Dynamic resource allocation in Radio-over-Fiber enabled dense cellular networks

Bart Post
Eindhoven University of Technology
Eindhoven, The Netherlands

Sem Borst
Eindhoven University of Technology
Eindhoven, The Netherlands

Ton Koonen
Eindhoven University of Technology
Eindhoven, The Netherlands

***Abstract*—Network densification has emerged as a powerful paradigm to boost spectral efficiency and accommodate the continual rise in demand for wireless capacity. The corresponding reduction in cell sizes also results however in greater spatial and temporal uncertainty and variation in traffic patterns and more extreme and unpredictable interference conditions. These features create unprecedented challenges for efficient allocation of spectral resources compared to conventional cellular networks. As a further challenge, the allocation of spectral resources needs to be jointly optimized with the assignment of wavelengths in the optical backhaul of Radio-over-Fiber (RoF) networks, which are increasingly used in dense deployments and indoor environments.**

Motivated by these issues, we develop online algorithms for joint radio frequency and optical wavelength assignment in RoF networks. The proposed algorithms rely on load measurements at the various access points, and involve configurable thresholds for triggering (re)assignment of spectral resources. We provide a detailed specification of a system implementation, and conduct extensive simulation experiments to examine the behaviour in various scenarios and assess the impact of key parameters. The results in particular demonstrate that the proposed algorithms are capable of maintaining adequate load levels for spatially heterogeneous and time-varying traffic conditions, while providing favourable throughput performance.

***Keywords*-Radio-over-Fiber, frequency and wavelength allocation, dense cellular networks, dynamic load balancing**

I. Introduction

A. Background and motivation

The proliferation of smartphones and other mobile devices is taxing wireless networks to the limit of their capacity. The deployment of pico access points yields a significant potential for capacity gains by reducing cell sizes and allowing higher spectral reuse and efficiency. However, physical constraints will typically make it hard to arrange pico access points in an ideal hexagonal pattern, which causes the coverage areas to significantly overlap, and the natural cell regions to be irregularly shaped. This brings significant new challenges for spectral (sub-band) allocations that need to be addressed in order to reach the 5G design objectives [1].

First of all, the nominal traffic loads will tend to exhibit more spatial variation than in traditional macro cell networks. To prevent severe performance degradation, the spectral resource allocations will have to take these load imbalances into account and direct capacity towards areas where it is most needed [2], [3].

Secondly, not only the spatial variations are more extreme, dense networks also experience more temporal load fluctuations. Load conditions are certain to change over time due to hourly and daily usage patterns. It is therefore crucial to devise sub-band allocation schemes that are able to dynamically react to changing load conditions.

Thirdly, wireless interference is less predictable and hence interference management is important to prevent performance degradation.

The above-mentioned challenges require new approaches and techniques for resource allocation in dense cellular networks [4], [5]. Furthermore, new spectral allocation schemes should be designed and optimized in an integrated framework with upcoming and developing technologies in future (wireless) networking.

One consequence of integrating state-of-the-art networking technologies is the applicability of centralized algorithms. Future (cellular) networks are envisioned to be operated through software-defined radio access networking (SD-RAN) technologies. SD-RAN decouples the access point (AP) intelligence from the physical antennas, allowing for more flexibility in system configurations (e.g. SoftAir [6]). A centralized control unit is running virtual APs, which are transmitting via a set of remote radio heads (RRHs). The RRHs are connected via a backhaul structure to the coordinated control unit (CCU), which allows for centralized allocation schemes. Having all intelligence at the CCU means that a lot of information on spectrum usage is available there, which can be used to better control the sub-band allocations.

Another aspect that needs to be accounted for is dealing with the allocation constraints imposed by an optical backhaul structure. Radio-over-Fiber (RoF, [7], [8]) is a specific backhaul structure which enables dense cellular networks to be operated by SD-RAN: the wireless signal is centrally generated at the software-defined AP and sent as optical signal through a fiber to a RRH. At the RRH, the optical signal is transformed to a wireless signal, and then transmitted to the destined user. RoF is a powerful enabling technology for SD-RAN since it potentially allows for several advanced techniques like beamforming and MIMO. The optical network can be quickly reconfigured to accommodate dynamic resource allocation schemes. Hence, RoF is a particularly promising backhaul architecture for dense cellular networks (pico-cells, femto-cells) [9]. However, the combination of dense small-cell networks with RoF technologies requires new and advanced resource allocation algorithms [4], [5], [10].

B. Contributions and results

In this paper, we develop load-aware algorithms for dynamic joint radio frequency and optical wavelength assignment in RoF networks. We study dense network scenarios where the APs are transmitting via RRHs, and where flow-level traffic dynamics and non-stationary hotspots influence the load conditions.

Generally there are two ways to balance loads in a system: bring users to locations where capacity is available, or bring capacity to where users are located. Typically in a wireless system, bringing users to APs that are further away decreases the sub-band efficiency, and results in lower service rates. Consequently, it is more efficient to try to bring available capacity (sub-bands) to an AP or RRH close to the user. However, since variations in user demands typically change on a faster timescale (milliseconds-seconds) than statistical differences in load conditions at RRHs (seconds-minutes), load-aware user association schemes should be used to deal with the variance in arriving user intensity, and load-aware resource allocation schemes should aim to deal with statistical changes in the average load conditions. We therefore focus on a decision timescale of multiple user activations, meaning in the order of seconds to minutes.

Inspired by local-search methods, we introduce a dynamic and load-aware heuristic: the Single Load Interval (SLI) algorithm. Sub-band or frequency allocation problems are typically NP-hard, as they mostly relate to - and generally extend - graph colouring problems, and many heuristics and approximation algorithms have been studied to come up with a near-optimal allocation. However, even such heuristics can be too slow for online operation in larger networks, where we envision decision making on a second to minute time-scale. In situations with a reasonably large network, we cannot afford to run a full optimization at every time step as it may take several minutes.

Based on a set of simple rules and the use of load proxies, the SLI-algorithm automatically detects changes in load conditions and adjusts the sub-band and wavelength allocation over time, to relieve highly loaded RRHs and guarantee favourable throughput performance under non-stationary load patterns. We furthermore provide extensive numerical results showing the behaviour of the SLI-algorithm. The numerical results demonstrate how several algorithm parameters can be used to reach desired effects, and how the load proxies can be tuned to adopt a certain decision timescale.

C. Discussion and related work

Resource allocation in wireless networks has been extensively studied already, but in most studies either only the wireless or only the optical domain is considered. Moreover, many existing dynamic channel assignment schemes treat the available spectrum as an infinitely divisible resource, while we purposefully stick to a discrete set of sub-bands, as also currently applied in 4G LTE systems and upcoming 5G networks.

Allocating sub-bands depending on loads may be seen as a variant of the multicolouring problem [11] (see Section II). The allocation of optical wavelengths to RRHs is related to the (1-dimensional) bin-packing problem with conflicts (see Section II) [12]. Depending on the objective, the combined resource allocation is therefore a combination of a multicolouring problem and a bin-packing problem, and hence NP-hard to solve in general as the multicolouring problem is NP-complete in general.

Zhang and Ansari [13] studied a combined wireless and optical resource allocation problem in a RoF pico-cellular network, but with some significant differences compared to the work presented in this paper. They do not explicitly model the optical fiber topology, and their proposed algorithms are not load-aware. In contrast, we focus on load-aware resource allocation schemes, and specifically include the optical backhaul topology in our model.

Klinkowski, Jaworski, and Careglio [14] have looked into the optical resource allocation in RoF networks and present many observations and system constraints for the optical domain. However, they do not consider the allocation of frequency bands, assuming these are already assigned in an interference-free manner. We on the other hand explicitly consider the combined wavelength and radio frequency allocations.

A dynamic resource allocation scheme which shows some resemblance with our SLI-algorithm is called channel-borrowing, first introduced by Engel and Peritsky [15].

Inter-Cell Interference Coordination (ICIC) [16], [17] is also concerned with resource allocation, but is more targeted at macro-pico cell interference avoidance by using time division schemes.

D. Organization of the paper

The remainder of this paper is organized as follows. In Section II we present a detailed model description, and then describe the optimization objective and introduce our dynamic SLI-algorithm in Section III. In Section IV we present extensive simulation experiments to gain insight in the behaviour of the SLI-algorithm, and show it outperforms static allocations in scenarios where load conditions change over time. Finally, in Section V we draw conclusions and provide suggestions for future research.

II. Model description

In this section we present a detailed model description and introduce some useful notation and terminology. We consider an area A with a wireless radio access system consisting of L remote radio heads (RRHs) which are connected to a centralized control unit by means of optical fiber. Let $\mathcal{L}$ denote the set of RRHs. The RRHs are divided into B disjoint subsets $\mathcal{L}_b$ (backbones), where each subset shares the optical fiber infrastructure (see Figure 2). This is a separation of optical dependence: in each backbone b, the lasers and fibers are shared and hence the optical resource allocation within the set $\mathcal{L}_b$ is dependent. However, across

different backbones the optical resource allocations are independent, as we assume that each backbone has its own set of lasers.

A. Wireless domain

We start by treating the wireless domain. We consider the downlink of the wireless network, where some RRHs may be physically co-located, but involve different radio access technologies (e.g. 4G LTE or 5G with multi-mode handsets). The spectral resources are divided into F sub-bands, all of equal bandwidth. Several studies [18], [19] have shown that the gain in adaptive power assignments on different sub-bands is marginal, and hence for control purposes it is preferred to emit with equal power on all sub-bands. For convenience we do not consider frequency-selective fading.

In dense cellular wireless networks it is crucial to accurately capture interference. Even though all RRHs emit with equal power on all sub-bands, irregular placements of the RRHs can result in varying interference conditions throughout the area. Where in macro cellular systems the cell regions were only marginally overlapping at the cell boundaries, in a dense cellular network cell regions may show huge overlaps. As a result, RRHs that are close together cannot be using the same sub-band simultaneously. This motivates the modelling of interference by means of an interference or conflict graph. Even though interference graphs are not as precise as SINR based interference models, several studies have shown that an interference graph is sufficiently accurate for capturing interference in this context [13], [20], [21]. The sub-band allocation problem will be a graph multi-colouring problem in the interference graph.

In the considered area where RRHs are placed, traffic flows are assigned to the RRHs based on - but not necessarily completely determined by - their Signal-to-Interference-Plus-Noise Ratio (SINR). The assignment of traffic flows generate a notion of offered traffic to a RRH, which in turn gives rise to a demand G_l at each RRH l.

Allocating sub-bands constrained to the interference graph means that we assume they are free of significant interference from other RRHs. This, together with the flat fading assumption across all sub-bands, implies that each sub-band f brings equal (long-term) capacity to a RRH. This allows us to interpret the demand G_l of RRH l as the fractional number of sub-bands needed by RRH l to support the average traffic volume. For example, if $G_l = 3.6$, then RRH l would need at least 4 sub-bands to sustain its load. Consequently, the long-term load of RRH l can be defined as

$$\rho_l = G_l / F_l, \tag{1}$$

where F_l is the number of sub-bands assigned to RRH l.

The capacities of sub-bands are not shared between different RRHs, but we do allow for re-use of sub-bands by RRHs. Lastly, denote by $F_{\max}$ the maximum number of sub-bands a RRH can operate simultaneously.

B. Connection to the optical domain

We will now describe the optical domain and make the connection with the wireless domain. We fix the bandwidth of sub-bands so that a sub-band occupies a fixed space on an optical wavelength. Hence Ω_j, the maximum number of wireless sub-bands that can be carried on an optical wavelength λ_j, is constant.

The RRHs are connected to the fiber ring by means of a wavelength add-drop multiplexer (λ-ADM), a photodiode and a frequency ADM (f-ADM), see Figure 1. The λ-ADM selects the optical wavelengths intended for the RRH. After that, the optical signal is converted to an electric signal by the photodiode. Then, the frequency ADM (f-ADM) selects - from the sub-bands encoded on the selected wavelengths - only those sub-bands that should be transmitted by the RRH. Thus, a RRH does not necessarily transmit all sub-bands encoded on the optical wavelengths it listens to.

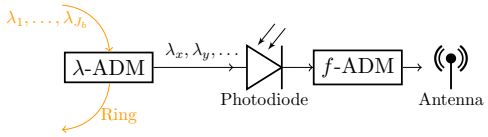


Figure 1: The optical and wireless filters at a RRH. λ_x, λ_y, ... are the wavelengths forwarded to the antenna.

The interference in the optical domain creates restrictions on the sub-band allocation. An optical wavelength can only carry a sub-band once, meaning that two RRHs which are allocated a common sub-band should receive those sub-bands through different wavelengths. In addition, the λ-ADM can select multiple wavelengths to forward to the transmitter. The optical receiver (photodiode) has no wavelength filter, and hence the incoming signal to the f-ADM can be thought of as all sub-bands encoded on the wavelengths that were selected by the λ-ADM. Consequently, each sub-band allocated to the RRH can only appear once on all received wavelengths combined, as (destructive) interference is caused otherwise.

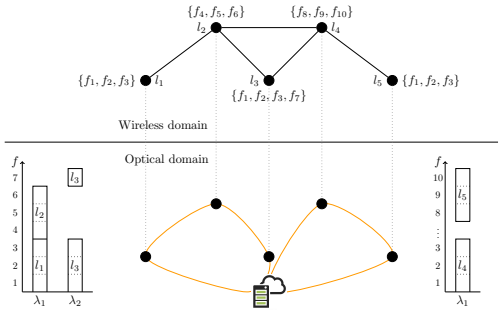


Figure 2: A feasible resource allocation, where $\Omega_j = 6$ for all j, and the black nodes represent RRHs. The left and right ring have wavelengths $\{\lambda_1, \lambda_2\}$ and $\{\lambda_1\}$ respectively.

III. Dynamic load balancing

In this section we state the load-balancing objective that we pursue, presented as a discrete optimization problem. Let $b(l)$ be the unique ring backbone b such that $l \in \mathcal{L}_b$. Let $z_{f,l,j} = 1$ if sub-band f is allocated to RRH l and is carried by wavelength λ_j; and 0 otherwise. Furthermore, let $\mathbf{z}$ be the $F \times L \times J$-vector containing all the $z_{f,l,j}$ variables. It follows that the number of sub-bands assigned to a RRH l at time t is $F_l = F_l(\mathbf{z}) = \sum_{f=1}^{F} \sum_{j=1}^{J} z_{f,l,j}$.

The objective is to minimize the highest load among all RRHs, thus balancing the loads among the RRHs. The sub-band and wavelength allocation with this objective can then be formulated as the following mixed integer linear programming (MILP) problem, where $\mathcal{J}_b$ is the set of wavelengths available in backbone b, and $U(\mathbf{z}(t))$ is the inverse of the highest load amongst all RRHs (which explains the maximization objective).

$$\max_{\mathbf{z}} \quad U(\mathbf{z}) \tag{2}$$

$$\text{sub:} \quad F_l = \sum_{f=1}^{F} \sum_{j=1}^{J} z_{f,l,j}, \qquad \forall l, \tag{3}$$

$$G_l^{-1} F_l \geq U(\mathbf{z}), \qquad \forall l, \tag{4}$$

$$\sum_{l \in \mathcal{L}_b} \sum_{f=1}^{F} z_{f,l,j} \leq \Omega_j, \qquad \forall b, \forall j, \tag{5}$$

$$F_l \leq F_{\max}, \qquad \forall l, \tag{6}$$

$$\sum_{l \in \mathcal{L}_b} z_{f,l,j} \leq 1, \qquad \forall b, \forall f, \forall j, \tag{7}$$

$$\sum_{j \in \mathcal{J}_{b(l)}} z_{f,l,j} \leq 1, \qquad \forall f, \forall l, \tag{8}$$

$$z_{f,l,j} + z_{f',l,j'} + \sum_{l' \in \mathcal{L}_{b(l)} \setminus \{l\}} z_{f,l',j'} \leq 2, \qquad \forall l, \forall f \neq f', \forall j \neq j', \tag{9}$$

$$\sum_{j=1}^{J} z_{f,l,j} + z_{f,l',j} \leq 1, \quad \forall (l,l') \in E. \tag{10}$$

In this optimization problem, constraint (5) specifies the wavelength capacity constraint, (6) represents the sub-band capacity per RRH, and (7) ensures that sub-bands are not shared between RRHs. Furthermore, (8) and (9) guarantee that a RRH receives a sub-band only once through all the wavelengths, and finally (10) is the interference constraint based on the interference graph $\mathcal{G}$. Moreover, the maximization induces a load $\rho_l < 1$ for all $l = 1, \ldots, L$ whenever possible, meaning that all RRHs receive enough capacity to serve their entire traffic demand. If the optimal solution has $\rho_l > 1$ for some l, it simply means that the current sub-bands and wavelengths are not sufficient to meet all the demand. We call a sub-band and wavelength allocation stable if it achieves $\rho_l < 1$ for all l.

The MILP problem (2)-(10) has a hidden time element which is coming from its dependence on the demands G_l. In a sense, the MILP formulation is a snapshot of the system, taken in a period where the the average traffic volume offered to a RRH does not change. However, in practice the offered traffic volume changes due to fluctuations in the arrival process, as well as due to statistical changes in the mean traffic intensity. So, over time the load characteristics at the RRHs change, and the sub-band and wavelength allocations should change with them. However, it is impractical and undesirable to find an optimal solution to (2)-(10) in each timestep, for several reasons we outline below.

The first reason stems from the hardness of the allocation problem. Using a reduction to a graph colouring problem, it can be shown that solving the optimization problem (2)-(10) is NP-hard for general interference graphs. As a consequence, advanced solvers like CPlex can take several minutes to find an optimal solution even for small systems with at most 20 RRHs. A computation time of several minutes is not feasible on the timescale we are considering.

Secondly, even if we could get optimal solutions in acceptable time, adopting a new allocation can mean a lot of switching in sub-bands. In other words, it is very well possible that a RRH has to give up all the sub-bands it currently has in use and adopt a whole new set of sub-bands. From a practical point of view, this is undesirable due to handover and reconfiguration times.

Considering the above stated issues, a dynamic resource allocation scheme should be based on (computationally) simple rules. The set of allocated sub-bands per RRHs should not change too drastically and should take load conditions into account. We envision local-search like methods: given the current allocation; choose a new allocation "close" to the current one.

A. Interval-based dynamic resource allocation

An intuitive local-search move is to allocate an extra sub-band to an RRH if its load exceeds a certain threshold. The intuition behind this move is twofold. On the one hand we wish to allocate extra resources where they are needed, and on the other hand we may not necessarily have to allocate extra sub-bands as long as the load of an RRH is not too high. A similar interval-based procedure is described by Velayos et al. [22], where instead of sub-bands, users are re-assigned to different access points. We remove a sub-band from an RRH if its load is below a certain (other) threshold to ensure frequencies are also released when RRHs have surplus capacity.

We introduce an interval-based dynamic allocation scheme, that we will refer to as the Single Load Interval (SLI) algorithm. In short, the concept is as follows: choose an interval $[\rho_{\min}, \rho_{\max}]$. For each RRH l, determine the load estimate $\rho_l(t_k)$ at decision time t_k and perform either one of the following three actions:

- If $\rho_l(t_k) \geq \rho_{\max}$, (try to) assign an extra sub-band to this RRH,
- If $\rho_l(t_k) \leq \rho_{\min}$, remove a sub-band from this RRH (as long as it does not end up with zero sub-bands),
- Do nothing if $\rho_l(t_k) \in [\rho_{\min}, \rho_{\max}]$.

The idea is to keep loads in the control interval. The upper bound of the interval guards against overloaded RRHs. The lower bound of the interval frees capacity when possible. The decision times are specified by deterministic intervals, and occur with a frequency of ν_{load}. The $\sigma_l(t_k)$ is defined as the percentage of time that RRH l was busy in the time interval $[t_{k-1}, t_k]$. Then, the load estimate $\rho_l(t_k)$ of RRH l is determined by $\rho_l(t_k) = (1-\varepsilon)\rho_l(t_{k-1}) + \varepsilon(\sigma_l(t_k))$, where $\varepsilon > 0$ determines the magnitude of the updates (or step sizes) and is typically small.

The SLI-algorithm has three elements influencing the reaction speed. First, there is the interval $[\rho_{\min}, \rho_{\max}]$. The upper bound of the interval determines when extra sub-bands are added. The lower this upper bound is, the sooner the algorithm will try to allocate extra sub-bands. Secondly, there is the frequency ν_{load} at which load measurements are done. A lower frequency will result in a more averaged view on the RRH loads, while a higher frequency will result in more instantaneous load values. Thirdly, the update size ε determines over how many estimation intervals the load is averaged. If ε is small, then many observations are needed before a statistical change in the loads is detected. However, if ε is big, a "too instantaneous" view on the system is obtained.

The eventual choice of the parameters ν_{load} and ε should ensure that the load estimates are not too sensitive to the temporal load variations, but that systematic changes in the underlying load parameters are detected sufficiently fast.

B. Practical implementation

We will now give a more detailed description of our practical implementation of the SLI-algorithm, based on the three actions stated above. Of specific concern is the action of assigning an extra sub-band to RRH l, which we implemented by the following steps:

1) Determine if RRH l has enough capacity to be assigned another sub-band, in view of constraint (6). If not, the SLI-algorithm reports a failed attempt.
2) Determine the set of available sub-bands by only looking at the wireless interference constraints (10). If no sub-bands are available, the SLI-algorithm reports a failed attempt F:W (Failure:Wireless).
3) From the set of available sub-bands, remove those which will cause interference on the wavelengths which are already feeding RRH l. If we are left with no wavelengths, the SLI-algorithm reports a failed attempt.
4) Among all wavelengths that are already feeding RRH l, we pick the first one (ordered by the identifier of the wavelength) that has enough residual capacity to carry another sub-band. We then assign the "smallest" sub-band and allocate it to the RRH, and let it be carried by the selected wavelength. If all wavelengths are full we proceed to the next step.
5) From the set of available sub-bands, we pick the "smallest" one. Then we go through the set of wavelengths (available on the backbone of RRH l) that are not yet feeding RRH l and select the first one that has enough capacity left and can carry the sub-band to the RRH without violating the constraints (2)-(10).
6) If no wavelengths can carry the sub-band, the sub-band is discarded and we return to step (5). The SLI-algorithm reports a failed attempt when all sub-bands have been discarded.

Notice that if Step 4 results in an allocation, the allocation satisfies all constraints of (2)-(10) by construction of the previous steps.

The above-described operations can be achieved by lookup and sorting operations, resulting in a running time (for a single RRH) of $\mathcal{O}(F^2|J_{b(l)}|L\log(F))$.

IV. Numerical results

In this section we present the results of the simulation experiments we conducted to gain insight in the performance and behaviour of the SLI-algorithm. We consider an area of $100m \times 50m$, where users appear uniformly at random according to a two-dimensional Poisson process with rate $\nu = 100$ (per unit of time; second). The file sizes of users are independent and exponentially distributed with a mean of $25Mb$. These specific arrival and service distributions are not essential for the implementation of the SLI-algorithm, but are mainly used for convenience in the simulation.

Let $R_{i,l}$ be the base rate at which user i can be served by RRH l. The base rate $R_{i,l}$ depends on the strength of the received signal of the RRH l. Since the capacity of the RRH is shared by the users assigned to it, the eventual realized service rate of that user i also depends on the load at the RRH. Suppose user i is assigned to RRH l. Then during a small time interval in which M_l - the number of users assigned to RRH l - does not change, the user i receives a service rate of $R_{i,l}/M_l$. This is in line with the currently used proportional fair schedulers.

The service rates $R_{i,l}$, in Mb/s, are calculated according to the 3GGP urban micro model defined in 3GPP 36.814 v9.0.0. Users are assigned to RRHs according to the Best-SINR algorithm: a user is assigned to RRH $l \in \arg\max_{l'} R_{i,l'}$. In case of tie, the user is assigned to the RRH with the largest id. The RRHs are grouped into 8 fiber loops based on their locations, as shown in Figure 3, and each wavelength j has a capacity of $\Omega_j = 40$ sub-bands.

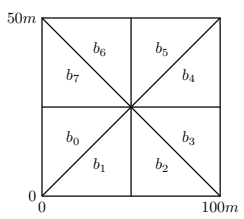


Figure 3: The eight areas that determine to which backbone b a RRH belongs.

We conducted experiments for three different scenarios. In Scenario 1 we have 26 RRHs located uniformly at

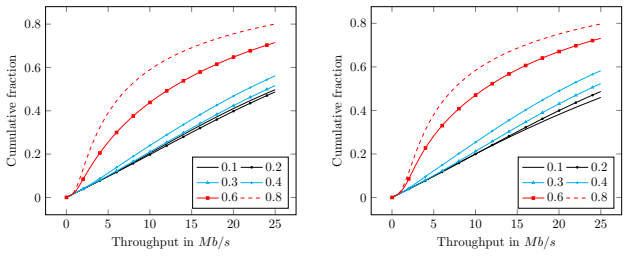
random. In Scenario 2, there are 30 RRHs; 26 are located uniformly at random, and an additional 4 are located uniformly at random in a rectangular $20m \times 10m$ area with its south-west corner at $(60, 30)$. Hence there is a zone where the RRH placement is extra dense. In Scenario 3: 26 RRHs are located uniformly at random. The arrival process of users in Scenario 3 has a HotSpot of $20m \times 10m$ which moves over time. It starts with its south-west corner at $(20, 10)$, then it moves to $(40, 10)$, $(60, 10)$, back to $(40, 10)$ to return to $(20, 10)$, after which this pattern repeats. The HotSpot has a relative arrival rate of 20 times the normal arrival rate.

The main performance measures in which we are interested are the user-perceived throughput and delay. The user-perceived delay is defined as the time it took for the user to receive its file, i.e. the time the user spent in service. The user-perceived throughput is defined as the user's file size divided by its perceived delay.

The plots presented in the following subsections each represent a single experiment of one random instance of a scenario. However, we conducted 100 numerical simulations for each combination of scenario and control parameter as named in the respective subsections, and confirmed that the displayed figures are representative for the 100 experiments conducted.

A. *Impact of the control interval*

The first experiments we conducted all used the same value for $\rho_{\max}$, but have different values for $\rho_{\min}$. All upper bounds $\rho_{\max}$ were set to $\rho_{\max} = 0.9$. For the lower bounds $\rho_{\min}$ we took $\rho_{min} \in \{0.1, 0.2, 0.3, 0.4, 0.6, 0.8\}$, and the frequency of load measurements is $\nu_{\text{load}} = 1$ measurements per unit of time (second), 100 times lower than the user arrival rate, and $\varepsilon = 0.01$. In Figures 4a and 4b we plotted the empirical cumulative distribution of the user-perceived throughput for Scenarios 1 and 2, based on random instances of the scenarios, with 5 000 000 generated users per instance and $\varepsilon = 0.01$.



(a) Scenario 1. (b) Scenario 2.

Figure 4: The user-perceived throughput, for different values of $\rho_{\min}$.

Notice that the throughput and delay performance are presented by empirical cumulative distributions. For throughput, that means that the lower the plot, the better, as more users will then have experienced a high throughput. With the delay plots it is the exact opposite: the graph should be as high as possible, quickly jumping to 1, as in that case more users will have experienced a low delay.

In Figures 4a and 4b, we see that in general a lower value of $\rho_{\min}$ results in better throughput and delay performance. This can be explained as follows. As the lower bound of the interval, $\rho_{\min}$ increases, RRHs give up extra allocated sub-bands under higher loads. This particularly affects the RRHs with moderate to low load conditions, as they apparently give up resources that end up unused otherwise.

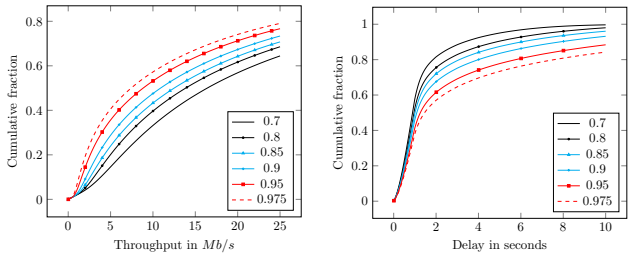
Over a set of 100 random instances, we kept track of the number of successful and unsuccessful allocation results returned by the SLI-algorithm. The numbers presented in Table I are based on 100 random instances of Scenario 1, with 500 000 generated users in each instance.

Result	$\rho_{\min} = 0.1$	$\rho_{\min} = 0.2$	$\rho_{\min} = 0.3$
None	11 434 644	11 129 799	10 891 111
F : W	490 980	387 847	192 292
+	10 983	11 612	16 618
Result	$\rho_{\min} = 0.4$	$\rho_{\min} = 0.6$	$\rho_{\min} = 0.8$
None	10 545 788	9 175 515	6 328 764
F : W	116 681	165 744	221 539
+	42 896	262 461	720 887

Table I: The realized number of successful (+), unsuccessful (F:W) allocation results.

Observe that for the two relatively high $\rho_{\min}$ values 0.6 and 0.8, the number of times that no reallocation was needed ("None"), and the number of failures to allocate an extra sub-band due to wireless constraints (F : W), decreased significantly. As RRHs give up resources under higher load conditions, more sub-bands are available when a RRH experiences severe load conditions. This can also be observed by the increase in successful allocations as $\rho_{\min}$ increases.

With the second set of experiments we tested the impact of $\rho_{\max}$ on the throughput and delay performance while keeping the lower bound fixed at $\rho_{\min} = 0.2$. The results are plotted in Figures 5a and 5b. It may come as no



(a) Perceived throughput. (b) Perceived delay.

Figure 5: The user-perceived throughput and delay, for different values of $\rho_{\max}$.

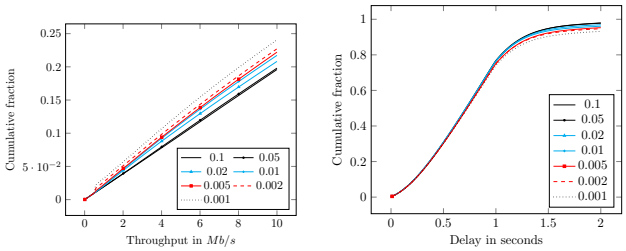
surprise that the smaller $\rho_{\max}$, the better the performance. From results in queueing theory we know that the mean number of users at a RRH l is given by $\rho_l/(1-\rho_l)$, if ρ_l can be considered constant, non-negative and smaller than one. Since we are dealing with a processor sharing discipline at the RRH, the more users in service, the

lower the user-perceived throughput will be. So, in a sense this implies that if $\rho_{\max}$ is smaller, the user-perceived throughputs and delays will be better.

The numerical experiments suggest that the lower bound $\rho_{\min}$ is preferably chosen quite small, and the upper bound $\rho_{\max}$ as well. However, $\rho_{\min}$ should stay strictly positive, to ensure resources are released and made available when load conditions are favourable.

B. Impact of load proxy parameters

We will now further investigate the impact of the load proxy parameters ε and ν_{load} on the performance of the SLI-algorithm. In Figures 6a and 6b we present the user-perceived throughput and delay performance for different values of ε. The control interval in these experiments was set to $[\rho_{\min}, \rho_{\max}] = [0.2, 0.8]$.



(a) Perceived throughput. (b) Perceived delay.

Figure 6: The user-perceived throughput and delay, for different values of ε.

In Figures 6a and 6b, we observe that larger values of ε result in a better throughput and delay performance. Indeed, in Section III-A we noted that the SLI-algorithm can detect changes in load conditions faster when ε is bigger.

Similar conclusions can be drawn from Figure 7, where we plotted the user-perceived throughput for different values of ν_{load} in a random instance of Scenario 2. Indeed, the higher the rate, the better the user-perceived throughput performance due to faster reaction times. Even

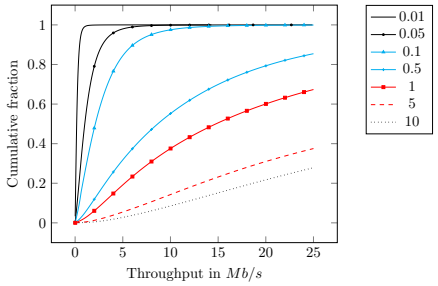
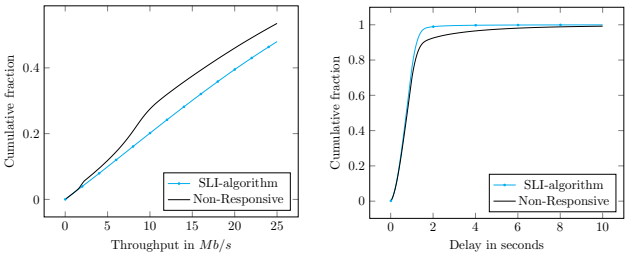


Figure 7: The user-perceived throughput, for different values of ν_{load}.

though the numerical results for the parameters ε and ν_{load} may suggest otherwise, bigger values of ε and ν_{load} are not necessarily better choices. Like we mentioned in the Introduction, the load proxy parameters should be chosen such that the SLI-algorithm can react fast enough to statistical load imbalances, but load imbalances that are a result of statistical variations in the user arrival process should preferably be handled by dynamic user association algorithms (which we intentionally do not cover in this paper). In view of this observation, ε and ν_{load} should not be too big, in order to provide a good "mean view" of the load conditions at RRHs.

C. Non-stationary HotSpot

We will now present simulation results for Scenario 3, where we tested the SLI-algorithm against a static, non-responsive sub-band and wavelength allocation. To obtain a reasonable static benchmark allocation, we discretized the area into $1m$ by $1m$ squares and calculated their long-term time-average traffic based on Scenario 3. The traffic of such a unit square is allocated to the RRH which offers the best rate to a user located at its center. This creates a demand at each RRH, and we use a heuristic to allocate sufficient sub-bands to each RRH such that it can cope with all offered time-average traffic (and hence keep long-term loads below 100%), adding more capacity (sub-band, wavelengths) to the system when necessary. Since the offered traffic is averaged over time and the HotSpot of users is moving over time, RRHs in a static allocation instance may find themselves switching between having (more than) sufficient capacity when the HotSpot is far away, and not having enough capacity when the HotSpot is close. The instances with a static allocation are in the plots referred to as Non-Responsive. In Figures 8a and 8b we present the user-perceived throughput and delay results for a single instance of Scenario 3, where the SLI-algorithm used control interval $[0.2, 0.8]$, $\nu_{\text{load}} = 1$ and $\varepsilon = 0.01$, based on a random instance of Scenario 1 with 5 000 000 generated users. Many experiments were conducted, and we confirmed that Figures 8a and 8b are representative for all these experiments.



(a) Perceived throughput. (b) Perceived delay.

Figure 8: The user-perceived throughput and delay, comparing the SLI-algorithm to a static allocation.

Figures 8a and 8b clearly show that the SLI-algorithm is better able to cope with the travelling HotSpot by moving resources to RRHs that are temporarily experiencing high load conditions.

V. Conclusion

In this paper we presented a dynamic sub-band allocation scheme, the SLI-algorithm, for dense cellular networks enabled by Radio-over-Fiber (RoF) technology.

The SLI-algorithm relies on load measurements at APs and uses the AP's loads a predefined tunable load interval $[\rho_{\min}, \rho_{\max}]$ to make re-allocations.

Extensive simulation experiments demonstrated the influence of key parameters on the throughput and delay performance of the SLI-algorithm. In particular, the choice of the interval $[\rho_{\min}, \rho_{\max}]$ has significant impact. Lower values of $\rho_{\min}$ utilize the sub-bands more efficiently, but more often result in a failure to allocate an extra sub-band to a RRH which is under high load. Lower values of $\rho_{\max}$ result in a better throughput and delay performance as RRHs claim resources more aggressively.

The load proxy parameters ν_{load} and ε also influence the performance of the SLI-algorithm. The higher the proxy rate ν_{load}, and the bigger ε, the more sensitive the SLI-algorithm becomes to variations in the offered traffic and the faster it reacts. However, it is undesirable for resource allocations to change on too fast a timescale: rather they should adapt to systematic changes in load conditions. We have shown how the load proxy parameters ν_{load} and ε can be used to achieve this.

In addition, we have tested the SLI-algorithm against a static sub-band and wavelength allocation, which was designed to sustain all long-term RRH loads. The numerical experiments clearly show that the SLI-algorithm is better suited to cope with a dynamic user environment with non-stationary hotspots.

This paper was only concerned with the (re)allocation of sub-bands, but load balancing can also be achieved through dynamic user association schemes. In particular, combining dynamic user association and dynamic resource allocation is an interesting next step towards optimizing resource management in 5G networks. Also, the SLI-algorithm could be improved by considering permutations on the sub-band and wavelength allocation, to make allocations "more efficient".

REFERENCES

[1] 5G-PPP. (2015) The 5G infrastructure public private partnership: the next generation of communication networks and services. [Online]. Available: https://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1.pdf

[2] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Comm.*, vol. 12, no. 6, pp. 2706–2716, 2013.

[3] J. G. Andrews, S. Singh, Q. Ye, X. Lin, and H. S. Dhillon, "An overview of load balancing in HetNets: Old myths and open problems," *IEEE Wir. Comm. Mag.*, vol. 21, no. 2, pp. 18–25, 2014.

[4] N. Ghazisaidi and M. Maier, "Fiber-wireless (FiWi) access networks: Challenges and opportunities," *IEEE Network*, vol. 25, no. 1, pp. 36–42, 2011.

[5] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: Insights and challenges," *IEEE Wir. Comm. Mag.*, vol. 22, no. 2, pp. 152–160, 2015.

[6] I. Akyildiz, P. Wang, and S. Lin, "SoftAir: A software defined networking architecture for 5G wireless systems," *Computer Networks*, vol. 85, pp. 1–18, 2015.

[7] M. Sauer, A. Kobyakov, and J. George, "Radio over fiber for picocellular network architectures," *J. Lightw. Tech.*, vol. 25, no. 11, pp. 3301–3320, 2007.

[8] U. R. Bhatt, A. Chhabra, and R. Upadhyay, "Fiber-wireless (Fi-Wi) architectural technologies: A survey," in *ICEEOT*. IEEE, 2016, pp. 519–524.

[9] A. M. J. Koonen and E. Tangdiongga, "Photonic home area networks," *J. Lightw. Tech.*, vol. 32, no. 4, pp. 591–604, 2014.

[10] N. Ghazisaidi, M. Maier, and C. M. Assi, "Fiber-wireless (FiWi) access networks: A survey," *IEEE Comm. Mag.*, vol. 47, no. 2, pp. 160–167, 2009.

[11] M. M. Halldórsson and G. Kortsarz, *Multicoloring: Problems and Techniques*. Springer Berlin Heidelberg, 2004, pp. 25–41.

[12] L. Epstein and A. Levin, "On bin packing with conflicts," *Lec. Notes Comp. Sc.*, vol. 4368, pp. 160–173, 2006.

[13] J. Zhang and N. Ansari, "On OFDMA resource allocation and wavelength assignment in OFDMA-based WDM radio-over-fiber picocellular networks," *IEEE J. Sel. Areas Comm.*, vol. 29, no. 6, pp. 1273–1283, 2011.

[14] M. Klinkowski, M. Jaworski, and D. Careglio, "Channel allocation in dense wavelength division multiplexing radio-over-fiber networks," in *ICTON 12*. IEEE, 2010, pp. 1–4.

[15] J. S. Engel and M. M. Peritsky, "Statistically-optimum dynamic server assignment in systems with interfering servers," *IEEE Trans. Veh. Tech.*, vol. 22, no. 4, pp. 203–209, 1973.

[16] A. L. Stolyar and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse for best-effort traffic through distributed inter-cell coordination," in *IEEE INFOCOM*. IEEE, 2009, pp. 1287–1295.

[17] A. Tall, Z. Altman, and E. Altman, "Self organizing strategies for enhanced ICIC (eICIC)," in *WiOpt'12*. IEEE, 2014, pp. 318–325.

[18] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks-part I: theoretical framework," *IEEE Trans. Wireless Comm.*, vol. 4, no. 2, pp. 614–624, 2005.

[19] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE J. Sel. Areas Comm.*, vol. 21, no. 2, pp. 171–178, 2003.

[20] M. M. Halldórsson and T. Tonoyan, "How well can graphs represent wireless interference?" in *Proc. 47th ACM Symp. Th. of Comp.*, ACM. ACM, 2015, pp. 635–644.

[21] X. Zhou, Z. Zhang, G. Wang, X. Yu, B. Y. Zhao, and H. Zheng, "Practical conflict graphs for dynamic spectrum distribution," in *SIGMETRICS Perf. Eval. Rev.*, vol. 41. ACM, 2013, pp. 5–16.

[22] H. Velayos, V. Aleo, and G. Karlsson, "Load balancing in overlapping wireless LAN cells," in *IEEE Int. Conf. Comm.*, vol. 7. IEEE, 2004, pp. 3833–3836.