Optimal and near-optimal alpha-fair resource allocation algorithms based on traffic demand predictions for optical network planning

TANIA PANAYIOTOU* and GEORGIOS ELLINAS
Department of Electrical and Computer Engineering and KIOS Research and Innovation Center of Excellence, University of Cyprus, Nicosia, Cyprus
*Corresponding author: panayiotou.tania@ucy.ac.cy

Received 8 September 2020; revised 28 December 2020; accepted 28 December 2020; published 5 February 2021 (Doc. ID 409536)

We examine proactive network optimization in reconfigurable optical networks based on traffic predictions. Specifically, the fair spectrum allocation (SA) problem is examined for a priori reserving resources aiming to achieve near-even minimum quality-of-service (QoS) guarantees for all contending connections. The fairness problem arises when greedy SA policies are followed, which are based on point-based maximum demand predictions, especially under congested networks, with some connections highly overprovisioned and others entirely blocked, resulting in highly uneven QoS connection guarantees. To address this problem, we consider predictive traffic distributions allowing the exploration of several combinations of possible SAs. To find a fair SA policy, we resort to an \( \alpha \)-fairness scheme, while QoS fairness is evaluated according to a game-theoretic analysis based on the coefficient of variations of the connections' unserved traffic metric, which measures the dispersion of unserved traffic for a connection around the mean of the unserved traffic over all connections. Non-contending (NC) and link-based contending (LBC) \( \alpha \)-fair SA integer linear programming algorithms are proposed, where the optimal NC approach encompasses global network contention information, while the near-optimal LBC approach encompasses partial contention information to reduce problem complexity. We show that as parameter \( \alpha \) increases, QoS fairness improves, along with connection blocking, resource utilization, and overprovisioning and underprovisioning. Further, LBC exhibits results close to the optimal, with a significant improvement in processing time. © 2021 Optical Society of America

https://doi.org/10.1364/JOCN.409536

1. INTRODUCTION

Resource allocation in optical networks traditionally deals with how to allocate bandwidth on arriving connection requests to form feasible lightpaths that efficiently utilize the network resources. Efficiency is usually evaluated according to connection blocking (CB) and/or resource utilization (RU) measures that are optimized as part of the resource allocation process. However, optimization of network efficiency lacks fairness, as the resources may not be fairly divided among the connections to achieve sufficient quality-of-service (QoS) levels for all network connections, especially under congested network situations.

Specifically, in reconfigurable optical networks, where network optimization is performed periodically (e.g., hourly, daily) and proactively based on projections of future traffic demands, it is important to guarantee a priori, before network reconfiguration takes place, the minimum achievable QoS levels that each connection is expected to attain, within the planning period of interest. In an optical network where traffic demand varies over time (i.e., fluctuates), this can be achieved by reserving a priori the spectrum resources required for servicing the maximum predicted traffic demand (i.e., temporal peak-rate demand) for each connection. Such a greedy spectrum allocation (SA) policy ensures that the highest QoS levels will always be met for all connections. However, in a network environment where the connections are likely to contend for the spectrum resources along their routes, a greedy policy may lead to highly uneven QoS levels among the connections, especially under congested network environments; that is, reserving for all connections the maximum predicted demand may not be possible, resulting in some connections that will enjoy their highest service level and in some connections with no QoS guarantees. Even though, after network reconfiguration takes place, the connections with no a priori reserved spectrum can be handled on demand dynamically (i.e., by taking advantage of the temporarily unutilized spectrum of other connections), such a practice does not provide any guarantee that these connections will be eventually served.

On this basis, in this work, we examine the proactive network optimization problem, with the objective of solving the
QoS-based fairness problem that may arise in elastic optical networks (EONs), when the allocation of the maximum predicted spectrum is not possible for all contending connections. Specifically, the objective is to find the SA policy that provides fair QoS guarantees for all contending connections, before network reconfiguration takes place. To achieve this objective, instead of considering during the optimization procedure the maximum demand predictions, we consider the distributional shape of traffic predictions within the planning period of interest. This leads to several possible SAs for each connection, and to several SA combinations among the connections. The fairest SA is the one that achieves near-even QoS levels for the contending connections. This, however, entails solving several instances of the conventional static SA problem, that was shown to be NP-complete [1].

Since examining all the possible combinations of SAs is not a computationally efficient option, in this work, we view the fair-SA problem through the lens of welfare economics. Specifically, we resort to an \( \alpha \)-fairness scheme based on the maximization of the social elasticity welfare function [2–4]. In welfare economics, this function ranks social states (i.e., complete descriptions of the society), as less desirable, more desirable, or indifferent, for every possible pair of social states. For our SA problem, this function is applied to rank the feasible SAs of each connection as less desirable, more desirable, or indifferent, for every possible pair of feasible SAs among two connections. Parameter \( \alpha \), also known as the inequality aversion parameter, controls fairness. It is known that as \( \alpha \) increases, fairer allocations are derived. In our SA problem, by increasing \( \alpha \), we aim at increasing the minimum amount of spectrum allocated to the “poorest” connection, i.e., the connection with the lowest demand tendency. Special cases of the \( \alpha \) parameter are known to be Pareto optimal (e.g., the fairest \( \alpha \) value), which means that connections will always reserve their maximum predicted demand, when the EON constraints are not violated. Hence, contention is resolved only when necessary.

To this end, the proactive SA problem is formulated according to an \( \alpha \)-fairness scheme that considers the distributional shape of traffic predictions (i.e., connections’ preferences over the possible SAs), subject to the EON constraints. Even though parameter \( \alpha \) is considered to be a natural measure of fairness, in this work, to directly interpret and measure QoS-based fairness, we follow a game-theoretic analysis utilizing the coefficient of variation (CV) [5] of the connections’ unserved traffic. Furthermore, to examine the impact of fairness on commonly considered efficiency evaluation measures, we also evaluate CB, RU, and connection overprovisioning (COP) and underprovisioning (CUP) measures, as \( \alpha \) increases. Overall, the evaluation of all these measures indicates the fairness–efficiency trade-offs, allowing a network operator to approximate the \( \alpha \)-fair SA that best meets the performance requirements of both the connections and the network environment.

The \( \alpha \)-fair SA problem is formulated according to an optimal and a near-optimal \( \alpha \)-fair SA algorithm, namely, the non-contending (NC) and the link-based contending (LBC) algorithms, respectively. The proposed \( \alpha \)-fair SA algorithms are based on two integer linear programming (ILP) formulations, with the optimal approach encompassing global network contention information and the near-optimal approach utilizing only partial contention information between the connections to reduce problem complexity. LBC optimality is evaluated according to the deviation of each measure of interest from the optimal NC values. The reader should note that the problem examined in this work refers to off-line (proactive) network planning for future network reconfigurations. Therefore, even though the reconfigurations are performed dynamically, the network planning problem is static between the pre-planned time intervals. On this basis, an exact problem formulation for the off-line problem (providing an optimal solution) is proposed, followed by a problem formulation based on the exact formulation but with some relaxations on the constraints (providing a near-optimal solution).

Preliminary results of this work were previously presented in [6]. In [6], the \( \alpha \)-fair SA problem was examined only according to the optimal ILP formulation, without specifically taking into consideration connection contention during the formulation. This work greatly extends [6] by formulating the \( \alpha \)-fair SA problem:

- According to an ILP-based algorithm (NC \( \alpha \)-fair algorithm) that finds the optimal \( \alpha \)-fair SA by decomposing the connections set according to a number of NC connection sets, aiming at partially addressing the scalability problem of the optimal ILP formulation of [6];
- According to an ILP-based algorithm (LBC \( \alpha \)-fair algorithm) that finds a near-optimal \( \alpha \)-fair SA by decomposing the connection set according to a number of LBC connection sets, aiming at further reducing the computational complexity as the problem size increases;
- Providing extensive performance evaluation and comparison of the NC- and LBC-based schemes on a larger problem size, consisting of more connections and possible SAs.

In general, this work shows that as parameter \( \alpha \) increases, fairer QoS guarantees are derived for the connections in both NC and LBC algorithms. Fairer QoS-based SAs are also associated with improved network efficiency, as RU is reduced by significantly improving both COP and CUP. Further, it is shown that the LBC algorithm operates near-optimal, with significant reduction in computational time.

The rest of the paper is organized as follows: Section 2 discusses related work, Section 3 is devoted to problem motivation, while Section 4 is devoted to the \( \alpha \)-fairness preliminaries. Section 5 provides the approach overview, and the problem formulations of NC- and LBC-based \( \alpha \)-fair routing and spectrum allocation (RSA) algorithms are given in Section 6. Evaluation measures are discussed in Section 7, while Section 8 presents the performance evaluation results. Concluding remarks, including avenues for future research, are given in Section 9.

2. RELATED WORK

The emergence of new types of networks, applications, and services, along with the daily spatiotemporal movement of the population has given rise to varying traffic fluctuations (i.e., the tidal traffic effect) that can be predicted, albeit with some uncertainty [7]. To address this effect, software-defined networking (SDN)-enabled reconfigurable optical networks have recently gained attention that are able to adapt, as closely
as possible, to the traffic demand flows, shifting from the traditional quasi-static optical networks [8–24].

The general SDN-enabled (re)configuration framework consists of modeling and predicting the traffic demand flows [8,10–13,15–18,21–24], and the predictions are then used for proactive (off-line) network optimization [8–21,24] between predefined (re)configuration time points. The general objective is to find a resource allocation policy that best fits the future traffic demand needs of the network and the connections as well.

In general, works related to the proactive resource allocation problem under time-varying traffic [8–24] focus on optimizing efficiency-related measures. Specifically, [8,16–20,23] focus on CB, [16–18,20,21,23] on RU, and [20] on the degree of defragmentation, without taking into account fairness-related measures. Consequently, resource allocation strategies may on the one hand optimize network efficiency, but on the other hand may lead to allocations that are highly uneven (unfair), especially with respect to the QoS (i.e., unserved traffic) that each connection is expected to enjoy after network reconfiguration takes place.

Moreover, resource allocation strategies are usually based on point estimates of the traffic (e.g., peak rate) [10–13,15,21,23,24], without considering the probabilistic shape (uncertainty) of traffic predictions. However, point-based estimates may lead to inappropriate resource allocation decisions. In particular, expected-rate estimates may lead to connection underprovisioning and peak-rate estimates may lead to connection overprovisioning. This outcome holds under the assumption of a highly asymmetrical traffic described by skewed distributions with heavy tails (e.g., Internet traffic) [25–27]. In particular, Internet traffic exhibits today a high divergence between the average to peak-rate demand [28], with peak rates rarely occurring. Even though traffic demands of probabilistic shape have been previously considered in [16–18], the fairness of the allocations, especially with respect to the achievable QoS of each connection, has not been examined. Specifically, in [16–18], a decentralized multi-agent reinforcement learning approach was applied, aiming at striking a balance between the aggregated COP and CUP effects. As such, only efficiency-related measures were examined.

To this end, in the presence of time-varying traffic (i.e., traffic fluctuations), the commonly considered optimization objectives along with the point-based traffic estimates, may lead to resource allocation decisions that are highly unfair, especially as it concerns the unserved traffic (i.e., QoS) each contending connection observes after network (re)configuration takes place. For these reasons, in this work, fairness considerations are explicitly taken into account with the objective of attaining fair QoS guarantees for the connections.

While various fairness measures exist in the literature (e.g., Jain’s index, entropy functions, etc.), the α-fairness measure—and its associated social elasticity welfare function—is the one that has gained attention in the networking research community. Specifically, α-fair resource allocation schemes have been previously examined in wireless networks [29], SDN IP networks [30], and Transmission Control Protocol (TCP) networks [3,4,31]. It should be noted though that α-fair resource allocation schemes have not been previously considered for optical networks. Nevertheless, fair resource allocation strategies in optical networks have been previously examined in [32–35]. In these works, however, fairness is examined only according to CB in dynamic optical networks, aiming at achieving as similar as possible blocking rates for the different classes of connections considered.

### 3. PROBLEM MOTIVATION

The general topic of proactive network optimization has been largely examined in the literature, considering various optimization objectives, assumptions, and optical network technologies, including wavelength division multiplexing (WDM) networks and EONs [8,10–13,15–18,21–24]. In most of these works, predictive traffic demand matrices assume point-based predictions, usually reflecting the predicted peak-rate demands within the future planning period of interest. The routing and bandwidth allocation problem is solved according to these point-based predictions, aiming at reserving for each connection just enough resources to accommodate the time-varying traffic fluctuations.

In EONs, the aim is to reserve the spectrum that is capable of accommodating the peak-rate demand, ensuring that bandwidth variable transponders (BVTs) will always achieve to adapt (i.e., expand the allocated spectrum) to the maximum requested; that is, reserved bandwidth will not be utilized at every time point, but it ensures that the connection will never be underprovisioned, enjoying at every time point the highest possible level of service. Therefore, the peak-rate demand assumption is appropriate for successfully handling the underprovisioning effect, but it may lead to connection overprovisioning, inefficient resource utilization, and even highly uneven levels of service between the connections, especially under congested network conditions. Specifically, uneven QoS may arise if reserving the peak-rate spectrum for all connections is not possible, leading to a situation where some connections will enjoy their highest possible QoS, while others will not be able to a priori reserve bandwidth; that is, some connections will be either entirely blocked or will be served according to a best-effort approach that dynamically allocates spectrum to the connections whenever possible (i.e., at time points where other connections do not fully utilize their reserved bandwidth).

To address the uneven QoS problem, in this work, instead of considering point-based predictions, we consider the probabilistic shape of traffic demand predictions, providing information on the traffic demand uncertainty within the planning period of interest. This consideration leads to several possible SAs for each connection, and we are interested in finding the SA policy that provides fair QoS guarantees for all contending connections.

To illustrate the aim of this work, a toy example is utilized, where we assume the presence of two virtual connections for which the predictive traffic distributions provide information on the virtual link spectrum utilization (i.e., in number of spectrum slots) within the future interval of interest (Fig. 1). As shown in Fig. 2, predictive traffic distributions allow us to extract useful information regarding a number of (discretized)
Possible SAs. Specifically, for each possible allocation, we extract information regarding the probability of the connection requiring such spectrum amounts during the future time interval, the connection underprovisioning probability (i.e., the probability of a temporal demand to be above the reserved resources), the expected demand, and the maximum demand, among others.

Given the information in Fig. 2, the RSA problem must be solved accordingly, and for simplicity, we assume that routing and SA problems are solved separately. We assume that for the routing problem, a scheme is applied that minimizes contention; nevertheless, in this example, contention arises between the optical paths of the connections (Fig. 3), leading to the following question. How much spectrum is it best to reserve for connection 1 and 2, given the limited spectrum capacity of the optical links (in this example, we assume that $M = 80$ spectrum slots are available per fiber link)?

To solve the SA problem, we need a policy that effectively utilizes the information in the predictive traffic tables (Fig. 2). First, let us assume that a SA policy is followed that is based on the maximum possible demand. According to this policy, 60 slots must be allocated to connection 1 and 70 slots to connection 2. However, this exceeds the available link bandwidth; therefore, we choose to serve the connection with the highest demand (connection 2) and block the other. This, however, means that connection 2 will enjoy the highest possible service level, while connection 1 will be 100% underprovisioned. Clearly, this scheme leads to a highly unfair SA between the two connections.

A second SA policy could also be investigated, where the expected demand for each connection is now considered; that is, 30 slots will be allocated to connection 1 and 30 slots to connection 2 (Fig. 2). Clearly, this policy leads to the admission of both connections, with the second connection being, however, highly underprovisioned, despite the fact that 20 slots are still available on the shared link. Again, this scheme does not seem to best utilize the available network resources. A third scheme can follow the maximum probability, which again leads to highly underprovisioning the connections even though 40 slots are still available.

For this toy example, by observation, it is obvious that by allocating 30 slots to the first connection and 50 to the second, both connections will enjoy the same level of service, as both connections are fairly degraded in regards to their respective traffic requirements, taking full advantage of the available spectrum. Specifically, for this example (Fig. 2), both connections are equally underprovisioned, since for both 12% of their requested spectrum is expected to be above their reserved spectrum (i.e., having equal levels of unserved traffic after reconfiguration). In other words, such allocation leads to fairly dividing the available spectrum among the contending connections. However, to find such a fair SA, we need a scheme that does not need to explore all the possible combinations of the possible SAs. Therefore, we resort to an $\alpha$-fairness scheme and formulate the static fair SA problem accordingly. Since the $\alpha$-fair SA entails solving several instances of the conventional static SA problem that was shown to be NP-complete [1], the $\alpha$-fair SA is also NP-complete. The problem becomes even harder as the problem size increases in number of connections, link capacity, and SA granularity. Considering jointly solving the routing and SA problems further increases problem complexity, and the same holds if we consider splitting the traffic demand into several lightpaths. Note that splitting the traffic into several lightpaths may reduce contention, but contention may still be present. Therefore, in this work, we start examining the simplest possible fairness SA problem by assuming that routing and SA problems are solved separately, and that one route is possible for each connection, with the routing problem solved according to Dijkstra’s algorithm [36].

The reader should note that while several routing schemes can be applied (e.g., $\kappa$-shortest paths algorithms [37]), the choice of the routing scheme does not affect the scope of this
work, i.e., to evaluate the effectiveness of the fair SA algorithms proposed. In particular, the routing scheme provides just the means of identifying the contention between the connections, which is passed as information to the fair SA algorithms.

Furthermore, in this work, we assume that the predictive traffic distributions provide information on the future virtual link utilization in number of spectrum slots (inferred by historical link utilization information). Therefore, we do not account for the various modulation formats during the proactive optimization problem, as this more concerns spectrum expansion/contraction policies followed after network configuration takes place as the traffic fluctuates [38,39]. At this stage of our work, the focus is on investigating the impact of \( \alpha \)-fairness considerations on the network performance and connections’ achievable QoS, and also on reducing the computational complexity of the exact fair-SA algorithm previously presented in [6]. Consideration of distance adaptive modulation formats, under the assumption that predictive traffic distributions are functions of the bit-rate requests, is planned for future work (i.e., by transforming the predictive bit-rate distributions to spectrum utilization information according to the modulation format). In this work, we assume the same modulation format for all connections in order for the unserved traffic (i.e., measured in number of unserved spectrum slots) to be comparable for all connections.

Finally, even though this work focuses on deriving SAs with fair QoS guarantees, the \( \alpha \)-fair SA framework proposed in this work is applicable, with the appropriate adjustments, to various use cases for different optical network applications. Indicatively, it can be applied under the network slicing context for managing connections with diverse traffic demand models (i.e., bit-rate trends) but similar service level agreements on systems (i.e., derived by the predictive traffic distributions) towards a SA that will be fairer compared to the utilitarian benchmark approach. Specifically, as we increase \( \alpha \), connections with lower spectrum demands (i.e., “poorest” connections) are expected to reserve more spectrum compared to the spectrum reserved for lower \( \alpha \) values, therefore increasing their achievable QoS. Upon convergence (i.e., an egalitarian allocation is approximated), the poorest connections are expected to increase their achievable QoS as much as possible (i.e., subject to the EON constraints), approximating near-equal QoS guarantees among the connections, provided that the problem is appropriately defined for capturing the proportions of connection satisfaction (i.e., level of service) over the utilities.

In general, function \( W(\alpha) \) derives fair allocations, as it exhibits a diminishing marginal welfare increase as utilities increase [2]. For example, if agent \( i \) is allocated fewer utilities than agent \( j \), then a marginal increase in the utilities of agent \( i \) would yield a higher welfare increase compared to a marginal increase in the utilities of agent \( j \). Hence, the marginal increase in the utilities of agent \( i \) is more desirable for the central decision management unit (i.e., for the network orchestrator).

Inequality aversion parameter \( \alpha \) controls the rate at which marginal increases diminish. Specifically, a utilitarian allocation is said to be derived for \( \alpha = 0 \) and corresponds to maximizing system efficiency (e.g., in IP networks, this is equivalent to maximizing the network throughout). For our SA problem, this means that a utilitarian allocation is obtained by preferring the maximum predicted demands (i.e., spectrum sizes analogous to the peak-rate demand). Therefore, in this work, \( \alpha = 0 \) is the benchmark approach, as it corresponds to the state-of-the-art SA schemes that consider during the optimization procedure the peak-rate traffic predictions.

As \( \alpha \to \infty \), the scheme converges to max-min fairness by maximizing min, \( u \) [41,42]. Specifically, as \( \alpha \) increases, the allocations are said to become fairer, and an egalitarian welfare allocation is derived upon convergence [43]. Formally, max-min fairness is said to be achieved by a utility allocation \( u \) if and only if the allocation is feasible (i.e., \( u \in U \)), and any increase in an agent’s allocated utility, \( u_j \), results in a decrease in some other agents’ allocated utility, \( u_i \), with equal or smaller allocation. In our SA problem, by gradually increasing \( \alpha \), we let the SA scheme explore the different SA options of each connection (i.e., derived by the predictive traffic distributions) towards a SA that will be fairer compared to the utilitarian benchmark approach. Specifically, as we increase \( \alpha \), connections with lower spectrum demands (i.e., “poorest” connections) are expected to reserve more spectrum compared to the spectrum reserved for lower \( \alpha \) values, therefore increasing their achievable QoS.

For our SA allocation problem, the utilities are the optical spectrum resources, and the agents are the connections defined by their source–destination nodes and their predictive traffic demand distributions. Note that hereafter, agents and connections will be used interchangeably in the paper, and the same holds for spectrum resources and utilities.

In general, function \( W(\alpha) \) derives fair allocations, as it exhibits a diminishing marginal welfare increase as utilities increase [2]. For example, if agent \( i \) is allocated fewer utilities than agent \( j \), then a marginal increase in the utilities of agent \( i \) would yield a higher welfare increase compared to a marginal increase in the utilities of agent \( j \). Hence, the marginal increase in the utilities of agent \( i \) is more desirable for the central decision management unit (i.e., for the network orchestrator).

Inequality aversion parameter \( \alpha \) controls the rate at which marginal increases diminish. Specifically, a utilitarian allocation is said to be derived for \( \alpha = 0 \) and corresponds to maximizing system efficiency (e.g., in IP networks, this is equivalent to maximizing the network throughout). For our SA problem, this means that a utilitarian allocation is obtained by preferring the maximum predicted demands (i.e., spectrum sizes analogous to the peak-rate demand). Therefore, in this work, \( \alpha = 0 \) is the benchmark approach, as it corresponds to the state-of-the-art SA schemes that consider during the optimization procedure the peak-rate traffic predictions.

As \( \alpha \to \infty \), the scheme converges to max-min fairness by maximizing min, \( u \) [41,42]. Specifically, as \( \alpha \) increases, the allocations are said to become fairer, and an egalitarian welfare allocation is derived upon convergence [43]. Formally, max-min fairness is said to be achieved by a utility allocation \( u \) if and only if the allocation is feasible (i.e., \( u \in U \)), and any increase in an agent’s allocated utility, \( u_j \), results in a decrease in some other agents’ allocated utility, \( u_i \), with equal or smaller allocation. In our SA problem, by gradually increasing \( \alpha \), we let the SA scheme explore the different SA options of each connection (i.e., derived by the predictive traffic distributions) towards a SA that will be fairer compared to the utilitarian benchmark approach. Specifically, as we increase \( \alpha \), connections with lower spectrum demands (i.e., “poorest” connections) are expected to reserve more spectrum compared to the spectrum reserved for lower \( \alpha \) values, therefore increasing their achievable QoS.

Upon convergence (i.e., an egalitarian allocation is approximated), the poorest connections are expected to increase their achievable QoS as much as possible (i.e., subject to the EON constraints), approximating near-equal QoS guarantees among the connections, provided that the problem is appropriately defined for capturing the proportions of connection satisfaction (i.e., level of service) over the utilities.

In general, utilitarian and egalitarian schemes are known to yield Pareto optimal allocations [44] (i.e., utilities cannot be reallocated in favor of one agent without reducing the allocation of at least another agent). Note, however, that Pareto optimality does not ensure the equality or fairness of the allocation. For our SA problem, it ensures that connections that do not contend along their routes with other connections will receive their maximum spectrum predicted, subject only to the link capacity constraints. Also, it ensures that contending connections can be allocated their maximum spectrum predicted, if such a SA does not violate the EON constraints. Pareto

4. ALPHA-FAIRNESS BACKGROUND

The \( \alpha \)-fairness scheme is based on the maximization of the constant elasticity welfare function \( W_\alpha \) parameterized by a scalar \( \alpha \geq 0 \) [2,40,41] and given by

\[
W_\alpha(u) = \begin{cases} 
\sum_{i=1}^{n} \frac{u_i^{1-\alpha}}{1-\alpha} & \text{for } \alpha \geq 0, \alpha \neq 1 \\
\sum_{i=1}^{n} \log(u_i) & \text{for } \alpha = 1
\end{cases},
\]

where \( u \in \mathbb{R}_{+}^{n} \) is a utility allocation, \( n \) is the number of agents contending for the system utilities, and \( u_i \) is the utility allocated to agent \( i \). The utility set \( U \subset \mathbb{R}_{+}^{n} \) is defined as the set of all feasible utility allocations, and the feasibility of the allocations is defined according to the agents’ system constraints. An \( \alpha \)-fair allocation is denoted by \( \pi(\alpha) \) and is given by

\[
\pi(\alpha) \in \text{argmax} \ W_\alpha(u), \quad u \in U
\]
optimality of the egalitarian case constitutes a nice property for our problem, as on the one hand, it allows for the maximum predicted spectrum to be allocated when possible, and on the other hand, it improves the QoS guarantees only for the contending connections that raise unfairness; that is, not all these connections can reserve their maximum predicted spectrum due to the EON constraints, and such a SA policy would yield connection blocking (i.e., not reserving a priori spectrum for some connections). For additional information on \( \alpha \)-fairness, the reader is referred to [2, 40–44].

5. APPROACH OVERVIEW

Given an EON with \( k \) links and \( M \) available frequency slots (FSs) at each network link, the objective is to find an \( \alpha \)-fair SA for \( n \) connections, subject to the feasibility constraints of EON technology (including spectrum continuity and contiguity, and no frequency overlap constraints). Optimal and near-optimal \( \alpha \)-fair SA algorithms are developed and compared, with the near-optimal approach addressing the increased computational and time complexity as the problem size increases (i.e., in number of connections, link capacity, etc.). The optimal algorithm, namely, the NC \( \alpha \)-fair SA, is based on a set of NC sets \( \Lambda_1, \ldots, \Lambda_v \), in which the connections/agents in each set contend with each other without, however, contending with agents in other sets. Therefore, the ILP for this algorithm is formulated to derive the optimal \( \alpha \)-fair SA for each NC set \( \Lambda_c \), with the solutions derived from all agent sets constituting the optimal \( \alpha \)-fair RSA for all agents in \( \Lambda = \bigcup_{c=1}^{v} \Lambda_c \). In the near-optimal algorithm, namely, the LBC \( \alpha \)-fair RSA, the ILP is formulated according to \( \Lambda_c \) sets of agents in which contention exists not only between the agents of the same set, but also between agents in different sets.

For both NC and LBC algorithms, the assumption is that the routes are pre-computed and given in the link utilization matrix \( P = \{ p_{ij} \} \in \mathbb{R}^{a \times k} \), indicating the contending agents for each link in the network (i.e., connections utilizing the same links). As the routing problem is solved a priori, the ILPs are formulated to solve the contention that arises during the utility allocation problem, with the utilities representing the available FSs at each network link. Therefore, the utilities are given by the SA matrix \( U = \{ u_{ij} \} \in \mathbb{R}^{a \times m} \), where \( u_{ij} \) is SA \( j \) for connection \( i \), measured in the number of FSs, and \( m \) is the number of possible SAs (actions) for each agent \( i \). Since \( M \) is the available link bandwidth, \( 0 \leq u_{ij} \leq M \) for all agents and SAs.

Given matrices \( P, U \), and the link capacity constraint \( M \), both the NC and LBC algorithms aim to maximize the social elasticity welfare function, \( W_\alpha(\hat{u}) \), parameterized by \( \alpha \geq 0 \) and defined for \( \hat{u} \in \mathbb{R}_+^a \), where \( \hat{u} = [\hat{u}_1, \ldots, \hat{u}_n] \in \mathcal{U} \) are the normalized utilities \( u = [u_1, \ldots, u_n] \in \mathcal{U} \), and \( \mathcal{U} \) is the normalized set of utilities derived from the feasible set of utilities \( \mathcal{U} \). Normalizing the utilities is an important consideration for this work, as by doing so, the \( \alpha \)-fair SA algorithms capture the fact that we are not interested in finding fair SA with respect to the amount of utilities derived by the agents; instead, we are interested in deriving SAs that will be fair with respect to the fraction of maximum predicted utilities the agents receive [45] (i.e., capturing the agent’s achievable level of services/preferences over the possible SA). The optimal \( \alpha \)-fair SA of the normalized utilities is thus given by

\[
\hat{\pi}(\alpha) = \arg\max_{\pi \in \mathcal{W}(\hat{u})} \pi(\alpha),
\]

with \( \pi \) being any feasible solution of the RSA problem. From which the optimal \( \alpha \)-fair allocation SA of the utilities \( \pi(\alpha) = \hat{u} \in \mathcal{U} \) can be obtained. For the near-optimal algorithm (and its corresponding ILP), an optimal \( \alpha \)-fair SA allocation is denoted by \( \hat{\pi}^*(\alpha) \). Note that sets \( \mathcal{U} \) and \( \mathcal{U} \) are derived during the execution of the algorithms.

The general \( \alpha \)-fair RSA framework is illustrated in Fig. 4. All operations within this framework are performed off-line, before the actual (re)configuration time points, and for each future planning period of interest. NC and LBC algorithms are executed for several \( \alpha \) values, and each \( \alpha \)-fair allocation \( \pi(\alpha) \) is evaluated in advance. Hence, the network operator is capable of selecting the \( \alpha \)-fair RSA that best meets specific performance requirements.

The measures used to evaluate the \( \alpha \)-fair SAs include the CV of unserved traffic, CB, RU, and COP and CUP. The evaluation of COP and CUP measures is based on the expected unserved traffic and expected excess utilities, \( u^- = [u_1^-, \ldots, u_n^-] \) and \( u^+ = [u_1^+, \ldots, u_n^+] \), respectively, estimated according to the \( a \) priori inferred traffic prediction models. Specifically, \( u^- \) and \( u^+ \) are estimated by the fluctuations around \( u \in \pi(\alpha) \) (or \( u \in \pi^*(\alpha) \)). It is important to mention that the evaluation of all these measures corresponds to a worst-case bound analysis, since all measures can potentially improve after network reconfiguration takes place, without, however, any guarantees. Specifically, after network reconfiguration, connections can expand/contract their spectrum, at least within their reserved resources derived by the \( \alpha \)-fair SA. Even though it may be possible for a connection to expand its spectrum above its reserved, dynamically, upon demand, this depends on the state of other in-service connections (i.e., depends on how other connections temporarily utilize their spectrum resources), without any guarantees that the network state will allow for such spectrum expansion. As we cannot \( a \) priori evaluate these measures according to the dynamic expansion/contraction policies followed (we cannot
know a priori the time points of possible fluctuations), we evaluate them according to the worst-case scenario, where connections cannot expand above their reserved spectrum. Note that this approach provides worst-case bounds for every measure. As an example, the unserved traffic measure provides the minimum QoS guarantees of each connection and can be evaluated a priori by simulating traffic fluctuations from the predictive traffic distributions.

Similar to [18], each distribution is assumed to be a function of the spectrum demand (i.e., link utilization predictions) measured according to the requested FSs. Even though the requested number of FSs depends on several factors, such as the transmission distance and the modulation format among others, in this work, for simplicity, it is assumed that the distributions directly reflect the requested number of FSs, an assumption that does not affect the scope of this work. Indicatively, the derivation of such a predictive traffic distribution is possible by monitoring the aggregated bit rates of a logical connection towards the inference of a predictive bit-rate distribution that can be afterwards converted into a predictive spectrum demand distribution, i.e., according to the distance of the pre-computed route and the appropriate modulation format selected [46]. Note, however, that in this work, we assume that a single modulation format is applied, since by doing so each FS has the same impact on all connections regarding the unserved traffic (i.e., degradation of bit rate resulting from the unallocated FSs). This work can be extended to consider the modulation format impact on the unserved traffic by redefining accordingly the utilities \( \hat{u} \) used as input to the \( \alpha \)-fair SA algorithms.

Furthermore, it is assumed that models \( F \) are inferred in local controllers by means of monitoring and analyzing the aggregated traffic of each source-destination pair forming a logical connection. After model inference, the relevant parameters (e.g., distribution, mean, and standard deviation) are communicated to the central controller, from which fluctuations \( f \) can be generated. Therefore, traffic inference is performed off-line based on historical information (i.e., bit-rate or link-utilization information). The reader should note that the traffic inference is out of the scope of this work. Works explicitly examining the traffic prediction problem in optical networks can be found in [10–12,22,24]. In these works, statistical methods [e.g., autoregressive integrated moving average (ARIMA)] or machine learning techniques have been applied, showcasing that the inference of sufficiently accurate traffic demand models is possible, with machine learning methods (e.g., deep neural networks) better capturing the nonlinear nature of time-varying traffic demand.

### 6. \( \alpha \)-Fair Routing and Spectrum Allocation

As previously mentioned, \( \alpha \)-fair SA algorithms are based on the maximization of \( W_\alpha(\hat{u}) \). First, matrices \( U \) and \( \hat{U} \), defining the set of possible SAs for each connection, are obtained as follows:

- **SA matrix** \( U = [u_{ij}] \in \mathbb{R}^n \times m \) is given by

\[
\hat{u}_{ij} = \frac{u_{ij}}{\text{max } f_i} \quad \text{if } 0 < u_{ij} \leq \text{max } f_i, \\
\hat{u}_{ij} = \frac{u_{ij}}{\epsilon} \quad \text{otherwise}
\]  

where \( \text{max } f_i \) is the peak-rate demand of connection \( i \) (i.e., according to the fluctuations \( f_i \sim F_c(i) \)), and \( \epsilon \) is a small positive value used for scoring the SAs that are less preferred compared to SAs with scores that increase as \( u_{ij} \) increases. Furthermore, Eq. (5) normalizes the utilities to capture the agent’s preferences over the possible utilities, with respect to the achievable level (fraction) of service. Hence, \( W_\alpha \) uses reward values \( \hat{u} \in [\epsilon, 1] \).

For both NC and LBC algorithms, it is assumed that the routing problem is solved using Dijkstra’s algorithm [36], and \( P = [p_{il}] \in [0, 1]^{n \times k} \). Then, the NC sets are defined according to the pre-computed routes; \( p_{il} = 1 \) if agent \( i \) utilizes link \( l \), and 0 otherwise.

#### A. NC \( \alpha \)-Fair SA Algorithm

For the NC algorithm, the agents’ set \( \Lambda = \{1, \ldots, n\} \) is decomposed into a number of NC sets \( \{\Lambda_i, i = 1, \ldots, v\} \) such that \( \Lambda_i \subseteq \Lambda \), \( \Lambda_i \cap \Lambda_{i'} = \emptyset \) and no agent in \( \Lambda_i \) is contending (i.e., shares a link) with any other agent in \( \Lambda_i' \) for all \( i \neq i' \). Therefore, the ILP can be formulated to independently process each \( \Lambda_i \) set without compromising the optimality of the \( \alpha \)-fair SA solution.

Specifically, to identify the NC sets \( \Lambda_i \), an auxiliary graph \( (AG) \) is formed in which each node is an agent \( i \in \Lambda \) and a link between two agents exists if the agents share at least one link along their routes (i.e., according to the link utilization matrix \( P \)). Then, the NC sets are defined according to the connectivity of the AG. Specifically, an agent \( i \) belongs to \( \Lambda_i \) if and only if there exists a path in \( AG \) that connects agent \( i \) with any other agent in \( \Lambda_i \). Note that the first agent in \( \Lambda_i \) can be randomly selected, from which the rest of the agents in \( \Lambda_i \) can be identified, and the process repeats until all agents in \( \Lambda \) belong exactly to one set \( \Lambda_i \). As an example, Fig. 5 illustrates an AG of seven agents forming NC sets \( \Lambda_1 = \{1, 2, 3, 5, 6\} \) and \( \Lambda_2 = \{4, 7\} \). Therefore, the ILP can independently find the optimal solution for each \( \Lambda_i \subseteq \Lambda \).

![Fig. 5. AG with NC agent sets.](image-url)
On this basis, the ILP formulation utilizes the following variables:

- $x_{ij}^c$: Boolean variable equal to one if SA $j$ is chosen for agent $i \in \Lambda_c$, and zero otherwise, for all $\{\Lambda_c | c = 1, \ldots, v\}$.
- $y_{ij}^c$: Boolean variable equal to one if FS $s$ is utilized by agent $i \in \Lambda_c$, and zero otherwise, for all $\{\Lambda_c | c = 1, \ldots, v\}$.
- $z_{ij}$: Boolean variable equal to one if FS $s$ is the first FS utilized by agent $i \in \Lambda$ among a set of contiguous FSs allocated to agent $i \in \Lambda$, and zero otherwise.

On this basis, the objective and constraints of the ILP are defined as follows:

**Objective:**

$$\text{Maximize: } W_o(\hat{u}) = \left\{ \begin{array}{ll}
\sum_{i,j} x_{ij}^c & \text{if } \alpha \geq 0, \alpha \neq 1 \\
\sum_{ij} x_{ij}^c \log(\hat{u}_{ij}) & \text{if } \alpha = 1
\end{array} \right.$$  \hspace{1cm}(6)

**Subject to:**

$$\sum_{ij} u_{ij} x_{ij}^c p_{il} \leq M, \quad \forall l = 1, \ldots, k,$$  \hspace{1cm}(7)

$$\sum_{j} x_{ij}^c = 1, \quad \forall i \in \Lambda_c,$$  \hspace{1cm}(8)

$$\sum_{j} x_{ij}^c = 0, \quad \forall i \notin \Lambda_c,$$  \hspace{1cm}(9)

$$\sum_{i} y_{ij}^c = \sum_{j} u_{ij} x_{ij}^c, \quad \forall i \in \Lambda,$$  \hspace{1cm}(10)

$$\sum_{i} y_{ij}^c p_{il} \leq 1, \quad \forall s = 1, \ldots, M, l = 1, \ldots, k,$$  \hspace{1cm}(11)

$$\sum_{i} z_{ij} \leq 1, \quad \forall i \in \Lambda,$$  \hspace{1cm}(12)

$$y_{ij}^c - y_{i(s-1)}^c \leq z_{ij}, \quad \forall i \in \Lambda, s = 2, \ldots, M,$$  \hspace{1cm}(13)

$$y_{ij}^c = 0 \quad \text{for } s = 1, \quad \forall i \in \Lambda.$$  \hspace{1cm}(14)

Given an $\Lambda_c$ set, the objective is to maximize $W_o(\hat{u})$, resulting in an $\alpha$-fair allocation $\pi(\alpha) = \arg\max_{\alpha \in \mathbb{R}} W_o(\hat{u})$, from which the actual $\alpha$-fair SA $\pi(\alpha)$ can be derived. Constraint (7) ensures that the SA decisions meet the link capacity constraints, constraint (8) ensures that one allocation is chosen for each agent in $\Lambda_c$, and constraint (9) ensures that the agents not in $\Lambda_c$ are not allocated resources (i.e., not considered in the optimization problem). Constraint (10) ensures that the number of allocated FSs is equal to the number of FSs indicated by the SA implemented, and constraint (11) ensures that the no-frequency overlapping constraint is met for all network links. Finally, constraints (12), (13), and (14) ensure that each connection is assigned contiguous FSs on all links of its route.

As the ILP is executed for all $\Lambda_c$ sets, the optimal $\pi(\alpha) = u \in \mathcal{U}$ SA solution is derived by considering all $\pi(\alpha)$ solutions. Specifically, $\pi(\alpha)$ is derived by $\chi^v = [x_{ij}^v] \in \mathbb{R}^{n \times m}$, where

### Algorithm 1. NC $\alpha$-Fair SA Algorithm

1: for $\alpha = 0$ to $B$ (where $B$ is a large number) do
2: for $\epsilon = 1$ to $v$ then
3: Solve the ILP for $\Lambda_c$ [Eqs. (6)–(14)]
4: Compute $\chi^v(\alpha)$, $\psi^v(\alpha)$
5: Derive from $\chi^v$ SA $\pi(\alpha) = u \in \mathcal{U}$
6: Return $\pi(\alpha)$, $\forall \alpha$

and with $\psi^v = [y_{ij}^v] \in \mathbb{R}^{n \times M}$, given by

$$\psi^v = \sum_{c=1}^{v} y_{ij}^v,$$  \hspace{1cm}(16)

indicating the FSs allocated to each agent. The pseudocode of the NC $\alpha$-fair SA is given in Algorithm 1.

**B. LBC $\alpha$-Fair SA Algorithm**

For the LBC algorithm, the aim is to further reduce the problem size by now decomposing the agents set $\Lambda$ into a larger number of smaller (in size) agent sets $\Lambda_c \subseteq \Lambda$, where $c = 1, \ldots, v$, and $v$ is the number of agent sets. Note that $\Lambda_c \cap \Lambda_{c'} = \emptyset$ for all $c \neq c'$, and $\cup_{c=1}^{v} \Lambda_c = \Lambda$. While for the NC algorithm, agent sets $\Lambda_c$ strictly do not contend with each other, for LBC, this constraint is relaxed, and it is assumed that an agent in set $\Lambda_c$ may contend with an agent in set $\Lambda_{c'}$ (i.e., an agent in set $\Lambda_c$ shares a link with an agent in set $\Lambda_{c'}$).

Specifically, LBC sets are identified according to an auxiliary graph $AG'$ that has as nodes all agents (connections) and all network links. Then, an agent node $i$ and a link node $j$ are connected in $AG'$ if $p_{ij} = 1 \in P$. An example of an $AG'$, indicating the link contention in a network consisting of four link nodes (denoted as $A$–$D$) and seven agent nodes (denoted as $1$–$7$), is shown in Fig. 6. As LBC is executed successively over the $\Lambda_c$ sets, the link nodes in $AG'$ are ordered in list $L$ in a descending order. Specifically, the maximum degree link node in $AG'$ (i.e., the link with the maximum contention) is placed first on the list. Thus, according to the example in Fig. 6, $L = \{B, A, C, D\}$. The first set ($\Lambda_1$) is then created according to the link node $j$ placed first in $L$ and consists of all the agent nodes adjacent to $j$. According to Fig. 6, the maximum link contention node is $B$, and hence, $\Lambda_1 = \{2, 3, 5, 6\}$. The second set ($\Lambda_2$) consists of all the agents connected with the second link node in $L$ (i.e., link node $A$ in Fig. 6), excluding the agent nodes already included in $\Lambda_1$. Hence, $\Lambda_2 = \{1\}$. Overall, $\Lambda_c = \{i \in C \setminus \cup_{c=1}^{v-1} \Lambda_{c'}\}$, where $C$ is the set of all agent nodes connected to the $c$th link node in $L$. Hence, the agent

![Fig. 6. $AG'$ indicating link contention.](image-url)
sets $\Lambda_c$ obtained from Fig. 6 are as follows: $\Lambda_1 = \{2, 3, 5, 6\}, \Lambda_2 = \{1\}, \Lambda_3 = \{4, 7\},$ and $\Lambda_4 = \emptyset.$

Figure 7 illustrates the AG of the $\Lambda_c$ sets created according to the example in Fig. 6. Clearly, for the LBC algorithm, the number of agent sets may increase, albeit by decreasing the number of agents in some sets, as opposed to the exact approach that assumes no contention between the various $\Lambda_c$ sets (Fig. 5). However, contention between the various sets may exist (e.g., agent 1 in $\Lambda_2$ contends with agents 3 and 5 in $\Lambda_1$).

For the LBC algorithm, the corresponding ILP is formulated utilizing the following additional variables:

- $\chi_{ij}':$ variable equal to one if $SA$ $j$ is chosen for agent $i \in U_{w=1}^w \Lambda_w$, and zero otherwise. Specifically,

\[
\chi_{ij}' = \sum_{w=1}^c \chi_{ij}^w,
\]

where $\chi_{ij}^w$ is the SA solution of the ILP for the $\Lambda_w$ set. Note that $\chi_{ij}^0 = 0 \forall i \in \Lambda, j = 1, \ldots, m$.

- $\psi_{ij}':$ variable equal to one if $FS$ $s$ is chosen for agent $i \in U_{w=1}^w \Lambda_w$, and zero otherwise. Specifically,

\[
\psi_{ij}' = \sum_{w=1}^c \psi_{ij}^w,
\]

where $\psi_{ij}^w$ is the FS solution of the ILP for the $\Lambda_w$ set. Note that $\psi_{ij}^0 = 0 \forall i \in \Lambda, s = 1, \ldots, M$.

The LBC algorithm is solved for each $\Lambda_c$ set for different $\alpha$ values, starting from the maximum (in size) $\Lambda_c$ set. As the solution of the ILP-based LBC algorithm is based on the partial contention information of the $\Lambda_c$ sets, the algorithm, apart from utilizing the solution from the previous agent sets (i.e., $\chi^{c-1}, \psi^{c-1}$), also utilizes information from the solutions found for the previous $\alpha$ value, denoted as $\alpha'$. This is done to guide the solution of the ILP-based algorithm towards an optimal $\alpha'$-fair SA. Specifically, preliminary results in [6] indicate that aggregated excess utilities (COP) and excess traffic (CUP) tend to decrease as $\alpha$ increases, and this information is now utilized in the ILP formulation. In particular, COP is likely to only decrease as $\alpha$ increases, and thus by utilizing this information in the ILP-based algorithm (as explained below), a near-optimal $\alpha'$-fair SA can be better approximated. On the other hand, CUP tends to decrease as $\alpha$ increases, without, however, any guarantees that it will not start increasing as $\alpha \to \infty$. Nevertheless, decreasing CUP is, in general, a preferable measure from a network operator’s perspective. Thus, constraining the CUP to only decrease as $\alpha$ increases within the ILP may impact the fairness of the SA solution, but this is done by favoring the aggregated unserved traffic. Note that in Section 8, it is shown how the solution is affected when the constraints on COP and CUP are removed from the ILP formulation, indicating that these constraints indeed contribute towards a near-optimal solution.

On this basis, the following are defined:

- $\text{unserved traffic matrix } U^- = [u^-_{ij}] \in \mathbb{R}^{n \times m},$ where $u^-_{ij}$ is the average unserved traffic of connection $i,$ when allocated $u_{ij}$ utilities, given by

\[
u^-_{ij} = \frac{1}{T} \sum_{t|u_{ij} < f_{it}} |(u_{ij} - f_{it})|,
\]

- $\text{excess utilities matrix } U^+ = [u^+_{ij}] \in \mathbb{R}^{n \times m},$ where $u^+_{ij}$ is the average excess utilities of connection $i,$ when allocated $u_{ij}$ utilities, given by

\[
u^+_{ij} = \frac{1}{T} \sum_{t|u_{ij} \geq f_{it}} (u_{ij} - f_{it}),
\]

where

\[
f_{it} \sim F_i(\cdot) \quad \forall t = 1, \ldots, T.
\]

Therefore, $u^+_{ij}$ and $u^-_{ij}$ are evaluated according to $T$ traffic demand fluctuations [Eq. (21)] around $u_{ij},$ quantifying the expected connection overprovisioning and underprovisioning effects of SA $u_{ij}.$ Note that the $a$ priori evaluation of $u^+_{ij}$ and $u^-_{ij}$ is possible due to the $a$ priori modeled traffic $F_i(\cdot).$

Furthermore, it is also examined how the solution is affected when the constraints on COP and CUP may slightly deviate from the results derived from the previous $\alpha'$ value. On this basis, the following are also defined:

- $\sigma(\alpha):$ a small constant value defined for each $\alpha,$ such that $\sigma(\alpha') \leq \sigma(\alpha)$ and $\alpha' < \alpha.$ This constant is used to control the acceptable deviation of the resulting COP and CUP as $\alpha$ increases. Note that by decreasing $\sigma(\alpha)$ as $\alpha$ increases, it is ensured that both COP and CUP will acquire a decreasing tendency as $\alpha$ increases.

- $\beta: \text{a constant denoting the difference between (the acceptable deviations) } \sigma(\alpha) \text{ and } \sigma(\alpha').$

- $\gamma: \text{a constant denoting the minimum acceptable } \sigma(\alpha) \text{ value.}$

To this end, the ILP formulation is now as follows:

**Objective:**

\[
\text{Maximize : } W_\alpha(\hat{u}).
\]

**Subject to:**

\[
\sum_i u_{ij} \chi_{ij}' p_{il} + \sum_i u_{ij} \chi_{ij}'^{-1} p_{il} \leq M, \quad \forall l = 1, \ldots, k,
\]

\[
\sum_j \chi_{ij}' = 1, \quad \forall i \in \Lambda_c,
\]

\[
\sum_j \chi_{ij}' = 0, \quad \forall i \notin \Lambda_c,
\]

Fig. 7. AG illustrating the agent sets for the example in Fig. 6.
\[
\sum_i y_{ij}^s = \sum_j u_{ij}x_{ij}, \quad \forall i \in \Lambda, \\
\sum_i y_{ij}^s p_{il} + \sum_i \psi_{ij}^{-1} p_{il} \leq 1 \\
\forall s = 1, \ldots, M, l = 1, \ldots, k, \\
\sum_i z_{ij} \leq 1, \quad \forall i \in \Lambda, \\
y_{ij}^s - y_{ij}^{(s-1)} \leq z_{ij}, \quad \forall i \in \Lambda, s = 2, \ldots, M, \\
y_{ij}^1 = 0 \quad \text{for } s = 1, \forall i \in \Lambda,
\]

Algorithm 2. LBC \(\alpha\)-Fair SA Algorithm

1: for \(\alpha = 0 \) to \(B\) (where \(B\) is a large number) do
2: if \(\alpha = 0\) then
3: Initialize \(\sigma(0)\)
4: else
5: \(\sigma(\alpha) = \max[y, \sigma(\alpha') - \beta]\)
6: for \(c = 1\) to \(v\) do
7: if \(c = 1\) then
8: \(\chi_{ij}^c = 0 \quad \forall i, j, \psi_{ij}^c = 0 \quad \forall i, s\)
9: Solve the ILP for \(\Lambda_c\) [Eqs. (22)–(32)]
10: if feasible solution found then
11: Calculate \(\chi^c\) and \(\psi^c\) [Eqs. (17) and (18)]
12: else
13: \(\chi^c(\alpha) = \chi^c(\alpha'), \quad \psi^c(\alpha) = \psi^c(\alpha')\)
14: Break
15: Set \(\alpha' = \alpha\)
16: Return \(\chi^c(\alpha), \quad \psi^c(\alpha)\)
17: Return \(\pi^c(\alpha) = u\)
18: Return \(\pi^c(\alpha), \quad \forall \alpha\)

7. EVALUATION MEASURES

Even though parameter \(\alpha\) is considered to be a natural measure of fairness, in this work, we introduce the CV of connections’ unserved traffic to directly interpret and evaluate the QoS-based fairness. As we are also interested in examining the impact of fairer SAs (i.e., impact of \(\alpha\)) on other measures commonly considered in optical networks, CB, RU, and aggregated COP and CUP are also evaluated over \(\alpha\). These measures, for convenience, are referred to in this work as efficiency-related measures. Fairness-related measures include parameter \(\alpha\) and the CV of unserved traffic. Note that the definitions of all measures examined is followed by a description of how different values of parameter \(\alpha\) are likely to affect each measure examined, irrespective of the distributional shape of the traffic predictions.

- **CV** is in general defined as the ratio of the standard deviation \(\sigma\) to the mean \(\mu\) [5], and it shows the extent of variability in relation to the mean of a population. Hence, \(CV = \sigma / \mu\). The CV is used as a quantitative measure of fairness for resource allocation in general, and in computer systems as well [5]. As an example, in any resource allocation problem, the CV of the agents’ utilities measures the variability of utilities allocated to the agents around the mean of the allocated utilities. Hence, by minimizing the CV of agents’ utilities, one derives an allocation that achieves as similar as possible amounts of utilities for each agent.

In a network environment where the preferences of the connections over the available spectrum resources may greatly vary (i.e., the mean and dispersion of their predictive traffic distributions may vary), deriving SAs that minimize the CV of connections’ utilities is not a representative measure of QoS-based fairness. As in this work the focus is on evaluating the similarity of QoS guarantees among the connections, we define the CV of \(u^-\) to be a measure of QoS-based fairness. Since the standard deviation \(\sigma\) of \(u^-\) is given by \(\sqrt{\frac{1}{n} \sum_{i=1}^n (u_{ij}^- - \mu)^2}\), where \(\mu = \bar{u}^- = \frac{1}{n} \sum_{i=1}^n u_{ij}^-\), then by substituting in \(CV = \sigma / \mu\), we derive...
where $u_i^+$ is the expected unserved traffic of agent $i$ when allocated $u_i \in \pi(\alpha)$ utilities. The $\alpha$-fair SA solution that minimizes the CV of $u^-$ is the one that best approximates the fairest QoS-based allocation. As $\alpha$ increases, the CV of $u^-$ is likely to decrease, as connections with lower demand tendency are expected to increase their allocated FSs, reducing the variability of unserved traffic in set $u^-$.

- **CB** measures the blocking percentage (i.e., number of blocked connections over the total number of connections to be provisioned). In general, CB($\alpha$) is likely to decrease as $\alpha$ increases, since fairer allocations tend to increase the minimum possible utilities allocated to a connection. The CB measure is examined to showcase that under congested network situations, attempting to reserve for all connections the spectrum that is analogous to their maximum possible demand may lead to connection blocking, resulting in highly uneven QoS guarantees. The latter is investigated in our $\alpha$-fair SA algorithms when $\alpha = 0$ (or near zero).

- **RU** measures the utilized FSs of the network. Specifically, RU is the sum of the FSs allocated to the connections along their routes. RU($\alpha$) is likely to decrease as $\alpha$ increases, since greater $\alpha$ values tend to decrease the utilities allocated to the connections with higher demand tendency, striving to increase the utilities allocated to the connections with lower demand tendency. Nevertheless, the overall behavior of RU($\alpha$) depends greatly on the number of links forming the routes of the connections.

- **COP** and **CUP** measure the aggregated excess utilities and unserved traffic, respectively, expected to emerge when the $\alpha$-fair SA is implemented [6]. Specifically,

$$\text{COP}(\alpha) = \sum_{i} u_i^+, \quad \text{CUP}(\alpha) = \sum_{i} u_i^-,$$

where $u_i^+$ and $u_i^-$ are given by Eqs. (20) and (19), respectively, for all $u_i \in \pi(\alpha)$.

In general, COP($\alpha$) is likely to decrease as $\alpha$ increases since greater $\alpha$ values tend to decrease the amount of utilities allocated to the contending connections with higher demand tendency, so as to favor the connections with lower demand tendency. Similarly, CUP($\alpha$) is also likely to decrease as $\alpha$ starts increasing beyond zero, as greater $\alpha$ values tend to allow the connections with lower demand tendency to receive more utilities, hence reducing their expected unserved traffic. However, it is possible that as $\alpha$ keeps increasing towards the fairest SA (i.e., $\alpha \rightarrow \infty$), CUP will start degrading; that is, striving for the fairest SA, connections may be highly degraded. Nevertheless, it is up to the network operator to select the $\alpha$-fair allocation that best meets the network’s COP/CUP performance targets, while also taking into consideration the users’ needs. The reader should note that in the LBC-based approach, both COP and CUP are constrained to only decrease as $\alpha$ increases.

### 8. PERFORMANCE EVALUATION

The NC- and LBC-based $\alpha$-fair SA algorithms were evaluated on the generic Deutsche Telekom (DT) network of Fig. 8. Each spectral slot in the network was set at 12.5 GHz with each fiber link utilizing $M = 180$ C-band slots. In total, $m = 91$ possible SAs were assumed. Three agent sets were examined with $n = 30, 40, 50$. Note, however, that $n = 40$ and $n = 50$ are used only for examining the impact of the number of connections on the $\Lambda_i$ sets and on the time required to solve the NC- and LBC-based algorithms. As will be shown next, for $n = 40$, NC exceeded the predefined running time, and thus comparative results on the performance evaluation metrics are given only for $n = 30$. The source–destination pair of each connection was randomly selected from the DT network nodes. For the LCB-based approach, two cases were examined that varied on the acceptable deviation $\sigma(\alpha)$. Specifically, the case where $\sigma(0) = 0$, $\beta = 0$, and $\gamma = 0$, and the case where $\sigma(0) = 5$, $\beta = 0.5$, and $\gamma = 0.1$ were examined. In the first case, no deviation was considered, while in the second case, a small deviation was acceptable even for the larger $\alpha$ values.

For evaluation of the excess utilities and unserved traffic matrices, $U^+$ and $U^-$, respectively, $T = 1000$ traffic demand fluctuations $f$ were drawn from a set of log-normal distributions $F$. Specifically, for each connection $i$, $f_i \sim F(\mu_i, \sigma_i^2)$, with each $f_i \in (0, M')$, where $M' \leq M$ is equal to the maximum achievable spectrum of the installed BVTs. Standard deviation parameters $\sigma_i^2$ were drawn from a uniform distribution as $\sigma_i^2 \sim \text{unif}(0, 1)$, and mean parameters $\mu_i$ were drawn from a uniform distribution as $\mu_i \sim \text{unif}(2.5, 4.5)$. For the simulations, fluctuations $f_i$ were scaled down by a factor of two to fit the link capacity constraints. Furthermore, the scaled fluctuations indicating a spectrum demand higher than $M$ were set to $f_i = M$. It is important to note that similar to [18], the log-normal distribution is opted for, since, just like the Internet traffic demand, it is characterized by skewness and heavy tails, as also experimentally shown [25–27]. Note that other distributions with similar characteristics could also be used (e.g., Student’s $t$ distribution), while the normal/Gaussian distribution is in general considered as not representative in describing Internet traffic demand [27]. Nevertheless, irrespective of the shape of the traffic demand distribution, parameter $\alpha$ is expected to similarly affect the performance evaluation metrics examined in this work. The impact of the $\alpha$ parameter on the different evaluation measures, without considering any specific distribution for traffic demand, is described in Section 7.

![Fig. 8. Deutsche Telekom network.](image-url)
reduced by approximately 97% for both LBC and $\sigma$ table, whereas both LBC-based approaches achieve tractable cases, while for examined. Specifically, for required to find a near-optimal $\alpha$ agent sets. Note that hereafter, the LBC-based algorithm when rendering the problem computationally inefficient for larger $\sigma$. Table 2 shows the processing time required to solve the NC- and LBC-based simulations. Table 2 contains information regarding the $\Lambda_c$ agent sets for $n = 30, 40, 50$, and for each algorithm examined. According to this table, with the NC-based approach, a smaller number of agent sets is processed, consisting of a greater number of agents, compared to the LBC-based technique, which allows contention between the agents of different sets. Furthermore, as $n$ becomes larger, the number of agents in the $\Lambda_c$ sets tends to increase for both NC- and LBC-based techniques, which is reasonable if we consider that contention (i.e., the number of links shared between the connections) is likely to increase. Regarding the total number of $\Lambda_c$ sets, clearly, NC tends to create fewer NC sets as $n$ increases, due to the increasing contention between the connections. For the LBC approach, due to the partial contention allowed, larger agent sets do not have as significant an impact on the total number of $\Lambda_c$ sets, which slightly increases as $n$ becomes larger.

Table 1. Information on $\Lambda_c$ Sets

| $\Lambda_c$ | NC $n = 30$ | LBC $n = 30$ | NC $n = 40$ | LBC $n = 40$ | NC $n = 50$ | LBC $n = 50$ |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| No. of $\Lambda_c$ sets $v$ | 8 | 17 | 6 | 18 | 3 | 19 |
| $\max(|\Lambda_1|, \ldots, |\Lambda_v|)$ | 12 | 4 | 19 | 7 | 46 | 7 |
| $\min(|\Lambda_1|, \ldots, |\Lambda_v|)$ | 1 | 1 | 1 | 1 | 1 | 1 |

Table 1 contains information regarding the $\Lambda_c$ agent sets for $n = 30, 40, 50$, and for each algorithm examined. According to this table, with the NC-based approach, a smaller number of agent sets is processed, consisting of a greater number of agents, compared to the LBC-based technique, which allows contention between the agents of different sets. Furthermore, as $n$ becomes larger, the number of agents in the $\Lambda_c$ sets tends to increase for both NC- and LBC-based techniques, which is reasonable if we consider that contention (i.e., the number of links shared between the connections) is likely to increase. Regarding the total number of $\Lambda_c$ sets, clearly, NC tends to create fewer NC sets as $n$ increases, due to the increasing contention between the connections. For the LBC approach, due to the partial contention allowed, larger agent sets do not have as significant an impact on the total number of $\Lambda_c$ sets, which slightly increases as $n$ becomes larger.

**A. Processing Time Evaluation**

To solve the proposed $\alpha$-fair SA algorithms, a MATLAB machine with a CPU at 2.60 GHz and 8 GB RAM was used. Inequality aversion parameter $\alpha$ was examined for the range of values $[0, 3]$, with an $\alpha$ step equal to 0.1, resulting in 30 $\alpha$-fair SAs for each algorithm evaluated. Note that for $\alpha$ values less than (but close to) three, the resulting SAs did not significantly change, an indicator that $\alpha = 3$ approximates an egalitarian SA. For each algorithm examined, three simulations were performed with the results averaged over all simulations. Table 2 shows the processing time required to solve the NC- and LBC-based $\alpha$-fair SA algorithms for $n = 30$ and $n = 40$. Larger values of $n$ were not examined, as for $n = 40$, the processing time exceeded the predefined limit of 100 h. Specifically, NC for $n = 40$ required 106 h up to $\alpha = 1$, rendering the problem computationally inefficient for larger agent sets. Note that hereafter, the LBC-based algorithm when $\sigma(0) = 0$ (i.e., no deviation is considered) is denoted as LBC, while the LBC-based algorithm when $\sigma(0) = 5$ is denoted as $\sigma$-LBC.

According to Table 2, LBC significantly reduces the time required to find a near-optimal $\alpha$-fair SA for all $\alpha$ values examined. Specifically, for $n = 30$, the processing time is reduced by approximately 97% for both LBC and $\sigma$-LBC cases, while for $n = 40$, NC becomes computationally intractable, whereas both LBC-based approaches achieve tractable computational times. In general, the results indicate that as the problem size increases in number of agents, the processing time significantly increases, rendering the application of NC impractical, especially for short time intervals between the planned (re)configurations. Note that the problem complexity is also affected by other network parameters, such as the link capacity, granularity of the SAs (i.e., possible SA options), and network topology. We opted not to evaluate the impact of these factors on the processing time, since the main focus is to compare the NC and LBC approaches that are differently affected by the number of agents and their contention, whereas the other factors are expected to similarly affect both approaches.

**B. Fairness and Efficiency Evaluation**

Both NC- and LBC-based algorithms were evaluated according to a number of fairness- and efficiency-related measures. For all measures, NC- and LBC-based algorithms are compared only for $n = 30$, as for larger problem sizes (i.e., more connections), the NC approach exceeded the predefined processing time of 100 h, and hence, the results of such problem sizes are not illustrated. To compare NC- and LBC-based algorithms, we evaluate the deviation of the LBC-based approach from the optimal approach (NC) as follows:

\[
\text{xGL}(\alpha) = \frac{x(\alpha) - x^*(\alpha)}{x(\alpha)},
\]

where $x(\alpha)$ may refer to any measure of interest resulting from the NC-based $\pi(\alpha)$ allocation (e.g., RU(\(\alpha\)), CV(\(\alpha\)), etc.), and $x^*(\alpha)$ refers to that measure of interest resulting from the LBC-based $\pi^*(\alpha)$ allocation. Note that for all measures of interest, a positive xGL(\(\alpha\)) denotes a gain (i.e., preferable by the network operator and/or agents), and a negative xGL(\(\alpha\)) denotes a loss. As an example, a positive deviation in CV(\(\alpha\)) of unserved traffic indicates that LBC achieves fairer QoS-based SAs compared to NC, which is preferable by a network operator striving to improve the QoS guarantees among the connections. Nevertheless, this measure is used only for evaluating the optimality of the LBC approach, and the observation of deviations stemming from an isolated fairness- or efficiency-related measure does not lead to the best $\alpha$-fair SA; that is, a network operator must observe the trade-offs among all measures to find the SA policy that best meets predefined performance targets.

Specifically, QoS-based fairness as $\alpha$ increases is illustrated in Fig. 9(a), with Fig. 9(b) illustrating the deviation of the CV of unserved traffic (the CV of $u^−$) of the LBC-based schemes from the optimal NC results. According to Fig. 9(a), the CV of $u^−$ tends to decrease for both NC- and LBC-based approaches as $\alpha$ increases, indicating that the SAs become fairer with respect to $u^−$ (i.e., SAs with fairer QoS guarantees can be approximated by an egalitarian SA as $\alpha \to \infty$). It is also noted that in the small region with $\alpha$ values ranging from zero to 0.2, the CV of $u^−$ increases with $\alpha$. This increase is caused due to the blocked connections in this range (Fig. 10(a)) that differently affect the CV of $u^−$ depending on the traffic demand behavior of the exact connections that were blocked.

Interestingly, according to Fig. 9(b), both LBC schemes result in a positive or a close to zero deviation from the NC
approach, especially for the fairer $\alpha$ values (i.e., for values greater than one). Note that in this case, LBC operates closer to NC compared to $\sigma$-LBC, with $\sigma$-LBC deriving fairer QoS-based SAs. The positive deviations of both LBC-based schemes are caused mainly by the partial contention information that LBC schemes utilize to find the SAs for each connection, compared to NC encompassing global contention information. Specifically, partial contention information does not allow the LBC SAs to optimally utilize the available spectrum, leading to SAs spanning a smaller spectrum range that in turn causes a reduction in the variability of the agents’ unserved traffic. Further, $\sigma$-LBC derives even fairer QoS-based SAs than LBC due to the $\sigma(\alpha)$ deviation used for handling COP and CUP as $\alpha$ increases. Specifically, $\sigma(\alpha)$ in LBC allows for slight increments in COP and CUP as $\alpha$ increases, compared to LBC with $\sigma(\alpha) = 0$ in which COP and CUP decrease [Eqs. (31) and (32)]. Therefore, in $\sigma$-LBC, higher levels of COP and CUP are possible compared to LBC, resulting in SAs spanning an even smaller spectrum range, resulting in fairer QoS-based SAs (i.e., due mainly to the higher CUP allowed). Note, however, that the latter depends on the $\sigma(\alpha)$ considered.

The rest of the results investigate how efficiency-related measures are affected for the sake of fairness. Specifically, Fig. 10(a) shows CB over $\alpha$ (i.e., connections not able to proactively reserve spectrum), clearly illustrating that the percentage of blocked connections reduces for both NC- and LBC-based approaches as $\alpha$ increases. As expected, the highest CB is observed for the utilitarian allocation ($\alpha = 0$). This is reasonable, as for $\alpha = 0$, the connections with higher demand tendencies are preferred, consequently blocking connections with lower demand tendencies. Note that even though zero CB is achieved for $\alpha = 0.2$, examining greater values of $\alpha$ is necessary so as to derive fairer SAs and eventually approximate an egalitarian SA achieving fairer QoS guarantees (i.e., in a network environment where the distributional shape of the agents’ traffic demands may greatly vary, accommodating all connections does not imply that the allocations are fair).

Figure 10(b) illustrates CB gain–loss (GL) for LBC and $\sigma$-LBC cases. According to the results, for both LBC-based schemes, only gain is observed. This is the case as CB is caused mainly at lower $\alpha$ values where the optimal approach tends to favor the greediest connections, while LBC, which does not have complete contention information, possibly accommodates first sets of agents with lower demand tendencies, thus consequently increasing the number of connections admitted into the network (i.e., proactively able to reserve resources).

The partial information of the LBC agent sets leads also to fewer aggregated resources allocated to the connections compared to the optimal NC approach. This is clearly demonstrated in Fig. 11(a) with the RU GL quantified in Fig. 11(b) (i.e., only RU gain is observed, ranging approximately between 3% and 7%). In general, for NC, as $\alpha$ increases (i.e., a fairer QoS-based SA is derived), RU tends to decrease. This is a reasonable outcome, considering that by increasing $\alpha$, an attempt is made to transfer resources from the agents with higher demand tendencies (“richer” agents) to the agents with lower demand tendencies (“poorer” agents). However, due to the lightpath feasibility constraints of the EON, the transfer of resources may not be one to one, leading to a decreased RU (i.e., fewer resources are transferred to the poorest agents compared to the resources taken from the richest agents). For LBC-based approaches, RU increases up to the $\alpha$ value that achieves zero CB (i.e., for $\alpha$ near 0.3), and for greater values of $\alpha$ RU only slightly varies. Note that $\sigma$-LBC achieves an RU($\alpha$) slightly closer to the NC values. This is due mainly to the small $\sigma(\alpha)$ deviation that $\sigma$-LBC allows, especially as it concerns the excess utilities $u^+$ (i.e., the COP effect), consequently allowing $\sigma$-LBC to derive SAs with larger $u$ values.

As previously discussed, it is advantageous to explore larger $\alpha$ values that obtain more efficient and fairer SAs. This is clearly demonstrated by observing COP and CUP results for larger $\alpha$ values [i.e., Figs. 12(a) and 13(a), respectively]. Specifically, according to Fig. 12(a), as $\alpha$ increases, COP decreases, for both NC- and LBC-based approaches, which is a desired effect considering that COP tends to waste resources. This is due to the fact that fairer SAs imply that fewer resources are allocated to the connections with higher demand tendencies (favoring connections with lower demand tendencies), consequently mitigating the COP effect. Importantly, both LBC-based heuristics result in less COP compared to the NC-based approach, with COP GL results in Fig. 12(b) showing...
only gain that varies between 5% and 10%. Further, \( \sigma \)-LBC operates closer to NC, due mainly to the small acceptable \( \sigma(\alpha) \) deviation on \( u^+ \) that allows \( \sigma \)-LBC agents to consume more resources compared to LBC agents, which are not allowed any deviation.

The positive COP GL results, however, are achieved at the expense of CUP, as the connections in LBC-based schemes are underprovisioned more compared to the NC-based approach (Fig. 13). Figure 13(a) shows that CUP tends to decrease for both NC- and LBC-based schemes as \( \alpha \) increases, while Fig. 13(b) shows that LBC-based schemes result in a CUP loss, especially for the fairer SAs (i.e., for \( \alpha \) values greater than 0.7). Specifically, for LBC, the highest CUP loss is approximately 25%, while for \( \sigma \)-LBC, it is approximately 15%, indicating that \( \sigma \)-LBC operates closer to NC. Again, this is the case as the acceptable \( \sigma(\alpha) \) deviation renders \( \sigma \)-LBC more flexible to approximate SAs that are closer to NC SAs. Specifically, the acceptable deviation on \( u^+ \) allows allocations of larger spectrum sizes compared to LBC, consequently resulting in less CUP. The reader should note that even though the deviation of \( \sigma \)-LBC on \( u^- \) does not have an observable benefit on the results, in general, it provides the heuristic with more flexibility and thus the possibility of finding a feasible solution for every \( \alpha \) increase (i.e., both COP and CUP can be further decreased after each \( \alpha \) iteration). On the other hand, when the LBC algorithm with no allowable deviation is utilized (i.e., \( \sigma(\alpha) = 0 \)), it may lead to the occurrence of more cases of \( \alpha \) values for which feasible solutions cannot be found, consequently failing to further reduce COP and CUP as \( \alpha \) increases.

Table 3 summarizes the xGL results (averaged and for \( \alpha = 3 \)) of the LBC-based schemes for all performance metrics considered. According to these results, LBC-based schemes operate close to the optimal NC algorithm, and in most cases, the deviations from optimal values result in performance gains. Only on the CUP effect a loss is observed; nevertheless, a maximum deviation of \(-15\%\) on CUP (for \( \sigma \)-LBC) can be considered an acceptable deviation from the optimal solution, especially considering that the average CUP deviation over all \( \alpha \) values is \(-7\%\), and for the \( \alpha \) value approximating an egalitarian SA (i.e., \( \alpha = 3 \)), the deviation is approximately \(-10\%\). Overall, given the positive gains in CB, RU, COP, and especially in the CV of \( u^- \) (i.e., \( 10\% \)), and the improvement in processing time (i.e., \( 97\% \)), LBC-based techniques can be considered as an efficient and practical alternative when the time interval between (re)configurations and/or the memory requirements of the problem size render NC impractical.

C. Impact of COP and CUP Constraints on LBC

This section provides in brief the results of the LBC-based approach when the constraints on COP and CUP effects [Eqs. (31) and (32)] are not considered. The LBC approach without these constraints is denoted as unc-LBC. By doing so, the impact of these constraints on the \( \alpha \)-fair SAs can be showcased, including the fact that these constraints indeed guide LBC-based schemes towards a near-optimal \( \alpha \)-fair SA. Indicatively, Fig. 14 illustrates CV(\( \alpha \)) of \( u^- \) for the unc-LBC and its GL when compared to the NC approach. Clearly, this metric does not decrease or even converge as \( \alpha \) increases, greatly deviating from the optimal \( \alpha \)-fair SAs. Similar behavior was also observed for other measures, such as the CUP and COP effects, on which the constraints were set (results omitted due to space requirements). Table 4 summarizes the xGL results of the unc-LBC algorithm, clearly demonstrating that some measures experience a negative deviation from the optimal value that is considerably high (e.g., \(-60\%\) in CUP and \(-50\%\) in the CV of \( u^- \) for the case of \( \alpha = 3 \)). Clearly, the constraints on COP and CUP effects considered in LBC-based techniques add useful information during the \( \alpha \)-fair SA computations, mitigating the partial contention information in the LBC agent sets and leading the techniques to operate near optimal.

![Fig. 12.](image-url) (a) Connection overprovisioning versus \( \alpha \) (COP(\( \alpha \)) in FSs. (b) COP(\( \alpha \)) gain–loss.

![Fig. 13.](image-url) (a) Connection underprovisioning versus \( \alpha \) (CUP(\( \alpha \)) in FSs. (b) CUP(\( \alpha \)) gain–loss.

![Fig. 14.](image-url) (a) Coefficient of variation of unserved traffic versus \( \alpha \) (CV(\( \alpha \)) of \( u^- \)) for NC and unconstrained LBC techniques. (b) CV(\( \alpha \)) of \( u^- \) gain–loss for the unconstrained LBC technique.

| \( \alpha \) | LBC | \( \sigma \)-LBC | LBC | \( \sigma \)-LBC |
|---|---|---|---|---|
| CV of \( u^- \) | +0.03 | +0.1 | −0.02 | +0.1 |
| CB | +0.03 | +0.02 | 0 | 0 |
| RU | +0.05 | +0.05 | +0.03 | +0.03 |
| COP | +0.07 | +0.07 | +0.05 | +0.05 |
| CUP | −0.14 | −0.07 | −0.16 | −0.1 |

Table 3. xGL Results for LBC-Based Heuristics
Table 4. xGL Results for the unc-LBC-Based Heuristic for All Metrics Considered

| X          | xGL Average | xGL for $\alpha = 3$ |
|------------|-------------|----------------------|
| CV of $u^-$ | $-0.11$     | $-0.5$               |
| CB         | $+0.02$     | $0$                  |
| RU         | $+0.04$     | $+0.06$              |
| COP        | $+0.06$     | $+0.07$              |
| CUP        | $-0.2$      | $-0.6$               |

9. CONCLUSION

This work demonstrated that $\alpha$-fairness schemes can derive SAs with fairer minimum QoS guarantees for the contending connections, when being aware of the distributional shape of traffic predictions, with the fairest QoS guarantees derived when approximating an egalitarian allocation. Specifically, in this work, optimal and near-optimal $\alpha$-fair SAs were developed and compared for proactive network optimization in reconfigurable optical networks. Both NC- and LBC-based approaches are based on the $\alpha$-fairness scheme associated with maximization of the social elasticity welfare function according to ILP formulations that consider the feasibility constraints of EONs. The main difference in the two approaches is that NC obtains the $\alpha$-fair SA on a number of agent sets that utilize global contention information, as opposed to the LBC-based schemes that utilize partial contention information of the agent sets that are processed in a sequential manner. By doing so, LBC-based schemes are able to efficiently reduce the problem size, achieving approximately near-optimal $\alpha$-fair SAs within a running time that is 97% less compared to the NC scheme.

Importantly, for LBC-based approaches, a number of efficiency- and fairness-related measures were examined, including CB, RU, COP, and the CV of $u^-$, and were shown to achieve a positive deviation (gain) as compared to the values derived by NC. The positive deviation of these measures, however, occurs at the expense of a negative CUP deviation that is on average $-7\%$ and close to $-10\%$ when an egalitarian allocation is considered (for the $\sigma$-LBC approach). Nevertheless, this is a small penalty to pay considering the positive deviations on all other performance measures considered, as well as the significant improvement in processing time. Overall, NC can be utilized when the time interval between the reconfigurations and memory capacity of the central management unit allows for derivation of the optimal solution; alternatively, LBC-based approaches can be used. Future research work includes the development of heuristic approaches to address even larger-sized problems, as well as the consideration of the modulation format during connection provisioning.

Funding. H2020 Spreading Excellence and Widening Participation (739551); Government of the Republic of Cyprus.

Acknowledgment. This work has been supported by the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 739551 (KIOS CoE) and by the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

REFERENCES

1. Y. Wang, X. Cao, and Y. Pan, “A study of the routing and spectrum allocation in spectrum-sliced elastic optical path networks,” in *IEEE INFOCOM* (2011).
2. D. Bertsimas, V. F. Farias, and N. Trichakis, “On the efficiency-fairness trade-off,” *Manage. Sci.* 58, 2234–2250 (2012).
3. A. Tang, J. Wang, and S. H. Low, “Counter-intuitive throughput behaviors in networks under end-to-end control,” *IEEE/ACM Trans. Netw.* 14, 355–368 (2006).
4. T. Lan, D. Kao, M. Chiang, and A. Sabharwal, “An axiomatic theory of fairness in network resource allocation,” in *IEEE INFOCOM* (2010).
5. R. K. Jain, D. Chiu, and W. Hawe, “A quantitative measure of fairness and discrimination for resource allocation in shared computer systems,” *Digital Equipment Corporation Technical Report* (1984).
6. T. Panayiotou and G. Ellinas, “Fair resource allocation in optical networks under tidal traffic,” arXiv:2005.05874 (2020).
7. Z. Zhong, N. Hua, M. Tornatore, Y. Li, H. Liu, C. Ma, Y. Li, X. Zheng, and B. Mukherjee, “Energy efficiency and blocking reduction for tidal traffic via stateful grooming in IP-over-optical networks,” *J. Opt. Commun. Netw.* 8, 175–189 (2016).
8. B. Yan, Y. Zhao, X. Yu, W. Wang, Y. Wu, Y. Wang, and J. Zhang, “Tidal-traffic-aware routing and spectrum allocation in elastic optical networks,” *J. Opt. Commun. Netw.* 10, 832–842 (2018).
9. S. Yan, R. Nejabati, and D. Simeonidou, “Data-driven network analytics and network optimisation in SDN-based programmable optical networks,” in *Optical Network Design and Modeling (ONDM)* (2018).
10. R. Alvizu, S. Troia, G. Maier, and A. Pattavina, “Matheuristic with machine-learning-based prediction for software-defined mobile metrocore networks,” *J. Opt. Commun. Netw.* 9, D19–D30 (2017).
11. F. Morales, M. Ruiz, L. Gifre, L. M. Contreras, V. Lopez, and L. Velasco, “Virtual network topology adaptability based on data analytics for traffic predictions,” *J. Opt. Commun. Netw.* 9, A35–A45 (2017).
12. N. Fernández, R. J. D. Barroso, D. Siracusa, A. Francescon, I. de Miguel, E. Salvadori, J. C. Aguado, and R. M. Lorenzo, “Virtual topology reconfiguration in optical networks by means of cognition: evaluation and experimental validation,” *J. Opt. Commun. Netw.* 7, A162–A173 (2015).
13. Y. Ohsita, T. Miyamura, S. Arakawa, S. Ata, E. Oki, K. Shimomoto, and M. Murata, “Gradually reconfiguring virtual network topologies based on estimated traffic matrices,” in *IEEE INFOCOM* (2007).
14. A. Caballero, R. Borkowski, I. de Miguel, R. J. Durán, J. C. Aguado, N. Fernández, T. Jiménez, I. Rodríguez, D. Sánchez, R. M. Lorenzo, D. Klonidis, E. Palkopoulou, N. P. Diamantopoulos, I. Tomkos, D. Siracusa, A. Francescon, E. Salvadori, Y. Ye, J. L. Vizcaíno, F. Pittalà, A. Tymecki, and I. T. Monroy, “Cognitive, heterogeneous and reconfigurable optical networks: the CHRON project,” *J. Lightwave Technol.* 32, 2308–2323 (2014).
15. X. Chen, J. Guo, Z. Zhu, A. Castro, R. Proietti, H. Lu, M. Shamsabardeh, and S. J. B. Yoo, “Leveraging deep learning to achieve knowledge-based autonomous service provisioning in broker-based multi-domain SD-EONs with proactive and intelligent predictions of multi-domain traffic,” in *European Conference on Optical Communication (ECOC)* (2017).
16. T. Panayiotou, K. Manousakis, S. P. Chatzis, and G. Ellinas, “On learning bandwidth allocation models for time-varying traffic in flexible optical networks,” in *IEEE Optical Network Design and Modeling (ONDM)* (2018).
17. T. Panayiotou and G. Ellinas, “Data-driven bandwidth allocation in EONs,” in *IEEE Photonics in Switching and Computing (PSC)* (2018).
18. T. Panayiotou, K. Manousakis, S. P. Chatzis, and G. Ellinas, “A data-driven bandwidth allocation framework with QoS considerations for EONs,” *J. Lightwave Technol.* 37, 1853–1864 (2019).
19. B. Yan, Y. Zhao, Y. Li, X. Yu, J. Zhang, Y. Wang, L. Yan, and S. Rahman, “Actor-critic-based resource allocation for multi-modal optical networks,” in *IEEE Globecom Workshops* (2018).
20. P. Kokkinos, P. Soumplis, and E. A. Varvarigos, “Pattern-driven resource allocation in optical networks,” IEEE Trans. Netw. Serv. Manage. 16, 489–504 (2019).
21. S. Troia, R. Alvizu, Y. Zhou, G. Maier, and A. Pattavina, “Deep learning-based traffic prediction for network optimization,” in IEEE International Conference on Transparent Optical Networks (ICTON) (2018).
22. M. Aibin, “Traffic prediction based on machine learning for elastic optical networks,” Opt. Switching Netw. 30, 33–39 (2018).
23. S. K. Singh and A. Jukan, “Machine-learning-based prediction for resource (re)allocation in optical data center networks,” J. Opt. Commun. Netw. 10, D12–D28 (2018).
24. G. Choudhury, D. Lynch, G. Thakur, and S. Tse, “Two use cases of machine learning for SDN-enabled IP/optical networks: traffic matrix prediction and optical path performance prediction,” J. Opt. Commun. Netw. 10, D52–D62 (2018).
25. I. Antoniou, V. V. Ivanov, V. V. Ivanov, and P. V. Zrelov, “On the log-normal distribution of network traffic,” Physica D 167, 72–85 (2002).
26. M. Kassim, M. Ismail, and M. Yusof, “Statistical analysis and modeling of internet traffic IP-based network for tele-traffic engineering,” ARPN J. Eng. Appl. Sci. 10, 1505–1512 (2015).
27. K. Fukuda, “Towards modeling of traffic demand of node in large scale network,” in IEEE International Conference on Communications (ICC) (2008).
28. Cisco Systems, “The zettabyte era: trends and analysis,” White Paper (2017).
29. E. Altman, K. Avrachenkov, and A. Garnaev, “Generalized α-fair resource allocation in wireless networks,” in IEEE Conference on Decision and Control (CDC) (2008).
30. Z. Allybokus, K. Avrachenkov, J. Leguay, and L. Maggi, “Multi-path alpha-fair resource allocation at scale in distributed software-defined networks,” IEEE J. Sel. Areas Commun. 36, 2655–2666 (2018).
31. L. Buzna and R. Carvalho, “Controlling congestion on complex networks: fairness, efficiency and network structure,” Sci. Rep. 7, 9152 (2017).
32. F. Callegati, L. H. Bonani, and W. Cerroni, “Service fairness in flexible optical networks,” in Optical Fiber Communication Conference (OFC) (2013).
33. S. Ma, B. Guo, X. Chen, Z. Chen, and Y. He, “An improved fairness-aware dynamic spectrum allocation scheme in elastic optical networks,” in Asia Communications and Photonics Conference (ACP) (2013).
34. S. Ma, Y. Wang, B. Guo, X. Chen, J. Li, Z. Chen, and Y. He, “A fairness-aware dynamic spectrum allocation scheme in elastic optical networks,” in 18th OptoElectronics and Communications Conference (OEC) (2013).
35. H. Chen, H. Yang, B. Luo, and L. Yan, “A joint fairness-aware and fragmentation-reduction spectrum allocation scheme in elastic optical networks,” in Opto-Electronics and Communications Conference (OECC) and Photonics Global Conference (FGC) (2017).
36. T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, “Section 24.3: Dijkstra’s algorithm,” in Introduction to Algorithms, 3rd ed. (MIT, 2009).
37. J. Y. Yen, “Finding the K shortest loopless paths in a network,” Manage. Sci. 17, 712–716 (1971).
38. K. Christodouloupolous, I. Tomkos, and E. Varvarigos, “Time-varying spectrum allocation policies and blocking analysis in flexible optical networks,” IEEE J. Sel. Areas Commun. 31, 13–25 (2013).
39. M. Klinkowski, M. Ruiz, L. Velasco, D. Careglio, V. Lopez, and J. Comellas, “Elastic spectrum allocation for time-varying traffic in flexgrid optical networks,” IEEE J. Sel. Areas Commun. 31, 26–38 (2013).
40. N. Barr, The Economics of the Welfare State (Oxford University, 1987).
41. A. Mas-Colell, M. D. Whinston, and J. R. Green, Microeconomic Theory (Oxford University, 1995).
42. E. Kalai and M. Smorodinsky, “Other solutions to Nash’s bargaining problem,” Econometrica 43, 513–518 (1975).
43. T. Bonald and L. Massoulié, “Impact of fairness on Internet performance,” in ACM SIGMETRICS (2001).
44. D. Bertsimas, V. F. Farias, and N. Trichakis, “The price of fairness,” Oper. Res. 59, 17–31 (2011).
45. H. Aziz, “Justifications of welfare guarantees under normalized utilities,” ACM SIGecom Exchanges 17, 71–75 (2020).