

Elastic traffic engineering subject to a fair bandwidth allocation via bilevel programming

Stefano Coniglio*, Luca G. Gianoli[†], Edoardo Amaldi[‡], and Antonio Capone[‡]

Abstract—The ability of TCP’s congestion control scheme to adapt the rate of traffic flows and fairly use all the available resources is one of the Internet’s pillars. So far, however, the elasticity of traffic has been disregarded in traffic engineering (TE) methodologies mainly because, only recently, the increase in access capacity has moved the bottlenecks from the access network to the operator network and hungry cloud-based applications have begun to use all the available bandwidth. We propose a new approach to TE with elastic demands which models the interaction between the network operator and the end-to-end congestion control scheme as a *Stackelberg game*. Given a set of elastic traffic demands only specified by their origin-destination pairs, the network operator chooses a set of routing paths (leader’s problem) which, when coupled with the fair bandwidth allocation that the congestion control scheme would determine for the chosen routing (follower’s problem), maximizes a network utility function. We present bilevel programming formulations for the above TE problem with two widely-adopted bandwidth allocation models, namely, max-min fairness and proportional fairness, and derive corresponding exact and approximate single-level mathematical programming reformulations. After discussing some key properties, we report on computational results obtained for different network topologies and instance sizes. Interestingly, even feasible solutions to our bilevel TE problems with large optimality gaps yield substantially higher network utility values than those obtained by solving a standard single-level TE problem and then fairly reallocating the bandwidth *a posteriori*.

Index Terms—Traffic engineering, elastic traffic, max-min fairness, proportional fairness, Stackelberg games, bilevel programming.

I. INTRODUCTION

MOST of the state-of-the-art approaches and methods for traffic engineering assume that, given a set of traffic demands with the corresponding origin-destination pairs and bandwidth requirements (*traffic matrix*), the network operator can freely choose the routing paths for each demand (*routing*) and the corresponding bandwidth allocation (or rate) which satisfy the bandwidth requirements. A typical objective is maximizing a suitable *utility function*, for instance the network throughput, or minimizing a *penalty function*, such as the network congestion or a measure of energy consumption [1]. Such

traffic demands are commonly referred to as *inelastic*, as their data rates are assumed to be fixed and known *a priori*, even though uncertainty in the traffic matrices can be accounted for with robust or stochastic optimization approaches.

The appropriateness of assuming inelastic demands is often challenged due to the very nature of the traffic generated by modern Internet applications, as well as due to the recent evolution of network infrastructures [2]. In the Internet, as known, most of the traffic is handled by TCP (Transmission Control Protocol), which, thanks to its *congestion control mechanism*, allows traffic flows to adapt their rate so to, potentially, use all the available capacity (new protocols which are rapidly gaining popularity, such a QUIC—Quick UDP Internet Connections, have a similar behavior). Given a set of routing paths, TCP autonomously determines a bandwidth allocation with the aim of delivering a *best-effort* service in which the available network resources are *fairly* shared among the different (concurrent) connections. Such traffic demands are commonly referred to as *elastic*.

As shown in [3]–[5], the bandwidth allocation determined by TCP and similar protocols can be suitably approximated by the solution to a well-defined (implicit) optimization problem. In many relevant cases, such as First-In-First-Out (FIFO) and per-flow fair queuing [5], it can be shown that such problem amounts to, given a routing, finding a bandwidth allocation which maximizes a certain measure of *fairness*.¹ In particular, it has been shown that, due to the interaction with different queuing policies and other congestion control mechanisms, the resource allocation that is carried out in existing TCP/IP networks can be well approximated by one of two frequently adopted fairness paradigms: *Max-Min-Fairness* (MMF) and *Proportional Fairness* (PF) [3], [5].²

Although the best-effort nature of TCP has been known for many years, Internet Service Providers have been exposed to the elastic nature of the Internet only recently. Indeed, on the one hand, many applications have recently become very bandwidth-hungry and capable of using all the available resources for a relatively long time and, on

¹In this work, we refer to *fairness maximization* rather than to *utility maximization*, which is more common in the literature, to avoid confusion with the network operator’s utility function.

²The behavior of TCP is influenced by a number of factors, including its version [4], the Round-Trip-Times (RTTs) of the different connections, upper layer protocols such as, e.g., HTTP [6], and the presence of queue management mechanisms such as Random Early Detection (RED) [7] or Explicit Congestion Notification (ECN) [8].

*Department of Mathematical Sciences, University of Southampton, Southampton, UK; [†]Département de génie électrique, École Polytechnique de Montréal, Montréal, Canada; [‡]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy.

the other hand, the capacity of the access networks has recently started to grow very quickly. No more restricted by the bottlenecks induced by the small capacity of the access network, traffic flows can now fully reveal their elastic nature to the operators, causing the network bottlenecks to shift from the access network to the operators' networks. The impact of a high-access capacity is particularly relevant when considering increasingly popular cloud-based applications, where multiple data centers interconnected through a geographical network host portions of the service platforms that need to communicate.

When applied to the case of elastic demands, classical approaches to traffic engineering developed for inelastic demands have been successful, so far, only because the traffic demands used to behave, *de facto*, as inelastic, as they were limited by a tight bottleneck at the access level (i.e., one outside of the network domain under consideration). Due to the rapid increase in capacity of the access network [9], [10], this situation is rapidly changing and it is bound to become radically different in the future. Even if one could use traffic shaping mechanisms at the network edges to limit the access rates, which would allow for the adoption of the classical methods, this would drastically limit the traffic demands and lead to a waste of resources. Ad hoc methods for *elastic traffic engineering* are, thus, necessary. Our work constitutes a first step in this direction.

II. CONTRIBUTIONS AND OUTLINE OF THE PAPER

We propose a new traffic engineering (TE) approach which models the **current** interaction between the network operator and the congestion control mechanism in a TCP/IP network as a *Stackelberg game* (i.e., a **two-stage game with two players, a leader and a follower, the former playing before the latter**). **The model captures the fact that the network operator plays first, choosing a set of routing paths which, when coupled with the fair bandwidth allocation that the congestion control scheme would determine for the routing chosen by the network operator (by maximizing an appropriate fairness measure), maximizes a network utility function. Equilibria in Stackelberg games are found by computing a strategy for the leader by which, taking into account the rational (i.e., optimal) way the follower would react to such strategy, the leader's utility is maximized. From an optimization perspective, their computation amounts to solving a bilevel programming problem.**

In this paper, we investigate the above **Stackelberg-equilibrium-finding bilevel-programming** problem, which we refer to as *Bilevel Traffic Engineering with a Fair Bandwidth Allocation* (BTE-FBA), both theoretically and computationally. For simplicity, we consider as network operator's utility function the weighted total throughput, but our approach is sufficiently general to allow for different choices. We focus on two specific versions: *Bilevel Traffic Engineering with an MMF Bandwidth Allocation* (BTE-FBA-MMF) and *Bilevel Traffic Engineering with a PF Bandwidth Allocation* (BTE-FBA-PF).

In line with the most popular flow-based routing protocols, such as MPLS (MultiProtocol Label Switching), **we assume that** a single **unsplittable** path is explicitly assigned to each demand.

Our approach is fundamentally different from previous works on fairness in traffic engineering which, to the best of our knowledge, only address the single-level problem of finding a set of routing paths and a bandwidth allocation which, *together*, maximize a measure of fairness.

The aim of the paper is twofold. On the one hand, we aim to solve BTE-FBA-MMF and BTE-FBA-PF with mathematical programming methods. On the other hand, we aim to quantify the improvement that solutions to BTE-FBA-MMF and BTE-FBA-PF can yield in terms of the network operator's utility function when compared to those obtained by ignoring the bilevel aspect of the problem. **In particular, we will propose *single-level reformulations of BTE-FBA-MMF and BTE-FBA-PF which can be tackled with state-of-the-art mathematical programming solvers.***

The paper is organized as follows.³ In Section III, we summarize previous work on traffic engineering and fairness in TCP/IP networks. In Section IV, we lay the basis for our bilevel TE approach and illustrate how to recast the second-level problem of BTE-FBA (determining an MMF or PF allocation for a given set of routing paths) in terms of compact sets of optimality conditions. In Section V, we quantify the worst-case difference in utility function between BTE-FBA-MMF and BTE-FBA-PF and **two standard single-level methods**. In Sections VI and VII, we derive single-level Mixed Integer Linear Programming (MILP) reformulations for BTE-FBA-MMF and BTE-FBA-PF, either exact or approximate. Computational results for different network topologies and instance sizes are reported and discussed in Section VIII. Section IX contains some concluding remarks.

III. PREVIOUS WORK

A. Traffic engineering in the inelastic case

Traffic engineering problems have been extensively studied in the literature. Depending on the protocol, the routing may involve, for instance, the adjustment of the link weights so as to determine a shortest path tree [15] or the explicit computation of the routing paths used for each demand [16]. For an overview, we refer the reader to [1] as well as to the aforementioned [15], [16].

B. Fairness in TCP/IP networks

In the last 20 years, a lot of attention has been devoted to the study of the different congestion control mechanisms implemented in IP networks, including TCP and its variants. The pioneering work in [3] is the first one to look at the bandwidth allocation achieved by a congestion control mechanism as to the solution to a suitably designed *implicit* optimization problem where a

³Preliminary versions of part of this work appeared in [11]–[14].

fairness function is maximized. Several definitions of fairness have been considered. Under suitable circumstances, it has been shown that the two most popular ones, *Max-Min Fairness* and *Proportional Fairness*, well approximate the bandwidth allocation determined by TCP [1], [3], [17]. For alternative definitions, among which those of *Balanced Fairness*, *Relative Fairness*, and *Maximal Fairness*, we refer the reader to [18], [19].

Informally speaking, a bandwidth allocation is Max-Min Fair (MMF) if there is no way to increase the rate of a connection without decreasing that of another one which is receiving a smaller share. Differently, a bandwidth allocation is Proportional Fair (PF) if it maximizes the sum of the logarithms of the rates of the different demands. Formal definitions are provided in Section IV.

C. Approximations of TCP via fairness maximization

We classify the nonalgorithmic works on fairness in networking and traffic engineering in two main categories: prescriptive and descriptive.

Prescriptive works typically focus on understanding which type of fairness could guarantee the highest network performance (e.g., which one could maximize the overall throughput). Their aim is providing guidelines for the design of efficient congestion control mechanisms. Examples can be found in [18], [20], [21]. A substantial body of work has been carried out for the case of a given single path routing with the aim of understanding the advantages of enforcing different types of fairness [18], [20], [22], [23] and their relationship to other performance measures such as the overall throughput or the average delay [18], [20], [21]. These works show that, by giving priority to connections with a smaller Round-Trip-Time (RTT), PF tends to achieve a slightly better throughput than MMF.

Descriptive works focus on deriving analytic ways to evaluate the level of fairness achieved by an existing congestion control mechanism. They show that, in the general case, a TCP-based bandwidth allocation can be quite well approximated by an MMF allocation [5], [24], provided that the RTTs for the concurrent connections are not too different. The speed at which the transmission window of each TCP connection grows is, in fact, determined by the time required by a packet to reach its destination node, plus that taken by the successive acknowledgment message to reach the original transmitting node. Accordingly, the smaller the RTT of a connection, the faster the corresponding demand can occupy the available bandwidth and, thus, the larger its rate can be. Empirical and analytic studies suggest that, in modern TCP/IP networks with highly heterogeneous RTTs and link capacities, the resource allocation may be better approximated by PF [3], [4].

The fairness function maximized by the different versions of TCP is studied in [4], while [6] addresses the case of HTTP-over-TCP. The work in [5] shows how the adoption of different queuing policies such as First-In-First-Out (FIFO), per-flow fair queuing, and Longest-Queue-First (LQF) may produce different types of allocations (PF,

MMF, or one which maximizes the throughput, respectively). Further studies on the role of fairness in other TCP variants or in the presence of other mechanisms such as Random Early Detection (RED) or Explicit Congestion Notification (ECN) are presented in [7], [8].

D. Fairness maximization in network routing

Algorithmic approaches to compute an MMF allocation for a set of elastic demands in a given network topology with link capacities have been extensively studied [1], [17], [25]–[29]. When the routing paths are not given, a substantial body of work has focused on the *simultaneous* determination of a routing *and* of a bandwidth allocation which are *together* MMF. See [1], [17] for the case of splittable routing and [26], [27] for that of unsplittable routing with a single path per demand. See also [28], [29] for methods to compute an MMF solution for general convex or nonconvex optimization problems. To the best of our knowledge, similar studies have not been carried out for the PF allocation of elastic demands. **In our setting where we assume an unsplittable single-path routing, we refer to the problem of determining a routing and a bandwidth allocation which are together MMF as *MMF overall* and, by extension, we refer to its counterpart for the PF case as *PF overall* (see Section V).**

Other relevant works on fairness in resource allocation include, among others, weaker definitions of fairness such as Upward Max-Min Fairness (UMMF) [30] whose allocations can be efficiently computed via a distributed algorithm, a study on how to obtain MMF solutions by optimizing suitably chosen nonlinear objective functions [31], and a variational-inequality-based approach leading to a form of fairness called Utility Proportional Fairness [32]. As to UMMF, a linear programming formulation is presented in [33] to jointly determine both a routing and a bandwidth allocation so to balance throughput and fairness in the case of splittable routing.

IV. BILEVEL TRAFFIC ENGINEERING WITH A FAIR BANDWIDTH ALLOCATION

In this section, we formalize **our Stackelberg model and the corresponding equilibrium-finding** bilevel TE problem, we give a formal definition of the second-level fair bandwidth allocation problem for both MMF and PF, and lay the basis for solution approach. After introducing a bilevel programming formulation for both versions of BTE-FBA, we illustrate how to reformulate its second-level problem via a compact set of optimality conditions, **which we will exploit in the next sections to derive equivalent single-level mathematical programming reformulations solvable with state-of-the-art mathematical programming algorithms.**

A. Problem definition

The **bilevel TE** problem that we address in this paper can be formally stated as follows:

Bilevel Traffic Engineering with a Fair Bandwidth allocation (BTE-FBA): *Given a directed graph $G =$*

(V, A) with vertex set V and arc (link) set A , a capacity $c : A \rightarrow \mathbb{R}^+ \setminus \{0\}$, and a set D of elastic demands with, for each $d \in D$, the corresponding origin-destination pair (s^d, t^d) , a weight $w^d \in \mathbb{R}^+$, and no predetermined bandwidth requirements, find an *unsplittable* single-path routing such that, after a fair bandwidth allocation $\phi = \{\phi^d\}_{d \in D}$ has been determined as an optimal solution to the second-level fairness maximization problem, the network utility function $U(\phi) = \sum_{d \in D} w^d \phi^d$ is maximized.

Although, in this paper, we will consider as utility function the total weighted throughput, it is worth emphasizing that our approach is sufficiently general to allow for any $U : \phi \mapsto \mathbb{R}^+$ that can be handled by a sufficiently efficient mathematical programming solver.

B. Bilevel programming formulation

Besides the allocation vector ϕ , let us introduce, for each elastic demand $d \in D$ and arc $(i, j) \in A$, a binary variable x_{ij}^d which takes value 1 if the demand $d \in D$ is routed over the arc $(i, j) \in A$ and 0 otherwise.

Let $F : \phi \mapsto \mathbb{R}^+$ be a *fairness measure*. After an *unsplittable* single-path routing $\mathbf{x} = \{x_{ij}^d\}_{(i,j) \in A, d \in D}$ has been chosen by the network operator, the congestion control mechanism determines a fair bandwidth allocation ϕ by solving the following second-level *fairness maximization* problem subject to resource allocation constraints:

$$\operatorname{argmax}_{\phi \geq \mathbf{0}} \left\{ F(\phi) : \sum_{d \in D: x_{ij}^d = 1} \phi^d \leq c_{ij} \quad \forall (i, j) \in A \right\}. \quad (1)$$

Note the dependency on \mathbf{x} in the summation over D , by which ϕ^d is taken into account only if $x_{ij}^d = 1$, i.e., only if the demand $d \in D$ is routed over the arc $(i, j) \in A$.

Relying on (1), the *Stackelberg-equilibrium-finding* bilevel TE problem tackled by the network operator can be cast as the following bilevel program:

$$\max_{\phi, \mathbf{x} \in \{0,1\}^{|A||D|}} \sum_{d \in D} w^d \phi^d \quad (2a)$$

$$\text{s.t. } \sum_{(i,j) \in A} x_{ij}^d - \sum_{(j,i) \in A} x_{ji}^d = \begin{cases} 1 & \text{if } i = s^d \\ -1 & \text{if } i = t^d \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in V, d \in D \quad (2b)$$

$$\sum_{(ij) \in A} x_{ij}^d \leq 1 \quad \forall i \in V, d \in D \quad (2c)$$

$$\sum_{(i,j) \in A: i, j \in S} x_{ij}^d \leq |S| - 1 \quad \forall S \subseteq V, d \in D \quad (2d)$$

$$\phi \in \operatorname{argmax}_{\phi \geq \mathbf{0}} \left\{ F(\phi) : \sum_{d \in D: x_{ij}^d = 1} \phi^d \leq c_{ij} \quad \forall (i, j) \in A \right\}. \quad (2e)$$

Constraints (2b), (2c), and (2d) are, respectively, flow conservation, degree, and subtour elimination constraints which, together with the binarity of x_{ij}^d , guarantee that each demand $d \in D$ be routed over a single s^d - t^d path P^d , with $P^d = \{(i, j) \in A : x_{ij}^d = 1\}$.

We remark that, as we will explain in Sections VI and VII, subtour elimination constraints (which are usually not necessary in single-level routing problems) must be explicitly introduced here.

As we will show in the next subsection, the problem underlying Constraints (2e) admits a unique optimal solution in both the MMF and PF cases. BTE-FBA-MMF and BTE-FBA-PF are, thus, two well-posed bilevel programming problems, not affected by the ambiguity arising when multiple optimal second-level solution are present.

The major challenge one has to face when solving Formulation (2) is posed by the *argmax operator* in Constraint (2e). To arrive at a single-level reformulation which is solvable with state-of-the-art mathematical programming methods, we describe *optimality conditions* for the second level problem (for both the MMF and PF versions) which will allow us to drop the *argmax operator*.

The results in the next subsection are adapted from the literature. Since they are instrumental for the original results presented in the following sections, we also report their proofs to make the paper self-contained.

C. Optimality conditions for an MMF allocation

Formally speaking, an allocation vector ϕ is MMF if only if it is lexicographically maximum for all permutations of its indices [17], i.e., if it is not possible to increase the allocation of any elastic demand $d \in D$ without decreasing that of any other demand $d' \in D$ with $\phi^{d'} \leq \phi^d$.

Let $\sigma : \mathbb{R}^{|D|} \rightarrow \mathbb{R}$ be a nondecreasing sorting operator such that $\sigma(\phi)^d \leq \sigma(\phi)^{d'}$ for all $d, d' \in D$ with $d' \leq d$. Applying σ it to ϕ , we have that ϕ is MMF if and only if it is impossible to increase the value of $\sigma(\phi)^d$ for some $d \in D$ without decreasing that of $\sigma(\phi)^{d'}$ for some $d' \in D$ with $d' < d$. By relying on this definition, an allocation vector ϕ is MMF if and only if it satisfies:

$$\phi \in \operatorname{argmaxlex}_{\phi \geq \mathbf{0}} \left\{ \sigma(\phi) : \sum_{d \in D: x_{ij}^d = 1} \phi^d \leq c_{ij} \quad \forall (i, j) \in A \right\}, \quad (3)$$

where the *argmaxlex* operator corresponds to maximizing the vector $\sigma(\phi)$ componentwise from $d = 1$ to $d = |D|$.

From an algorithmic perspective, an MMF allocation vector ϕ can be computed in polynomial time using the Water Filling Algorithm [25]. Given an *unsplittable* single-path routing \mathbf{x} , the algorithm constructs an MMF allocation ϕ by, at each iteration, simultaneously increasing by the same amount the allocation of all the demands whose arcs have a strictly positive residual capacity and discarding the demands whose paths have at least one saturated arc.

As a consequence of the Water Filling Algorithm, we can characterize MMF allocations in terms of bottleneck arcs: an allocation vector ϕ is MMF if and only if the path for each demand $d \in D$ contains a *bottleneck arc*, i.e., an arc which i) is saturated and on which ii) the allocation ϕ^d is at least as large as that of any other demand $d' \in D$ [25].

This characterization allows for recasting Problem (3) via a set of Mixed Integer Linear Programming constraints

and variables. We report them here in a slightly modified version w.r.t. the original one:

Proposition 1 ([14], [34]). *A vector $\phi \geq \mathbf{0}$ is the unique MMF allocation for a given **unsplittable** single-path routing \mathbf{x} if and only if there is a vector $\mathbf{y} \in \{0, 1\}^{|A| \times |D|}$ such that the triple $(\phi, \mathbf{x}, \mathbf{y})$ satisfies:*

$$\sum_{(i,j) \in A: x_{ij}^d = 1} y_{ij}^d \geq 1 \quad \forall d \in D \quad (4a)$$

$$\sum_{d' \in D: x_{ij}^{d'} = 1} \phi^{d'} \geq c_{ij} y_{ij}^d \quad \forall (i, j) \in A, d \in D \quad (4b)$$

$$\phi^d \geq \phi^{d'} - c_{ij}(1 - y_{ij}^d) \quad \forall (i, j) \in A, d \in D, d' \in D \quad (4c)$$

$$\sum_{d \in D: x_{ij}^d = 1} \phi^d \leq c_{ij} \quad \forall (i, j) \in A. \quad (4d)$$

Proof. Due to Constraints (4a)–(4c), the binary variable y_{ij}^d takes value 1 if and only if arc $(i, j) \in A$ is bottleneck for demand $d \in D$. Constraints (4a) guarantee that there is at least a bottleneck arc per elastic demand, associated with one of the arcs contained in its path.⁴ Constraints (4b) guarantee that bottleneck arcs are saturated. Constraints (4c) impose, for each demand $d \in D$ and arc $(i, j) \in A$ which is bottleneck for it, an allocation at least as large as that of any other demand routed over it.

Unicity follows as a consequence of the (strict) concavity of the problem's objective function, see [25]. \square

D. Optimality conditions for a PF allocation

By definition, an allocation vector ϕ is PF if and only if it maximizes the sum of the logarithms of the allocations:

$$\phi \in \operatorname{argmax}_{\phi \geq \mathbf{0}} \left\{ \sum_{d \in D} \log(\phi^d) : \sum_{d \in D: x_{ij}^d = 1} \phi^d \leq c_{ij} \forall (i, j) \in A \right\}. \quad (5)$$

Since Problem (5) features a set of linear constraints, constraint qualification is satisfied everywhere in the feasible region and, since the problem is convex, the Karush-Kuhn-Tucker conditions (KKTs) are both necessary and sufficient for the optimality of any of its feasible solutions. We deduce the following:

Proposition 2 ([3]). *A vector $\phi \geq \mathbf{0}$ is the unique PF allocation for a given **unsplittable** single-path routing \mathbf{x} if and only if there is a vector $\lambda \geq \mathbf{0}$ such that $(\phi, \mathbf{x}, \lambda)$ satisfies:*

$$\frac{1}{\phi^d} = \sum_{(i,j) \in A: x_{ij}^d = 1} \lambda_{ij} \quad \forall d \in D \quad (6a)$$

$$\left(\sum_{d \in D: x_{ij}^d = 1} \phi^d - c_{ij} \right) \lambda_{ij} = 0 \quad \forall (i, j) \in A \quad (6b)$$

$$\sum_{d \in D: x_{ij}^d = 1} \phi^d \leq c_{ij} \quad \forall (i, j) \in A. \quad (6c)$$

⁴We assume w.l.o.g. that more links can be simultaneously bottlenecked for the same demand. We consider this version as, with it, the final formulation performs computationally better.

Proof. Let λ_{ij} and μ^d be the multipliers of, respectively, the capacity and nonnegativity constraints. The KKTs of Problem (5) are the same as those in Constraints (6) with the extra complementarity condition $-\phi^d \mu^d = 0$, for all $d \in D$, and with $\frac{1}{\phi^d} = \sum_{(i,j) \in A: x_{ij}^d = 1} \lambda_{ij} - \mu^d$, for all $d \in D$, substituted for Constraint (6a).

Note that if, for all $d \in D$, $P^d = \{(i, j) \in A : x_{ij}^d = 1\}$ contains an s^d - t^d path, then $\phi > \mathbf{0}$ in any optimal solution to Problem (5). **This is because**, if P^d contains an s^d - t^d path for all $d \in D$, there is a feasible allocation with $\phi^d > 0$ for all $d \in D$ and, thus, an optimal solution of value $\sum_{d \in D} \log(\phi^d) > -\infty$. Therefore, any ϕ with $\phi^d = 0$ for some $d \in D$ cannot be optimal. The claim follows as $\phi > \mathbf{0}$ implies $\mu = \mathbf{0}$.

Unicity is **due to the** the convexity of the problem, which calls for the maximization of a strictly concave function subject to polyhedral constraints. The result follows since **the** Hessian matrix of $\sum_{d \in D} \log(\phi^d)$ is $H_\phi = \operatorname{diag} \left(-\left(\frac{1}{\phi^1}\right)^2, \dots, -\left(\frac{1}{\phi^{|D|}\right)^2 \right) < 0$ whenever $\phi > \mathbf{0}$. \square

Note that, for each $d \in D$, Constraints (6) impose that the allocation ϕ^d be inversely proportional to the sum of λ_{ij} for all the arcs which are i) contained in the path P^d and ii) saturated (due to the complementarity constraints in (6b), as $\sum_{d \in D: x_{ij}^d = 1} \phi^d < c_{ij}$ implies $\lambda_{ij} = 0$). This is in line with the MMF conditions in Proposition 1 as, in both cases, a fair flow allocation is solely determined by the arcs which are saturated.

E. Multi-session extension

Under the assumption that elastic demands with the same origin-destination pair are routed over the same path, the following holds true:

Corollary 1. *In an MMF or PF bandwidth allocation, all elastic demands which, given an **unsplittable** single-path routing \mathbf{x} , are routed over the same path, have the same bandwidth.*

Proof. For the MMF case, the result follows from the correctness of the Water Filling algorithm. Indeed, as all demands routed over the same path share the same set of arcs, their residual capacities are the same throughout the iterations of the algorithm. For the PF case, the result follows from Constraints (6a) in Proposition 2. Indeed, for every $d \in D$, ϕ^d is a function only of the multipliers λ_{ij} associated with the arcs contained in the corresponding path and, therefore, all the demands carried by that path must have the same allocation. \square

According to Corollary 1, one can *bundle* demands with the same origin-destination pair into a single demand d' , whose allocation $\phi^{d'}$ corresponds to the total bandwidth allocated to the bundle. We refer to each demand in the bundle as a *session*. With $\Delta^{d'}$ sessions per demand, each receives a share equal to $\frac{1}{\Delta^{d'}} \phi^{d'}$. Given the weights of a

set of sessions $d \in D'$, the weight of the demand in which they are bundled is $w^{d'} = \sum_{d \in D'} \frac{\phi^{d'}}{\Delta^{d'}}$.

V. ON THE DIFFERENCE BETWEEN BILEVEL AND SINGLE-LEVEL TRAFFIC ENGINEERING

The current practice for network operators when facing traffic engineering problems with elastic demands is to ignore the bilevel (two-player) aspect of the problem, assuming that they can **maximize their network utility function** by simultaneously determining both the routing and the bandwidth allocation.⁵ This corresponds to solving the following problem:

Single-level Traffic Engineering (STE):

Given a directed graph $G = (V, A)$ with vertex set V and arc (link) set A , a capacity function $c : A \rightarrow \mathbb{Z}^+ \setminus \{0\}$, and a set D of elastic demands with, for each $d \in D$, the corresponding origin-destination pair (s^d, t^d) , a weight $w^d \in \mathbb{R}^+$ and no predetermined bandwidth requirements, find an **unsplittable single-path routing** and a bandwidth allocation $\phi = \{\phi^d\}_{d \in D}$ which maximize the network utility function $U(\phi) = \sum_{d \in D} w^d \phi^d$.

STE ignores that, with elastic demands, the bandwidth allocation is independently determined by the congestion control mechanism. **Therefore**, it is only when the routing found by solving STE is implemented in the network that the network operator can realize that a different bandwidth allocation has been enforced on the chosen routing paths, leading to a different utility-function value.

In the following, we refer as STE+MMF and STE+PF to the two-step approach in which, after an **unsplittable single-path routing** \mathbf{x} has been found by solving STE, an appropriate MMF or PF allocation is recomputed according to the congestion control protocol at hand.

An important question which naturally arises is quantifying the **worst-case** difference in terms of utility function value between solutions to the bilevel problems BTE-FBA-MMF or BTE-FBA-PF and those provided by STE+MMF or STE+PF. The following result highlights the danger of neglecting the bilevel aspect of traffic engineering with elastic demands.

Theorem 1. *In the worst case, the loss in utility-function value determined by choosing, when tackling either BTE-FBA-MMF or BTE-FBA-PF, a routing that is optimal for STE and then recomputing an MMF or PF allocation for that routing a posteriori can be arbitrarily large and it can increase at least linearly w.r.t. the size and capacity of the network.*

Proof. To prove the result, we consider the instance depicted in Figure 1. Note that all demands but (v, w) , namely, demands $(v, a_1), (a_1, a_2), \dots, (a_{k-1}, w)$ and demands $(v, b_1), (b_1, b_2), \dots, (b_{\ell-1}, w)$, can be routed over a

⁵From a bilevel programming perspective, this is the same as solving the so-called *high-point relaxation* of a bilevel problem, in which the argmax constraint accounting for the optimality of the second-level solution is discarded.

unique path consisting of a single arc. Demand (v, w) can be routed over the upper path (we refer to it as “up”) or over the lower path (we refer to it as “down”).

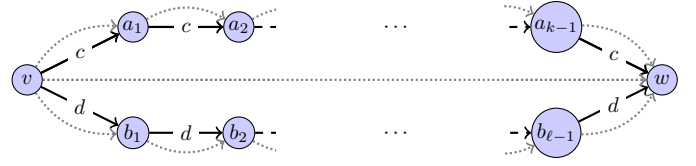


Fig. 1. A parametric instance with $k + \ell$ nodes, $k + \ell$ arcs (thick), and $k + \ell + 1$ demands (dotted), with capacities c and d .

First, we need to determine the bandwidth allocations for the two possible routings of demand (v, w) . Since these routings are topologically symmetrical, it suffices to look for an MMF or a PF bandwidth allocation for the substance reported in Figure 2.



Fig. 2. A parametric substance with $k + 1$ nodes, k arcs (thick), and $k + 1$ demands (dotted): $(v, a_1), (a_1, a_2), \dots, (a_k, w)$, and (v, w) .

According to the analysis of Section IV, we deduce that, as a consequence of the Water Filling algorithm, the unique MMF allocation for the substance in Figure 2 is:

$$\phi^{vw} = \frac{1}{2}c \quad \phi^{ij} = \frac{1}{2}c \quad \forall (i, j) \in A. \quad (7)$$

As to the PF case, according to Proposition 2 a bandwidth allocation $\phi \geq \mathbf{0}$ satisfying the capacity constraints is PF if and only if there is a vector $\lambda \geq \mathbf{0}$ such that:

$$\phi^{vw} + \phi^{ij} = c \quad \forall (i, j) \in A \quad (8a)$$

$$\phi^{ij} = \frac{1}{\lambda_{ij}} \quad \forall (i, j) \in A \quad (8b)$$

$$\phi^{vw} = \frac{1}{\sum_{(ij) \in A} \lambda_{ij}}. \quad (8c)$$

Equations (8a) correspond to Constraints (6c) in Proposition 2 imposed as equations (this is correct as any ϕ not saturating all arcs in Figure 2 cannot maximize $\sum_{d \in D} \log(\phi^d)$ —by increasing the allocation on all the nonsaturated arc by some $\epsilon > 0$, a larger $\sum_{d \in D} \log(\phi^d)$ would be obtained). From Equations (8a) we deduce $\phi^{ij} = c - \phi^{vw}$, which, due to Equations (8b), implies $\lambda_{ij} = \frac{1}{c - \phi^{vw}}$. Substituting $\frac{1}{c - \phi^{vw}}$ for λ_{ij} in Equation (8c), one obtains $\phi^{vw} = \frac{1}{k \frac{1}{c - \phi^{vw}}} = \frac{c - \phi^{vw}}{k}$ which, after collecting ϕ^{vw} , shows that $\phi^{vw} = \frac{1}{k+1}c$. The latter, combined with Equations (8a), allows us to conclude $\phi^{ij} = \frac{k}{k+1}c$. Overall, we have the following unique PF allocation:

$$\phi^{vw} = \frac{1}{k+1}c \quad \phi^{ij} = \frac{k}{k+1}c \quad \forall (i, j) \in A. \quad (9)$$

We remark that (9) illustrates a typical property of PF allocations: demands with longer routing paths such as

(v, w) are likely to receive a smaller bandwidth [24]. This is not the case for MMF allocations, as suggested by (7).

From (7) and (9), we deduce the following closed-form bandwidth allocations:

Problem	path	(v, w)	(a_i, a_{i+1})	(b_i, b_{i+1})
BTE-FBA-MMF	<i>up</i>	$\frac{1}{2}c$	$\frac{1}{2}c$	d
BTE-FBA-MMF	<i>down</i>	$\frac{1}{2}d$	c	$\frac{1}{2}d$
BTE-FBA-PF	<i>up</i>	$\frac{1}{k+1}c$	$\frac{k}{k+1}c$	d
BTE-FBA-PF	<i>down</i>	$\frac{1}{\ell+1}d$	c	$\frac{\ell}{\ell+1}d$

The total throughput with each of the two routings (*up* or *down*) and two fairness paradigms (MMF or PF) is thus:

$$up_{\text{MMF}} = \frac{1}{2}c + \frac{1}{2}kc + \ell d \quad (10a)$$

$$down_{\text{MMF}} = \frac{1}{2}d + kc + \frac{1}{2}\ell d \quad (10b)$$

$$up_{\text{PF}} = \frac{1}{k+1}c + \frac{k^2}{k+1}c + \ell d \quad (10c)$$

$$down_{\text{PF}} = \frac{1}{\ell+1}d + kc + \frac{\ell^2}{\ell+1}d. \quad (10d)$$

Let $\ell = c - 2$, $d = \frac{c}{10} + 5$, $k = d - 1 = \frac{c}{10} + 4$. With few algebraic manipulations, we obtain the following two utility function differences for the two fairness paradigms:

$$up_{\text{MMF}} - down_{\text{MMF}} = \frac{17}{20}c - \frac{15}{2} \quad (11a)$$

$$down_{\text{PF}} - up_{\text{PF}} = \frac{9c^3 + 193c^2 - 2500c + 7500}{10c^2 + 490c - 500}. \quad (11b)$$

Since both functions are strictly positive for $c \geq 9$, we deduce that, for $c \geq 9$, *up* is the unique optimal routing for BTE-FBA-MMF and *down* is the unique optimal routing for BTE-FBA-PF.⁶

In STE, since $\phi^{vw} = 0$ is feasible we have the following bandwidth allocations:

problem	path	(v, w)	(a_i, a_{i+1})	(b_i, b_{i+1})
STE	<i>up</i>	0	c	d
STE	<i>down</i>	0	c	d

Both routings *up* and *down* are optimal for STE. When coupled with an *a posteriori* fair allocation, their total throughput is $up_{\text{MMF}}, down_{\text{MMF}}$ for the MMF case and $up_{\text{PF}}, down_{\text{PF}}$ for the PF case. If *down* is selected as an optimal solution to STE, in the MMF case after the reallocation we have a throughput of $down_{\text{MMF}}$ which, when compared to the optimal throughput obtained with BTE-FBA-MMF, leads to a difference of $up_{\text{MMF}} - down_{\text{MMF}}$ as in Equation (11a). If *up* is selected, in the PF case after the reallocation we have a throughput of up_{PF} which, we compared to BTE-FBA-PF, leads to a difference of $down_{\text{PF}} - up_{\text{PF}}$ as in Equation (11b). To prove the claim of the theorem, it now suffices to observe that $\lim_{c \rightarrow \infty} up_{\text{MMF}} - down_{\text{MMF}} = \infty$ and $\lim_{c \rightarrow \infty} down_{\text{PF}} - up_{\text{PF}} = \infty$. \square

⁶For $c = \{30, 40, \dots, 90\}$ and $(c, d, k, \ell) \in \mathbb{N}^4$, it can be shown, empirically, that our choice of parameters also maximizes $\min\{up_{\text{MMF}} - down_{\text{MMF}}, down_{\text{PF}} - up_{\text{PF}}\}$. We omit the details for the sake of brevity.

For completeness, we report the following result concerning the worst-case difference in terms of network utility function between solutions to BTE-FBA-MMF or BTE-FBA-PF and solutions to MMF overall or PF overall, i.e., the problems where one looks for both a routing x and an allocation ϕ which are together MMF or PF.

Theorem 2. *The difference in terms of utility between solutions to BTE-FBA-MMF or BTE-FBA-PF and optimal solutions to MMF overall or PF overall can be arbitrarily large and it can increase with the number of demands.*

Proof. Consider the example in Figure 3. The graph contains $k = |D|$ origin-destination pairs, capacity 1 for the arcs (a_i, b_i) , $i = 2, \dots, k$, capacity $k(1 + \delta)$ for the arc (a_1, b_1) , with $\delta > 0$, and an arbitrarily large capacity for the other arcs. It is easy to verify that $\phi = (1, \dots, 1, k(1 + \delta))$, with a total throughput of $k(1 + \delta) + k - 1$, is an optimal solution to both BTE-FBA-MMF and BTE-FBA-PF. It can also be verified that $\phi = (1 + \delta, \dots, 1 + \delta)$, with a total throughput of $k(1 + \delta)$, is an optimal solution to both MMF overall and PF overall. Indeed, in the MMF-overall case, it is clear that, by routing any set of demands with indices $\ell \in D \setminus \{1\}$ over the corresponding paths $s_\ell - t_\ell$, such demands would receive an allocation smaller than $1 + \delta$ —and, thus, the solution would not be optimal. A similar observation holds also for the PF-overall case for a sufficiently large k . The difference in throughput between the optimal solution to BTE-FBA-MMF and BTE-FBA-PF and that to MMF overall and PF overall is equal to $k - 1$. The claim is obtained by driving k to infinity. \square

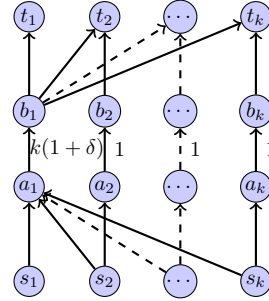


Fig. 3. Instance used to prove Theorem 2.

VI. SOLVING BTA-FBA-MMF

In this section, we derive a single-level Mixed-Integer Linear Programming (MILP) reformulation for BTE-FBA-MMF which allows for solving the problem with state-of-the-art MILP solvers.

A. Single level MILP formulation

To arrive at a computationally manageable single-level reformulation for BTE-FBA with an MMF allocation, we first reformulate the MMF optimality conditions in Proposition 1 so as to remove, from Constraints (4b) and (4d), the dependence on x_{ij}^d from the indices over which the summations run.

For every $(i, j) \in A$ and $d \in D$, we introduce the auxiliary flow variable f_{ij}^d and a set of constraints stemming from the well-known McCormick's envelope [35], Constraints (12e)–(12g), which guarantee that f_{ij}^d be equal to ϕ^d if $x_{ij}^d = 1$ and to 0 otherwise. This allows us to recast the MMF optimality conditions in Proposition 1 with linear constraints and mixed-integer variables:

Corollary 2. *A vector $\phi \geq \mathbf{0}$ is the unique MMF allocation for a given **unsplittable** single-path routing \mathbf{x} if and only if there are two vectors $\mathbf{y} \in \{0, 1\}^{|A| \times |D|}$, $\mathbf{f} \geq \mathbf{0}$ such that $(\phi, \mathbf{x}, \mathbf{y}, \mathbf{f})$ satisfies:*

$$\sum_{(i,j) \in A} y_{ij}^d \geq 1 \quad \forall d \in D \quad (12a)$$

$$\sum_{d' \in D} f_{ij}^{d'} \geq c_{ij} y_{ij}^d \quad \forall (i, j) \in A, d \in D \quad (12b)$$

$$\phi^d \geq \phi^{d'} - c_{ij}(1 - y_{ij}^d) \quad \forall (i, j) \in A, d, d' \in D \quad (12c)$$

$$\sum_{d \in D} f_{ij}^d \leq c_{ij} \quad \forall (i, j) \in A \quad (12d)$$

$$f_{ij}^d \leq c_{ij} x_{ij}^d \quad \forall (i, j) \in A, d \in D \quad (12e)$$

$$f_{ij}^d \leq \phi^d \quad \forall (i, j) \in A, d \in D \quad (12f)$$

$$f_{ij}^d \geq \phi^d - \max_{(i,j) \in A} \{c_{ij}\} (1 - x_{ij}^d) \quad \forall (i, j) \in A, d \in D. \quad (12g)$$

Proof. Constraints (12b), (12d) are a reformulation of Constraints (4b), (4d) written in terms of f_{ij}^d instead of ϕ^d and x_{ij}^d . Constraints (12e)–(12g), together with the nonnegativity of ϕ , are a slightly tighter version of the McCormick's envelope constraints applied to $f_{ij}^d = \phi^d x_{ij}^d$. They differ from the standard ones as Constraints (12e) are a tightened version of the original McCormick constraints $f_{ij}^d \leq \max_{(i,j) \in A} \{c_{ij}\} x_{ij}^d$. They are obtained by introducing the tighter upper bound c_{ij} on x_{ij}^d in lieu of the weaker upper bound $\max_{(i,j) \in A} \{c_{ij}\}$ on f_{ij}^d . \square

B. Subtour elimination for BTE-FBA-MMF

It is crucial to remark that Proposition 1 and Corollary 2 hold only under the assumption that \mathbf{x} encode an **unsplittable** single-path routing. Formally speaking, any routing which is feasible for Constraints (2b) and (2c) can be decomposed, for each demand $d \in D$, into an s^d – t^d path P^d carrying flow $f_{ij}^d = \phi^d$ for all $(i, j) \in P^d$, plus a (possibly nonempty) set of subtours S_1^d, \dots, S_τ^d , for some $\tau \in \mathbb{N} \cup \{0\}$. Each subtour S^d is a circuit $(v_1, v_2), \dots, (v_{k-1}, v_1)$ also carrying, due to Constraints (12e)–(12g), a flow $f_{ij}^d = \phi^d$ for all $(i, j) \in S^d$.

The example depicted in Figure 4 illustrates that, if a routing contains subtours, a bandwidth allocation satisfying the conditions in Proposition 1 and Corollary 2 does not need to be MMF.

Example 1. *The instance in Figure 4 admits a unique **unsplittable** single-path routing \mathbf{x} (without subtours) whose MMF allocation vector is $\phi_{\text{MMF}} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, with a total throughput of $\frac{3}{2}$. Assume that no subtour elimination constraints are imposed on \mathbf{x} and that $\bar{c} \in (0, \frac{1}{2})$. Consider demand $d = 3$ from node 1 to node 3. Assume that \mathbf{x}^3*

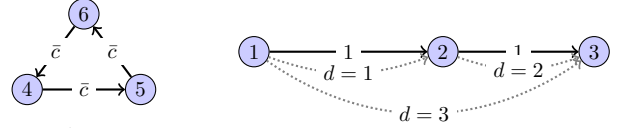


Fig. 4. An instance containing a subtour which, if not prevented from the routing \mathbf{x} , allows for an allocation ϕ which is not MMF in spite of satisfying the conditions in Proposition 1 or Corollary 2.

contains the 1–3 path given by $x_{12}^3 = x_{23}^3 = 1$, together with the subtour given by $x_{45}^3 = x_{56}^3 = x_{64}^3 = 1$, and let $\phi^3 = \bar{c}$, so that all subtour arcs are saturated. Let either arc (4, 5), (5, 6), or (6, 4) be bottleneck for demand $d = 3$, i.e., let either $y_{45}^3 = 1$, $y_{56}^3 = 1$, or $y_{64}^3 = 1$. This way, the same capacity \bar{c} is subtracted from the arcs (1, 2) and (2, 3), leaving a residual capacity of $1 - \bar{c}$ on both of them. Since $y_{12}^3 = y_{23}^3 = 1$, such capacity is fully allocated to demands $d = 1$ and $d = 2$. Overall, we obtain $\phi = (1 - \bar{c}, 1 - \bar{c}, \bar{c})$, with a total throughput of $2 - \bar{c}$, which is larger than $\frac{3}{2}$. Clearly ϕ is not MMF as its smallest allocation, of value \bar{c} , is smaller than the smallest allocation, of value $\frac{1}{2}$, of ϕ_{MMF} .

The example shows that, due to the bilevel nature of BTE-FBA-MMF, if subtours are not explicitly prevented one could include them in the routing \mathbf{x} so to force the second level problem towards a solution which, while having a higher utility function value, is not MMF in spite of satisfying the conditions in Proposition 1 and Corollary 2. This is in contrast with most of the classical single-level traffic engineering problems where subtours, due to consuming arc capacity without carrying any extra flow from the origins to the destinations, have no impact neither on the feasibility nor on the optimality of the solution and, hence, can be discarded in post-processing.

To prevent subtours, instead of adding Constraints (2d), whose number grows exponentially with the number of nodes, we derive a compact (polynomial size) formulation by introducing an additional set of variables and constraints. We achieve this by adapting the multicommodity flow extended formulation proposed in [36] for the Traveling Salesman Problem (TSP). Denoting by P^d the path used for demand $d \in D$, the idea is preventing subtours from arising by forcing the origin s^d to send an auxiliary *multicommodity flow* to each intermediate node visited by P^d . Let $V^d := V \setminus \{s^d\}$, let the binary variable ψ_h^d be equal to 1 if and only if path P^d visits node $h \in V^d$, and let the continuous variable q_{ijh}^d represent the auxiliary flow from s^d to $h \in V^d$ (equal to 0 if $\psi_h^d = 0$). Consider the following constraints, with $\mathbf{q} \in [0, 1]^{|D| \times |A| \times |V|}$ and $\psi \in \{0, 1\}^{|D| \times |V|}$:

$$\sum_{(i,h) \in A} x_{ih}^d = \psi_h^d \quad \forall d \in D, h \in V^d \quad (13a)$$

$$\sum_{(i,j) \in A} q_{ijh}^d - \sum_{(j,i) \in A} q_{jih}^d = \begin{cases} \psi_h^d & \text{if } i = s^d \\ -\psi_h^d & \text{if } i = h \\ 0 & \text{otherwise} \end{cases} \quad \forall d \in D, h \in V^d, i \in V \quad (13b)$$

$$q_{ijh}^d \leq x_{ij}^d \quad \forall d \in D, (i, j) \in A, h \in V^d \quad (13c)$$

$$q_{ijh}^d \leq \psi_h^d \quad \forall d \in D, (i, j) \in A, h \in V^d. \quad (13d)$$

Constraints (13a) guarantee that ψ_h^d be equal to 1 if and only if the path for demand $d \in D$ visits node $h \in V^d$. Constraints (13b) are flow conservation constraints on the multicommodity flow q_{ijh}^d from s^d to $h \in V^d$, carrying one unit of flow from s^d to $h \in V^d$ if $\psi_h^d = 1$. Constraints (13c) prevent the flow vector $\mathbf{q}_h^d = (q_{ijh}^d)_{(i,j) \in A}$ from using any arcs not present in the routing specified by the routing vector \mathbf{x}^d for the corresponding demand $d \in D$. Constraints (13d) prevent the existence of any multicommodity flow from s^d to $h \in V^d$ if $\psi_h^d = 0$, i.e., whenever the path P^d does not visit node h in the routing encoded by \mathbf{x} .

With this extended formulation, we have all the ingredients for a polynomial-size single-level MILP formulation for BTE-FBA-MMF.

C. Complete single-level MILP reformulation

To tighten the linear programming relaxations of our single-level reformulation for BTE-FBA-MMF, we introduce the following valid (in)equalities:

$$\sum_{(i,j) \in A} f_{ij}^d - \sum_{(j,i) \in A} f_{ji}^d = \begin{cases} \phi^d & \text{if } i = s^d \\ -\phi^d & \text{if } i = t^d \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in V, d \in D \quad (14a)$$

$$\phi^d \geq \frac{\min_{(i,j) \in A} \{c_{ij}\}}{|D|} \quad \forall d \in D. \quad (14b)$$

Constraints (14a) are flow conservation constraints which are clearly valid for \mathbf{f} due to the latter being componentwise equal to $\phi^d x_{ij}^d$ and \mathbf{x} itself being subject to flow conservation constraints. Constraints (14b) are valid inequalities imposing a lower bound on f_{ij}^d tighter than 0, achieved in the worst case where all the demands in D are simultaneously routed over the same arc $(i, j) \in A$.

To reduce the overall number of constraints without hindering the tightness of the formulation, we **introduce $\mathbf{u} \geq \mathbf{0}$ and reformulate Constraints (12c) as:**

$$u_{ij} \geq f_{ij}^d \quad \forall (i, j) \in A, d \in D \quad (15a)$$

$$f_{ij}^d \geq u_{ij} - c_{ij}(1 - y_{ij}^d) \quad \forall (i, j) \in A, d \in D, \quad (15b)$$

where u_{ij} is a new continuous variable accounting, for each $(i, j) \in A$, for the flow of the largest demand carried by link $(i, j) \in A$. The inequalities are obtained by multiplying the left and right-hand sides of Constraint (12c) by x_{ij}^d and substituting f_{ij}^d for $\phi^d x_{ij}^d$. This way, constraint $\phi^d \geq \phi^{d'} - c_{ij}(1 - y_{ij}^d)$ is restated as $f_{ij}^d \geq f_{ij}^{d'} - c_{ij}(1 - y_{ij}^d)$ for all $d, d' \in D, (i, j) \in A$. This allows for a smaller formulation: while the number of Constraints (12c) is quadratic in the number of demands, the number of Constraints (15a) and (15b) is only linear in the number of demands.

The complete single-level MILP formulation for BTE-FBA-MMF consists of (2a)–(2c), (12a)–(12g), (13a)–(13d), (14a)–(14b), with (12c) substituted for (15a)–(15b). We report it in its entirety in Appendix A.

VII. SOLVING BTE-FBA-PF

In this section, we show how to derive a single-level bilinear reformulation for BTE-FBA-PF and, then, an approximate, linearized MILP reformulation.

A. Reformulation of the PF optimality conditions

To bring the optimality conditions for a PF allocation (Proposition 2) into a more suitable form, we proceed as for BTE-FBA-MMF by introducing a variable f_{ij}^d for each $(i, j) \in A$ and $d \in D$, and a set of constraints guaranteeing $f_{ij}^d = \phi^d x_{ij}^d$ in order to remove the dependence on x_{ij}^d from the index sets in Constraints (6a)–(6c). We recast Proposition 2 as follows:

Corollary 3. *A vector $\phi \geq \mathbf{0}$ is the unique PF allocation for a given **unsplittable** single-path routing \mathbf{x} if and only if there are two vectors $\lambda, \mathbf{f} \geq \mathbf{0}$ such that $(\phi, \mathbf{x}, \lambda, \mathbf{f})$ satisfies:*

$$\sum_{(i,j) \in A} f_{ij}^d \lambda_{ij} = 1 \quad \forall d \in D \quad (16a)$$

$$\left(\sum_{d \in D} f_{ij}^d - c_{ij} \right) \lambda_{ij} = 0 \quad \forall (i, j) \in A \quad (16b)$$

$$\sum_{d \in D} f_{ij}^d \leq c_{ij} \quad \forall (i, j) \in A \quad (16c)$$

$$f_{ij}^d \leq c_{ij} x_{ij}^d \quad \forall (i, j) \in A, d \in D \quad (16d)$$

$$f_{ij}^d \leq \phi^d \quad \forall (i, j) \in A, d \in D \quad (16e)$$

$$f_{ij}^d \geq \phi^d \max_{(i,j) \in A} \{c_{ij}\} (1 - x_{ij}^d) \quad \forall (i, j) \in A, d \in D. \quad (16f)$$

Proof. Constraints (16a) and (16b) are obtained from Constraints (6a) and (6b) by restating the latter as, respectively, $\phi^d \left(\sum_{(i,j) \in A} x_{ij}^d \lambda_{ij} \right) = 1$ and $\left(\sum_{d \in D} x_{ij}^d \phi^d - c_{ij} \right) \lambda_{ij} = 0$ and, then, substituting f_{ij}^d for $\phi^d x_{ij}^d$. Constraints (16c) are a reformulation of Constraints (6c). Constraints (16d)–(16f), together with the nonnegativity of \mathbf{f} , are McCormick envelope constraints guaranteeing $f_{ij}^d = \phi^d x_{ij}^d$. Due to x_{ij}^d being a binary variable, these constraints yield an exact reformulation of the bilinear constraint $f_{ij}^d = \phi^d x_{ij}^d$. \square

Note that the optimality conditions in Corollary 3 are nonlinear and nonconvex due to the presence of the bilinear products $f_{ij}^d \lambda_{ij}$ in Constraints (16a) and (16b). Before elaborating on how to handle this nonlinearity, let us discuss on the impact of subtours.

B. Subtour elimination in BTE-FBA-PF

As for BTE-FBA-MMF, Proposition 2 and Corollary 3 hold only under the assumption that \mathbf{x} encode an **unsplittable** single-path routing. Thus, a correct BTE-FBA-PF formulation must explicitly prevent subtours. Consider the following example.

Example 2. *The instance in Figure 4 admits a unique **unsplittable** single-path routing \mathbf{x} (without subtours) whose PF allocation is $\phi_{PF} = (1 - \frac{1}{3}, 1 - \frac{1}{3}, \frac{1}{3})$, with a total throughput of $2 - \frac{1}{3}$ (see Appendix C). Let $\bar{c} < \frac{1}{3}$. Consider*

demand $d = 3$ from node 1 to node 3. Assume that \mathbf{x}^3 contains the 1–3 path given by $x_{12}^3 = x_{23}^3 = 1$, together with the subtour given by $x_{45}^3 = x_{56}^3 = x_{64}^3 = 1$. Due to the conditions in Proposition 2 (see Appendix C), we deduce the allocation $\phi = (1 - \bar{c}, 1 - \bar{c}, \bar{c})$, with a larger total throughput $2 - \bar{c} > 2 - \frac{1}{3}$. Such allocation is not PF though, as $\log(1 - \bar{c}) + \log(1 - \bar{c}) + \log(\bar{c}) < \log(1 - \frac{1}{3}) + \log(1 - \frac{1}{3}) + \log(\frac{1}{3})$ for $c < \frac{1}{3}$.

Although we could adapt to BTE-FBA-PF the same subtour elimination technique we used for BTE-FBA-MMF, we propose here an alternative (and almost surprising) way of preventing subtours which entirely relies on the bilevel nature of BTE-FBA-PF. For the purpose, we introduce the following alternative version of Problem (5) (the PF allocation problem):

$$(\phi, \mathbf{f}) \in \underset{\phi, \mathbf{f} \geq 0}{\operatorname{argmax}} \left\{ \begin{array}{l} \sum_{d \in D} \log(\phi^d) : \\ \sum_{(i,j) \in A} f_{ij}^d - \sum_{(j,i) \in A} f_{ji}^d = \begin{cases} \phi^d & \text{if } i = s^d \\ -\phi^d & \text{if } i = t^d \\ 0 & \text{otherwise} \end{cases} \\ \forall i \in V, d \in D \\ \sum_{d \in D} f_{ij}^d \leq c_{ij} \quad \forall (i, j) \in A \\ f_{ij}^d \leq c_{ij} x_{ij}^d \quad \forall (i, j) \in A, d \in D \end{array} \right. \quad (17)$$

This problem is a modified version of Problem (5) in which the second-level asks for determining not only ϕ but also \mathbf{f} . It is crucial to observe that, in this version, only three of the original four McCormick constraints are imposed on \mathbf{f} . While Constraints (16d) and the nonnegativity of \mathbf{f} are imposed explicitly and Constraints (16e) are implied by the new flow balance constraints on \mathbf{f} , Constraints (16f) are not imposed at all. This way, while f_{ij}^d is still constrained to be 0 if $x_{ij}^d = 0$ and to be equal to ϕ^d on the $s^d - t^d$ path P^d for all $(i, j) \in A$ and $d \in D$, by not imposing Constraints (16f) its value is not forced to be equal to ϕ^d on each subtour arc. This leads to the following result:

Proposition 3. *Given a routing \mathbf{x} possibly containing subtours, any optimal solution to Problem (17) ϕ coincides with the unique PF allocation for the unsplittable single-path routing \mathbf{x}' obtained by removing all subtours from \mathbf{x} .*

Proof. The claim is clearly satisfied if $f_{ij}^d = 0$ for every (i, j) belonging to a subtour. Assume that, for some $d \in D$, $f_{ij}^d > 0$ for some arc (i, j) belonging to a subtour S^d . Two cases can arise. If $x_{ij}^{d'} = 0$ for all $d' \in D \setminus \{d\}$, i.e., the arc (i, j) is only present in the subtour S^d , we can set $f_{ij}^d = 0$ for all arcs in S^d without affecting ϕ since subtours carry no additional flow from s^d to t^d . If $x_{ij}^{d'} = 1$ for some $d' \in D \setminus \{d\}$, i.e., the arc (i, j) is contained in some path $P^{d'}$, the value of $\phi^{d'}$ is limited by a tighter residual capacity on another arc $(i', j') \in P^{d'}$. This is because, if not, $f_{ij}^{d'}$ limits $\phi^{d'}$ to $c_{ij} - f_{ij}^{d'}$ and, by reducing the value of $f_{ij}^{d'}$ and, consequently, the flow on subtour S^d (always possible, as subtours carry no additional flow

from s^d to t^d), $f_{ij}^{d'}$ and $\phi^{d'}$ can be increased, leading to a larger value of $\sum_{d \in D} \log(\phi^d)$, thus contradicting the optimality of the solution to Problem (17). Therefore, if a subtour is present, it can be removed without affecting the bandwidth allocation ϕ . The argument can be easily extended to the case with multiple subtours. \square

It follows that if, for a given demand d , $f_{ij}^d > 0$ for all arcs (i, j) of a subtour, then the subtour is disjoint from any path $P^{d'}$ for all $d' \setminus \{d\}$ and, hence, we can set $f_{ij}^d = 0$ for all the subtour arcs without loss of generality.

C. Approximate MILP reformulation

In principle, one could state the KKT conditions for Problem (17) and derive from them an exact single-level reformulation for BTE-FBA-PF. Such a reformulation, however, would contain many bilinear terms involving continuous variables. Preliminary experiments revealed that, with this approach, only instances with up to 5 nodes are computationally tractable.

To tackle larger-size instances, we propose an approximate MILP reformulation for BTE-FBA-PF based on a piecewise-affine approximation of the term $\log(\phi^d)$ appearing in the second-level objective function $\sum_{d \in D} \log(\phi^d)$. We consider a piecewise-affine approximation with $|H|$ pieces, with index set H . For each affine piece $h \in H$, let α_h denote its slope and β_h its offset. See Figure 5 for an illustration.

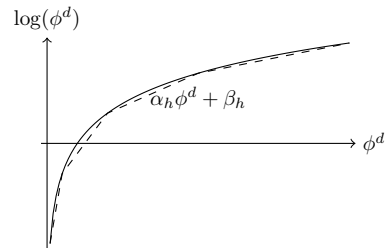


Fig. 5. Piecewise-affine approximation of $\log(\phi^d)$.

First, we approximate Problem (17) by imposing that (ϕ, \mathbf{f}) be an optimal solution to the following Linear Programming (LP) problem:

$$\max_{\eta} \sum_{d \in D} \eta^d \quad (18a)$$

$$\mathbf{f}, \phi \geq 0 \quad \text{s.t.} \quad \eta^d \leq \alpha_h \phi^d + \beta_h \quad \forall d \in D, h \in H \quad (18b)$$

$$\sum_{(i,j) \in A} f_{ij}^d - \sum_{(j,i) \in A} f_{ji}^d = \begin{cases} \phi^d & \text{if } i = s^d \\ -\phi^d & \text{if } i = t^d \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in V, d \in D \quad (18c)$$

$$\sum_{d \in D} f_{ij}^d \leq c_{ij} \quad \forall (i, j) \in A \quad (18d)$$

$$f_{ij}^d \leq c_{ij} x_{ij}^d \quad \forall (i, j) \in A, d \in d. \quad (18e)$$

For each $h \in H$, Constraints (18b) ensure that η^d be upper bounded by the affine function $\alpha_h \phi^d + \beta_h$ corresponding

to the h -th piece. Since, in any optimal solution, $\eta^d = \min_{h \in H} \{\alpha_h \phi^d + \beta_h\}$ for each d , the minimum of $\sum_{d \in D} \eta^d$ corresponds to that of $\sum_{d \in D} \phi^d$ up to the approximation error which is due to the linearization.

We derive the corresponding optimality conditions:

Proposition 4. *For $M \rightarrow \infty$, $(\boldsymbol{\eta}, \mathbf{f}, \boldsymbol{\phi})$ with $\mathbf{f}, \boldsymbol{\phi} \geq \mathbf{0}$ is optimal for Problem (18) if and only if there are six vectors $\boldsymbol{\lambda}, \boldsymbol{\kappa}, \boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \mathbf{v} \geq \mathbf{0}$ such that the following set of linear inequalities are satisfied:*

$$\sum_{d \in D} \eta^d + M \sum_{(i,j) \in A} \sum_{d \in D} \gamma_{ij}^d \geq$$

$$\sum_{d \in D} \sum_{h \in H} \beta_h \lambda_h^d + \sum_{(i,j) \in A} c_{ij} \mu_{ij} + \sum_{(i,j) \in A} \sum_{d \in D} c_{ij} v_{ij}^d \quad (19a)$$

$$\eta^d \leq \alpha_h \phi^d + \beta_h \quad \forall d \in D, h \in H \quad (19b)$$

$$\sum_{(i,j) \in A} f_{ij}^d - \sum_{(j,i) \in A} f_{ji}^d = \begin{cases} \phi^d & \text{if } i = s^d \\ -\phi^d & \text{if } i = t^d \quad \forall i \in V, d \in D \\ 0 & \text{otherwise} \end{cases} \quad (19c)$$

$$\sum_{d \in D} f_{ij}^d \leq c_{ij} \quad \forall (i,j) \in A \quad (19d)$$

$$f_{ij}^d \leq c_{ij} x_{ij}^d + \gamma_{ij}^d \quad \forall (i,j) \in A, d \in D \quad (19e)$$

$$\sum_{h \in H} \lambda_h^d \geq 1 \quad \forall d \in D \quad (19f)$$

$$\kappa_i^d - \kappa_j^d + \mu_{ij} + \theta_{ij}^d \geq 0 \quad \forall (i,j) \in A, d \in D \quad (19g)$$

$$\kappa_{td}^d - \kappa_{sd}^d \geq \sum_{h \in H} \alpha_h \lambda_h^d \quad \forall d \in D \quad (19h)$$

$$v_{ij}^d \leq M x_{ij}^d \quad \forall (i,j) \in A, d \in D \quad (19i)$$

$$v_{ij}^d \leq \theta_{ij}^d \quad \forall (i,j) \in A, d \in D \quad (19j)$$

$$v_{ij}^d \geq \theta_{ij}^d + M(x_{ij}^d - 1) \quad \forall (i,j) \in A, d \in D. \quad (19k)$$

Proof. We establish the optimality of $(\boldsymbol{\eta}, \mathbf{f}, \boldsymbol{\phi})$ by strong LP duality. In order to do this, we first construct the LP dual of Problem (18). Let, for the corresponding indices, $\lambda_h^d, \kappa_i^d, \mu_{ij}$, and θ_{ij}^d be the dual variables of, respectively, Constraints (18b), (18c), (18d), and (18e). The LP dual of Problem (18) reads:

$$\min_{\substack{\boldsymbol{v}, \boldsymbol{\lambda} \\ \boldsymbol{\kappa}, \boldsymbol{\mu}, \boldsymbol{\theta} \geq \mathbf{0}}} \sum_{d \in D} \sum_{h \in H} \beta_h \lambda_h^d + \sum_{(i,j) \in A} c_{ij} \mu_{ij} + \sum_{(i,j) \in A} \sum_{d \in D} c_{ij} x_{ij}^d \theta_{ij}^d \quad (20a)$$

$$\text{s.t.} \quad \sum_{h \in H} \lambda_h^d \geq 1 \quad \forall d \in D \quad (20b)$$

$$\kappa_i^d - \kappa_j^d + \mu_{ij} + \theta_{ij}^d \geq 0 \quad \forall (i,j) \in A, d \in D \quad (20c)$$

$$\kappa_{td}^d - \kappa_{sd}^d \geq \sum_{h \in H} \alpha_h \lambda_h^d \quad \forall d \in D. \quad (20d)$$

where Constraints (20b), (20c), and (20d) are, for the appropriate indices, the dual constraints of, respectively, variables f_{ij}^d, ϕ^d , and η^d .

Since both x_{ij}^d and θ_{ij}^d are variables, the last term of the dual objective function is bilinear. To avoid this nonlinearity, we first introduce the dual constraints $\theta_{ij}^d \leq M$ for all $(i,j) \in A, d \in D$. Their effect is threefold: new primal slack variables γ_{ij}^d are introduced for each $(i,j) \in A$ and $d \in D$, each primal constraint $f_{ij}^d \leq c_{ij} x_{ij}^d$ is relaxed into

$f_{ij}^d \leq c_{ij} x_{ij}^d + \gamma_{ij}^d$ (see Constraints (19e)), and the new slack variables γ_{ij}^d are penalized in the primal objective function via the additive term $M \sum_{(i,j) \in A} \sum_{d \in D} \gamma_{ij}^d$. This way, the larger M , the larger the penalty on γ_{ij}^d for all $(i,j) \in A$ and $d \in D$, and the less likely it is for any of constraints $f_{ij}^d \leq c_{ij} x_{ij}^d$ to be violated in an optimal solution.

With θ_{ij}^d upperbounded by M , we now replace each bilinear product $x_{ij}^d \theta_{ij}^d$ with a new variable v_{ij}^d , imposing $v_{ij}^d = x_{ij}^d \theta_{ij}^d$ for each $(i,j) \in A, d \in D$ via the McCormick envelope constraints, namely, Constraints (19i)–(19k), together with the nonnegativity of v_{ij}^d .

The claim is proven by equating the objective functions of the primal problem (extended after the introduction of γ_{ij}^d) and of the dual problem (with v_{ij}^d in lieu of $x_{ij}^d \theta_{ij}^d$) and adopting a sufficiently large value for M , thereby guaranteeing a penalty sufficiently large on $M \sum_{(i,j) \in A} \sum_{d \in D} \gamma_{ij}^d$ to drive γ_{ij}^d to 0 for all $(i,j) \in A$ and $d \in D$. \square

The complete approximate MILP formulation for BTE-FBA-PF, which we report in its entirety in Appendix B, consists of: (2a)–(2c), and (19).

VIII. COMPUTATIONAL RESULTS

We present a set of computational results aimed at assessing the solvability of our MILP reformulations for BTE-FBA-MMF and BTE-FBA-PF as well as the advantages in terms of utility function when solving these problems *in lieu* of the traditional single-level approach discussed in Section V, which is oblivious of the bilevel nature of the TE problem. **Without loss of generality**, all the results are obtained by solving the **extension** of BTE-FBA-MMF and BTE-FBA-PF **introduced in Subsection IV-E**, with Δ^d sessions associated with each origin-destination pair $d \in D$.

A. Instances and setup

The instances in our testbed are based on 3 network topologies taken from the widely used SNDLIB library [37]: **polska** ($|V| = 12$ and $|A| = 36$), **nobel-us** ($|V| = 14$ and $|A| = 42$), and **nobel-ger** ($|V| = 17$ and $|A| = 52$). In the library, each instance is associated with a set of *edge* nodes $E \subseteq V$. For each topology, we sample 8 subsets $E' \subseteq E$ of edge nodes of increasing cardinality and introduce an origin-destination pair for every pair of nodes $i, j \in E'$ which have been sampled, thus obtaining instances with an increasing number of demands. We assign 3 different sets (cap1, cap2, cap3) of random capacities to the arcs, each taking a value (in Gbps) among 2, 2.4, 5, 8, sampled with a probability of, respectively, 0.2, 0.2, 0.3, and 0.3. We assign, with each instance, a number Δ^d of sessions per origin-destination pair, computed by sampling from a discrete uniform distribution over $[1, 2^{tr}]$, with values of tr in 1, 4, 7, 9, 10. Thus, we generate an instance for each value of tr , for each topology, and for each set of capacities. The utility function weights w^d take values among 1, 2, 3 with a probability equal to,

respectively, 0.25, 0.5, 0.25. Overall, we consider a testbed of 270 instances. In all the experiments, we set $M = 10000$.

The experiments are carried out on a Dell PowerEdge Quad Core Xeon 2.0 GHz with 16 GB of RAM, with a time limit of 2 hours per instance. We employ AMPL as modeling language, relying on CPLEX 12.6 as MILP solver. When needed (as described in the following), we use IPOPT 3.11 to determine, *a posteriori* and for a given routing, a PF allocation by solving Problem (5).

B. Solution approaches

Based on the analysis and discussion in Section V, we compare the results obtained with four different methods. We adopt the notation **F+R**, with **F** as in “formulation” and **R** as in “reallocation”. The latter denotes if the bandwidth is *reallocated a posteriori* according to either the MMF or the (exact) PF paradigm.

Overall, we consider the following four methods:

- **BTE-FBA-MMF**: exact MILP reformulation for BTE-FBA-MMF; no reallocation needed;
- **BTE-FBA-APF**: approximate MILP reformulation for BTE-FBA-PF, without exact reallocation;
- **BTE-FBA-APF+PF**: approximate MILP reformulation for BTE-FBA-PF with an exact PF reallocation (computed *a posteriori* via IPOPT);
- **STE+MMF** and **STE+PF**: MILP formulations for the single-level traffic engineering problem (STE) with an exact MMF or PR reallocation.

Note that, due to the linearization that we use, the allocation of **BTE-FBA-APF** may not be exactly PF, which justifies computing an exact PF allocation *a posteriori*.

C. Tuning the linearization in BTE-FBA-APF

It is clear that a good piecewise-affine approximation of $\sum_{d \in D} \log(\phi^d)$ is crucial in order to obtain good approximations of BTE-FBA-PF with **BTE-FBA-APF**.

In our computational experiments, we employ a piecewise-affine approximation with $|H| = 20$ pieces constructed in an iterative way. First, we solve the LP relaxation of **BTE-FBA-APF** with a piecewise-affine approximation with $|H| = 10$ pieces with predetermined parameters α_h, β_h for all $h \in H$. From its solution, we let ϕ_0 be equal to the smallest value taken by ϕ^d for all $d \in D$, and ϕ_{20} to the largest one. We then partition the interval $[\phi_0, \phi_{20}]$ into $|H| = 20$ equally-sized subintervals $[\phi_{h-1}, \phi_h]$, for all $h \in H$, and compute for each subinterval a pair of parameters (α_h, β_h) such that the affine function $\alpha_h \phi + \beta_h$ satisfies $\alpha_h \phi_{h-1} + \beta_h = \log(\phi_{h-1})$ and $\alpha_h \phi_h + \beta_h = \log(\phi_h)$.

Overall, when applying (*a posteriori*, with IPOPT) an exact PF allocation to solutions obtained with **BTE-FBA-APF** (solved with CPLEX), we register a loss in utility function smaller than 3% on average. We also observe that the fairness measure $\sum_{d \in D} \log(\phi^d)$ is almost identical for the solutions obtained with **BTE-FBA-APF** and **BTE-FBA-APF+PF**, with a loss in $\sum_{d \in D} \log(\phi^d)$ of less than 0.6% on average. This suggests that the accuracy of our piecewise approximation is quite high.

D. Results

Table I summarizes the results obtained with the different methods for the three topologies, as a function of the number of edge nodes $|E|$. In particular, it reports, for each topology and value of $|E|$, the average over the 15 instances with that topology and number of edge nodes in the data set. The full set of results is reported, without aggregation, in Appendix D, in Tables II, III, and IV.

TABLE I
RESULTS AGGREGATED BY TOPOLOGY AND NUMBER OF EDGE NODES $|E|$, AVERAGES OVER 15 INSTANCES.

inst.	$ E $	BTE-FBA-MMF	BTE-FBA-APF	BTE-FBA-MMF	STE+MMF	BTE-FBA-APF+PF	STE+PF
		gap	gap	utility	utility	utility	utility
Polksa	7	0.06	0.09	12197.6	8301.5	11537.2	8485.8
	8	0.06	0.08	18987.9	10896.4	18129.3	11293.6
	9	0.06	0.10	23395.0	10768.2	21953.2	11240.3
	10	0.09	0.08	32491.3	18433.8	31612.9	18693.3
	11	0.11	0.13	40653.9	19988.8	39168.0	20009.7
	12	0.13	0.18	43167.6	21458.8	40489.9	22526.1
avg		0.08	0.11	28482.2	14974.6	27148.4	15374.8
Nobel-us	7	0.06	0.03	20207.2	15534.4	19932.8	15830.5
	9	0.18	0.09	26902.3	20782.5	27546.3	19787.9
	11	0.61	0.25	33371.9	27454.9	40925.2	27163.9
	12	0.69	0.29	35980.7	27519.8	44944.6	27884.2
	13	0.69	0.28	35973.8	27519.8	45409.5	27884.2
	14	0.86	0.37	34984.0	24904.9	44916.2	25854.6
avg		0.51	0.22	31236.6	23952.7	37279.1	24067.6
Nobel-ger	8	0.05	0.04	23443.8	17461.5	23166.1	17816.7
	10	0.06	0.08	38516.1	26865.5	36855.7	27212.1
	12	0.11	0.15	42231.8	29274.2	39074.0	29681.0
	14	0.29	0.21	42523.3	23540.7	42844.9	24509.9
	16	0.32	0.26	50850.9	30471.3	52082.7	31134.9
	17	0.32	0.29	56167.3	37464.9	54852.7	38346.8
avg		0.19	0.17	42288.9	27513.0	41479.3	28116.9
avg		0.26	0.17	34002.6	22146.8	35302.3	22519.7

The third and fourth columns of Table I indicate the average optimality gaps obtained with CPLEX within the time limit when solving the **BTE-FBA-MMF** or the **BTE-FBA-APF** reformulations. The columns show that, in the time limit, they can be solved to within an optimality gap of, on average, 26% for **BTE-FBA-APF** and 17% for **BTE-FBA-MMF**. The optimality gap increases with $|E|$ and, consequently, with the number of origin-destination pairs. The largest gaps are registered for the **Nobel-US** instance with $|E| = 14$, where they reach 86% for **BTE-FBA-MMF** and 37% for **BTE-FBA-APF**. In particular, among the two MILP reformulations, **BTE-FBA-MMF** turns out to be the hardest to solve. This is likely due to two factors. On the one hand, it requires subtour elimination via the extended formulation, which is not the case in **BTE-FBA-APF**, as shown in Proposition 3. On the other hand, in **BTE-FBA-MMF**, the optimality conditions for the second-level MMF problem involve binary variables, differently from those for the PF problem in **BTE-FBA-APF**, which only involves continuous variables.

As to solution quality, Table I shows that the difference in utility between a single-level and a bilevel solution can be extremely large in spite of the optimality gap. Indeed, **BTE-FBA-MMF** yields, in the MMF case, routings which

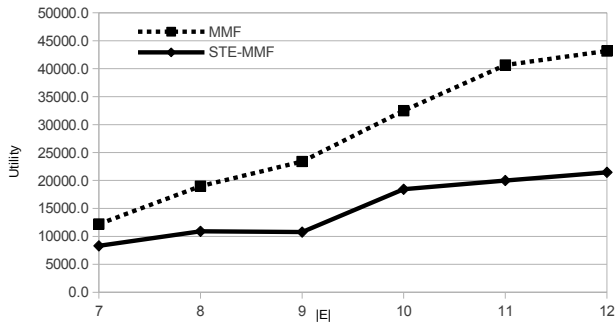


Fig. 6. Difference in utility between solutions obtained with BTE-FBA-MMF and STE+MMF for the Polska instances, as a function of $|E|$.

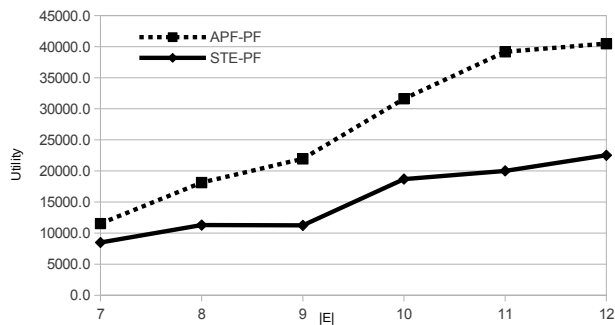


Fig. 7. Difference in utility between solutions obtained with BTE-FBA-APF+PF and STE+PF for the Polska instances, as a function of $|E|$.

are, on average, 54% better than those obtained with STE+MMF. In the PF case, the improvement is of 57%. In particular, the improvement measured on instances of a given topology can be substantially larger such as, e.g., on the Polska instances, where it is, on average, of 90% in the MMF case and of 77% in the PF case. For an illustration, see Figures 6 and 7.

Overall, the results show that, by explicitly accounting for elastic traffic and for the Stackelberg nature of the corresponding TE problem, the bilevel programming approaches that we proposed in this paper significantly outperform the single-level ones which ignore the impact of TCP as a second-level player. It is worth pointing out that, even for instances where the optimality gap is quite large (such as Nobel-us, where the average gap is above 60% with BTE-FBA-MMF and above 24% with BTE-FBA-APF), the solutions obtained by solving our reformulations of the BTE-FBA-MMF and BTE-FBA-PF problems have much higher utility function values (respectively higher by, on average, 30% and 58%) than those found with the single-level methods STE+MMF and STE+PF. This shows the large potential of the bilevel method, suggesting that better quality solutions (with smaller gaps) could yield even higher total utility gains.

IX. CONCLUDING REMARKS

We have proposed and investigated a new approach to traffic engineering with elastic demands where the interaction between the network operator and the congestion control mechanism is modeled as a Stackelberg game. For the corresponding equilibrium-finding bilevel-optimization problems BTE-FBA-MMF and BTE-FBA-PF, we have presented exact and approximate single-level MILP reformulations which can be solved with state-of-the-art mathematical programming algorithms. We have quantified the worst-case difference in utility function between the solutions to our bilevel traffic-engineering problems and those obtained with standard single-level methods oblivious of the bilevel nature of the problem. The computational results obtained for different network topologies and instance sizes indicate that even feasible solutions to our bilevel traffic engineering problems with an optimality gap as high as 80% can yield substantially higher utility function values (weighted network throughput) than those obtained by solving a standard single-level traffic engineering problem and *a posteriori* fairly reallocating the bandwidth according to the MMF or PF paradigms. This suggests that our Stackelberg-game approach and bilevel programming methods constitute a promising step towards the development of traffic engineering techniques for elastic demands.

REFERENCES

- [1] M. Pióro and D. Medhi, *Routing, Flow and Capacity Design in Communication and Computer Networks*. Morgan Kaufman, 2004.
- [2] J. Garcia-Dorado, A. Finamore, M. Mellia, M. Meo, and M. Munafò, “Characterization of ISP traffic: Trends, user habits, and access technology impact,” *IEEE Transactions on Network and Service Management*, vol. 9, no. 2, pp. 142–155, 2012.
- [3] F. Kelly, A. Maulloo, and D. Tan, “Rate control for communication networks: shadow prices, proportional fairness and stability,” *Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, Apr. 1998.
- [4] S. Low, “A duality model of TCP and queue management algorithms,” *IEEE/ACM Transactions on Networking*, vol. 11, no. 4, pp. 525–536, Aug. 2003.
- [5] L. Massoulié and J. Roberts, “Bandwidth sharing: objectives and algorithms,” *IEEE/ACM Transactions on Networking*, vol. 10, no. 3, pp. 320–328, 2002.
- [6] C. Chang, Z. Liu, and S. Member, “A bandwidth sharing theory for a large number of HTTP-like connections,” in *Proc. of Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2, no. 5. IEEE, 2004, pp. 667–675.
- [7] E. Altman, C. Barakat, E. Laborde, P. Brown, and D. Collange, “Fairness analysis of TCP/IP,” in *Decision and Control, 2000. Proc. of the 39th IEEE Conference on*, 2000, pp. 61–66.
- [8] G. Hasegawa and M. Murata, “Survey on fairness issues in TCP congestion control mechanisms,” *IEICE Transactions on Communications*, 2001.
- [9] “ITU-T G.984.1 Gigabit-capable passive optical networks (GPON): General characteristics,” <http://www.itu.int/rec/T-REC-G.984.1-200803-I/en>, March 2008.
- [10] D. Breuer and F. Geilhardt, “Opportunities for next-generation optical access,” *IEEE Communications Magazine*, vol. 2, no. 49, pp. 16–24, 2011.
- [11] E. Amaldi, S. Coniglio, and L. Taccari, “Maximum throughput network routing subject to fair flow allocation,” in *International Symposium on Combinatorial Optimization*. Springer, 2014, pp. 1–12.

- [12] E. Amaldi, A. Capone, S. Coniglio, and L. Gianoli, "Energy-aware traffic engineering with elastic demands and MMF bandwidth allocation," in *2013 IEEE 18th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*. IEEE, sep 2013, pp. 169–174.
- [13] E. Amaldi, S. Coniglio, L. Gianoli, and C. Ileri, "On single-path network routing subject to max-min fair flow allocation," *Electronic Notes in Discrete Mathematics*, vol. 41, pp. 543–550, 2013.
- [14] E. Amaldi, A. Capone, S. Coniglio, and L. Gianoli, "Network optimization problems subject to max-min fair flow allocation," *Communications Letters, IEEE*, vol. 17, no. 7, pp. 1463–1466, 2013.
- [15] A. Altin, B. Fortz, M. Thorup, and H. Ümit, "Intra-domain traffic engineering with shortest path routing protocols," *4OR*, vol. 7, no. 4, pp. 301–335, dec 2009.
- [16] N. Wang, K. Ho, G. Pavlou, and M. Howarth, "An overview of routing optimization for internet traffic engineering," *IEEE Communications Surveys & Tutorials*, vol. 10, no. 1, 2008.
- [17] D. Nace and M. Pioro, "Max-min fairness and its applications to routing and load-balancing in communication networks: a tutorial," *IEEE Communications Surveys & Tutorials*, vol. 10, no. 4, pp. 5–17, 2008.
- [18] T. Bonald, L. Massoulié, A. Proutière, and J. Virtamo, "A queueing analysis of max-min fairness, proportional fairness and balanced fairness," *Queueing Systems*, vol. 53, no. 1-2, pp. 65–84, 2006.
- [19] S. Sarkar and L. Tassiulas, "Fair allocation of utilities in multirate multicast networks: a framework for unifying diverse fairness objectives," *IEEE Transactions on Automatic Control*, vol. 47, no. 6, pp. 931–944, 2002.
- [20] T. Bonald and L. Massoulié, "Impact of fairness on internet performance," in *Proc. of the 2001 ACM SIGMETRICS international conference on Measurement and modeling of computer systems - SIGMETRICS '01*. New York, New York, USA: ACM Press, 2001, pp. 82–91.
- [21] L. Massoulié and J. Roberts, "Bandwidth sharing and admission control for elastic traffic," *Telecommunication Systems*, vol. 15, no. 1-2, pp. 185–201, 2000.
- [22] T. Bonald and A. Proutière, "Insensitive bandwidth sharing in data networks," *Queueing Systems*, vol. 44, no. 1, pp. 69–100, 2003.
- [23] L. Massoulié, "Structural properties of proportional fairness: Stability and insensitivity," *The Annals of Applied Probability*, pp. 809–839, 2007.
- [24] J. Boudec, "Rate adaptation, congestion control and fairness: A tutorial," *Web page*, vol. November, 2005.
- [25] D. Bertsekas, R. Gallager, and P. Humblet, "Data networks," *Networks*, 1987.
- [26] P. Nilsson, "Fairness in communication and computer network design," Ph.D. dissertation, Lund University, 2006.
- [27] W. Ogryczak, M. Pioro, and A. Tomaszewski, "Telecommunications network design and max-min optimization problem," *Journal of telecommunications and informaton technology*, vol. 3, pp. 1–14, 2005.
- [28] B. Radunovic and J. Le Boudec, "A unified framework for max-min and min-max fairness with applications," *IEEE/ACM Transactions on Networking*, vol. 15, no. 5, pp. 1073–1083, Oct. 2007.
- [29] M. Pioro, "Fair routing and related optimization problems," *15th International Conference on Advanced Computing and Communications (ADCOM 2007)*, vol. 1, pp. 229–235, Dec. 2007.
- [30] E. Danna, A. Hassidim, H. Kaplan, and A. Kumar, "Upward max min fairness," in *IEEE INFOCOM*, 2012, pp. 1–9.
- [31] A. Coluccia, A. D'Alconzo, and F. Ricciato, "On the optimality of max-min fairness in resource allocation," *Annals of Telecommunications*, vol. 67, no. 1-2, pp. 15–26, Mar. 2011.
- [32] T. Harks and T. Poschwatta, "Congestion control in utility fair networks," *Computer Networks*, vol. 52, no. 15, pp. 2947–2960, 2008.
- [33] E. Danna, S. Mandal, and A. Singh, "A practical algorithm for balancing the max-min fairness and throughput objectives in traffic engineering," *Proc. IEEE Infocom*, pp. 846–854, 2012.
- [34] A. Tomaszewski, "A polynomial algorithm for solving a general max-min fairness problem," *European Transactions on Telecommunications*, vol. 16, no. 3, pp. 233–240, May 2005.
- [35] G. P. McCormick, "Computability of global solutions to factorable nonconvex programs: Part I-convex underestimating problems," *Mathematical programming*, vol. 10, no. 1, pp. 147–175, 1976.
- [36] A. Claus, "A new formulation for the travelling salesman problem," *SIAM Journal on Algebraic Discrete Methods*, vol. 5, no. 1, pp. 21–25, 1984.
- [37] S. Orlowski, R. Wessäly, M. Pióro, and A. Tomaszewski, "SNDlib 1.0 survivable network design library," *Networks*, vol. 55, no. 3, pp. 276–286, 2010.



Stefano Coniglio is Lecturer of Operational Research at the University of Southampton, UK, which he joined in 2016 after receiving a PhD in Information Technology from Politecnico di Milano, Italy, in 2011 and spending three years as postdoc at RWTH Aachen University, Germany. His research interests are in integer programming (linear and nonlinear), combinatorial optimization, and bilevel programming, with applications to operations research, game theory, and data science. He serves as PC member of international conferences such as AAAI, IJCAI, ICCL, and EvoCOMNET.



Luca G. Gianoli obtained a Ph.D. in electrical engineering at both Politecnico di Milano and Polytechnique Montreal, Canada, in 2014. After two years as a postdoc at Polytechnique, where he conducted research on cloud optimization and communication for swarm robotics, he is now ICT Lead Scientist at Humanitas Solutions, Montreal. His research interests include network optimization, ad-hoc networks, operations research, robust optimization, cloud computing, and networking.



Edoardo Amaldi is Full Professor of Operations Research at DEIB, Politecnico di Milano, Italy. His research interests are in discrete optimization, with applications in telecommunications, data mining, energy, transportation, and computational finance. He received two IBM Faculty Awards in 2005 and 2011. He has been on the editorial boards of *International Transactions in Operational Research* and *Neural Processing Letters*, and he is an editor of the *AIRO Springer Series in Operations Research*. He serves as PC member of conferences such as INOC, ISCO, SEA, ESA, and ESANN, and is on the board of the European Network Optimization Group (ENOG) of EURO.



Antonio Capone is Full Professor at Politecnico di Milano, where he is the Dean of the School of Engineering and director of the ANTLab (Advanced Network Technologies Lab). He serves on the TPCs of major conferences in networking, is currently an editor of *IEEE Transactions on Mobile Computing and Computer Communications*, and he has served in the past as editor of *ACM/IEEE Transactions on Networking and Computer Networks*. He is a Fellow of the IEEE.

The following appendices are to be intended as an online-only extension of the paper.

APPENDIX A

COMPLETE FORMULATION FOR BTE-FBA-MMF

For completeness, we report the complete single-level MILP reformulation for BTE-FBA-MMF. It consists of (2a)–(2c), (12a)–(12g), (13a)–(13d), (14a)–(14b), with (12c) substituted for (15a)–(15b):

$$\begin{aligned}
& \max \sum_{d \in D} w^d \phi^d \\
& \text{s.t. } \sum_{(i,j) \in A} x_{ij}^d - \sum_{(j,i) \in A} x_{ji}^d = \begin{cases} 1 & \text{if } i = s^d \\ -1 & \text{if } i = t^d \\ 0 & \text{else} \end{cases} \quad \forall i \in V, d \in D \\
& \sum_{(i,j) \in A} f_{ij}^d - \sum_{(j,i) \in A} f_{ji}^d = \begin{cases} \phi^d & \text{if } i = s^d \\ -\phi^d & \text{if } i = t^d \\ 0 & \text{else} \end{cases} \quad \forall i \in V, d \in D \\
& \sum_{d \in D} f_{ij}^d \leq c_{ij} \quad \forall (i,j) \in A \\
& f_{ij}^d \leq c_{ij} x_{ij}^d \quad \forall (i,j) \in A, d \in D \\
& \sum_{(i,j) \in A} y_{ij}^d \geq 1 \quad \forall d \in D \\
& y_{ij}^d \leq x_{ij}^d \quad \forall d \in D, (i,j) \in A \\
& \sum_{d' \in D} f_{ij}^{d'} \geq c_{ij} y_{ij}^d \quad \forall (i,j) \in A, d \in D \\
& u_{ij} \geq f_{ij}^d \quad \forall (i,j) \in A, d \in D \\
& f_{ij}^d \geq u_{ij} - \max_{(i,j) \in A} \{c_{ij}\} (1 - y_{ij}^d) \quad \forall (i,j) \in A, d \in D \\
& \phi^d \geq \frac{\min_{(i,j) \in A} \{c_{ij}\}}{|D|} \quad \forall d \in D \\
& \sum_{(i,h) \in A} x_{ih}^d = \psi_h^d \quad \forall d \in D, h \in V^d \\
& \sum_{(i,j) \in A} q_{ijh}^d - \sum_{(j,i) \in A} q_{jih}^d = \begin{cases} \psi_h^d & \text{if } i = s^d \\ -\psi_h^d & \text{if } i = h \\ 0 & \text{otherwise} \end{cases} \quad \forall d \in D, h \in V^d, i \in V \\
& q_{ijh}^d \leq x_{ij}^d \quad \forall d \in D, (i,j) \in A, h \in V^d \\
& q_{ijh}^d \leq \psi_h^d \quad \forall d \in D, (i,j) \in A, h \in V^d \\
& y_{ij}^d \in \{0, 1\} \quad \forall (i,j) \in A, d \in D \\
& u_{ij} \geq 0 \quad \forall (i,j) \in A \\
& x_{ij}^d \in \{0, 1\} \quad \forall (i,j) \in A, d \in D \\
& f_{ij}^d \geq 0 \quad \forall (i,j) \in A, d \in D \\
& \phi^d \geq 0 \quad \forall d \in D \\
& q_{ijh}^d \in [0, 1] \quad \forall d \in D, (i,j) \in A, h \in V^d \\
& \psi_h^d \in \{0, 1\} \quad \forall d \in D, h \in V^d.
\end{aligned}$$

APPENDIX B

COMPLETE FORMULATIONS FOR BTE-FBA-PF

The complete approximate single-level MILP reformulation for BTE-FBA-PF consists of B, consists of: (2a)–(2c), and (19):

$$\begin{aligned}
& \max \sum_{d \in D} w^d \phi^d \\
& \text{s.t. } \sum_{(i,j) \in A} x_{ij}^d - \sum_{(j,i) \in A} x_{ji}^d = \begin{cases} 1 & \text{if } i = s^d \\ -1 & \text{if } i = t^d \forall i \in V, d \in D \\ 0 & \text{else} \end{cases} \\
& \sum_{d \in D} \eta^d + M \sum_{(i,j) \in A} \sum_{d \in D} \gamma_{ij}^d \geq \\
& \sum_{d \in D} \sum_{h \in H} \beta_h \lambda_h^d + \sum_{(i,j) \in A} c_{ij} \mu_{ij} + \sum_{(i,j) \in A} \sum_{d \in D} c_{ij} v_{ij}^d \\
& \eta^d \leq \alpha_h \phi^d + \beta_h \quad \forall d \in D, h \in H \\
& \sum_{(i,j) \in A} f_{ij}^d - \sum_{(j,i) \in A} f_{ji}^d = \begin{cases} \phi^d & \text{if } i = s^d \\ -\phi^d & \text{if } i = t^d \forall i \in V, d \in D \\ 0 & \text{otherwise} \end{cases} \\
& \sum_{d \in D} f_{ij}^d \leq c_{ij} \quad \forall (i,j) \in A \\
& f_{ij}^d \leq c_{ij} x_{ij}^d + \gamma_{ij}^d \quad \forall (i,j) \in A, d \in D \\
& \sum_{h \in H} \lambda_h^d \geq 1 \quad \forall d \in D \\
& \kappa_i^d - \kappa_j^d + \mu_{ij} + \theta_{ij}^d \geq 0 \quad \forall (i,j) \in A, d \in D \\
& \kappa_{td}^d - \kappa_{sd}^d \geq \sum_{h \in H} \alpha_h \lambda_h^d \quad \forall d \in D \\
& v_{ij}^d \leq M x_{ij}^d \quad \forall (i,j) \in A, d \in D \\
& v_{ij}^d \leq \theta_{ij}^d \quad \forall (i,j) \in A, d \in D \\
& v_{ij}^d \geq \theta_{ij}^d + M(x_{ij}^d - 1) \quad \forall (i,j) \in A, d \in D \\
& \eta^d, f_{ij}^d, \phi^d, \nu_{ij}^d \geq 0 \quad \forall (i,j) \in A, d \in D \\
& \lambda_h^d, \kappa_i^d, \mu_{ij}, \theta_{ij}^d \geq 0 \forall h \in H, d \in D, i \in V, (i,j) \in A \\
& x_{ij}^d \in \{0, 1\} \quad \forall (i,j) \in A, d \in D \\
& f_{ij}^d \geq 0 \quad \forall (i,j) \in A, d \in D \\
& \lambda_{ij} \geq 0 \quad \forall (i,j) \in A \\
& \phi^d \geq 0 \quad \forall d \in D \\
& \eta^d \geq 0 \quad \forall (i,j) \in A, d \in D.
\end{aligned}$$

APPENDIX C

CALCULATIONS FOR EXAMPLE 2

A. PF allocation without subtour

Assuming $x_{ij}^d = 0$ for all $d \in D$ and for all arcs (i,j) on the subtour, Proposition 2 implies that an allocation ϕ satisfying the capacity constraints is a PF allocation if and only there are $\lambda_{12}, \lambda_{23} \geq 0$ such that the following system is satisfied:

$$\begin{aligned}
\frac{1}{\phi^1} &= \lambda_{12} \\
\frac{1}{\phi^2} &= \lambda_{23}
\end{aligned}$$

$$\begin{aligned}\frac{1}{\phi^3} &= \lambda_{12} + \lambda_{23} \\ (1 - \phi^1 - \phi^3)\lambda_{12} &= 0 \\ (1 - \phi^2 - \phi^3)\lambda_{23} &= 0.\end{aligned}$$

Since $\phi^1, \phi^2 < \infty$ implies $\lambda_{12}, \lambda_{23} > 0$, we deduce:

$$\begin{aligned}\frac{1}{\phi^1} &= \lambda_{12} \\ \frac{1}{\phi^2} &= \lambda_{23} \\ \frac{1}{\phi^3} &= \lambda_{12} + \lambda_{23} \\ 1 - \phi^1 - \phi^3 &= 0 \\ 1 - \phi^2 - \phi^3 &= 0.\end{aligned}$$

This system implies:

$$\begin{aligned}\phi^1 &= 1 - \phi^3 \\ \phi^2 &= 1 - \phi^3 \\ \phi^3 &= \frac{1}{\frac{1}{1-\phi^3} + \frac{1}{1-\phi^3}}.\end{aligned}$$

From the third equation, we obtain:

$$\phi^3 \left(\frac{1}{1-\phi^3} + \frac{1}{1-\phi^3} \right) = 1,$$

which implies $\frac{2\phi^3}{1-\phi^3} = 1$, i.e., $\phi^3 = \frac{1}{3}$. Hence, $\phi^1 = \phi^2 = 1 - \phi^3 = \frac{2}{3}$ and, overall, we have the allocation $\phi = (1 - \frac{1}{3}, 1 - \frac{1}{3}, \frac{1}{3})$, with throughput $2 - \frac{1}{3}$.

B. PF allocation with subtour

Assume that \mathbf{x}^3 contains the 1-3 path given by $x_{12}^3 = x_{23}^3 = 1$, together with the subtour given by $x_{45}^3 = x_{56}^3 = x_{64}^3 = 1$.

Proposition 2 implies the following conditions on ϕ , subject to capacity constraints on all the arcs and to the nonnegativity of $\lambda_{12}, \lambda_{23}, \lambda_{45}, \lambda_{46}, \lambda_{64}$:

$$\begin{aligned}\phi^1 &= \frac{1}{\lambda_{12}} \\ \phi^2 &= \frac{1}{\lambda_{23}} \\ \phi^3 &= \frac{1}{\lambda_{12} + \lambda_{23} + \lambda_{45} + \lambda_{56} + \lambda_{64}} \\ (1 - \phi^1 - \phi^3)\lambda_{12} &= 0 \\ (1 - \phi^2 - \phi^3)\lambda_{23} &= 0 \\ (\bar{c} - \phi^3)\lambda_{45} &= 0 \\ (\bar{c} - \phi^3)\lambda_{56} &= 0 \\ (\bar{c} - \phi^3)\lambda_{64} &= 0.\end{aligned}$$

We show that this system admits a solution with $\lambda_{56} = \lambda_{64} = 0$ and $\lambda_{12}, \lambda_{23}, \lambda_{45} > 0$. When $\lambda_{56} = \lambda_{64} = 0$, we have:

$$\begin{aligned}\phi^1 &= \frac{1}{\lambda_{12}} \\ \phi^2 &= \frac{1}{\lambda_{23}}\end{aligned}$$

$$\begin{aligned}\phi^3 &= \frac{1}{\lambda_{12} + \lambda_{23} + \lambda_{45}} \\ 1 - \phi^1 - \phi^3 &= 0 \\ 1 - \phi^2 - \phi^3 &= 0 \\ \bar{c} - \phi^3 &= 0.\end{aligned}$$

This implies:

$$\begin{aligned}\phi^1 &= 1 - \bar{c} = \frac{1}{\lambda_{12}} \\ \phi^2 &= 1 - \bar{c} = \frac{1}{\lambda_{23}} \\ \phi^3 &= \bar{c} = \frac{1}{\frac{1}{\phi^1} + \frac{1}{\phi^2} + \lambda_{45}}.\end{aligned}$$

Removing ϕ^3 from the system, we obtain:

$$\begin{aligned}\phi^1 &= 1 - \bar{c} = \frac{1}{\lambda_{12}} \\ \phi^2 &= 1 - \bar{c} = \frac{1}{\lambda_{23}} \\ \bar{c} &= \frac{1}{\frac{1}{\phi^1} + \frac{1}{\phi^2} + \lambda_{45}}.\end{aligned}$$

As $\phi^1, \phi^2 > 0$ and $\lambda_{45} \geq 0$ imply $\frac{1}{\phi^1} + \frac{1}{\phi^2} + \lambda_{45} > 0$, the system can be rewritten as:

$$\begin{aligned}\lambda_{12} &= \frac{1}{1 - \bar{c}} \\ \lambda_{23} &= \frac{1}{1 - \bar{c}} \\ \lambda_{45} &= \frac{1}{\bar{c}} - \lambda_{12} - \lambda_{23} = \frac{1}{\bar{c}} - \frac{1}{1 - \bar{c}} - \frac{1}{1 - \bar{c}}.\end{aligned}$$

When imposing $\lambda_{12}, \lambda_{23}, \lambda_{45} > 0$, we deduce $\bar{c} < 1$ from the first two rows and $\frac{1}{\bar{c}} > \frac{2}{1-\bar{c}}$ from the third row. The latter implies $1 - \bar{c} > 2\bar{c}$, i.e., $\bar{c} < \frac{1}{3}$. Thus, the system is satisfied for any $\bar{c} \in (0, \frac{1}{3})$, implying that $\phi = (1 - \bar{c}, 1 - \bar{c}, \bar{c})$, with throughput $2 - \bar{c}$, is a feasible allocation.

APPENDIX D COMPLETE RESULTS

We report, in Tables II, III, and IV, the complete set of results obtained over the whole data set without aggregation, with a table per network topology.

TABLE III
COMPLETE RESULTS ON THE NOBEL-US INSTANCES.

		Nobel-us, E = 7						Nobel-us, E = 9						Nobel-us, E = 11							
		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF			
		gap	utility	gap	utility	gap	utility	gap	utility	gap	utility	gap	utility	gap	utility	gap	utility	gap	utility		
tr	cap	0.01	0.00	0.00	17950.0	16873.3	18858.6	0.06	0.10	30520.0	22563.5	26170.5	21627.2	0.60	0.37	33383.4	23329.8	0.39	0.76	22846.6	
1	1	0.02	0.01	0.02	15394.0	16178.1	16178.1	0.20	0.03	27032.6	20820.9	29697.1	20082.6	0.51	0.29	34125.0	28440.3	0.36	0.58	27821.9	
1	1	0.10	0.05	0.10	11788.9	16747.8	11253.5	0.41	0.25	20311.1	16472.2	16443.5	15815.4	0.42	0.28	38426.5	28023.7	0.35	0.88	28199.6	
4	1	0.01	0.01	0.01	18856.8	21801.4	19616.9	0.08	0.07	30097.6	23601.5	29473.0	22306.1	0.95	0.18	27559.2	24724.8	0.95	0.18	24194.4	
4	2	0.02	0.01	0.02	15358.0	22028.0	16111.2	0.21	0.04	26906.1	21828.6	30578.0	20945.1	0.62	0.29	31689.3	28569.5	0.62	0.29	28086.7	
4	3	0.10	0.01	0.01	16993.9	16974.6	11791.9	0.44	0.14	19909.1	17152.2	24893.3	16243.0	0.88	0.30	28975.3	29733.3	0.88	0.30	29854.7	
7	1	0.01	0.03	0.03	22219.9	19019.4	21312.3	19804.5	0.07	0.09	30306.5	23800.6	28156.8	22481.2	0.72	0.11	30891.6	24741.2	0.72	0.11	24168.1
7	2	0.02	0.02	0.02	22065.2	15496.7	22076.5	16210.2	0.05	0.06	30879.4	22023.4	29985.2	21114.7	0.53	0.27	33631.7	28561.0	0.53	0.27	28104.8
7	3	0.22	0.08	0.08	15515.0	12528.6	16989.5	11851.1	0.31	0.11	21789.9	17247.8	25326.1	16317.2	0.35	0.17	40252.1	29676.6	0.35	0.17	29844.2
9	1	0.01	0.02	0.02	22219.5	19031.8	22048.0	19818.2	0.12	0.07	29110.0	23831.1	29295.3	22500.7	0.65	0.21	32579.1	24761.0	0.65	0.21	24186.4
9	2	0.04	0.02	0.02	1729.8	15504.6	22068.9	16214.0	0.03	0.06	31474.7	22034.2	30458.2	21129.0	0.45	0.25	35533.3	28542.3	0.45	0.25	28097.6
9	3	0.15	0.08	0.08	16331.4	12534.4	17167.8	11856.2	0.25	0.13	22797.9	17250.5	24865.9	16315.9	0.55	0.38	35051.3	29703.9	0.55	0.38	29880.8
10	1	0.01	0.02	0.02	22219.4	19032.1	21969.7	19819.6	0.14	0.02	28480.9	23830.5	31369.8	22498.9	0.56	0.18	34250.6	24764.6	0.56	0.18	24190.1
10	2	0.03	0.02	0.02	22000.6	15509.8	22188.6	16218.6	0.04	0.03	31265.9	22032.4	31141.4	21127.6	0.50	0.30	34285.3	28541.9	0.50	0.30	28098.1
10	3	0.15	0.08	0.08	16329.8	11855.1	17156.3	11855.1	0.26	0.10	22652.4	17248.2	25340.3	16314.1	0.82	0.17	29945.0	29709.5	0.82	0.17	29885.0
avg		0.06	0.03	0.03	20207.2	15534.4	19932.8	15830.5	0.18	0.09	26902.3	20782.5	27546.3	19787.9	0.61	0.25	33371.9	27454.9	0.61	0.25	27163.9

		Nobel-us, E = 13						Nobel-us, E = 14													
		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF					
		gap	utility	gap	utility	gap	utility	gap	utility	gap	utility	gap	utility	gap	utility	gap	utility				
tr	cap	0.83	0.36	0.36	33879.6	25894.1	41667.8	0.83	0.36	33879.6	25894.1	41667.8	26003.2	0.99	0.31	33875.9	19963.9	0.99	0.31	21056.7	
1	1	0.90	0.29	0.29	30988.3	25223.2	42519.9	25191.4	0.97	0.29	29942.1	25223.2	42519.9	25191.4	0.96	0.33	31614.9	25158.2	0.96	0.33	26428.5
1	3	0.67	0.24	0.24	35685.5	24177.1	45409.7	25409.6	0.64	0.24	36315.0	24177.1	45409.7	25409.6	0.79	0.38	35160.2	21133.4	0.79	0.38	22623.7
4	1	0.95	0.30	0.30	31593.5	26461.3	45111.9	27085.3	0.95	0.26	31633.4	26461.3	46433.6	27085.3	1.22	0.36	30612.6	22066.0	1.22	0.36	23040.0
4	2	0.48	0.31	0.31	39931.5	27892.1	42889.5	27859.3	0.48	0.31	39929.5	27892.1	42889.5	27859.3	0.98	0.32	31344.3	29454.1	0.98	0.32	30040.2
4	3	0.71	0.34	0.34	34905.3	29098.0	43285.7	29746.6	0.71	0.34	34905.3	29098.0	43285.7	29746.6	0.93	0.42	32663.9	24201.8	0.93	0.42	25265.1
7	1	0.82	0.31	0.31	33891.3	26096.0	46202.6	26733.1	0.82	0.31	33891.3	26096.0	46202.6	26733.1	1.09	0.35	32347.1	22106.4	1.09	0.35	23044.0
7	2	0.72	0.28	0.28	34338.4	28436.4	44361.0	28325.3	0.67	0.28	35327.8	28436.4	44361.0	28325.3	0.71	0.33	36261.6	30276.8	0.71	0.33	30814.9
7	3	0.54	0.20	0.20	38741.1	30042.5	47334.4	30502.3	0.54	0.20	38645.1	30042.5	47816.9	30502.3	0.64	0.46	38526.9	24722.4	0.64	0.46	25846.5
9	1	0.61	0.27	0.27	38200.5	26093.5	46591.4	26737.5	0.61	0.27	38200.5	26093.5	46591.4	26737.5	1.10	0.34	32059.7	22099.1	1.10	0.34	23044.4
9	2	0.41	0.26	0.26	41852.9	28485.6	45139.9	28364.6	0.44	0.12	41112.8	28485.6	45160.1	28364.6	0.96	0.45	31560.3	30354.0	0.96	0.45	30887.6
9	3	0.49	0.22	0.22	39918.0	30144.7	47033.2	30593.0	0.49	0.20	39918.0	30144.7	48147.2	30593.0	0.41	0.51	44710.5	24774.9	0.41	0.51	25880.4
10	1	1.10	0.35	0.35	29634.0	26089.3	44922.2	26734.8	1.10	0.35	29634.0	26089.3	44922.2	26734.8	0.47	0.22	45162.0	22098.2	0.47	0.22	23043.1
10	2	0.54	0.15	0.15	38199.6	28497.6	50565.2	28375.0	0.54	0.19	38322.0	28497.6	48304.4	28375.0	0.75	0.27	35489.4	30376.6	0.75	0.27	30909.3
10	3	0.57	0.43	0.43	37950.4	30165.6	41030.2	30602.1	0.57	0.43	37950.4	30165.6	41030.2	30602.1	0.89	0.49	33370.4	24757.7	0.89	0.49	25894.7
avg		0.69	0.29	0.29	35980.7	27519.8	44944.6	27884.2	0.69	0.28	35973.8	27519.8	45409.5	27884.2	0.86	0.37	34984.0	24904.9	0.86	0.37	25854.6

TABLE IV
COMPLETE RESULTS ON THE NOBEL-GER INSTANCES.

		Nobel-ger, E = 8						Nobel-ger, E = 10						Nobel-ger, E = 12							
		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF			
		gap	utility	gap	utility	gap	utility	gap	utility	gap	utility	gap	utility	gap	utility	gap	utility	gap	utility		
tr	cap	STE+MMF		APF+PF		STE+MMF		APF+PF		STE+MMF		APF+PF		STE+MMF		APF+PF		STE+MMF			
1	1	0.04	0.01	18642.4	9927.8	18218.4	11132.4	0.06	0.00	33893.8	25563.3	27339.0	25660.1	0.13	0.13	36035.0	29057.4	34191.3	29000.6		
1	2	0.02	0.01	23489.1	17867.5	23139.7	18636.4	0.02	0.13	40842.6	25334.3	36304.2	25804.3	0.17	0.13	42896.2	28077.7	35949.1	28695.1		
1	3	0.02	0.03	28972.0	24710.0	28349.3	24921.0	0.03	0.07	43233.2	32093.9	40749.3	33435.5	0.07	0.14	47652.3	32131.3	41469.0	32601.1		
4	1	0.05	0.06	18410.4	10616.2	18220.8	11123.5	0.06	0.05	33842.1	25173.4	33581.9	25044.3	0.16	0.14	34934.3	28410.5	35266.8	28334.0		
4	2	0.03	0.03	23424.0	17677.4	23346.6	18159.8	0.02	0.11	40903.6	24290.1	37317.9	24710.9	0.10	0.16	44299.0	26801.8	41432.2	27449.7		
4	3	0.03	0.01	28889.0	24354.5	28324.6	24440.8	0.24	0.05	36063.4	30661.9	41775.4	31364.4	0.06	0.10	48721.4	32515.2	44795.0	33162.5		
7	1	0.12	0.07	17500.8	10583.4	18395.5	11020.0	0.09	0.09	33159.4	25180.1	31813.6	24995.7	0.14	0.01	35587.7	28243.5	33444.8	28160.6		
7	2	0.03	0.04	23403.5	17682.5	23148.1	17971.4	0.04	0.11	40121.2	24352.2	37363.9	24658.9	0.09	0.22	44766.8	26634.9	40005.7	27216.3		
7	3	0.03	0.03	28725.0	24057.7	28347.1	24036.6	0.04	0.07	43010.7	30402.8	40864.8	31078.0	0.09	0.13	47235.9	32478.0	44153.3	33172.8		
9	1	0.05	0.08	18649.6	10571.4	17647.8	11010.7	0.08	0.10	33236.2	25184.5	32271.4	24990.9	0.16	0.17	35130.1	28247.1	34145.0	28159.2		
9	2	0.03	0.02	23457.0	17682.8	23414.5	17954.9	0.05	0.05	40121.7	24379.2	39768.0	24660.4	0.12	0.28	43588.2	26646.4	37548.5	27206.9		
9	3	0.04	0.05	28348.9	23979.8	27815.8	23954.7	0.04	0.06	42838.5	30405.8	41114.6	31072.4	0.09	0.13	47111.4	32493.4	44824.3	33201.1		
10	1	0.13	0.06	17499.9	10573.0	18191.7	11010.0	0.07	0.07	33282.4	25187.5	33210.8	24991.8	0.11	0.18	36385.1	28247.5	33841.7	28159.0		
10	2	0.03	0.06	23456.9	17679.9	22630.7	17947.4	0.02	0.12	40913.2	24376.9	37274.5	24653.2	0.12	0.20	43534.8	26642.8	40192.7	27201.5		
10	3	0.03	0.04	28788.9	23958.0	28300.9	23930.6	0.06	0.05	42279.9	30396.5	42086.2	31060.5	0.13	0.12	45598.6	32486.3	44850.6	33194.6		
avg		0.05	0.04	23443.8	17461.5	23166.1	17816.7	0.06	0.08	38516.1	26865.5	36855.7	27212.1	0.11	0.15	42231.8	29274.2	39074.0	29681.0		
		Nobel-ger, E = 14						Nobel-ger, E = 16						Nobel-ger, E = 17							
		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF		BTE-FBA-MMF	
		gap	utility	gap	utility	gap	utility	gap	utility	gap	utility	gap	utility	gap	utility	gap	utility	gap	utility	gap	utility
tr	cap	STE+MMF		APF+PF		STE+MMF		APF+PF		STE+MMF		APF+PF		STE+MMF		APF+PF		STE+MMF		APF+PF	
1	1	0.14	0.01	41605.0	24950.6	33246.9	24815.3	0.20	0.27	50093.4	33212.9	46460.8	32939.8	0.26	0.27	50051.7	35553.2	48974.1	35795.6		
1	2	0.13	0.35	49912.4	25970.0	40701.6	27055.0	0.48	0.39	45149.0	29333.9	45878.6	30526.3	0.32	0.33	56343.7	36774.9	53278.8	37469.6		
1	3	0.67	0.15	35497.0	24810.9	47073.3	26872.8	0.24	0.22	59286.6	31674.9	58421.0	33524.1	0.33	0.01	63385.6	41497.8	58869.5	42754.0		
4	1	0.19	0.26	39586.3	22928.6	36990.8	23159.1	0.31	0.31	46215.2	33683.1	44987.7	33440.7	0.24	0.31	50778.4	35208.2	47035.3	35396.1		
4	2	0.17	0.22	48768.0	24167.5	43973.5	25210.2	0.46	0.31	46105.7	28149.7	48707.8	29479.7	0.33	0.21	55954.3	36067.3	58963.1	36773.6		
4	3	0.40	0.24	42119.3	23377.1	47499.3	25035.5	0.17	0.19	62737.9	29148.9	59313.7	30469.4	0.36	0.29	62026.9	40836.2	62011.1	42744.8		
7	1	0.24	0.23	38274.2	22330.4	36835.0	22662.2	0.21	0.20	49040.0	33673.4	48819.0	33412.1	0.29	0.25	49102.7	35166.1	49594.9	35292.5		
7	2	0.36	0.30	41939.9	23787.1	43977.1	24671.3	0.24	0.17	54275.8	27942.6	56031.7	29137.9	0.25	0.24	59751.7	35873.3	57761.8	36564.4		
7	3	0.11	0.20	53302.8	22896.7	48696.6	24570.1	0.38	0.18	53266.2	29011.6	61849.1	29746.0	0.35	0.29	62609.1	40941.6	63947.8	42864.6		
9	1	0.31	0.27	36064.8	22289.2	40444.3	22627.4	0.24	0.17	48798.3	33693.1	48610.6	33433.3	0.37	0.45	46239.3	35171.0	43016.9	35300.9		
9	2	0.44	0.27	46054.8	23811.3	43381.6	24646.0	0.43	0.30	46915.6	27952.7	49828.1	29135.0	0.45	0.25	51475.6	35892.6	57085.5	36580.6		
9	3	0.45	0.32	41001.7	22852.7	44434.0	24536.1	0.41	0.20	52238.2	28974.0	60194.7	29625.9	0.36	0.33	62025.0	40970.0	61357.7	42892.2		
10	1	0.33	0.15	35474.3	22284.3	39585.7	22622.2	0.26	0.39	47823.6	33693.6	42239.5	33433.7	0.26	0.47	50248.7	35167.2	41742.6	35298.1		
10	2	0.23	0.25	46238.9	23808.3	44316.2	24637.3	0.41	0.18	47738.3	27949.5	55418.2	29122.3	0.36	0.39	54717.3	35887.9	52049.3	36574.8		
10	3	0.41	0.14	42009.9	22845.6	51516.9	24527.2	0.38	0.34	53079.6	28976.0	54479.4	29596.6	0.25	0.23	67800.0	40965.7	67102.0	42900.5		
avg		0.29	0.21	42523.3	23540.7	42844.9	24509.9	0.32	0.26	50850.9	30471.3	52082.7	31134.9	0.32	0.29	56167.3	37464.9	54852.7	38346.8		