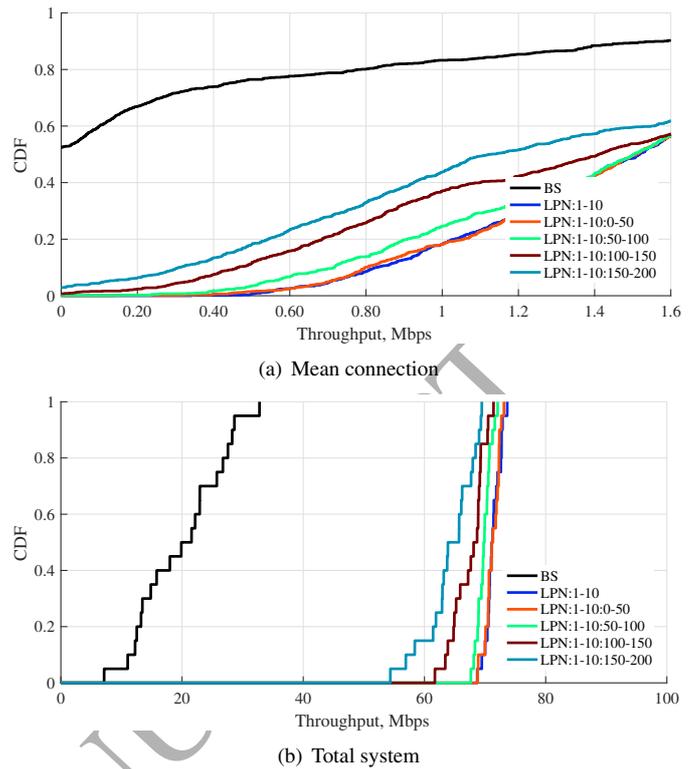


Figure 30: Throughput for 10 small cells (ISD 2000m, frequency 5GHz), non-optimal coordinates.

6. Conclusion

In this paper, we have considered a problem of finding optimal deployment configuration for low power nodes, such as small cells and relays. We developed an analytical model that takes macro level parameters (ISD, transmission power, etc) and finds the optimal network configuration for lower power nodes including their coordinates, connection scheme and frame resource partitioning. Extensive dynamic system level simulations have confirmed that analytically calculated coordinates are close to optimal as deviations from optimal solutions result in declined system performance. Even though our dynamic system level simulations mostly consider sub-urban environments with ISD of 1500...2000m, we do strongly believe that the proposed model would also work in urban scenarios.

It should be understood that network configurations obtained with the proposed model are optimal only under certain considerations and assumptions we make, so changing the overall system load, path-loss, shadowing or introduction of more advanced receivers may change the final outcome. Nevertheless, even if we specify accurately all the macro level parameters then it still does not mean that an operator will have a possibility to deploy a low power node at calculated locations. In that sense, results of this research could be used as a guidance or a recommendation regarding where the lower power nodes should be, while their actual coordinates may be further constrained by other real-life factors. In fact, we presented in our simulations that relatively small deviations from the optimal co-

ordinates can be tolerated without a big loss in performance.

One of the most important outcomes of our research is that an operator should consider carefully whether to deploy two-hop or multi-hop relays. As presented by extensive simulation results, it is not always the case that multi-hop relays provide better performance as one might initially anticipate. Since multi-hop relays take additional resources for the relay zone, fewer symbols remain for the access zone immediately impacting achievable throughput. Thus, multi-hop relays should be considered only for those scenarios where coverage is indeed an issue; otherwise two-hop relays will most likely provide better performance. In addition, we also showed that optimal frame resource partitioning plays a very crucial role in ensuring balance between relay and access link capacity. Once this balance is broken, the overall performance declines significantly irrespective of how good and optimal relay coordinates are.

Apart from tasks connected to real life deployments, obtained results can serve as input to various relay simulations and simulation methodologies. It is often the case that simulation methodologies either do not have an agreed set of coordinates for low power nodes, or they are chosen semi-randomly. As showed in this paper, non-optimal coordinates can lead to worse performance, which in some simulations may be construed wrongly as a sign of a particular technology or a solution not meeting target requirements. Thus, by providing optimal coordinates for low power nodes the research community can construe better achievable gains.

Our next research steps are to tackle more complicated non-homogeneous deployment scenarios for which our model will have to handle more base stations and more low power nodes. Even though the proposed model can be in principle applied to such scenarios, we will face much longer algorithm execution times. For this reason, additional research has to be done in order to reduce time complexity of the solution and to accelerate the computation process. Even though computation complexity is not a big issue when the operator searches for optimal low power node coordinates during the initial radio planning phase, it will become a more critical problem if the proposed model is applied to moving relays (e.g. the ones mounted on trains or buses), for which the network will have to decide quickly how to connect a particular relay node.

References

- [1] 3GPP TR 38.913, Study on scenarios and requirements for next generation access technologies (Jun 2016).
- [2] A. Sayenko, M. Zolotukhin, T. Hämäläinen, On optimal placement of low power nodes for improved performance in heterogeneous networks, in: IEEE/IFIP Network Operations and Management Symposium, 2016, pp. 349–357.
- [3] A. Sayenko, M. Zolotukhin, T. Hämäläinen, Multi-hop relays for high frequency next generation wireless systems, in: 24th International Conference on Telecommunication, 2017.
- [4] S. Baroudi, Y. R. Shayan, Performance enhancement of multi-hop wireless systems by optimizing the number and location of relays, in: International Wireless Communications and Mobile Computing Conference (IWCMC), 2015, pp. 465–468.
- [5] C. Chang, C. Chang, M. Li, C. Chang, A novel relay placement mechanism for capacity enhancement in IEEE 802.16j WiMAX networks, in: IEEE International Conference on Communications (ICC), 2009, pp. 1–5.
- [6] F. Guerriero, V. Loscri, P. Pace, R. Surace, Neural networks and SDR modulation schemes for wireless mobile nodes: a synergic approach, *Ad Hoc Networks* 54 (2017) 17–29.
- [7] Y. Yang, S. Murphy, L. Murphy, Planning base station and relay station locations in IEEE 802.16j multi-hop relay networks, in: Consumer Communications and Networking Conference (CCNC), 2008, pp. 922–926.
- [8] I. Siomina, D. Yuan, Optimization approaches for planning small cell locations in load-coupled heterogeneous LTE networks, in: Proceedings of the 24th IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 2013, pp. 2904–2908.
- [9] M. Qutqut, H. Abouzeid, H. Hassanein, A. Rashwan, F. AlTurjman, Dynamic small cell placement strategies for LTE heterogeneous networks, in: Proceedings of IEEE Symposium on Computers and Communications (ISCC), 2014, pp. 1–6.
- [10] Y. Pak, K. Min, S. Choi, Performance evaluation of various small-cell deployment scenarios in small-cell networks, in: Proceedings of the 18th IEEE International Symposium on Consumer Electronics, 2014, pp. 1–2.
- [11] Y. Kim, M. L. Sichitiu, Multihop Relay-Enhanced WiMAX Networks, Quality of Service and Resource Allocation in WiMAX, *InTech*, 2012. doi:10.5772/28698.
- [12] S.-J. Kim, S.-Y. Kim, B.-B. Lee, S.-W. Ryu, H.-W. Lee, C.-H. Cho, Multi-hop relay based coverage extension in the IEEE 802.16j based mobile WiMAX systems, in: Proceedings of the 4th International Conference on Networked Computing and Advanced Information Management, Vol. 1, 2008, pp. 516–522.
- [13] Z. Zhao, J. Wang, S. Redana, B. Raaf, Downlink resource allocation for LTE-advanced networks with type1 relay nodes, in: Proceedings of IEEE Vehicular Technology Conference (VTC Fall), 2012, pp. 1–5.
- [14] L. Dong, X. Zhu, Y. Huang, Optimal asymmetric resource allocation for dual-hop multi-relay LTE-advanced systems in the downlink, in: Proceedings of IEEE International Conference on Communications (ICC), 2013, pp. 4625–4629.
- [15] A. Sayenko, M. Zolotukhin, T. Hämäläinen, Simulation and performance analysis of frame structures for multi-hop relay systems, in: Proceedings of IEEE Conference on Standards for Communication and Networking, 2017.
- [16] Air interface for broadband wireless access systems: Multihop relay specification, IEEE Standard 802.16j (Jun 2009).
- [17] S. W. Peters, R. W. Heath, The future of WiMAX: multihop relaying with IEEE 802.16j, *IEEE Communications Magazine*.
- [18] 3GPP TR 36.806, Relay architecture for E-UTRA (LTE-Advanced) (Sep 2009).
- [19] 3GPP TS 36.300, Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description; stage 2, v12.0.0 (Mar 2014).
- [20] C. Shannon, Communication in the presence of noise 37 (1949) 10–21.
- [21] R. Sharapov, A. Lapshin, Optimal feature selection for support vector machines, *Pattern Recognition and Image Analysis* 16 (3) (2006) 392–397.
- [22] T. Mitchell, *Machine Learning*, 1st Edition, McGraw-Hill, Inc., New York, NY, USA, 1997.
- [23] A. Sayenko, O. Alanen, H. Martikainen, V. Tykhomyrov, O. Puchko, V. Hytönen, T. Hämäläinen, WINSE: WiMAX NS-2 Extension, Special Issue of Simulation: Software Tools, Techniques and Architectures for Computer Simulation 87 (2011) 24–44.
- [24] Y. Oda, K. Tsunekawa, M. Hata, Advanced LOS path-loss model in micro-cellular mobile communication, in: IEEE Transactions on Vehicular Technology, Vol. 49, 2000, pp. 2121–2125.
- [25] Multi-hop relay system evaluation methodology (channel model and performance metrics), IEEE 802.16 Broadband Wireless Access Working Group (Feb 2007).
- [26] A. Sayenko, T. Hämäläinen, Performance comparison of retransmission mechanisms for multi-hop relay networks, in: IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, 2015, pp. 189–194.
- [27] 3GPP TR 36.826, Evolved universal terrestrial radio access (E-UTRA); relay radio transmission and reception (Jul 2013).

Mikhail Zolotukhin received Ph.D. degree on Computer Science from University of Jyväskylä in 2014. At present, he is working in Faculty of Information Technology of University of Jyväskylä as a post-doctoral researcher. His research interests include data mining, cyber security, software-defined networking, multi-hop networks, game theory.

Alexander Sayenko received Ph.D. degree on Telecommunications from University of Jyväskylä in 2005. At present, he is working in Samsung Electronics as a principal engineer. His research interests include wireless communication standardization and research 4G/LTE and 5G/NR wireless systems 3GPP RAN WG and RAN TSG delegate.

Timo Hämäläinen received Ph.D. degree in Telecommunication from University of Jyväskylä in 2002. At present, he is working in Faculty of Information Technology of University of Jyväskylä as a professor. His research interests include QoS monitoring and management solutions for wireless and wired networks, all-IP network mobility and resource management, and network planning, network security.

ACCEPTED MANUSCRIPT



(a) Mikhail Zolotukhin (b) Alexander Sayenko (c) Timo Hämäläinen

Figure 1: Author photo.

ACCEPTED MANUSCRIPT