

COMPETITION IN CONGESTED SERVICE NETWORKS
WITH APPLICATION TO AIR TRAFFIC CONTROL PROVISION IN EUROPE

Nicole Adler, Hebrew University of Jerusalem, Israel. nicole.adler@mail.huji.ac.il

Eran Hanany, Tel Aviv University, Israel. hananye@tau.ac.il

Stef Proost, KU Leuven and Transport Mobility Leuven, Belgium. Stef.Proost@kuleuven.be

September 2020

Accepted for publication at *Management Science*

ABSTRACT

We analyze congested network-based markets and their impact on competition, equilibrium charges and efficiency. Several strategies are explored including price caps, mergers and investments in new technologies. We find that congested networks served by collaborating (serial) and competing (parallel) firms may lead to excessive prices. Additionally, oligopolists may only serve captive demand, leading to inefficiently low flows. Perhaps surprisingly, permitting a firm with market power to horizontally integrate with a competitor may improve efficiency. We also show that price caps in congested networks are ineffective due to their failure to signal the existence of scarce resources. Instead, partial vertical integration may prove beneficial by creating incentives to expand capacity through technology adoption, provided the price cap regime is dropped. The model is subsequently illustrated with a case study of air traffic control provision in Western Europe, in which it is shown that substantial changes in the regulation are required in order to create a more cost efficient sector with increased capacity.

INTRODUCTION

Service providers face the continuous challenge of adapting capacities to the fluctuations in demand patterns of congestion sensitive customers and of maintaining competitive advantage through congestion pricing policies. As an example, air traffic demand is estimated to double over the next two decades across the globe¹. The growth will only be served if sophisticated technologies to navigate aircraft accurately and safely through the skies are adopted. In network-based, international, air traffic control (ATC) markets, each section of the network is managed by a single service provider. Since flights cross multiple countries, service providers must collaborate to create a seamless service. Given the multiple potential routes from origin to destination, it is clear that service providers are also competing. An important research question therefore draws from the implications of the tradeoff between the benefits of collaboration and the need to compete in serving the market.

The international ATC market is one example in which service providers may compete over a geographically congested network. The service provider is both a monopolist setting capacities and

¹ <https://www.boeing.com/> and <https://www.airbus.com/>

charging for use of their own airspace and a competitor with their neighbors for transit traffic. International routes may have a choice of paths (i.e. service providers) as compared to domestic routes, which represent captive demand for the ATC providers. The setting becomes more complicated because the network users are non-atomistic in that each has market power by generating a non-negligible fraction of the total demand. The airlines maximize their profits by choosing the cheapest route to fly whilst internalizing self-imposed congestion costs. The extent to which airlines internalize congestion has been a subject of debate in both the theoretical (Brueckner, 2002; Brueckner and Van Dender, 2008) and empirical (Morrison and Winston, 2007; Rupp, 2009) literature. The major management issues caused by this complex market include high costs due to fragmentation of ATC service provision, slow technology adoption, lack of standardization in services across air control centers and inefficient scale of operation (Baumgartner and Finger, 2014). The main reasons for these inefficiencies in Europe are the relatively large number of service providers, each procuring their own systems, mostly training their own staff, creating their own operating procedures, providing services in a small airspace and failing to adopt the most performant technologies. As a consequence, the European Union passed laws in 2004 to create horizontal mergers, which have not been implemented, and in 2009 to set up a price cap system, which has proven to be a weak instrument. In the UK, the ATC service is a public-private partnership since 2000, and private companies today serve terminal ATC in the UK, Germany, Spain and Sweden. Consequently, it is important to understand which market design decisions, such as privatization, price caps, horizontal integration or investment in technology, could lead to greater efficiency.

Additional examples of competition across geographical service networks include rail, road and shipping services. In the rail sector, the infrastructure managers are the service providers and the railway companies serving passengers and freight are the non-atomistic customers. Similar management issues and regulatory policies may be considered, namely setting track access charges, harmonization of technical standards across the track network and efficient scale of operation. Consequently, the modeling approach developed in this paper could be applied to analyze the rail sector. Road network infrastructure is also divided into regional providers but the customers are independent and atomistic i.e. cars and trucks, choosing routes and frequencies assuming congestion levels as given. Efficiency issues in the road sector include high truck distance charges in Europe caused by the monopolistic power of each region (Mandell and Proost, 2016). In the US, regions use their monopoly power via state diesel taxes. Our model is relevant provided all road customers are affected by congestion in the same manner because they are atomistic. Similarly, in the shipping industry, the port managers are the service providers and the shipping companies serve the cargo market by choosing the frequency of port visits.

More generally, competition between service providers may be modelled using a directed flow network in which each arc represents service provided by a specific supplier and each customer demands path flows connecting specified origin-destination (OD) pair(s). In this research, we model such settings within a two-stage network congestion game. In a preliminary stage, the regulator sets the rules. In the first stage, each provider sets charges on the services (arcs) they provide in order to maximize profits.

Peak and off-peak pricing is also considered. In the second stage, each customer chooses their flows in order to minimize the sum of service charges and congestion-dependent operating costs, including the possibility of partial flows or not using any service when the associated costs are too high. The network structure matters because (1) decisions of one customer will impact the congestion levels of the other customers via network flows hence impact their choices and the size of the subsequent market; and (2) the equilibrium price setting will depend on network effects related to the relative size of the flexible (or non-captive) demand with a choice between first stage service providers.

Contributions

To the best of our knowledge, the modeling approach developed in this research is the first to consider general networks, elastic demand for multiple OD pairs and oligopolistic markets in both stages of the game. This enables us to model competition between service providers, which cannot be investigated using the existing approaches that consider only a single service provider. The network problems we study go beyond the simple 2-link serial and parallel networks considered in most of the literature. In particular, our network structure includes partly competing supply chains where one link may enjoy monopoly power whereas others are in competition. The game also models competition between providers who serve partly captive demand. Customers have an outside option thus their demand need not be fully satisfied, and are non-atomistic with market power in the second stage. Modeling customers with market power is important in industries such as aviation where an airline controls all of its flights. Our model is therefore better suited for such markets than existing approaches that analyze atomistic customers based on Wardrop equilibrium. The conclusions that we draw generally depend on the number of customers competing for resources and their relative market power.

Our network congestion model shows that service providers engage in competition selectively as a function of demand levels and network structure. Engaging in competition in some part of the network is worthwhile only when the demand level is sufficiently high, and this choice will exhibit interdependencies across different parts of the network. Equilibrium service charges are affected by the level of competition and congestion. Given this behavior, we analyze whether competition between service providers may lead to efficiency in the sense of minimizing total social costs and/or lower service charges for the benefit of the customers. Our setting enables an evaluation of multiple market design scenarios including the impact of deregulation, incentive based price caps, different forms of co-operation between players and the introduction of new technology inducing capacity expansion, which in turn reduces variable costs for the actors.

Key policy insights

The first policy insight emerging from our analysis is that introducing competition does not necessarily improve efficiency in the presence of network effects and congestion. In supply chains with both collaborating (serial) and competing (parallel) service providers, flows are efficient for relatively low demand levels. However, when demand is sufficiently high, competition may trigger inefficiencies.

The serial provider sets a monopoly margin and the parallel downstream providers charge congestion fees that disregard the serial provider's mark-up. As a result, total charges are too high despite the competition between the parallel providers. Charges will be lower and flows will become more efficient were the monopolist to integrate with a competing link in the supply chain. This goes against mainstream policy thinking that pleads for complete separation between the monopolistic component of service provision and the competing providers.

The second policy insight is that price caps in congested networks tend to be ineffective as they often fail to account for the role of congestion pricing and do not incentivize technology adoption. Technology adoption requires capacity investments that are essential to tackle congested networks.

The third policy insight deals with parallel suppliers that each have captive demand for their services and simultaneously compete for flexible demand. The relative size of the captive demand is crucial because the parallel providers may prefer to ignore the flexible demand, instead fully exploiting the willingness-to-pay of the captive demand. Whilst there is competition in principle, it is completely ineffective. In such markets, horizontal integration would lead to more demand being served. We note that much research has focused on flexible demand but it is the existence of captive demand that may completely change the outcome.

Finally, we observe that a merger between parallel providers may reduce costs for the providers but not necessarily the charges to the customers. In markets with non-atomistic demand, coordinating technology adoption between one important customer and one of the parallel suppliers may create beneficial incentives for investment on behalf of both parties. This fourth policy insight holds as long as there is no price cap.

Understanding the likely behavior of service providers and their customers and their implications for social welfare may help guide regulatory initiatives and highlight potential future institutional processes that will in turn lead to improved market conditions.

Related literature

Since the pioneering work of Pigou (1920), there has been a substantial and well established literature analyzing the efficiency of congested service systems, including network congestion games. The standard approaches to analyze such settings include Wardrop equilibria (Wardrop 1952) and the potential game approach (Rosenthal 1973, Monderer and Shapley 1996), both of which consider identical customers in the face of exogenous latency/congestion cost functions, with Wardrop equilibria additionally assuming atomistic customers, each demanding an infinitesimal flow. A different approach assumes that competing customers are non-atomistic and demonstrate market power in that each customer controls a non-negligible fraction of the total flow (e.g., Brueckner 2002, Cominetti et al. 2009). The difference between the two approaches asymptotically vanishes as the number of non-atomistic customers increases (Haurie and Marcotte 1985).

Congestion and contracting in competitive service industries has also been addressed within the operations management literature (e.g., Cachon and Harker 2002, Netessine and Shumsky 2005, Allon and Federgruen 2007, Johari et al. 2010). Competition between service providers in the presence of congestion costs was analyzed in depth by Acemoglu and Ozdaglar (2007a, 2007b) and Perakis and Sun (2014). Acemoglu and Ozdaglar (2007a) consider a two-stage game in which profit-maximizing oligopolists compete by setting prices for travel on each of several alternative and parallel routes, all connecting the same OD pair, in the first stage. Atomistic and identical users choose one of these routes to minimize travel and congestion costs in the second stage. Acemoglu and Ozdaglar (2007b) extend this analysis to parallel-serial networks with a single OD pair, i.e. each parallel route may include several serial links. Instead of Bertrand competition, Perakis and Sun (2014) consider differentiated Cournot competition in the first stage and multimodal general Wardrop equilibrium (Dafermos 1982) in the second stage. Although not formulated using a network, the model of Perakis and Sun (2014) is analogous to a simple network having a single OD pair connected by parallel routes, as in Acemoglu and Ozdaglar (2007a). Consequently the network effects we identify are absent from these papers.

Economic based research on the topic of ATC capacity includes Morrison and Winston (2008), Winston (2013) and Borenstein and Rose (2014). These papers acknowledge that the FAA has not used pricing instruments to address congestion issues in airspace, rather relied on capacity expansions. Zou and Hansen (2012) argue that given the substantial investments required to develop new technologies known as the NextGen system, a cost-benefit assessment based on equilibrium outcomes is of critical importance. Through the computation of supply-demand equilibrium, they show that classical cost-benefit analyses distort delay savings estimates and potentially demand estimates too. In Lulli and Odoni (2007), it is pointed out that air traffic flow management of en-route sectors in Europe is highly congested, particularly in the central and western sectors. The authors also demonstrate that issues of efficiency and equity in European airspace are far more complicated as compared to that of the US. Whilst in the US, there is one nationwide FAA which allocates resources across 21 air control centers and most airport tower services in addition to developing and adopting NextGen technologies, in Europe the fragmentation into 32 regional monopolies leads to a multi-agent problem.

Economic based research on the topic of ATC regulation to date analyzes individual providers and thus ignores some of the complications of the decentralized system. Castelli et al. (2013) argue that EU regulation² removes the requirement that ATC service providers simply cover their full costs thus potentially generating a more commercial approach to the supply of such services. Accordingly, they develop a Stackelberg game in which a single ATC service provider sets a charge in order to maximize profit and subsequently individual flights are routed in order to minimize costs. The authors argue that there is sufficient flexible traffic in the European system, that a single unit price is tempered by the interplay between captive and flexible flights. Castelli et al. (2011) argue that a slot allocation system

² Regulation 1191/2010 of the European Union

could be organized by a central planner to handle congestion such that a flight receiving an earlier slot is charged accordingly in order to compensate an alternative, delayed flight. This is one of the first attempts to consider an economic mechanism rather than an administrative approach for handling delays. Jovanović et al. (2014) develop a Stackelberg game with a single network planner and an airline with multiple flights, and argue that a congestion based charge with rebates would help to better balance demand in the European airspace.

Our two-stage game of price competition between service providers in the presence of congestion is the first to consider general networks with oligopolistic markets in both stages of the game, i.e. allows for non-atomistic heterogeneous customers with market power in the second stage who react to the first stage competitive pricing. Sub-game perfect equilibria (Selten 1975) allow customers to consider self-imposed congestion across the various routes, potentially leading to interior point flows that do not occur with atomistic Wardrop equilibria. This ensures the existence of equilibria in the two-stage game when customers are heterogeneous, hence impacts the comparative conclusions we can draw from the analysis. We apply our network congestion game with multiple ATC providers and airline customers in order to ascertain whether these conclusions hold and under what conditions could the West European ATC system be encouraged to adopt technologies that would increase capacity. We find that both horizontal and vertical integration could lead to greater efficiency, encourage technology adoption of new technologies and achieve the goals of a Single European Sky.

The paper is organized as follows: We develop the modeling approach in Section II and discuss several analytic results derived from the model in Section III. We present a case study of ATC in the West European airspace in Section IV, with numerical results presented in Section V. Section VI draws conclusions and identifies potential future directions of research. Formal versions of all theorems and proofs are presented in the Appendix.

MODELING APPROACH

We consider a two-stage network congestion game in which the service providers set their charges in stage one and congestion sensitive customers choose their providers via peak/off-peak path flows in stage two. The main focus of the model is to shed light on how first stage service providers choose to compete, which ultimately impacts the preferred market design. We develop two best response problems, one describing the simultaneous decisions of the first stage service providers and another defining those of the second stage customers, which represent a best response to each other and to the choices of the upstream market. The set-up is summarized in Table 1. The customers (firms) want to satisfy demand for specific OD pairs. Paths connecting OD pairs may consist of one to many arcs, where each arc is served by a single service provider. A set of arcs creating a path may therefore be served by a single service provider or require the collaboration of multiple providers. When one provider serves all arcs in all paths connecting an OD we refer to this as *captive* demand, otherwise it is *flexible* demand. Hence, an arc may serve both captive and flexible demand simultaneously.

Prior to establishing an equilibrium outcome in the market, the regulator sets rules with respect to institutional form and price regulation in a preliminary stage. The regulator may also enforce legislation with respect to horizontal integration across service providers. Finally, an a-priori vertical agreement between service providers and customers with respect to technology adoption may be signed, which is expected to reduce costs via capacity expansion. Decisions at this level are considered exogenous to the game defined hence create multiple model parameter scenarios to be assessed.

TABLE 1: Description of the model set-up

Preliminary stage	Regulator chooses rules of the game for the service providers anticipating the outcome of stages 1 and 2, possibly allowing horizontal or vertical agreements to be signed between service providers or between providers and customers, respectively.
First stage	Service providers set charges for each arc, anticipating the behavior of the customers and taking the behavior of other service providers as given.
Second stage	Customers choose the least cost path, including peak/off-peak or no service decisions, given charges over arcs and congestion created by all customers.

The network underlying the congestion game is composed of a set of origin, transit and destination nodes, and a set of arcs representing services offered. We use the following network definitions:

P	finite set of origin/destination nodes with indices o, d
T	finite set of transit nodes
N	set of all nodes, $N = P \cup T$, with indices i, j
A_s	set of arcs, $A_s \subseteq N \times N$, owned by service provider s
A	set of all arcs, $A = \bigcup_s A_s$, with index $a = (i, j)$
ϑ_a	distance of arc a
W	set of time windows with index $w = 1$ for peak and $w = 2$ for off-peak

As an example, in the ATC application P represents the set of airports to be served, T represents the connection between two service providers, ϑ_a is the kilometers of path a , and $w = 1$ the slot constrained hours at an airport.

For the service providers and customers we use the following definitions:

L	finite set of customers with index l
S	finite set of service providers, with index s
S^k	subset of service providers, $S^k \subseteq S$, with customer specific usage agreements during the peak
K_{la1}	maximal flow during the peak for customer l over arc a owned by $s(a) \in S^k$
$s(a)$	service provider owning arc a
D_{lod}	potential demand of customer l for service from origin o to destination d

In the ATC application, L represents the set of airlines, S the set of ATC providers, S^k is the set of terminal ATC providers for which K_{la1} is the grandfathered slot constraints per specific airline and airport in the peak window, and D_{lod} the potential demand between any origin and destination per airline.

We use the following definitions of costs and charges:

C_s^S	service cost per unit distance of arc controlled by service provider s
C_s^F	fixed cost per service provider s
C_{la}^O	operating cost per unit distance for customer l over arc a
C_{ls}^G	congestion cost per flow unit per unit distance for customer l served by provider s

C_{law}^R	consumer surplus loss per unit distance for customer l over arc a during time window w
C_{lod}^T	outside option cost for customer l per flow unit of non-service from origin o to destination d
τ_{sw}^0	price cap per unit distance for service provider s during time window w

In the ATC application, unit distance is one km flown and flow unit is one flight. Accordingly, C_s^S represents the production cost for providing ATC services per km flown, and C_s^F is their fixed cost including investment in technology. For airline l , each km flown generates an operating cost C_{la}^O , congestion cost $C_{ls(a)}^G$ caused by delayed flights, a cost associated with lower revenue C_{la2}^R when flying in the off-peak, and an outside option cost C_{lod}^T per flight which will be preferred if the total costs of flying between origin and destination are too high. Finally, τ_{sw}^0 represents a price cap on ATC providers, should a regulator determine this to be necessary.

The model includes the following sets of decision variables:

τ_{saw}	service provider s 's charge per unit distance over arc a during time window w
f_{lodaw}	customer l 's flow via arc a within origin-destination (o, d) during window w
f_{lod}^T	customer l 's non-flow from origin o to destination d

In the ATC application, each airline l pays an ATC charge $\tau_{s(a)aw}$ to the relevant ATC provider $s(a)$, and chooses how often to fly, f_{lodaw} , between origin and destination per path in the peak and off-peak, or not, f_{lod}^T .

We model the first-stage service providers as profit maximizers choosing charges τ_{saw} . Each service provider s best responds to the choices of its competitors, taking as given the equilibrium service flows f_{lodaw}^* that will be chosen in the second stage of the game, thus leading to a sub-game perfect equilibrium. For service provider s , model (2.1) includes the profit function and a set of constraints in which the charges are price capped, to be included where relevant. Fixed costs related to technology adoption affect the equilibrium indirectly via the model parameters, and are relevant when comparing outcomes across different scenarios.

$$\begin{aligned}
& \text{Max}_{\tau_{saw}} \sum_{a \in A_s} \sum_w (\tau_{saw} - C_s^S) \partial_a \sum_l \sum_{od} f_{lodaw}^* - C_s^F \\
& \text{s.t.} \\
& 0 \leq \tau_{saw} \leq \tau_{sw}^0 \quad \forall a \in A_s, w \in W
\end{aligned} \tag{2.1}$$

In the ATC application, we are interested in understanding the likely equilibrium outcome as a function of the ATC providers' objective. For example, what would happen if the ATC providers were privatized? We also examine the outcome under cost recovery, i.e. when revenues equal the sum of variable and fixed costs.

When applying the general model to the ATC case study, we note that the airline customers are limited players in this game because of the assumption that their fleet and the revenue per flight, which is modelled as the outside option cost C_{lod}^T in each OD market considered, are given. The assumed fixed outside option costs abstract away from determining airfares simultaneously with flows. Equilibrium behavior would become more complex if the outside option cost were allowed to depend on flows,

however the network effects we identify and our general conclusions continue to hold. As a result of our assumptions, the airlines are defined as cost minimizers rather than profit maximizers, which avoids the issue of modeling passengers. Additionally, the airline's fixed costs relating to fleet and overhead are not affected by the flight path and peak/off-peak decisions modelled. However, increases in congestion and/or air traffic control charges may encourage airlines to move to the off-peak, less congested times or to cancel flights. The reduced form model attempts to approximately capture this behavior in order to understand the market at the strategic level. Consequently, a day has been separated into two timeframes, and we include in the generalized cost function a revenue loss to airlines moving flights from the peak to off-peak, and an outside option cost per flight cancelled in each OD market. This formulation balances the desire to avoid congestion and reduce costs yet meet the most valuable passenger demand. The peak capacity has been limited according to the airport infrastructure. The model could account for airline and passenger preferences in greater detail, but this is not necessary in order to understand the implications of intervention policies such as price caps or changes in ownership form on the behavior of ATC providers.

The customer cost function, equation (2.2), which is modelled in the second stage of the game with linear latency costs³ (Roughgarden and Tardos, 2002), is composed of several categories, all of which are impacted to some degree by the choices in the preliminary and first stages. We note that congestion costs $C_{ls(a)}^G$ are multiplied by the total arc flow including that of competing customers. Hence the customer cost function is quadratic due to the congestion cost, which increases with total arc flow per unit distance. Additionally, in order to account for elastic demand, there exists an outside option flow, f_{lod}^T , which represents the choice to reduce the service demanded below the potential level D_{lod} , with cost C_{lod}^T per flow unit.

$$\psi_l \equiv \sum_w \sum_{a \in A} [C_{la}^O + C_{ls(a)}^G (\sum_{l' \in A} f_{l'odaw}) + \tau_{s(a)aw} + C_{law}^R] \partial_a \sum_{od} f_{lodaw} + \sum_{od} C_{lod}^T f_{lod}^T \quad (2.2)$$

We compare scenarios according to total social costs to be minimized in order to search for the most appropriate outcomes considering both sets of actors. The social cost function sums all customer costs minus service provider profits. Since service provider revenues cancel out, we are left with the customer costs plus service provider costs.

Two alternative solutions are modelled for the second stage: either a system optimal outcome as described in equations (2.3) to (2.6), or a user equilibrium outcome in which the objective function (2.3) is adapted as shown in (2.3') whereby each user minimizes only its own costs. In the system optimal

³ Although congestion is generally highly non-linear when flows are close to the capacity, objective function (2.2) assumes that linear congestion costs are a reasonable approximation for equilibrium flows sufficiently far from the capacity levels of the arcs. We also note that the reduced form cost function of a pure bottleneck congestion model is also linear in the ratio flow over capacity (Arnott et al.1993).

approach, a central planner chooses the service paths and timing (peak or off-peak) for all customers simultaneously to minimize total customer costs, including operating and congestion costs, service provision charges, losses from off-peak service and lost surplus for non-realized demand. The system optimal solution achieves efficiency in terms of minimizing total social costs when the service provider charge equals the service cost (per unit distance).

$$\text{Min}_{f_{lodaw}, f_{lod}^T} \sum_l \psi_l \quad (2.3)$$

s.t.

$$\sum_w [\sum_{j|(o,j) \in A} f_{lod(o,j)w} - \sum_{j|(j,o) \in A} f_{lod(j,o)w}] + f_{lod}^T = D_{lod}, \quad \forall l \in L, \forall o, d$$

$$\sum_w [\sum_{j|(j,d) \in A} f_{lod(j,d)w} - \sum_{j|(d,j) \in A} f_{lod(d,j)w}] + f_{lod}^T = D_{lod}, \quad \forall l \in L, \forall o, d \quad (2.4)$$

$$\sum_{j|(j,i) \in A} f_{lod(j,i)w} - \sum_{j|(i,j) \in A} f_{lod(i,j)w} = 0, \quad \forall l \in L, w \in W, o, d, i \in N (i \neq o, d)$$

$$\sum_{od} f_{loda1} \leq K_{la1}, \quad \forall l \in L, \forall a: s(a) \in S^k \quad (2.5)$$

$$f_{lodaw} \geq 0, f_{lod}^T \geq 0, \quad \forall l \in L, o, d \in N, a \in A, w \in W. \quad (2.6)$$

$$\text{Min}_{f_{lodaw}, f_{lod}^T} \psi_l \quad (2.3')$$

Constraints (2.4) sum the incoming less the outgoing flows to be equal to the (negative) demand at the (origin) destination and zero when using a transit point. The total flows are reduced by those that have been dropped via the outside option of not being served. Constraints (2.5) restrict the level of flows during the peak window on a per customer basis for service providers with customer specific usage agreements. This restriction may be removed if unnecessary, for example because congestion is not an issue. In other words, if all demand could be served in the peak window, constraint (2.5) would not be necessary. However, in many systems, agreements are drawn between firms and customers in order to constrain peak demand levels, for example through slot controls at an airport which are given to designated airlines at a specific hour every week per season. Constraints (2.6) ensure non-negativity of the flows and non-flow.

The distinction between the user equilibrium and system optimum approaches may be intuitively understood as follows. In a user equilibrium, each customer chooses paths and time windows taking into account its own costs alone and taking the flows of the other customers as given. Specifically, each customer l considers only its own congestion costs and ignores the external congestion costs imposed on the other customers. Hence the flows may be less balanced than those of the system optimal approach in which a system wide planner minimizes the sum of second stage costs including congestion. When congestion is sufficiently severe so that some demand opts out, low charges generate flows that are too high as compared to the system optimal approach, with the opposite effect with high charges. An efficient outcome that minimizes total social costs occurs in a user equilibrium only if the service provider charge is equal to the service cost plus external congestion costs (per unit distance). Recall that service providers may have market power depending on the network structure and customer demand. A monopoly service provider may set charges sufficiently above the service cost (per unit distance) such that flows align to efficient levels from the point of view of minimizing total social costs of first and second stages. In

contrast, depending on the level of competition generated by the network, competing service providers do not consider the entire system when choosing their charges. Consequently, their increasing reaction functions, which emerge from the differentiated Bertrand competition created by congestion, may lead their equilibrium charges to be higher than those set by a monopolist, which in turn would generate inefficiently low flows in some parts of the network. Our results in the following sections elaborate on the above intuition.

Existence of equilibrium

Our first result establishes the existence of an equilibrium. The equilibrium strategies involve randomization over the choices in the first stage and deterministic choices in the second stage.

Proposition I: *For any instance of the game, there exists a sub-game perfect equilibrium in mixed pricing strategies in the first stage and pure flow strategies in the second stage.*

Although an equilibrium with pure pricing (charges) strategies may fail to exist, we show that such equilibria do exist in the networks analyzed in this research. In the simplified network analyzed in the next section, this equilibrium is unique due to the cost symmetry assumed in both stages. In the more general network used in the air traffic control case study presented in Sections IV and V, uniqueness is verified by implementing the numerical solution method using several starting solutions.

SERIAL-PARALLEL NETWORK ANALYSIS

In this section we draw several initial conclusions with respect to the sub-game perfect equilibrium outcomes given a simplified network under various regulation policies. We model three service providers and n customers on a serial-parallel network with multiple origins and destinations, as shown in Figure 1, with symmetric cost functions and a single time window.

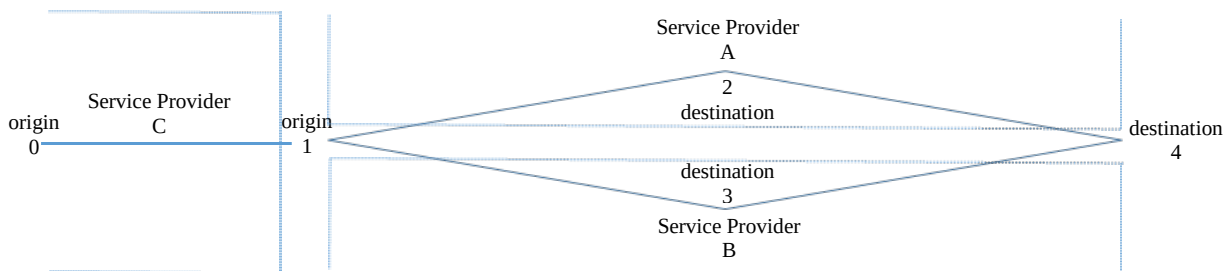


Figure 1: Simplified network

This simplified network enables an analysis of network effects beyond the simple parallel or serial competition cases that have been widely studied. The network effects generate new insights with respect to policy decisions such as privatization, price caps, horizontal integration and technology investment. Our more general setting enables us to analyze two network effects. The first network effect, named ‘asymmetric service collaboration’, consists of service providers with asymmetric market power and customers demanding a single OD pair (0-4) utilizing a monopoly service (provider C) and then choosing between competing parallel options (providers A or B). The second network effect, named ‘heterogeneous customers’, occurs when demand is both captive (OD pairs 1-2 and 1-3) and flexible

(OD pair 1-4). In the former case, the customers have no choice but the latter creates competition between the two parallel providers.

Both network effects arise in decentralized, geographical, service networks, for example in the ATC case study considered in Sections IV and V. Both network effects are absent from the existing analysis in the literature without a network (Brueckner, 2002; Perakis and Sun, 2014), which is equivalent to assuming a single OD pair connected only by multiple parallel links. Our analysis using the simplified network emphasizes the insights emerging in the presence of the network effects. Even for our stylized network, the equilibrium charge setting will already be complex, because it depends on the two types of network effects, namely asymmetric service collaboration and heterogeneous customers. Our conclusions in this section continue to hold whenever these effects are present separately or simultaneously, as demonstrated by the more involved ATC case study. The results are presented below and summarized in Table 2.

TABLE 2: Summary of results based on the simplified network

<i>Connecting service providers through simple network structures</i>	
Theorem I	Simple serial and simple parallel service providers
<i>Asymmetric service collaboration: customers from origin 0 to destination 4 using either the service collaboration of C and A or of C and B</i>	
Theorem II	Unregulated competition under asymmetric service collaboration as a hybrid of simple network structures
Corollary I	Price cap regulation with no congestion pricing is inefficient and discourages technology adoption
Theorem III	Horizontal integration of collaborating providers (C,A) competing with provider B changes charge structure
Corollary II	Horizontal integration may coordinate charges and improve efficiency
<i>Heterogeneous customers: flexible demand from origin 1 to destination 4, and captive demand from 1 to 2 and from 1 to 3</i>	
Theorem IV	Unregulated competition with heterogeneous customers is inefficient
Theorem V	Horizontal integration of parallel providers A and B improves efficiency
Corollary III	Merger cost savings do not lead to lower charges
Theorem VI	Vertical integration of provider A and customer 1 incentivizes technology adoption

Connecting service providers through simple network structures

We start by showing that competition may lead to higher prices as compared to integration due to the network effects arising from simple serial or simple parallel service providers. As a result, flows under competition are less efficient from the point of view of minimizing total social costs. We recall that efficiency is obtained from the system optimal solution when the service provider charge equals the service cost.

Consider the simple serial service provider case where the potential demand is equal to D per customer, for services from origin 0 to destination 4 using collaborating service providers C and A, i.e. B does not exist in this case. Consider also the simple parallel service provider case where the customers demand services from origin 1 to destination 4 using the competing services of A or B, i.e. C does not exist in this case. In both cases, we assume that the outside option cost, C^T , is sufficiently high such that

all demand is served when the service provider charges are sufficiently low. The assumed symmetry in costs allows us to find the unique, pure strategy solution outcome by analyzing the optimality conditions.

Theorem 1: Simple serial and simple parallel service providers. *There exists a unique user equilibrium, with a potential demand D threshold that determines two cases:*

Case 1 for serial providers, for D below the threshold: (i) the serial providers set high charges including a congestion component that decreases in D . The negative slope decreases in absolute value as the number of customers, n , increases, given fixed total potential demand nD ; and (ii) flows are efficiently increasing in D .

Case 1 for parallel providers, for D below the threshold: (i) the parallel providers set the same low charge including a congestion component that first increases and then decreases in D . The slope in absolute value decreases as the number of customers, n , increases; and (ii) flows are efficiently increasing in D .

Case 2 for serial and parallel providers, for D above the threshold: charges and flows are constant and flows are lower than the efficient levels.

The threshold for the parallel providers is higher than that of the serial providers. The thresholds increase with the outside option cost, C^T , and the number of customers, n . The thresholds decrease as the service cost, C^S , customer operating cost, C^O , and congestion cost, C^G , increase.

Under integration of serial providers compared to no integration, the threshold increases, charges are lower and flows are higher hence closer to the efficient levels. Under integration of the parallel providers as compared to no integration, the outcome is unchanged except in Case 1 where the charges are higher and only decreasing in D .

In the unique user equilibrium of unregulated competition, the service provider charges and customer flows in equilibrium are defined according to two cases. In case 1, the potential demand D is strictly lower than a threshold. In this case, the service providers serve the entire demand and given the symmetry of our simplified network, this means that the parallel providers each serve half the market. Their charge is a function of the service cost and increases in growing demand due to congestion. Serial service providers use their monopoly power to set a high charge, which decreases in demand and the congestion cost parameter, C^G , in order to keep the service sufficiently attractive given the outside option cost, C^T . In contrast, the charge set by parallel providers increases in demand because the providers will only keep their market share by preserving comparable delay levels. Moreover, the user equilibrium outcome is efficient in case 1.

With very low congestion costs, serial providers would set a monopolistic charge and parallel providers would set their charges according to the differentiated Bertrand equilibrium outcome i.e. based on their service costs alone. However, given that congestion does exist in equilibrium, the service provider charges will internalize congestion costs to a degree, as a function of the flow. The extent of internalization depends on the customers' market power, i.e. the number of customers. Under Wardrop equilibria, in which each infinitesimal flow is chosen independently i.e. there are infinitely many small atomistic customers (as n approaches infinity), the charge would include one half of the congestion costs. In our oligopolistic, non-atomistic game with customers exhibiting market power, parallel providers will always charge more than half of the congestion costs. For two customers, the charge would include three-

quarters of the congestion costs. Analogously, serial providers will always reduce the charge by more than half of the congestion costs.

In case 2, where potential demand, D , lies above the threshold, it is preferable for customers to partly forego the service due to excessive congestion. Therefore the threshold increases with the outside option cost and decreases with the various costs associated with the service and the customers' market power (i.e., when each customer's contribution to the total demand becomes smaller as n increases). The congestion is such that demand served will remain constant at the threshold level, thus the service provider charges also remain constant. Now, efficiency implies only serving partial demand. The monopolistic serial providers set a relatively high charge, consequently equilibrium flows are strictly lower than the efficient flows and congestion is inefficiently low too. In contrast, the competing parallel providers set a relatively low charge, however still too high due to customer market power as discussed in the previous paragraph. Moreover, total equilibrium flows are closer to the efficient levels as compared to serial providers.

The effect of integration also depends on the network structure. Integrated serial providers are able to coordinate by lowering their charges and generate efficient flows. In contrast, integration of parallel providers has no effect on flows. The integrated provider uses its monopoly power to increase charges, however only in the case of low demand. For high demand, the charges remain unchanged in order to attract customers despite the relatively high congestion.

Relying on the two simple network structures as a benchmark, we now turn to the two network effects discussed at the beginning of this section.

Asymmetric service collaboration: customers from origin 0 to destination 4 using either the service collaboration of C and A or of C and B

We show that competition may lead to higher prices as compared to horizontal integration (A+C) due to the network effects arising from asymmetric service collaboration (C+A or C+B). As a result, flows under asymmetric competition continue to be less efficient from the point of view of minimizing total social costs. Consider the case where the potential demand is equal to D per customer, for services from origin 0 to destination 4 using either collaborating service providers C+A or C+B (later we consider the case of captive demand). We still assume that the outside option cost, C^T , is sufficiently high such that all demand is served when the service provider charges are sufficiently low.

Theorem II: Unregulated competition under asymmetric service collaboration as a hybrid of simple network structures. *There exists a unique user equilibrium, with a potential demand D threshold, whereby compared to simple network structures (as in Theorem I):*

Case 1, for D below the threshold: (i) the charges of parallel providers A and B increase in D , i.e. qualitatively the same as the simple parallel providers; (ii) the charge of monopolist C decreases in D , i.e. qualitatively the same as the simple serial providers; and (iii) flows are efficiently increasing in D .

Case 2, for D above the threshold: charges and flows are constant and flows lie in-between the two simple network structures thus are lower than the efficient levels.

The threshold lies in-between the thresholds of the two simple network structures.

In the unique user equilibrium of unregulated competition, in case 1 the demand (from origin 0 to destination 4) is strictly lower than the threshold for the potential demand D . In this case, the competing service providers A and B each serve half the market. Their charge increases with growing demand because the service provider benefits from charging for congestion. Service provider C plays a different strategy as it can use its monopoly power to set a high charge, which decreases with growing demand and congestion costs in order to keep the service sufficiently attractive given the outside option cost, C^T . As in the analysis of the simple network structures, the user equilibrium outcome is efficient in case 1.

In case 2, above the threshold for the potential demand D , the demand served and the service provider charges remain constant. Again, efficiency implies only serving partial demand. The generated outcome is a hybrid of the corresponding outcomes in the two simple network structures. Network effects emerge in case 2 because monopolist C sets a relatively high charge. This leads service providers A and B, who have increasing reaction functions due to the differentiated Bertrand competition created by congestion, to also set inefficiently high prices. Consequently, total equilibrium flows are strictly lower than the efficient flows and congestion is inefficiently low too. Additionally, we have the following corollary.

Corollary I: Price cap regulation with no congestion pricing is inefficient and discourages technology adoption

If price cap regulation is imposed such that the price cap is almost equal to the service provider cost, and does not include a congestion cost element, then the price cap will always be lower than the unrestrained equilibrium charges. Consequently, the service provider charges in the user equilibrium will equal the price caps, leading to excessive congestion and inefficient flows from the point of view of minimizing overall social costs. Moreover, the service providers will not be able to cover the set-up cost of technology adoption, leading to the inefficient outcome of no technology adoption.

Next, we consider the outcome under partial horizontal integration of the service providers, in comparison to the fully competitive outcome.

Theorem III: Horizontal integration of collaborating providers (C,A) competing with provider B changes charge structure. *There exists a unique user equilibrium, with a potential demand D threshold, whereby compared to no integration (as in Theorem II):*

Case 1, for D below the threshold: charges and flows are unchanged.

Case 2, for D above the threshold: (i) charges are lower for parallel providers A and B and higher for monopolist C; (ii) flows are higher for integrated providers C and A and lower for B; and (iii) overall flows are higher and lie in-between the two simple network structures thus are closer to the efficient levels.

The potential demand threshold is higher under horizontal integration as compared to no integration of providers C and A, and lies in-between the thresholds of the two simple network structures.

In the unique user equilibrium under horizontal integration of the asymmetrically collaborating service providers A and C, we permit individual charge levels hence there are again two cases. In case 1, when the potential demand D is strictly lower than the threshold, the results are the same as in Theorem II. Providers A and C use the monopoly position of C to extract a maximum surplus from the flow. The

threshold is higher than that of Theorem II. In case 2, when D is above the threshold, we have an asymmetric equilibrium with constant service flows at the threshold level and constant charges. Given Theorem I, integration is potentially beneficial due to the serial component of the hybrid network. Indeed, the network effect of asymmetric service collaboration implies that compared to the no integration results, service provider C increases its charge, consequently the providers A and B reduce their charges, with an even lower charge for A that attracts a higher fraction of the partial demand served. The total charges are lower, hence the total flow is higher. Thus despite remaining inefficient, the total flow is higher compared to no horizontal integration, i.e. closer to the efficient flows. Consequently, we have the following corollary.

Corollary II: Horizontal integration may coordinate charges and improve efficiency

Integrated collaborating service providers are able to coordinate charges to attract more customers, thus leading to improved efficiency in the sense of minimizing overall social costs.

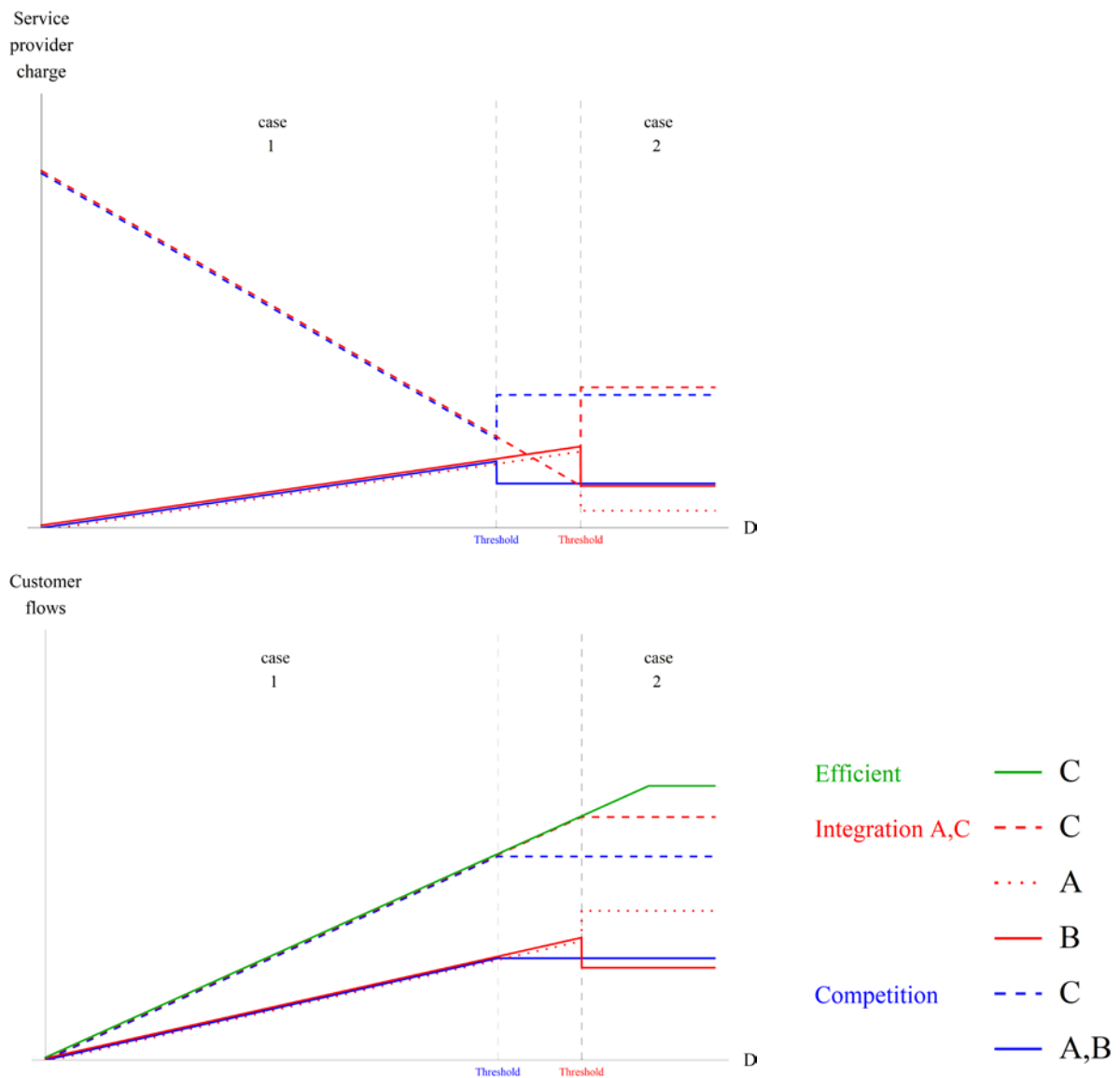


Figure 2: Charges and flows vs. demand

Figure 2 illustrates the charges of service providers and the resulting flows as the demand increases. The blue lines (solid and dashed) represent the results when all three service providers compete, and the red lines (solid, dashed and dotted) represent the results for the merger of service providers A and C. The green line, representing efficient system optimal flows, lies on the 45° line (i.e. customer flows equal potential demand) until it is socially preferable not to serve further demand due to excessive congestion.

Heterogeneous customers: flexible demand from origin 1 to destination 4, and captive demand from 1 to 2 and from 1 to 3

We now show how network effects arising from the combination of captive and flexible demand lead service providers to choose whether or not to compete. Consider the case where the customers demand services from origin 1 to destinations 2 and 3, defined as D_{cap} per customer per OD pair, and flexible services from origin 1 (rather than 0) to destination 4, using either service provider A or B, defined as D_{flex} per customer. Note that provider C is no longer serving the market. We assume that the outside option cost, C^T , is sufficiently high such that all captive demand is served. The results are presented in the three theorems below.

Theorem IV: Unregulated competition with heterogeneous customers is inefficient. *There exists a unique user equilibrium, with a potential flexible demand D_{flex} threshold that determines two cases:*

Case 1, for D_{flex} below the threshold: (i) for very low D_{flex} , providers A and B set high monopoly charges decreasing in D_{cap} hence flows are lower than the efficient levels and no flexible demand is served; and (ii) for higher D_{flex} , either A or B sets a lower monopoly charge than the other provider, consequently serving all D_{flex} and achieving efficient flows.

Case 2, for D_{flex} above the threshold: behavior is similar to competition with a single OD pair demand (as in Theorem II), thus (i) for lower D_{flex} , charges and flows increase then decrease as congestion increases and flows are efficient; and (ii) for the highest D_{flex} levels, charges and flows remain constant and flows are lower than the efficient levels.

In the unique user equilibrium of an unregulated duopoly, flexible demand (from origin 1 to destination 4) will be split between both service providers if and only if it is sufficiently large in comparison to the captive demand, i.e. above the threshold. Below the threshold, two cases are possible. For sufficiently low flexible demand, case 1(i), the service provider charges are extremely high such that only captive demand is served. Given a single service charge level, the service providers A and B each prefer to capture the profits of their own high value demand (the captive demand) rather than reduce the charge in order to capture the flexible demand. The charges decrease with growing demand as a function of the cost, C^T , at which point the customer is indifferent to using the route or foregoing the service. In this case, since the flexible demand is not served, the outcome is inefficient from the point of view of minimizing total social costs. The special case where D_{cap} equals zero is equivalent to the parallel providers in Theorem I, and to the existing analysis in the literature without a network (Brueckner, 2002; Perakis and Sun, 2014). We see here that the network matters because the results of the model are qualitatively different. Only when D_{cap} is positive do we arrive at the threshold below which the service

providers set charges sufficiently high that the flexible demand will not be served, which never happens in existing analyses.

In case 1(ii) however, above a flexible demand threshold, one of the service providers reduces their charges so that the entire flexible flow will use their service. The remaining service provider continues to set a high charge. There is specialization between the two providers, which leads to efficient flows.

In case 2, above the threshold, service provider charges and customer flows in equilibrium are defined according to two subcases. In case 2(i), service provider charges are lower, both serve the flexible demand, and given the symmetry of our simplified network, they each serve half the market. As in Theorem II with low demand, the charge increases with growing demand due to congestion. There is also a range in which both service providers continue to serve the entire flexible market but flows are sufficient such that the service provider charges decrease as flow increases because both providers set the charge such that the customers are indifferent to being served or not. In case 2(ii), the congestion is sufficiently severe that not all flexible demand will be carried and the service provider charge remains constant.

Efficiency implies serving all flexible demand in cases 1 through 2(i) and serving it only partially in case 2(ii). Despite the partial internalization of congestion costs by the customers, the service provider charges in case 2(ii) are too high and flows are inefficiently low. Consequently, the network creates relatively weak competitive effects across service providers.

We now move to the analysis under horizontal integration of the service providers, and compare it to the competitive outcome.

Theorem V: Horizontal integration of parallel providers A and B improves efficiency. *There exists a unique user equilibrium, with a potential flexible demand D_{flex} threshold, whereby compared to no integration (as in Theorem IV):*

Case 1, for D_{flex} below the threshold: a larger interval of D_{flex} suffers a high monopoly charge and no flexible demand is served hence flows are lower than the efficient levels;

Case 2, for D_{flex} above the threshold: charges are weakly higher and flows remain unchanged thus are weakly lower than efficient levels.

The unique user equilibrium under horizontal integration between two competing parallel service providers will lead to the same results as Theorem IV except in cases 1(ii) and 2(i) when the monopolist charges are higher than those of the duopoly case. In other words, as before, the monopolist sets the charge such that customers are indifferent to being served or not. Furthermore, in case 1(ii) an integrated service provider setting a single charge earns lower profits because it cannot replicate the two charge competitive outcome. For high potential flexible demand D_{flex} , the charge remains constant and a decreasing percentage of the flexible traffic is served, again with flows too low from the point of view of minimizing overall social costs.

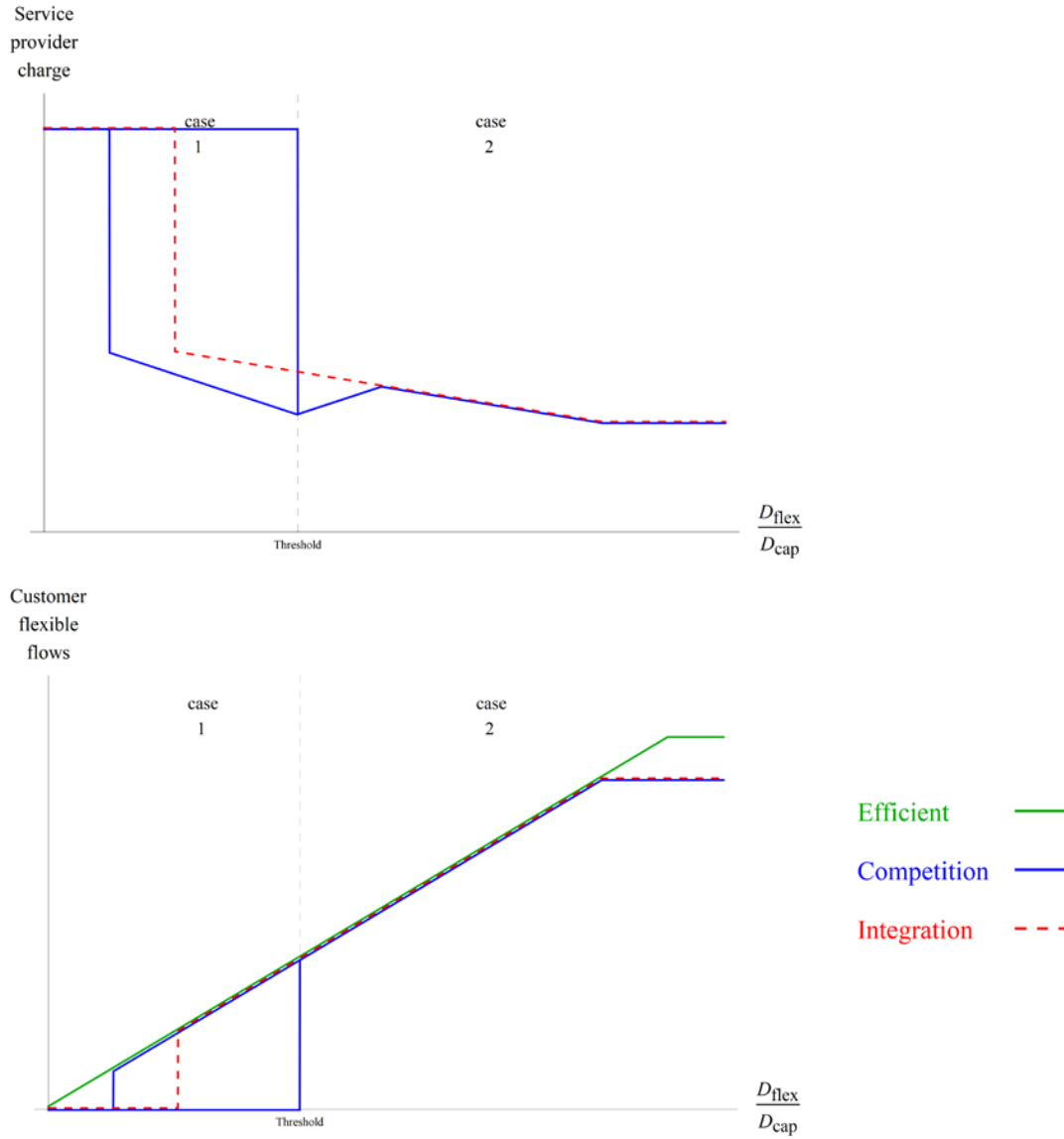


Figure 3: Charges and flows vs. demand ratio

Figure 3 illustrates the charges of the service providers and the resulting flows as the flexible demand increases in comparison to that of captive demand. The blue line represents the results when service providers A and B compete, the dotted red line represents the results for the merger, and the green line represents the efficient, system optimal flows. We see that for the lowest flexible demand, charges are equivalent for the monopolist and duopolist markets both of which serve only captive demand. In case 1(ii), one service provider continues to set the monopolist charge and the other chooses a charge slightly lower than that of the horizontally integrated case, thus carrying all flexible demand. Above the threshold for potential flexible demand, the charges of the duopolists are lower than that of the monopolist, and these charges increase with growing congestion. In case 2(ii), the charges are the same with or without integration.

Corollary III: Merger cost savings do not lead to lower charges

Although the pressure to merge service providers may exist in order to encourage lower service costs due to economies of scale, the service costs of a single server do not impact the charge to the customers (except in case 2(ii) where flexible demand is relatively very high and only served partially). Hence, lower costs will not lead to lower charges without the presence of regulation. Moreover, the merger may not improve the efficiency of the flows from the point of view of minimizing overall social costs.

Finally, in order to examine the incentives to encourage technology adoption, we analyze the case of partial vertical integration. We assume that customer 1 and service provider A reach an agreement whereby a new, capacity increasing technology is adopted exclusively by both parties. The new technology increases the service provider's fixed costs, however reduces variable costs for both the provider and the customer, such that $C_A^S < C^S$ and $C_{1A}^O < C^O$.

Theorem VI: Vertical integration of provider A and customer 1 incentivizes technology adoption. *There exists a unique user equilibrium, whereby compared to parallel competition without technology adoption (as in Theorem IV):*

- (i) *charges decrease by a fraction of the service provider cost saving, $C^S - C_A^S$, with a larger reduction for provider A compared to B; and*
- (ii) *charges change by a fraction of the customer operating cost saving, $C^O - C_{1A}^O$, such that A increases charges whereas B decreases equivalently, and this fraction decreases as the number of customers, n , increase.*

In the unique user equilibrium under duopoly competition in a parallel network with partial vertical integration in which service provider A and customer 1 adopt new technologies jointly, due to the service cost saving for A, the charges set by both service providers decrease, with a higher impact on the charge of A. Furthermore, due to the operating cost saving of customer 1, the charge of service provider A increases whereas service provider B's charge decreases. Even without any service cost savings, the equilibrium charge of A increases in comparison to the outcome without technology adoption and the increase equals a fraction of customer 1's operating cost savings (one sixth when the number of customers $n = 2$). The service provider purchases new technologies, which increase their fixed costs, but this may be offset by the increase in their charges. In order to compete, service provider B will adjust their charges downwards accordingly. We note that for as long as service provider charges are capped at their cost levels, no provider will be interested in signing either horizontal or vertical agreements because cost recovery from the improved service is not possible.

CASE STUDY: AIR TRAFFIC CONTROL IN EUROPE

In the European Union, air traffic control (ATC) is provided by 32 national providers and each have a national monopoly over their territory⁴. The current provision of ATC appears to be 34% more costly when compared to the US system (FAA and Eurocontrol report to the Performance Review Commission, 2013). The ATC company ownership in Europe ranges from state agencies belonging to the Department of Transport, to government-owned corporations, to semi-private firms with for-profit or not-for-profit

⁴ More institutional detail can be found in Baumgartner and Finger (2014)

mandates. Adler et al. (2020) find evidence that public-private ownership with stakeholder involvement achieves statistically significantly higher productive and cost efficiencies than other ownership forms. In this case study, we analyze the potential effects of different regulatory and institutional interventions on the service provision equilibrium.

In the remainder of this section we describe the West European airspace, the six ATC service providers and the five airlines modelled in the case study. Subsequently, we discuss the set of scenarios tested, including a base-run which reproduces the equilibrium outcome in 2011, and the methods applied to solve the model numerically.

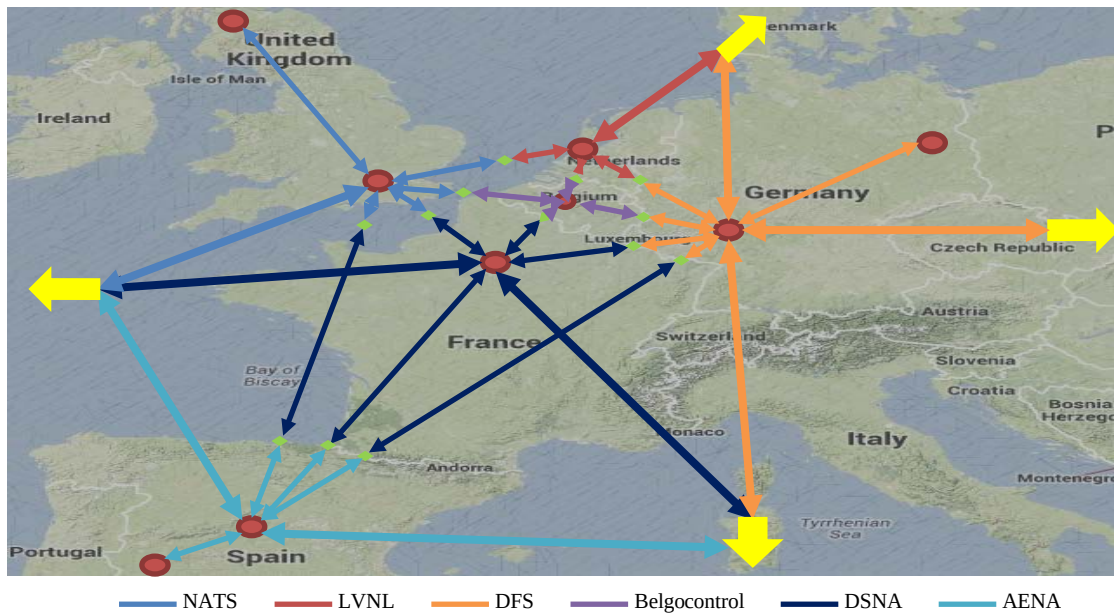


Figure 4: West European air traffic control network case study

IV.1 Network

The set of arcs is partitioned into air traffic control en-route sectors and the airspace above airports. The length of the arcs is measured in km (airport arcs are of distance 1km in order to count movements, i.e., arrivals/departures). The network analysed is depicted in Figure 4 and includes six ATC providers, represented by the coloured arcs, six major airports in each of the six regions, three regional airports and four additional nodes to aggregate flights to and from the region. Despite this being a clear simplification of reality, the network game is sufficiently rich as to enable us to understand the two types of network effects described in Section III, namely asymmetric service collaboration and heterogeneous customers, and how the players will react to changes in institutional or regulatory rules.

IV.2 The ATC service providers

We focus on six ATC providers, including NATS (UK), LVNL (Netherlands), DFS (Germany), Belgocontrol (Belgium), DSNA (France), and AENA (Spain). The activities of the Maastricht Upper Airspace Control Centre (MUAC) has been added to the services provided in the individual countries, namely the Netherlands, Belgium and Germany, in order to reduce the number of players. In 2011,

according to the ACE 2011 Benchmarking Report⁵, these ATC providers were responsible for 48.9% of European traffic (in terms of flight hours controlled) and 52.3% of total en-route ATM/CNS costs. Eurocontrol's performance review unit also publishes the en-route ATFM delay minutes per ATC provider and their costs, which are based on the Cook and Tanner study (2011). Out of the total European ATM system, 62.3% of the delay minutes were attributed to the ATC providers in this case study. Consequently, the total costs to the airlines flying in the relevant airspace as a result of these delays amounted to €933 million per year which mostly draws from additional fuel burn and crew costs. The parameters applied in the network congestion model concerning the ATC providers were drawn from the ACE report and are summarized in tables 3 and 4. Staff and other operating costs constitute the variable costs whereas depreciation, capital and exceptional items were classified as fixed costs.

Table 3: 2011 En-route ATC Service Provision Data

	Revenues (000 €)	Variable Costs (000 €)	Fixed Costs (000 €)	Total Distance (km)	Average Charge per km (τ)	Variable Cost per km (C^S)
NATS	651,366	368,015	153,001	707,474,135	0.921	0.520
LVNL	169,365	131,399	14,302	191,563,198	0.884	0.686
DFS	739,112	658,401	167,398	1,007,485,777	0.734	0.654
Belgocontrol	155,805	111,422	17,331	166,751,138	0.934	0.668
DSNA	1,167,138	804,653	113,876	1,463,618,011	0.797	0.550
AENA	794,710	498,756	135,599	859,175,623	0.925	0.581

The ATC terminal providers cover the nine airports included in Figure 3, however the ACE report publishes aggregated country level data as shown in Table 4. Consequently, the fixed costs for countries in the case study were disaggregated based on the airports' relevant proportions of activities.

The data presented in tables 3 and 4 suggest that cross-subsidizing occurs in specific countries where the service provider produces both en-route and airport terminal services, for example in Belgium. This is likely to impact the choice of investments to be implemented. A sensitivity analysis with respect to the variable cost per movement parameter, C^S , is presented in Section V.

Table 4: 2011 Terminal Air Traffic Control Data

Country	Fixed Costs (000 €)	IFR Airport Movements	Income From Charges per Movement (τ)	Variable Cost per Movement (C^S)
UK	9,863,000	1,746,362	115	87
Netherlands	5,313,000	485,525	113	99
Germany	41,208,000	2,059,372	101	86
Belgium	9,208,000	380,572	69	130
France	31,704,000	1,892,868	122	127
Spain	49,253,000	1,854,896	116	117

IV.3 The airlines as customers of the ATC service providers

Hundreds of airlines fly over European airspace providing both scheduled and charter services. For the sake of simplicity, we aggregate the airlines into three groups which best represent the structure of

⁵ Published in 2013, <https://www.eurocontrol.int/ACE/>

commercial aviation today. The groups include airline alliances, low cost carriers and non-aligned carriers. The aligned group is represented by three airlines: Lufthansa-Brussels (LH), British Airways-Iberia (BA) and Air France-KLM (AF), the main European airlines in the three airline alliances that exist today. Each alliance carrier is modelled with a two-hub system. LH utilizes Frankfurt and Brussels, BA utilizes London and Madrid whilst AF utilizes Paris and Amsterdam. For the purposes of this case, the low cost carrier group is represented by Easyjet (EJ) because the airline was ranked fourth in terms of seat capacity in Western Europe in 2013. Emirates airline was chosen as the representative carrier for the non-aligned carrier group (Rest). The Dubai based airline was ranked first among world airlines in terms of available seat kilometers in 2013 and Europe was their largest market by seat capacity.

We next describe our justifications with respect to the airline related parameters when applying the model to this case study. First, the potential demand D_{load} for flights between each OD pair is set per airline based on their scheduled timetable in 2011, and an airline chooses to fly in the peak or off-peak or not to fly. Given the assumed network, all OD pairs are considered as flexible demand, except for domestic flights⁶ which represent captive demand. Since ATC charges are price capped, and represent a relatively small component of total airline operating costs, we assume that demand elasticity with respect to ATC charges is relatively low. Therefore we set the opportunity cost of not flying, the outside option C_{load}^T per flight, to equal the airline's revenue per available seat kilometer (RASK) in 2011 multiplied by the km distance for the relevant OD. This is sufficiently high to ensure that no flights are cancelled in the base case scenario.

The airline groups achieve different cost levels which are a direct function of the level of service, output, network, average stage length and the employment costs of the airline's country of registration. The variable cost per available seat kilometer (CASK) in 2011 for BA, LH, AF, LC and Rest was 5.7, 10.5, 7.2, 4.2 and 4.5 euro cents, respectively, of which 85% are assumed to be operating costs with the remainder covering congestion and ATC charges (Swan and Adler, 2006). For purposes of simplification, we assume that all airlines fly 150 seat, short-haul aircraft and estimate the operating cost per km C_{la}^O for each airline accordingly.

Congestion impacts the cost categories to varying degrees. To be specific, the more indirect the flight path, the higher the fuel and staff operating costs for the airline. Based on conversations with airline management, and in order to approximately fit the reported CASK in 2011, we assume that the congestion cost parameter C_{ls}^G is 0.1 euro cent per flight per km en-route and 0.07 euro cent per airport movement. Indeed, the greater the delay in airspace, the higher the congestion costs for the airlines, which are frequently more substantial than the ATC service charges (Ball et al. 2010, Cook and Tanner, 2011). Congestion in air transport is caused in part by limited airport capacity, due to runway and terminal handling restrictions, and limited air traffic control capacity in the air. We assume that airport capacity is allocated efficiently across airlines using grandfathered slots. Slots are given to a designated airline

⁶ Domestic flights include the OD pairs LHR-MAN, FRA-BER and MAD-PMI.

and, provided they are utilized 80% of the time per season, continue into perpetuity. This better represents air traffic control congestion everywhere outside the US⁷ where aircraft are served on a first come, first served basis which creates higher demand for air traffic control in the peak period.

Sengupta and Wiggins (2014) discuss the difference in airfares and load factors between peak and off-peak periods. As a first test, we set a revenue loss parameter C_{la2}^R of 20% of the airline's revenue per available km. This ensures a preference to fly in the peak in the base case scenario. Airlines are restricted to a maximal flow in the peak, K_{la1} , over the airport arcs, which is set at 80% of the airline's total movements in 2011 at the six slot restricted airports, based on the data in this case study.

Finally, it is standard practice for airline dispatchers to choose the flight path by balancing all the costs. Table 5 summarizes all the airline cost parameters used in the case study. Sensitivity analyses with respect to the potential demand D_{lod} and cost parameters are presented in Section V.

Table 5: Airline case study cost parameters for 2011

	Airline operating costs in €/km C_{la}^O	Revenue loss from moving to off-peak in €/km C_{law}^R	Congestion cost C_{ls}^G	Outside option cost per flight in €/km C_{lod}^T / distance
BA	7.3	2.3	0.1 € cents / flight / km en-route	11.6
LH	13.4	3.8		19.1
AF	9.2	2.7		13.7
LC	5.3	1.7	0.07 € cents / movement at airport	8.6
Rest	5.7	2.0		9.9
Sources	CASK: financial reports of airlines, Swan and Adler 2006	Sengupta and Wiggins 2014	Ball et al. 2010, Cook and Tanner 2011	RASK: financial reports of airlines

IV.4 Scenarios

In order to analyse the potential impact of changes in institutional or regulatory arrangements, we study four groups of scenarios. The first group is referred to as the base case, in which we reproduce the 2011 equilibrium outcome for the case study depicted in Figure 4. In scenario group 2, we highlight the potential impact of horizontal integration of neighboring service providers. We assume that there will be no savings in labor costs or reduction in air control centers due to the power of the labor unions and the politics of sovereign protection but savings are possible in the sum of the fixed costs due to joint purchasing power and standardization of processes. In scenario group 3 we analyze the potential impact of technology on the equilibrium outcomes by modeling the expected costs and benefits of new technologies to both the ATC providers and the airlines. We note that all parameters in these scenarios draw from the 2012 ATM Masterplan⁸. In scenario group 4, under vertical integration, an ATC provider and its hubbing customers are assumed to adopt new technology and via the best-equipped best-served scheduling rule are able to achieve the benefits of the technology locally.

Within each group of scenarios, we analyse four sub-cases including the system optimum, user equilibrium with cost recovery constraint, and user equilibrium with and without upper bounds on

⁷ In the US, only three airports are currently slot controlled (Adler and Yazhemsky, 2018).

⁸ <https://www.atmmasterplan.eu/>

charges. We recall that system optimal sub-cases assume that a central planner organizes flight paths by minimizing the sum of costs of all airlines. User equilibrium refers to sub-cases in which each airline chooses flight paths taking as given the flows of other airlines. Under cost recovery constraints, we assume that the ATC charges are equal to the current price caps. Current price caps are set such that the ATC providers earn low profit levels, which are expected to be invested in capacity, and ensure that the providers cover their fixed costs and can continue should there be exogenous shocks in the future.

We note that all results presented here were also analyzed using Eurocontrol's demand forecast for the years 2020 and 2030. The results are qualitatively the same, and for reasons of brevity, we do not show most of these results here. For those interested in greater detail, additional results can be found in Adler et al. (2015) ⁹.

IV.5 Numerical solution method

In the numerical analysis developed to search for sub-game perfect equilibrium solutions, the second stage quadratic, convex problem is solved using CPLEX (version 12-9-0), and the first stage is solved using a grid search algorithm. We initiate the algorithm by setting the charges equal to preliminary values¹⁰. We start with the first service provider by testing higher and lower peak and off-peak charges in a 5*5 grid within an initial 25% radius. The charges tested always lie weakly below their respective price caps, which is set to infinity in the unregulated scenarios. Given each pair of charges, the algorithm solves the second stage cost minimization problems to optimality per customer in an iterative process. The iterations continue until all customers keep their chosen path flows compared to the previous iteration. The algorithm then updates the first stage provider's charges to the tested pair of values provided the profit is improved by more than a 2% numerical accuracy threshold. This process continues over the grid within the given radius. The algorithm then moves to the second provider and repeats the process. An entire cycle is completed when all service providers have been analyzed. The search radius is halved to 12.5% when a cycle is completed without any changes in the values of the tested charges or after 15 cycles. The process continues while gradually halving the grid search radius down to 1.5% and then returns to the initial 25% radius repeatedly until convergence, whereby an entire cycle across all radii is completed with no service provider or customer changing any of their decision variables. Computation runtimes for scenarios with price caps were approximately 1 minute using a 3.4GHz processor, as they typically require solving only the second stage of the game. Without price caps, computation runtimes were 1 to 2 days, requiring between 1 to 3 cycles across radii. The algorithm scales linearly in the number of service providers and customers. Additionally, repeating the computation with up to double the initial charges or with numerical accuracy thresholds improved to 0.01% did not affect the outcome. We also changed the grid search to a random search without impacting the outcome but

⁹ Accelerating Change Deliverable D4, <https://www.tmleuven.be/en/project/acchange/projectfile/102> under Horizon 2020 project no. E.02.31.

¹⁰ We started the runs with initial values equal to the 2011 price caps.

adding to computational burden. This suggests the existence of a unique sub-game perfect equilibrium in pure pricing and flow strategies.

CASE STUDY RESULTS

In this section, we first discuss the base run (scenario 1), which replicates the results of 2011. We then discuss horizontal integration through the concept of functional airspace blocks (scenario 2) and the adoption of two levels of technology provision (scenario 3). Finally, we analyze the regional forerunner concept in which a geographically aligned airline, ATC provider and relevant airports reach agreement to jointly adopt new technology (scenario 4).

Scenario Group 1: Base Case

In the base run scenario, the solution closest to the 2011 real world outcome is the user equilibrium with cost recovery constraint. As shown in the Scenario 1 Table, the ATC providers' output (km served), revenues and profits covering the 6 countries included in the analysis represent a close approximation to the outcome for 2011, as detailed in tables 3 and 4. We note that all service providers covered their costs in 2011 except for the German provider. However, in the ATC terminal sector, charges do not fully recover their costs for half the airports hence the losses are either covered through alternative airport revenues (for example on the commercial side) or from the provider's en-route profits, depending on the institutional arrangements that differ across countries. Additionally, each airline's schedule was fully met, with around 80% of the flights occurring during the peak window, and the costs per available seat kilometer (CASK) per airline also matching those that occurred in 2011. The revenue loss column presents the loss due to service in the off-peak.

Scenario 1 Table: User equilibrium cost recovery outcome

Airline	CASK	Annual Costs (000 €)	Revenue Losses (000 €)	ATC en-route	Prices per km	Annual Revenues (000 €)	Annual Profits (000 €)
BA	0.060	5,603,009	259,145	NATS	0.921	780,449	186,804
LH	0.103	8,857,017	403,336	LVNL	0.884	262,947	44,593
AF	0.076	4,429,533	188,759	DFS	0.734	511,924	-111,603
LC	0.047	9,758,897	215,780	Belgocontrol	0.934	127,645	19,022
Rest	0.047	6,920,377	89,710	DSNA	0.797	1,217,587	263,469
				AENA	0.925	417,433	19,641
Total:		35,568,833	1,156,730			3,317,984	421,926

ATC terminal	Prices per movement	Annual Revenues (000 €)	Annual Profits (000 €)
LHR	115	75,135	12,614
AMS	113	53,536	2,034
FRA	101	195,470	19,282
BRU	69	36,921	-38,297
CDG	122	167,714	-19,201
MAD	116	79,599	-12,088
MAN	115	96,710	22,653
BER	101	36,054	493
PMI	116	30,146	-5,044
Total:		771,285	-17,554

In the system optimum, we assume that an organization such as Eurocontrol chooses the flight paths in order to minimize overall social costs. They manage to save a moderate 0.01%. This is in line with Theorem II of Section III, which predicts that charges with a congestion cost component can generate approximately efficient user equilibrium flows. However, flows change slightly such that more flights are funneled through Belgian and German airspace at the expense of the French and Spanish airspace. Therefore, it is unlikely that the airlines and the ATC providers would prefer to leave the choice of flight paths to a central planner because the system optimal approach may result in some actors gaining at the expense of others.

In the user equilibrium price cap approach, charges were limited to 20% higher values than those charged in 2011. Although the ATC providers could opt to charge less and acquire a larger share of en-route traffic, in the sub-game perfect equilibrium they charge according to their upper limit, leading to profits approximately three times higher than those achieved in 2011. These results are in line with Corollary I of Section III. Therefore, the ATC providers could collect additional revenues to fund new technology were this deemed necessary without impacting demand dramatically. The results suggest that the airlines would continue to fly despite a slight increase in their cost per available seat km. The peak and off-peak price cap approach required the providers to set two separate charges, in order to internalize congestion whilst ensuring revenue neutrality from the ATC perspective. We note that the providers will only set two separate charges were the government(s) to require them to do so by enforcing two individual price caps. Under a setting in which prices could increase by up to 20% in the peak but were reduced equivalently in the off-peak, airline costs increase by a range of 1 to 2%. We also note that the slot allocations prevent airlines from increasing movements in the peak and the higher peak charge does not induce airlines to move to the off-peak. In other words, the 20% higher ATC charges are counter balanced by the lower airfares in the off-peak, leading airlines to continue to serve the peak market where possible.

In the user equilibrium profit maximization without price caps, the ATC providers are free to set charges such that they maximize their profits. The results show that the charges would increase up to sixfold and profits accordingly with approximately 60% of flights cancelled. Interestingly, this is the only case in which the ATC providers distinguish between peak and off-peak pricing endogenously which is in accordance with Theorem II whereby the ATC charges of a duopoly partly internalize the external congestion costs. However, the airlines' costs per available seat kilometer increase by 20% and almost all off-peak flights are dropped. Consequently, we arrive at the conclusion that there is insufficient competition across ATC providers in order to justify the removal of price regulation as has occurred in the airline industry globally and in the airport industry in the UK and Australia.

Comparing the outcome of the different cases, efficiency in terms of minimizing total social costs, i.e. the system optimal outcome in which the service provider charge equals the service cost, implies cancelling some of the flights. Consequently, due to ATC charges that are too low, the user equilibria with either cost recovery or price caps generate high levels of congestion as compared to the efficient

flows. In contrast, under user equilibrium without price caps, the high charges generate very low congestion effects as compared to the efficient flows. This is in line with Theorems II and IV, which show that competition may fail to promote efficiency in the presence of the two types of network effects, namely asymmetric service collaboration and heterogeneous customers. In summation, the network inhibits competition such that strong efficiency is not achieved.

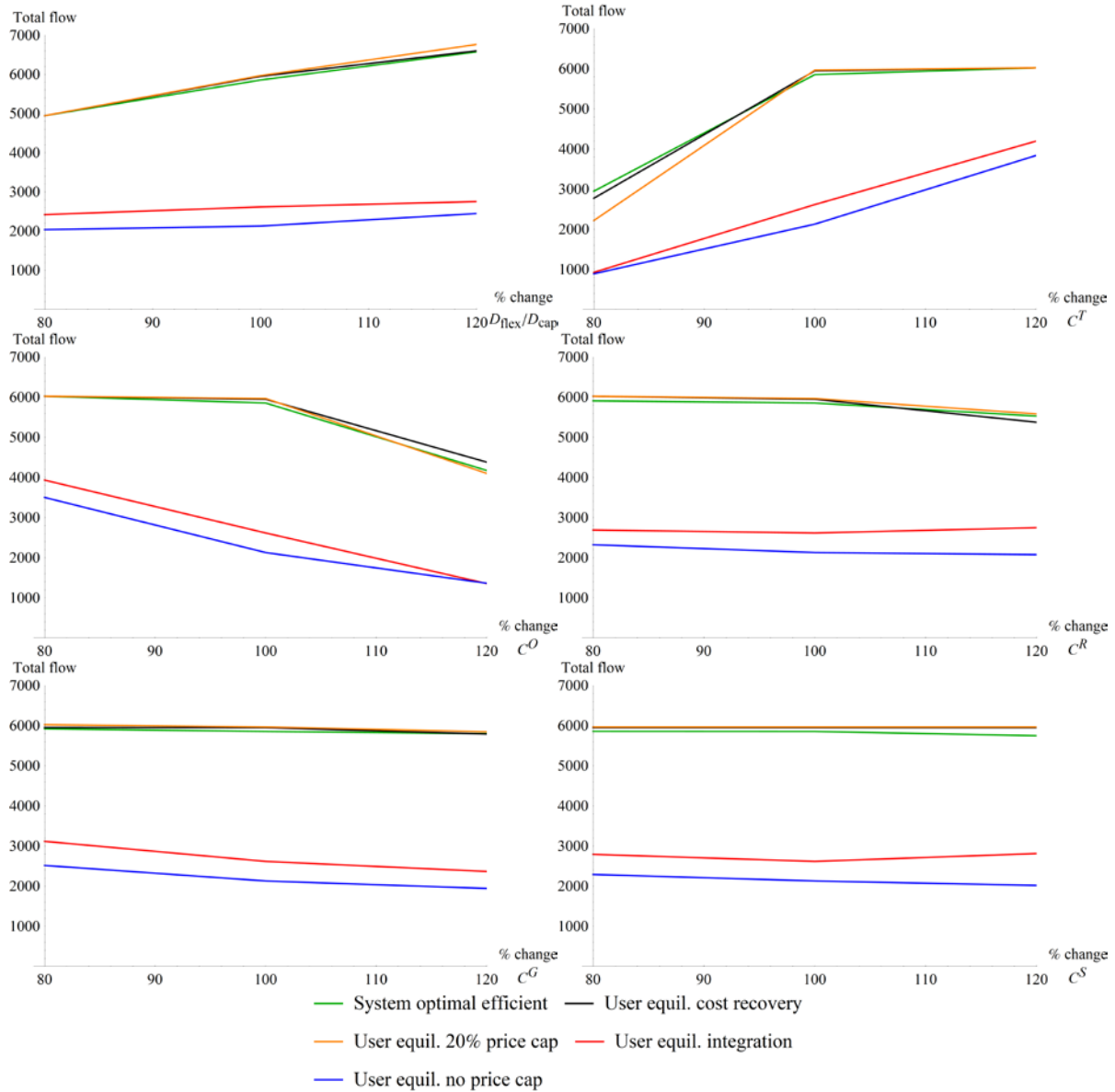


Figure 5: Effect of changes in demand and cost parameters

For all cases, peak demand is limited to 80% of the flows (constraint (3)) based on an analysis of data drawn from the CODA database¹¹. The airlines may also choose to fly off-peak, which would induce revenue losses from lower airfares. For sensitivity analysis, we tested all cases with 50% lower revenue losses with no notable change in airline flight patterns. The results suggest that all airlines prefer to fly in the peak given current ATC charge levels because the cost savings from flying off-peak are insufficient

¹¹ <https://www.eurocontrol.int/articles/coda-publications>

to counter balance the likely revenue losses. With the available data we were able to model a simple peak/off peak differentiation of ATC charges however we note that air traffic congestion is probably more of the bottleneck congestion type¹².

Finally, we examine the effect of changes in each of the demand and cost parameters used in the case study. The total flow results are depicted in Figure 5 and support the findings detailed above¹³. We find similar effects when analyzing the costs and profits. Additionally, total flow increases with increases in the ratio of flexible to captive potential demand D_{lod} or the outside option cost C_{lod}^T , and generally decreases in the airline operating cost C_{la}^O , revenue loss in the off-peak C_{la2}^R , congestion cost parameter C_{ls}^G and ATC service cost C_s^S .

Scenario Group 2: Horizontal integration of service providers

In 2004, the European Union passed a law creating nine functional airspace blocks (FABs) through cross-border merges that were meant to be in place by 2012. Scenario group 2 analyses the possibility that providers' horizontal integration may lead to technology adoption and a reduction in fixed costs. We assume that there will be no changes in labor costs and any cost savings will draw from the ability to purchase equipment jointly, resulting in a 30% saving in fixed costs through co-operation. The parameter is based on a 2012 report from the COOPANS alliance¹⁴ which jointly purchases technologies.

An important question that arises in this scenario is how the merger charge level under price cap regulation would compare to the current price caps. Setting a single rate per unit operation could force harmonization and lead to the use of more direct flight paths. This would appear to be the view of most ATC regulators when conceiving the idea of FABs. We set the price cap on charges per km to the weighted average of the 2011 prices according to the level of activity of each provider. According to Corollary III of Section III, in the user equilibrium price cap outcome, the ATC provider will have no incentives to decrease charges when costs decrease hence they will continue to charge according to the price caps.

In case 2a, we assume that the Dutch and German providers cooperate, resulting in a weighted average charge of 0.758 which increases the costs of flying through German airspace but substantially reduces the price to fly over the Netherlands. The result under the cost recovery constraint (Scenario 2 Table) is an increase in costs for Lufthansa and for the low cost carriers but lower costs for the other carriers. In case 2b we assume that two large service providers cooperate, namely Germany and France, with a weighted average charge of 0.771, which increases costs in Germany and lowers costs in France. As a

¹² In a pure bottleneck, where all users want to use the facility at the same moment and have the same values of time and delay, pricing by the minute is much more powerful than simple peak/ off peak price differentiation. In theory, this would convert all queuing into additional revenues, while the costs for the airlines would not increase as they would merely see their queuing costs converted to ATC charges.

¹³ Of the five policy options presented, the user equilibrium integration without price cap is discussed in Scenario Group 2.

¹⁴ The COOPANS alliance was initiated in 2006 by three European ATC providers together with one technology provider and today consists of six ATC partners. <http://www.coopans.com/About-Coopans/COOPANS-Value> accessed 26/5/2020.

result, all airlines are worse off with their costs increasing by 0.15% to 0.66%. Finally, in the case of scenario 2c, we assume that the Netherlands, Germany and Belgian airspaces cooperate and the weighted average price equals 0.779 cents per km, which increases the cost of German airspace whereas the Netherlands and Belgian airspaces are cheaper for the airlines. As a result, all airlines except for Lufthansa are better off. Consequently, unless some of the cost savings are passed on to the airlines through lower or differentiated ATC charges, at least one or more airlines are worse off as a result of such cooperation, which may explain why the airline industry has not pushed harder for the implementation of the Single European Skies approach. This is an illustration of Corollary III, and this result is also in line with the findings of Castelli et al. (2005).

Scenario 2 Table: Horizontal integration

Base Case		Case 2a		Case 2b		Case 2c	
Airlines	Annual Costs (000 €)	Annual Costs (000 €)	% Change	Annual Costs (000 €)	% Change	Annual Costs (000 €)	% Change
BA	5,603,009	5,601,302	-0.03%	5,618,299	0.27%	5,600,139	-0.05%
LH	8,857,017	8,858,961	0.02%	8,876,560	0.22%	8,857,570	0.01%
AF	4,429,533	4,407,969	-0.49%	4,436,205	0.15%	4,394,666	-0.79%
LC	9,758,897	9,763,962	0.05%	9,796,471	0.39%	9,756,148	-0.03%
Rest	6,920,377	6,898,962	-0.31%	6,965,893	0.66%	6,904,208	-0.23%
Total/Avg	35,568,833	35,531,156	-0.11%	35,693,428	0.35%	35,512,731	-0.16%

ATC	Charge	Annual Revenues (000 €)	Annual Profit	ATC	Charge	Annual Revenues (000 €)	% Change	Annual Profits (000 €)	% Change
<i>Case 2a</i>									
LVNL	0.884	262,947	44,593						
DFS Germany	0.734	511,924	-111,603						
Total:		774,870	-67,009	FAB	0.758	753,684	-2.73%	-28,754	57.1%
<i>Case 2b</i>									
DSNA France	0.797	1,217,587	263,469						
DFS Germany	0.734	511,924	-111,603						
Total:		1,729,511	151,867	FAB	0.771	1,080,816	-37.5%	54,037	-64.4%
<i>Case 2c</i>									
LVNL	0.884	262,947	44,593						
DFS Germany	0.734	511,924	-111,603						
Belgocontrol	0.934	127,645	19,022						
Total:		902,515	-47,987	FAB	0.779	882,209	-2.25%	-4,555	90.5%

Horizontal integration is only likely to occur if the costs to the merged ATC provider are reduced sufficiently that the savings outweigh the reduction in revenues, which would require a minimum reduction in fixed costs of 40% in this case study. Alternatively, FABs should be permitted to differentiate charges on flight legs according to the relevant cost base. This would require a Boiteux-Ramsey mark-up on top of the marginal service costs such that providers' revenues cover their total costs. This merits further exploration but would be a revolution in an industry where average cost pricing is the rule and cooperation among service providers has proven difficult.

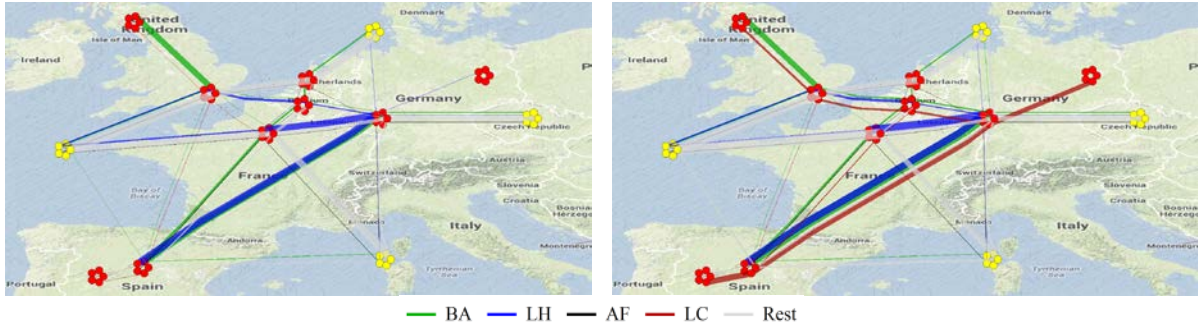


Figure 6: Airline flows for (left) base case and (right) horizontal integration of France & Germany; both without price caps

The charge set by the horizontally integrated provider under user equilibrium without price caps is lower than the pre-integration scenario. The change impacts the charges of all ATC providers remaining in the market, with the British provider reducing the off-peak charge in particular and the Belgian provider halving both charges. We depict the second stage flows of the pre-integration scenario in Figure 6(left) and the results of the French-German integration in Figure 6(right). We see stronger low cost carrier flows in many regions and flows between the UK and Germany being diverted through Belgian airspace. To the benefit of the integrated ATC provider, we also see substantially higher flows from Frankfurt through France to Madrid. Consequently, congestion is higher and closer to the efficient flows. This is in line with Corollary II, which shows that in the presence of network effects generated by asymmetric service collaboration, integrated collaborating service providers are able to coordinate charges and improve efficiency. This general conclusion is robust to changes in all the parameters used in the case study, as shown in Figure 5 for case 2b.

Scenario Group 3: Technology adoption

We analyze the potential impact of technology implementation based on two technology packages: the pilot common project (PCP) and the first step of SESAR as defined in the 2012 ATM Masterplan. The PCP consists of technology adoption approximately equivalent to 10% of the full Step 1 process¹⁵. We note that all parameters in these scenarios draw from the ATM Masterplan. The PCP is expected to cost approximately €2.5 billion of which the service providers cover 65%, the airlines 16% and the airports the remainder. Congestion en-route is estimated to be reduced by 8.7% and the operational costs to the airlines drop by a relatively minor 0.633%, after accounting for the trade-off between the costs of the PCP and the savings from more direct flights which reduce fuel usage. The en-route providers are expected to achieve a reduction in variable operating costs due to improved productivity per ATCO of 8.4% but fixed costs increase by 22% due to the investment in PCP technology. The terminal ATC providers are expected to achieve a 4% increase in airport capacity, a reduction of 8.4% in variable costs and an increase of 104% in fixed costs. The ATM Masterplan estimates assume that the providers will not change their charge levels.

¹⁵ We note that Steps 2 and 3 of the 2012 ATM Masterplan, leading to trajectory based ATC, were defined but not expected to be in place before 2030.

The results are presented in the Scenario 3 Table and show that the overall savings to the airlines outweigh the investment costs and all airlines are slightly better off, with average cost savings of about 1%. Most providers are better off, in particular the smaller providers, but Spain is worse off hence would be unlikely to willingly participate. Based on a sensitivity analysis, allowing the providers to increase their charges by 10% would incentivize participation in the PCP such that the airlines and providers all gain from this effort.

Scenario 3 Table: Adopting SESAR technologies

Base Case			PCP		SESAR Step 1		Base 2030	SESAR Step 1 2030	
Airlines	CASK	Annual Costs (000 €)	Annual Costs (000 €)	% Change	Annual Costs (000 €)	% Change	Annual Costs (000 €)	Annual Costs (000 €)	% Change
BA	0.060	5,603,009	5,562,829	-0.72%	5,530,475	-1.29%	7,629,816	7,625,342	-0.06%
LH	0.103	8,857,017	8,790,824	-0.75%	8,772,368	-0.96%	10,959,094	11,925,258	8.82%
AF	0.076	4,429,533	4,410,945	-0.42%	4,380,460	-1.11%	4,818,975	5,349,012	11.0%
LC	0.047	9,758,897	9,697,369	-0.63%	9,700,925	-0.59%	11,951,931	12,793,221	7.04%
Rest	0.047	6,920,377	6,858,055	-0.90%	6,789,358	-1.89%	9,691,116	9,482,388	-2.15%
Total:		35,568,833	35,320,023	-0.70%	35,173,586	-1.11%	45,050,934	47,275,220	4.72%

Base Case			PCP		SESAR Step 1		Base 2030	SESAR Step 1 2030	
ATC en-route	Price	Annual Profits (000 €)	Annual Profits (000 €)	% Change	Annual Profits (000 €)	% Change	Annual Profits (000 €)	Annual Profits (000 €)	% Change
NATS UK	0.921	186,804	190,516	2%	80,653	-57%	258,271	192,115	-26%
LVNL	0.884	44,593	58,587	31%	36,577	-18%	66,981	59,758	-11%
DFS	0.734	-111,603	-109,902	2%	-211,133	-89%	-97,547	-190,498	-95%
Belgocontrol	0.934	19,022	22,878	20%	8,048	-58%	24,832	19,839	-20%
DSNA	0.797	263,469	309,757	18%	189,191	-28%	371,806	318,276	-14%
AENA	0.925	19,641	11,874	-40%	-70,454	-459%	77,947	-11,643	-115%
Total:		421,926	483,709	15%	32,882	-92%	702,291	387,846	-45%

Base Case			PCP		SESAR Step 1		Base 2030	SESAR Step 1 2030	
ATC terminal	Price	Annual Profits (000 €)	Annual Profits (000 €)	% Change	Annual Profits (000 €)	% Change	Annual Profits (000 €)	Annual Profits (000 €)	% Change
LHR	115	12,614	11,533	-9%	-20,671	-264%	18,588	-14,165	-176%
AMS	113	2,034	1,191	-41%	-24,726	-1316%	4,364	-21,922	-602%
FRA	101	19,282	23,366	21%	-36,225	-288%	22,361	-27,977	-225%
BRU	69	-38,297	-38,336	0%	-68,749	-80%	-40,115	-77,427	-93%
CDG	122	-19,201	-17,294	10%	-88,427	-361%	-19,288	-88,957	-361%
MAD	116	-12,088	-17,202	-42%	-78,270	-547%	-12,340	-77,804	-531%
MAN	115	22,653	27,902	23%	17,810	-21%	25,783	25,368	-2%
BER	101	493	-1,985	-503%	-28,154	-5816%	2,573	-25,926	-1108%
PMI	116	-5,044	-7,465	-48%	-32,889	-552%	-5,144	-32,713	-536%
Total:		-17,554	-18,272	-4%	-360,301		-3,218	-341,522	

Step 1 of SESAR is expected to cost approximately €30 billion by 2030, of which the providers cover 16% and the airlines 50% according to the ATM Masterplan. It is expected that en-route congestion will be reduced by approximately 27% and the operational costs to the airlines increase by a relatively small 0.1%, after accounting for the technology investments less the savings from reductions in fuel usage. The en-route providers are expected to achieve a reduction in variable operating costs due to improved

productivity per ATCO of 8.4% but fixed costs increase by 64% due to the estimated technology investments. The terminal ATC providers are expected to achieve 14% increase in airport capacity, a reduction of 8.4% in variable costs and an increase of 600% in fixed costs. The ATM Masterplan also expects all providers to reduce their charge levels to the airlines by 6.1%. The user equilibrium cost recovery results, with price caps that are equal to the charges in 2011, show that the airlines' costs will decrease by approximately 1.1% overall, hence the airlines should be willing to invest. The results suggest that the reduction in congestion afforded by the new technology and procedure adoption is necessary if the forecasted increase in demand of 38.7% from 2011 to 2030 is to be accommodated. The ATC service providers, whether en-route or terminal, are all worse off after investing in SESAR step 1 projects, although this would be somewhat tempered were demand to increase as expected.

The user equilibrium price cap approach suggests that were the providers permitted to increase their charges by an upper limit of 20%, both the airlines and the providers would be in a position to gain from the new technologies, although the impact on the airlines would now be rather marginal. Relaxing the price caps is justified in order to encourage technology adoption, but the price caps cannot be abolished as long as there is insufficient competition, as suggested in Theorem II and Corollary I of section III.

Scenario Group 4: Regional forerunner

In scenario group 4, we test whether the vertical cooperation between a service provider, a local airline and an airport enables the adoption of the PCP program in a “regional forerunner” approach. We assume that the provider invests in the PCP technology and achieves higher levels of output per controller and that the participating airline achieves slightly lower operating costs and congestion levels, but only on the flight paths associated with the relevant airspace. A useful example of this type of cooperation would be FRAMaK, a Free Route Airspace Project run by a consortium of airspace users and providers (MUAC, the Karlsruhe Upper Area Control Centre and Lufthansa). 298 new direct routes were implemented in 2012, which increased the number of direct cross border routes in the area to a total of 656. The development of cross border routes by FRAMaK created an advantage for Lufthansa, which is the largest airspace user in the Maastricht-Karlsruhe area. Moreover, this led to additional user preferred, cross border routes, under pressure from European airlines and Eurocontrol. By 2014, at least 16 of the 64 European ACCs implemented various new Free Route Operations and savings have been estimated in the range of 150,000 tons of CO₂ equivalent to 37 million Euros¹⁶.

In scenario 4a we analyze a potential German regional forerunner such that DFS, LH, FRA and a secondary German airport cooperate. In scenario 4b we analyze a potential French regional forerunner with DSNA, AF and CDG cooperating and in scenario 4c we analyze a Spanish regional forerunner such that BA-Iberia, AENA, Madrid and a secondary Spanish airport cooperate. From the airline perspective, all airline carriers should be willing to cooperate as their costs are expected to decrease in the region of 1 to 2% in the user equilibrium price cap outcome presented in the Scenario 4 Table. Indeed, the incentive

¹⁶ Safety of Air Navigation, Eurocontrol, Belgium, July 8, 2013.

is likely to be underestimated because the airline's market share will probably increase due to a reduction in congestion thanks to the best-equipped best-served rule, which is not accounted for within the current modeling approach. DFS and DSNÄ are also likely to enjoy incentives from such cooperation, with DFS gaining 1.4% higher profits and DSNÄ gaining a 17% advantage. Indeed the AF-DSNÄ-CDG vertical integration would appear to be particularly positive. These results are in line with Theorem VI of Section III. For the German co-operation to occur, the smaller airports would need to be compensated for their investments. However, the Spanish regional forerunner is less likely because en-route and terminal ATC providers will lose from such cooperation.

Scenario 4 Table: Vertical integration

Base Case			Case 4a		Case 4b		Case 4c	
Airlines	CASK	Annual Costs (000 €)	Annual Costs (000 €)	% Change	Annual Costs (000 €)	% change	% Change	% Change
LH	0.103	8,857,017	8,756,389	-1.14%				
AF	0.076	4,429,533			4,368,634	-1.37%		
BA	0.060	5,603,009					5,560,650	-0.76%

Base Case		Case 4a		Case 4b		Case 4c	
ATC en-route	Annual Profits (000 €)	Annual Profits (000 €)	% Change	Annual Profits (000 €)	% Change	Annual Profits (000 €)	% Change
DFS	-111,603	-110,038	1.40%				
DSNA	263,469			309,599	17.5%		
AENA	19,641					11,814	-39.8%

Base Case		Case 4a		Case 4b		Case 4c	
ATC terminal	Annual Profits (000 €)	Annual Profits (000 €)	% Change	Annual Profits (000 €)	% Change	Annual Profits (000 €)	% Change
FRA	19,282	23,204	20.3%				
BER	493	-1,985	-503%				
CDG	-19,200			-17,302	9.9%		
MAD	-12,088					-17,202	-42.3%
PMI	-5044					-7,465	-48.0%

Summary and conclusions from the case study

An initial lesson learnt is that there is insufficient competition across flight paths in the case study to permit the removal of price regulation since all regions demonstrate strong spatial monopoly power. Air traffic control charges could increase by a magnitude of six beyond current prices were price regulation to be dropped. In the airport industry, the UK removed price regulation from all but one of their airports, arguing that there is sufficient competition for catchment areas and across hubs. This could occur in the ATC sector if and only if there are sufficient alternative flight paths between origin and destination. Consequently, ATC competition is only likely to arise when ATC providers are in a position to compete for services over the same set of flight paths as a function of new trajectory-based technologies.

Second, horizontal integration improves efficiency when congestion is sufficiently high. Although there is an obvious gain from merging operations of neighboring service zones, horizontal integration is unlikely to happen as long as the current practice of setting standard service rates based on the average

costs of the merged providers is applied. The current price cap system, combined with incomplete financial integration, implies that one of the service providers is likely to lose revenue from the merger. Furthermore, the cost of standardizing equipment in the shorter term will likely require subsidies or higher prices, which is in direct opposition to current price cap policy.

Third, there are important cost-efficiency gains if new technologies and more standardized equipment are introduced. This will mainly benefit the airlines that will receive improved service. However, it is the ATC providers that are required to finance the additional equipment costs for the most part. Consequently, there are almost no incentives to introduce these new technologies as long as the service providers are bound to the current price cap policies.

Fourth, a regional forerunner approach, where a large airline combines equipment efforts with its major hub airport and en-route ATC supplier in order to improve the efficiency of its operations, could benefit both the ATC and the local airline which would trigger competition among major hubs and airlines. It would appear that regional forerunners involving a service provider and their largest airline customer may be more successful in achieving the ultimate goal of a Single European Sky than a top-down regulated approach.

CONCLUSIONS

This research is the first to consider general networks, elastic demand for multiple OD pairs and oligopolistic markets in both stages of a game. We analyze a congested network served by providers with monopolistic control over a specific set of arcs. The customers of the service are also non-atomistic hence are likely to internalize at least a part of the congestion in the network.

We are able to draw several theoretical conclusions that depart from known results due to the presence of network effects. We then demonstrate their importance in explaining the fragmented West European air traffic control network. The first insight is that, in the presence of network effects and congestion, introducing competition does not necessarily improve efficiency. In supply chains with serial monopolistic links and parallel, competitive links, flows are only efficient for relatively low demand levels. For higher demand levels, the parallel providers will set high charges including congestion charges and disregard the high charges on the monopolistic links hence total charges will be excessive despite the competition. Charges will be lower and flows will become more efficient should the serial monopolist be permitted to integrate with another part of the supply chain. This goes against mainstream policy thinking that argues for complete separation between the monopolistic service provider and the parallel competitors.

We also show that price caps in congested networks tend to be ineffective as the regulators often fail to consider either the role of congestion pricing or the incentives for technology adoption which require capacity investments that are essential to tackle congested networks. Avoiding the need for price caps would require a change in ownership form in the air traffic control market analyzed here. For example, introducing a time-limited auction process may help, as has begun to occur in the terminal ATC market

in Sweden, the UK, Germany and Spain. This has also proven to be a useful tool in the telecommunications and electricity generation markets.

Another important insight is that when two parallel service suppliers each possess captive demand for their services but compete for flexible demand, the relative importance of captive demand is crucial. The service suppliers may choose to focus on fully exploiting the willingness to pay of the captive demand rather than serving any of the flexible demand. There is competition in principle but it is completely ineffective. This is often forgotten in network analyses.

Furthermore, a merger between parallel providers is shown to be ineffective. However, with non-atomistic demand, coordinated technology adoption by one relatively important customer and one of the competing suppliers may be a beneficial game changer, as long as there is no price cap.

Through the development of a case study of air traffic control in Western Europe, this research has shown that interregional transport operations are seriously hampered by local service monopolies that have little to no incentives to adopt better technologies. These monopolies are particularly strong in scheduled services like air, rail and bus markets that are mainly controlled by public agencies. Privatization of monopolies can only improve overall efficiency if there is a smart regulatory system in place that controls prices and capacities.

ACKNOWLEDGMENTS

This research was part of the results of the Horizon2020 WP-E Research and Long Term Innovation project entitled ACCHANGE (SESAR WP-E.02.31). We thank the consortium partners Transport & Mobility Leuven (specifically Thomas Blondiau and Eef Delhaye), CORE-INVEST and Moving Dot for their insights and Hadar Israeli, Avigail Lithwick and Oren Petraru for their research assistance. We thank Serguei Netessin, an Associate Editor and three anonymous reviewers for their comments which helped to improve the paper immeasurably. We also acknowledge participants at the following conferences for constructive criticism: ATRS (2014), INFORMS (2014), OPTION (2015), the 3rd European Aviation Conference, the 11th USA/Europe ATM Seminar and the Hong Kong Polytechnic on-line Seminar (2020). Nicole Adler also thanks the Recanati Foundation for partial funding of this research.

REFERENCES

- D. Acemoglu and A. Ozdaglar. "Competition and Efficiency in Congested Markets." *Mathematics of Operations Research*, vol. 32 (1), pp. 1-31, 2007a.
- D. Acemoglu and A. Ozdaglar. "Competition in Parallel-Serial Networks." *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 6, pp. 1180-1192, 2007b.
- N. Adler, E. Delhaye, A. Kivel, and S. Proost, "Motivating air navigation service provider performance." *Transportation Research Part A: Policy and Practice*, 2020.
- N. Adler and E. Yazhemsky. "The value of a marginal change in capacity at congested airports." *Transportation Research Part A: Policy and Practice*, vol. 114, pp.154-167, 2018.
- G. Allon and A. Federgruen, "Competition in service industries." *Operations Research*, vol. 55(1), pp. 37–55, 2007.
- R. Arnott, A. de Palma, and R. Lindsey. "Economics of a bottleneck with elastic demand." *American Economic Review*, vol. 83, pp. 161–179, 1993.

- M.O. Ball, C. Barnhart, M. Dresner, M. Hansen, K. Neels, A. Odoni, E. Peterson, L. Sherry, A.A. Trani and B. Zou. "Total delay impact study: a comprehensive assessment of the costs and impacts of flight delay in the United States." UC Berkeley Transportation Library, 2010.
- M. Baumgartner and M. Finger, "The Single European Sky gridlock: a difficult 10-year reform process." *Utilities Policy*, vol. 31, pp. 289-301, 2014.
- S. Borenstein and N.L. Rose. "How airline markets work... or do they? Regulatory reform in the airline industry." In *Economic Regulation and Its Reform: What Have We Learned?*, pp. 63-135. University of Chicago Press, 2014.
- J. Brueckner, "Airport congestion when carriers have market power." *American Economic Review*, vol. 92(5), pp. 1357-1375, 2002.
- J.K. Brueckner and K. Van Dender, "Atomistic congestion tolls at concentrated airports? Seeking a unified view in the internalization debate." *Journal of Urban Economics*, 64(2), pp.288-295, 2008.
- G.P. Cachon and P.T. Harker, "Competition and outsourcing with scale economies." *Management Science*, vol. 48(10), pp. 1314-1333, 2002.
- L. Castelli, P. Debels and W. Ukovich, "Route-charging policies for a central European cross-border upper airspace." *Journal of Air Transport Management*, vol. 11, pp. 432-441, 2005.
- L. Castelli, M. Labbé and A. Violin, "A network pricing formulation for the revenue maximization of European air navigation service providers." *Transportation Research part C: Emerging Technologies*, vol. 33, pp. 214-226, 2013.
- L. Castelli, R. Pesenti and A. Ranieri, "The design of a market mechanism to allocate air traffic flow management slots." *Transportation Research part C: Emerging Technologies*, vol. 19 (5), pp. 931-943, 2011.
- R. Cominetti, J.R. Correa and N.E. Stier-Moses, "The Impact of Oligopolistic Competition in Networks." *Operations Research*, vol. 57 (6), pp. 1421-1437, 2009.
- A.J. Cook and J. Tanner. "European airline delay cost reference values." University of Westminster Working Paper, 2011.
- S. Dafermos. "The general multimodal network equilibrium problem with elastic demand." *Networks*, vol. 12(1), pp. 57-72, 1982.
- EUROCONTROL Long-Term Forecast – Flight Movements 2010 to 2030, <https://www.eurocontrol.int>, 2010.
- EUROCONTROL, "Performance Review Report: An assessment of air traffic management in Europe during the calendar year 2012." Performance Review Commission, 2013.
- EUROCONTROL/FAA, Performance Review Commission. US / Europe comparisons of ATM operational and economic performance, 2013.
- European Commission, Single Sky: Commission acts to unblock congestion in Europe's airspace, MEMO/13/525, June 2013.
- I.L. Glicksberg. "A Further Generalization of the Kakutani Fixed Point Theorem, with Application to Nash Equilibrium Points." *Proceedings of the American Mathematical Society* vol. 3 (1), pp. 170-174, 1952.
- A. Haurie and P. Marcotte, "On the relationship between Nash—Cournot and Wardrop equilibria." *Networks*, vol. 15 (3), pp. 295-308, 1985.
- R. Johari, G.Y. Weintraub and B. Van Roy, "Investment and market structure in industries with congestion." *Operations Research*, vol. 58(5), pp. 1303-1317, 2010.
- R. Jovanović, V. Tošić, M. Čangalović and M. Stanojević, "Anticipatory modulation of air navigation charges to balance the use of airspace network capacities." *Transportation Research part A: Policy and Practice*, vol. 61, pp. 84-99, 2014.
- H.W. Kuhn, "Nonlinear programming: a historical view." In *Traces and Emergence of Nonlinear Programming*, pp. 393-414. Springer Basel, 2014.
- G. Lulli and A. Odoni, "The European Air Traffic Flow Management Problem." *Transportation Science*, vol. 41 (4), pp. 431-443, 2007.
- S. Mandell and S. Proost, "Why truck distance taxes are contagious and drive fuel taxes to the bottom." *Journal of Urban Economics*, vol. 93(C), pp. 1-17, 2016.
- D. Monderer and L. Shapley, "Potential games." *Games and Economic Behavior*, vol. 14, pp. 124-143, 1996.

- S.A. Morrison and C. Winston, "Another Look at Airport Congestion Pricing." *American Economic Review*, 97 (5): 1970-1977, 2007.
- S.A. Morrison and C. Winston, "The effect of FAA expenditures on air travel delays." *Journal of Urban Economics*, vol. 63 pp. 669–678, 2008.
- S. Netessine and R.A. Shumsky, "Revenue Management Games: Horizontal and Vertical Competition." *Management Science*, vol. 51(5), pp. 813–831, 2005.
- G. Perakis and W. Sun, "Efficiency Analysis of Cournot Competition in Service Industries with Congestion." *Management Science*, vol. 60(11), pp. 2684–2700, 2014.
- A. C. Pigou, "The Economics of Welfare." Macmillan, London, UK, 1920.
- R.W. Rosenthal, "A class of games possessing pure-strategy Nash equilibria." *International Journal of Game Theory*, vol. 2, pp. 65-67, 1973.
- N.G. Rupp, "Do carriers internalize congestion costs? Empirical evidence on the internalization question." *Journal of Urban Economics* 65, pp. 24-37, 2009.
- T. Roughgarden and É. Tardos, "How bad is selfish routing? " *Journal of the ACM (JACM)*, vol. 49(2), pp. 236-259, 2002.
- R. Selten, "Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games." *International Journal of Game Theory*, vol. 4, pp. 25-55, 1975.
- A. Sengupta and S.N. Wiggins, "Airline Pricing, Price Dispersion, and Ticket Characteristics On and Off the Internet." *American Economic Journal: Economic Policy*, vol. 6(1), pp. 272-307, 2014.
- W. Swan and N. Adler. "Aircraft trip cost parameters: A function of stage length and seat capacity." *Transportation Research part E: Logistics and Transportation Review*, vol. 42 (2), pp. 105-115, 2006.
- J.G. Wardrop, "Some theoretical aspects of road traffic research." *Proceedings of the Institute of Civil Engineering*, vol. 1, pp. 325–378, 1952.
- C. Winston, "On the Performance of the U.S. Transportation System: Caution Ahead." *Journal of Economic Literature*, vol. 51 (3), pp. 773–824, 2013.
- B. Zou and M. Hansen, "Flight delays, capacity investment and social welfare under air transport supply-demand equilibrium." *Transportation Research Part A: Policy and Practice*, vol. 46 (6), pp. 965-980, 2012.

A Appendix

Proof of Proposition I. The set of continuous, convex, quadratic customer cost objective functions with non-empty, compact, convex feasible sets generated by the linear constraints ensures the existence of a pure strategy equilibrium in the second stage of the game (Glicksberg 1951) given any choices in the first stage. Such an equilibrium is obtained by solving the Karush Kuhn Tucker conditions simultaneously (Kuhn 2014). Furthermore, since the first stage service provider charges affect these conditions additively, the second stage flows are continuous, piecewise linear functions of the first stage charges. Consequently, after substituting the second stage equilibrium actions as a function of the charges, the first stage objectives become continuous, piecewise concave functions of the charges. This continuity, together with the non-empty, compact sets of feasible charges, ensure for the first stage of the game that an equilibrium in mixed pricing strategies always exists. ■

In the remainder of the appendix we provide formal versions and proofs of Theorems I through VI in the main text. Since these results do not involve price-cap regulation, we may assume throughout that the price-cap is sufficiently high so that the corresponding constraint may be ignored. For Theorems I through III we also assume that the outside option cost, C^T , is sufficiently high such that all demand, D , is served when the service provider charges are sufficiently low, specifically

$$C^T \geq 3(C^O + C^S).$$

Theorem A.1 (formal version of Theorem I) *For simple serial service providers, there exists a unique user equilibrium, with a potential demand D threshold that determines two cases: Case 1, low demand, $D < \frac{32(C^T - 3C^O - 3C^S)}{288(n+1)C^G}$ under unregulated competition and $D < \frac{48(C^T - 3C^O - 3C^S)}{288(n+1)C^G}$ under integration:*

$$\begin{aligned}\tau_C^{\text{comp}} &= C^S + \frac{1}{2}[C^T - 3C^O - 3C^S - 6\frac{n+1}{2n}C^G(nD)] \\ \tau_A^{\text{comp}} &= C^S + \frac{1}{4}[C^T - 3C^O - 3C^S - 6\frac{n+1}{2n}C^G(nD)] \\ f_{l,s}^{\text{comp}} &= f_{l,s}^{\text{eff}} = D, \forall l, s.\end{aligned}$$

Case 2, excessive demand, $D \geq \frac{32(C^T - 3C^O - 3C^S)}{288(n+1)C^G}$ under unregulated competition:

$$\begin{aligned}\tau_C^{\text{comp}} &= C^S + \frac{96}{288}(C^T - 3C^O - 3C^S) \\ \tau_A^{\text{comp}} &= C^S + \frac{48}{288}(C^T - 3C^O - 3C^S) \\ f_{l,s}^{\text{comp}} &= \frac{96(C^T - 3C^O - 2\tau_A - \tau_C)}{288(n+1)C^G} = \frac{32(C^T - 3C^O - 3C^S)}{288(n+1)C^G} < \frac{48(C^T - 3C^O - 3C^S)}{288nC^G} = f_{l,s}^{\text{eff}}, \forall l, s,\end{aligned}$$

and $D \geq \frac{48(C^T - 3C^O - 3C^S)}{288(n+1)C^G}$ under integration:

$$\begin{aligned}\tau_C^{\text{int}} &= C^S + \frac{72}{288}(C^T - 3C^O - 3C^S) < \tau_A^{\text{comp}} \\ \tau_A^{\text{int}} &= C^S + \frac{36}{288}(C^T - 3C^O - 3C^S) > \tau_C^{\text{comp}} \\ f_{l,s}^{\text{int}} &= \frac{96(C^T - 3C^O - 2\tau_A - \tau_C)}{288(n+1)C^G} = \frac{48(C^T - 3C^O - 3C^S)}{288(n+1)C^G} < \frac{48(C^T - 3C^O - 3C^S)}{288nC^G} = f_{l,s}^{\text{eff}}, \forall l, s.\end{aligned}$$

For simple parallel service providers, there exists a unique user equilibrium, with a potential demand D threshold that determines two cases:

Case 1, low demand, $D < \frac{144(C^T - 2C^O - 2C^S)}{288(n+1)C^G}$:

$$\begin{aligned}\tau_A^{\text{comp}} &= \tau_B^{\text{comp}} = \begin{cases} C^S + 2\frac{n+1}{2n}C^G(nD) & \text{if } D \leq \frac{96(C^T - 2C^O - 2C^S)}{288(n+1)C^G} \\ C^S + \frac{1}{2}[C^T - 2C^O - 2C^S - 2\frac{n+1}{2n}C^G(nD)] & \text{if } D > \frac{96(C^T - 2C^O - 2C^S)}{288(n+1)C^G} \end{cases} \\ f_{l,A}^{\text{comp}} &= \frac{1}{2}D + \frac{\tau_B - \tau_A}{2(n+1)C^G} = \frac{1}{2}D = f_{l,A}^{\text{eff}}, \forall l \\ f_{l,B}^{\text{comp}} &= D - f_{l,A}^{\text{comp}} = \frac{1}{2}D = f_{l,B}^{\text{eff}}, \forall l.\end{aligned}$$

Case 2, excessive demand, $D \geq \frac{144(C^T - 2C^O - 2C^S)}{288(n+1)C^G}$:

$$\begin{aligned}\tau_A^{\text{comp}} &= \tau_B^{\text{comp}} = C^S + \frac{1}{4}(C^T - 2C^O - 2C^S) \\ f_{l,A}^{\text{comp}} &= \frac{144(C^T - 2C^O - 2\tau_A)}{288(n+1)C^G} = \frac{72(C^T - 2C^O - 2C^S)}{288(n+1)C^G} < \frac{72(C^T - 2C^O - 2C^S)}{288nC^G} = f_{l,A}^{\text{eff}}, \forall l \\ f_{l,B}^{\text{comp}} &= \frac{144(C^T - 2C^O - 2\tau_B)}{288(n+1)C^G} = \frac{72(C^T - 2C^O - 2C^S)}{288(n+1)C^G} < \frac{72(C^T - 2C^O - 2C^S)}{288nC^G} = f_{l,B}^{\text{eff}}, \forall l.\end{aligned}$$

Under integration, the outcome is unchanged except that the charges in Case 1 are always as in when $D > \frac{96(C^T - 2C^O - 2C^S)}{288(n+1)C^G}$.

Proof. Simple parallel service providers is the special case of $D_{\text{cap}} = 0$ in Theorems A.4 and A.5, thus for the remainder of the proof we concentrate on serial service providers. The equilibrium is constructed using a folding back procedure. Under unregulated competition, in the second stage of the game customer l takes as given the choices of other customers and solves the problem

$$\begin{aligned}\min \quad & \left(C^O + C^G(f_l + \sum_{l' \neq l} f_{l'}) + \tau_C \right) f_l + 2 \left(C^O + C^G(f_l + \sum_{l' \neq l} f_{l'}) + \tau_A \right) f_l + C^T(D - f_l) \\ \text{s.t.:} \quad & 0 \leq f_l \leq D.\end{aligned}$$

Since there exists a symmetric equilibrium by the assumed symmetry, any index l may be omitted. As the objective function is convex and the constraints are linear, the solution is obtained using the first order conditions

$$0 = 3C^O + 3(n+1)C^G f + \tau_C + 2\tau_A - C^T - \lambda + \mu,$$

where the Lagrange multipliers are μ for the demand D constraint and λ for the non-negativity constraint.

Considering all possible cases for the KKT conditions, we find the solution

$$f^{comp} = \begin{cases} D & \text{if } 0 \leq \tau_C + 2\tau_A \leq C^T - 3C^O - 6\frac{n+1}{2n}C^G(nD), \\ \frac{C^T - 3C^O - \tau_C - 2\tau_A}{3(n+1)C^G} & \text{if } C^T - 3C^O - 6\frac{n+1}{2n}C^G(nD) \leq \tau_C + 2\tau_A \leq C^T - 3C^O, \\ 0 & \text{if } \tau_C + 2\tau_A \geq C^T - 3C^O. \end{cases}$$

Next we analyze the first stage of the game. Each service providers s , taking the charge of all other service provider as given, sets the charge τ_s to maximize their own profit, i.e.,

$$(\tau_C - C^S)n f^{comp} \text{ and } (\tau_A - C^S)2n f^{comp}$$

for $s = C, A$, respectively. We analyze the best response maximal profit depending on the charge regions for f^{comp} . For $\tau_C + 2\tau_A \geq C^T - 3C^O$, $f^{comp} = 0$, so the profit is 0 for all s . When $f^{comp} > 0$, there are two potential solutions:

(a) Given that τ_C, τ_A are in the range for which $f^{comp} = \frac{C^T - 3C^O - \tau_C - 2\tau_A}{3(n+1)C^G}$, maximizing simultaneously the concave profit functions $(\tau_C - C^S)n(\frac{C^T - 3C^O - \tau_C - 2\tau_A}{3(n+1)C^G})$ and $2(\tau_A - C^S)n(\frac{C^T - 3C^O - \tau_C - 2\tau_A}{3(n+1)C^G})$ with respect to the relevant charge, the first order conditions imply $\tau_C^{comp} = C^S + \frac{96(C^T - 3C^O - 3C^S)}{288}$ and $\tau_A^{comp} = C^S + \frac{48(C^T - 3C^O - 3C^S)}{288}$, consequently $f^{comp} = \frac{32(C^T - 3C^O - 3C^S)}{288}$. This is possible if and only if $D \geq \frac{32(C^T - 3C^O - 3C^S)}{288} \geq 0$, in order to have $0 \leq f^{comp} \leq D$.

(b) Given that τ_C, τ_A are in the range for which $f^{comp} = D$, the profit function of service provider s is increasing in τ_s , thus the τ_s attain their maximal value in this region consistently with the case (a), i.e. $\tau_C^{comp} = C^S + \frac{1}{2}[C^T - 3C^O - 3C^S - 6\frac{n+1}{2n}C^G(nD)]$ and $\tau_A^{comp} = C^S + \frac{1}{4}[C^T - 3C^O - 3C^S - 6\frac{n+1}{2n}C^G(nD)]$.

Thus we consider two cases depending on the value of D :

(Case 2) $D \geq \frac{32(C^T - 3C^O - 3C^S)}{288} \geq 0$. In this case, by concavity, the only relevant solution is (a). The profits of each service provider is $\frac{n}{n+1} \frac{(C^T - 3C^O - 3C^S)^2}{27C^G}$.

(Case 1) $0 \leq D < \frac{32(C^T - 3C^O - 3C^S)}{288}$. In this case, the only possible solution is (b). The profit of each service provider is $\frac{1}{2}[C^T - 3C^O - 3C^S - 6\frac{n+1}{2n}C^G(nD)]nD$.

Finally, the system optimal solution flows are obtained by considering the charge dependent user equilibrium flow expressions in f_l^{comp} and replacing the term $n+1$ everywhere by the term $2n$, representing full internalization of congestion costs. The efficient flows f_l^{eff} that minimize total social costs are then derived when setting the charge $\tau_s = C^S$ for each s .

Under integration, the analysis of the second stage of the game is unchanged. In the first stage, the integrated service provider sets their charge to maximize their total profit, i.e.,

$$(\tau_C + 2\tau_A - 3C^S)n f^{int}.$$

As in the competition case, when $\tau_C + 2\tau_A \geq C^T - 3C^O$, $f^{int} = 0$ and the profit is 0. When $f^{int} > 0$, there are two potential solutions:

(a') Given that τ_C, τ_A are in the range for which $f^{int} = \frac{C^T - 3C^O - \tau_C - 2\tau_A}{3(n+1)C^G}$, maximizing simultane-

ously the concave profit function $(\tau_C + 2\tau_A - 3C^S)2n \frac{C^T - 3C^O - \tau_C - 2\tau_A}{3(n+1)C^G}$ with respect to the charges, the first order conditions imply $\tau_C^{int} = C^S + \frac{72(C^T - 3C^O - 3C^S)}{288}$ and $\tau_A^{int} = C^S + \frac{36(C^T - 3C^O - 3C^S)}{288}$, consequently $f^{int} = \frac{48(C^T - 3C^O - 3C^S)}{288}$. This is possible if and only if $D \geq \frac{48(C^T - 3C^O - 3C^S)}{288} \geq 0$, in order to have $0 \leq f^{int} \leq D$.

(b') Given that τ_C, τ_A are in the range for which $f^{int} = D$, the profit is increasing in τ_C, τ_A , thus the τ_s attain their maximal value in this region consistently with the case (a'), i.e. the same as in the competition case.

Thus we consider several cases depending on the values of D :

(Case 2) $D \geq \frac{48(C^T - 3C^O - 3C^S)}{288} \geq 0$. In this case, by concavity, the only relevant solution is solution (a'). The profit of the integrated service provider is $\frac{n}{n+1} \frac{(C^T - 3C^O - 3C^S)^2}{12C^G}$.

(Case 1) $0 \leq D < \frac{48(C^T - 3C^O - 3C^S)}{288}$. In this case, the only possible solution is (b'). The profit is $[C^T - 3C^O - 3C^S - 6\frac{n+1}{2n}C^G(nD)]nD$. ■

Theorem A.2 (formal version of Theorem II) *There exists a unique user equilibrium, with a potential demand D threshold that determines two cases:*

Case 1, low demand, $D < \frac{54(C^T - 3C^O - 3C^S)}{288(n+1)C^G}$:

$$\begin{aligned}\tau_A^{\text{comp}} &= \tau_B^{\text{comp}} = C^S + 2\frac{n+1}{2n}C^G(nD) \\ \tau_C^{\text{comp}} &= C^S + [C^T - 3C^O - 3C^S - 8\frac{n+1}{2n}C^G(nD)] \\ f_{l,A}^{\text{comp}} &= f_{l,A}^{\text{eff}} = \frac{1}{2}D + \frac{\tau_B - \tau_A}{2(n+1)C^G}, \forall l \\ f_{l,B}^{\text{comp}} &= f_{l,B}^{\text{eff}} = D - f_{l,A}^{\text{comp}}, \forall l \\ f_{l,C}^{\text{comp}} &= f_{l,C}^{\text{eff}} = f_{l,A}^{\text{comp}} + f_{l,B}^{\text{comp}} = f_{l,A}^{\text{eff}} + f_{l,B}^{\text{eff}} = D, \forall l.\end{aligned}$$

Case 2, excessive demand, $D \geq \frac{54(C^T - 3C^O - 3C^S)}{288(n+1)C^G}$:

$$\begin{aligned}\tau_A^{\text{comp}} &= \tau_B^{\text{comp}} = C^S + \frac{36}{288}(C^T - 3C^O - 3C^S) \\ \tau_C^{\text{comp}} &= C^S + \frac{108}{288}(C^T - 3C^O - 3C^S) \\ f_{l,A}^{\text{comp}} &= \frac{72(C^T - 3C^O - 3\tau_A + \tau_B - \tau_C)}{288(n+1)C^G} = \frac{27(C^T - 3C^O - 3C^S)}{288(n+1)C^G} < \frac{36(C^T - 3C^O - 3C^S)}{288nC^G} = f_{l,A}^{\text{eff}}, \forall l \\ f_{l,B}^{\text{comp}} &= \frac{72(C^T - 3C^O + \tau_A - 3\tau_B - \tau_C)}{288(n+1)C^G} = \frac{27(C^T - 3C^O - 3C^S)}{288(n+1)C^G} < \frac{36(C^T - 3C^O - 3C^S)}{288nC^G} = f_{l,B}^{\text{eff}}, \forall l \\ f_{l,C}^{\text{comp}} &= f_{l,A}^{\text{comp}} + f_{l,B}^{\text{comp}} = \frac{54(C^T - 3C^O - 3C^S)}{288(n+1)C^G} < \frac{72(C^T - 3C^O - 3C^S)}{288nC^G} = f_{l,C}^{\text{eff}}, \forall l.\end{aligned}$$

Proof. The equilibrium is constructed using a folding back procedure. Under the simplified network, in the second stage of the game customer l takes as given the choices of other customers

and solves the problem

$$\begin{aligned}
\min \quad & \left(C^O + C^G[f_{l,A} + f_{l,B} + \sum_{l' \neq l} (f_{l',A} + f_{l',B})] + \tau_C \right) (f_{l,A} + f_{l,B}) \\
& + 2 \left(C^O + C^G[f_{l,A} + \sum_{l' \neq l} f_{l',A}] + \tau_A \right) f_{l,A} + 2 \left(C^O + C^G[f_{l,B} + \sum_{l' \neq l} f_{l',B}] + \tau_B \right) f_{l,B} \\
& + C^T(D - f_{l,A} - f_{l,B}) \\
\text{s.t.} \quad & f_{l,A} + f_{l,B} \leq D \\
& f_{l,A}, f_{l,B} \geq 0.
\end{aligned}$$

Since there exists a symmetric equilibrium by the assumed symmetry, any index l may be omitted. As the objective function is convex and the constraints are linear, the solution is obtained using the first order conditions

$$\begin{aligned}
0 &= 3C^O + (n+1)C^G(3f_A + f_B) + 2\tau_A + \tau_C - C^T - \lambda_A + \mu \\
&= 3C^O + (n+1)C^G(f_A + 3f_B) + 2\tau_B + \tau_C - C^T - \lambda_B + \mu,
\end{aligned}$$

where the Lagrange multipliers are μ for the first constraint and λ_A, λ_B for the non-negativity constraints.

Considering all possible cases for the KKT conditions, we find the solution

$$f_B^{comp} = \begin{cases} D & \text{if } 0 \leq \tau_B \leq \tau_A - 2\frac{n+1}{2n}C^G(nD) \text{ and } \tau_B \leq \frac{1}{2}[C^T - 3C^O - \tau_C - 6\frac{n+1}{2n}C^G(nD)], \\ \frac{1}{2}D + \frac{\tau_A - \tau_B}{2(n+1)C^G} & \text{if } \tau_A - 2\frac{n+1}{2n}C^G(nD) \leq \tau_B \leq \tau_A + 2\frac{n+1}{2n}C^G(nD) \text{ and } \tau_B \leq C^T - 3C^O - 4\frac{n+1}{2n}C^G(nD) - \tau_A - \tau_C, \\ \frac{C^T - 3C^O - 2\tau_B - \tau_C}{3(n+1)C^G} & \text{if } \tau_B \geq \frac{1}{2}[C^T - 3C^O - \tau_C - 6\frac{n+1}{2n}C^G(nD)] \text{ and } \tau_B \leq \frac{1}{2}(C^T - 3C^O - \tau_C) \text{ and } \tau_B \leq 3\tau_A - (C^T - 3C^O - \tau_C), \\ \frac{C^T - 3C^O + \tau_A - 3\tau_B - \tau_C}{4(n+1)C^G} & \text{if } \tau_B \leq \frac{1}{3}(C^T - 3C^O + \tau_A - \tau_C) \text{ and } \tau_B \geq C^T - 3C^O - 4\frac{n+1}{2n}C^G(nD) - \tau_A - \tau_C \text{ and } \tau_B \geq 3\tau_A - (C^T - 3C^O - \tau_C), \\ 0 & \text{if } \tau_B \geq \frac{1}{2}(C^T - 3C^O - \tau_C) \text{ or } \tau_B \geq \tau_A + 2\frac{n+1}{2n}C^G(nD) \text{ or } \tau_B \geq \frac{1}{3}(C^T - 3C^O + \tau_A - \tau_C), \end{cases}$$

where f_A^{comp} have the same expressions except for swapping everywhere the indices A and B .

Next we analyze the first stage of the game. Each service providers s , taking the charge of all other service provider as given, sets the charge τ_s to maximize their own profit, i.e.,

$$(\tau_A - C^S)2nf_A^{comp}, (\tau_B - C^S)2nf_B^{comp} \text{ and } (\tau_C - C^S)n(f_A^{comp} + f_B^{comp})$$

for $s = A, B, C$, respectively. We analyze the best response maximal profit depending on the charge regions for f_s^{comp} . Since there exists a symmetric equilibrium by the assumed symmetry of service providers A, B , we may consider only cases where $\tau_A = \tau_B$. For $\tau_A = \tau_B \geq \frac{1}{2}(C^T - 3C^O - \tau_C)$, $f_s^{comp} = 0$, so the profit is 0 for all s . When $f_s^{comp} > 0$, there are two potential solutions:

- (a) Given that τ_A, τ_B are in the range for which $f_A^{comp} = \frac{C^T - 3C^O + \tau_B - 3\tau_A - \tau_C}{4(n+1)C^G}$ and $f_B^{comp} = \frac{C^T - 3C^O + \tau_A - 3\tau_B - \tau_C}{4(n+1)C^G}$, maximizing simultaneously the concave profit functions $(\tau_A - C^S)2n(\frac{C^T - 3C^O + \tau_B - 3\tau_A - \tau_C}{4(n+1)C^G})$, $(\tau_B - C^S)2n(\frac{C^T - 3C^O + \tau_A - 3\tau_B - \tau_C}{4(n+1)C^G})$ and $(\tau_C - C^S)2n(\frac{C^T - 3C^O - \tau_B - \tau_A - \tau_C}{4(n+1)C^G})$ with respect to the relevant charge, the first order conditions imply $\tau_A^{comp} = \tau_B^{comp} = C^S + \frac{36(C^T - 3C^O - 3C^S)}{288}$ and $\tau_C^{comp} = C^S + \frac{108(C^T - 3C^O - 3C^S)}{288}$, consequently $f_A^{comp} = f_B^{comp} = \frac{27(C^T - 3C^O - 3C^S)}{288}$. This is possible if and only if $D \geq \frac{54(C^T - 3C^O - 3C^S)}{288} \geq 0$, in order to have $0 \leq f_A^{comp} + f_B^{comp} \leq D$.
- (b) Given that τ_A, τ_B are in the range for which $f_A^{comp} = \frac{1}{2}D + \frac{\tau_B - \tau_A}{2(n+1)C^G}$ and $f_B^{comp} = \frac{1}{2}D + \frac{\tau_A - \tau_B}{2(n+1)C^G}$, the profit function of service provider C is increasing in τ_C because $f_A^{comp} + f_B^{comp} = D$, thus τ_C attains its maximal value $C^T - 3C^O - 4\frac{n+1}{2n}C^G(nD) - \tau_A - \tau_B$ in this region. Under this restriction, maximizing simultaneously the service providers' concave profit functions $(\tau_A - C^S)n(D + \frac{\tau_B - \tau_A}{(n+1)C^G})$ and $(\tau_B - C^S)n(D + \frac{\tau_A - \tau_B}{(n+1)C^G})$, the first order conditions imply $\tau_A^{comp} = \tau_B^{comp} = C^S + 2\frac{n+1}{2n}C^G(nD)$ and $\tau_C^{comp} = C^S + [C^T - 3C^O - 3C^S - 8\frac{n+1}{2n}C^G(nD)]$, consequently $f_A^{comp} = f_B^{comp} = \frac{1}{2}D$.

Thus we consider two cases depending on the value of D :

(Case 2) $D \geq \frac{54(C^T - 3C^O - 3C^S)}{288} \geq 0$. In this case, by concavity, the only relevant solution is (a). The profits are $\frac{n}{n+1} \frac{486(C^T - 3C^O - 3C^S)^2}{20736C^G}$ for service providers A, B and $\frac{n}{n+1} \frac{1458(C^T - 3C^O - 3C^S)^2}{20736C^G}$ for C .

(Case 1) $0 \leq D < \frac{54(C^T - 3C^O - 3C^S)}{288}$. In this case, the only possible solution is (b). The profits are $2\frac{n+1}{2n}C^G(nD)^2$ for service providers A, B and $[C^T - 3C^O - 3C^S - 8\frac{n+1}{2n}C^G(nD)]nD$. When $D = \frac{54(C^T - 3C^O - 3C^S)}{288}$, the profit of service provider C in solution (a) is strictly higher than that of solution (b), so C dictates solution (a) by setting the charge accordingly.

Finally, the system optimal solution flows are obtained by considering the charge dependent user equilibrium flow expressions in $f_{l,s}^{comp}$ and replacing the term $n+1$ everywhere by the term $2n$, representing full internalization of congestion costs. The efficient flows $f_{l,s}^{eff}$ that minimize total social costs are then derived when setting the charge $\tau_s = C^S$ for all s . ■

Theorem A.3 (formal version of Theorem III) *There exists a unique user equilibrium, with a potential demand D threshold, whereby compared to no integration (as in Theorem A.2):*
Case 1, low demand, $D < \frac{64(C^T - 3C^O - 3C^S)}{288(n+1)C^G}$: the outcome is as in case 1 of Theorem A.2.

Case 2, excessive demand, $D \geq \frac{64(C^T - 3C^O - 3C^S)}{288(n+1)C^G}$.

$$\begin{aligned}
\tau_A^{\text{int } A,C} &= C^S + \frac{16}{288}(C^T - 3C^O - 3C^S) < \tau_A^{\text{comp}} \\
\tau_B^{\text{int } A,C} &= C^S + \frac{32}{288}(C^T - 3C^O - 3C^S) < \tau_B^{\text{comp}} \\
\tau_C^{\text{int } A,C} &= C^S + \frac{112}{288}(C^T - 3C^O - 3C^S) > \tau_C^{\text{comp}} \\
f_{l,A}^{\text{int } A,C} &= \frac{72(C^T - 3C^O - 3\tau_A + \tau_B - \tau_C)}{288(n+1)C^G} = \frac{40(C^T - 3C^O - 3C^S)}{288(n+1)C^G} > f_{l,A}^{\text{eff}}, \forall l \\
f_{l,B}^{\text{int } A,C} &= \frac{72(C^T - 3C^O + \tau_A - 3\tau_B - \tau_C)}{288(n+1)C^G} = \frac{24(C^T - 3C^O - 3C^S)}{288(n+1)C^G} < f_{l,B}^{\text{eff}}, \forall l \\
f_{l,C}^{\text{int } A,C} &= f_{l,A}^{\text{int } A,C} + f_{l,B}^{\text{int } A,C} = \frac{64(C^T - 3C^O - 3C^S)}{288(n+1)C^G} < f_{l,C}^{\text{eff}}, \forall l.
\end{aligned}$$

Proof. The equilibrium is constructed using a folding back procedure. Under the simplified network, the analysis of the second stage of the game is the same as in Theorem A.2. In the first stage, the integrated service provider A, C and service provider B , taking the charge of the opponent as given, each set their charge to maximize their own profit, i.e.,

$$(\tau_A - C^S)2nf_A^{\text{int } A,C} + (\tau_C - C^S)n(f_A^{\text{int } A,C} + f_B^{\text{int } A,C}) \text{ and } (\tau_B - C^S)2nf_B^{\text{int } A,C},$$

respectively. As in Theorem A.2, when τ_A or τ_B are weakly higher than $\frac{1}{2}(C^T - 3C^O - \tau_C)$, $f_s^{\text{int } A,C} = 0$ and the profit is 0. When $f_s^{\text{int } A,C} > 0$, there are two potential solutions:

(a') Given that τ_A, τ_B are in the range for which $f_A^{\text{int } A,C} = \frac{C^T - 3C^O + \tau_B - 3\tau_A - \tau_C}{4(n+1)C^G}$ and $f_B^{\text{int } A,C} = \frac{C^T - 3C^O + \tau_A - 3\tau_B - \tau_C}{4(n+1)C^G}$, maximizing simultaneously the concave profit functions $(\tau_A - C^S)2n\frac{C^T - 3C^O + \tau_B - 3\tau_A - \tau_C}{4(n+1)C^G} + (\tau_C - C^S)2n\frac{C^T - 3C^O - \tau_B - \tau_A - \tau_C}{4(n+1)C^G}$ and $(\tau_B - C^S)2n\frac{C^T - 3C^O + \tau_A - 3\tau_B - \tau_C}{4(n+1)C^G}$ with respect to the relevant charge, the first order conditions imply $\tau_A^{\text{int } A,C} = C^S + \frac{16(C^T - 3C^O - 3C^S)}{288}$, $\tau_B^{\text{int } A,C} = C^S + \frac{32(C^T - 3C^O - 3C^S)}{288}$ and $\tau_C^{\text{int } A,C} = C^S + \frac{112(C^T - 3C^O - 3C^S)}{288}$, consequently $f_A^{\text{int } A,C} = \frac{40(C^T - 3C^O - 3C^S)}{288}$ and $f_B^{\text{int } A,C} = \frac{24(C^T - 3C^O - 3C^S)}{288}$. This is possible if and only if $D \geq \frac{64(C^T - 3C^O - 3C^S)}{288} \geq 0$, in order to have $0 \leq f_A^{\text{int } A,C} + f_B^{\text{int } A,C} \leq D$.

(b') Given that τ_A, τ_B are in the range for which $f_A^{\text{int } A,C} = \frac{1}{2}D + \frac{\tau_B - \tau_A}{2(n+1)C^G}$ and $f_B^{\text{int } A,C} = \frac{1}{2}D + \frac{\tau_A - \tau_B}{2(n+1)C^G}$, the profit function of service provider C is increasing in τ_C because $f_A^{\text{int } A,C} + f_B^{\text{int } A,C} = D$, thus τ_C attains its maximal value $C^T - 3C^O - 4\frac{n+1}{2n}C^G(nD) - \tau_A - \tau_B$ in this region. Under this restriction, maximizing simultaneously the service providers' concave profit functions $(\tau_A - C^S)n(D + \frac{\tau_B - \tau_A}{(n+1)C^G}) + (\tau_C - C^S)nD$ and $(\tau_B - C^S)n(D + \frac{\tau_A - \tau_B}{(n+1)C^G})$, the first order conditions imply $\tau_A^{\text{int } A,C} = \tau_B^{\text{int } A,C} = C^S + 2\frac{n+1}{2n}C^G(nD)$ and $\tau_C^{\text{int } A,C} = C^S + [C^T - 3C^O - 3C^S - 8\frac{n+1}{2n}C^G(nD)]$, consequently $f_A^{\text{int } A,C} = f_B^{\text{int } A,C} = \frac{1}{2}D$.

Thus we consider several cases depending on the values of D :

(Case 2) $D \geq \frac{64(C^T - 3C^O - 3C^S)}{288} \geq 0$. In this case, by concavity, the only relevant solution is solution (a'). The profits are $\frac{n}{n+1} \frac{2112(C^T - 3C^O - 3C^S)^2}{20736C^G}$ for the integrated service provider A, C and $\frac{n}{n+1} \frac{384(C^T - 3C^O - 3C^S)^2}{20736C^G}$ for B .

(Case 1) $0 \leq D < \frac{64(C^T - 3C^O - 3C^S)}{288}$. In this case, the only possible solution is (b'). The profits are $2\frac{n+1}{2n}C^G(nD)^2 + [C^T - 3C^O - 3C^S - 8\frac{n+1}{2n}C^G(nD)]nD$ for the integrated service provider

A, C and $2\frac{n+1}{2n}C^G(nD)^2$ for B . When $D = \frac{54(C^T - 3C^O - 3C^S)}{288}$, the profit of A, C in solution (a') is strictly higher than that of solution (b'), so C dictates solution (a') by setting the charge accordingly. ■

For the last three theorems we assume that the outside option cost, C^T , is sufficiently high such that all captive demand is served, specifically

$$0 \leq D_{\text{cap}} \leq \frac{C^T - C^O - C^S}{2(n+1)C^G}. \quad (\text{A.1})$$

Theorem A.4 (formal version of Theorem IV) *There exists a unique user equilibrium, with a potential flexible demand D_{flex} threshold that determines two cases:*

Case 1, $D_{\text{flex}} < \sqrt{D_{\text{cap}}(\frac{C^T - C^O - C^S}{(n+1)C^G} - D_{\text{cap}})} - D_{\text{cap}}$:

Case 1(i), lowest flexible demand:

$$\begin{aligned} \tau_A^{\text{comp}} = \tau_B^{\text{comp}} &= C^S + [C^T - C^O - C^S - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})] \\ f_{l,\text{flex},A}^{\text{comp}} = f_{l,\text{flex},B}^{\text{comp}} &= 0 < \frac{1}{2}D_{\text{flex}} = f_{l,\text{flex},A}^{\text{eff}} = f_{l,\text{flex},B}^{\text{eff}}, \forall l \\ f_{l,\text{cap},A}^{\text{comp}} = f_{l,\text{cap},B}^{\text{comp}} &= D_{\text{cap}} = f_{l,\text{cap},A}^{\text{eff}} = f_{l,\text{cap},B}^{\text{eff}}, \forall l. \end{aligned}$$

Case 1(ii), low flexible demand leads to asymmetry:

$$\begin{aligned} \tau_A^{\text{comp}} \text{ or } \tau_B^{\text{comp}} &= C^S + \frac{1}{2}[C^T - 2C^O - 2C^S - 2\frac{n+1}{2n}C^G n(D_{\text{cap}} + 2D_{\text{flex}})] \\ f_{l,\text{flex},A}^{\text{comp}} = D_{\text{flex}} \text{ and } f_{l,\text{cap},B}^{\text{comp}} &= 0 \text{ or } f_{l,\text{flex},A}^{\text{comp}} = 0 \text{ and } f_{l,\text{cap},B}^{\text{comp}} = D_{\text{flex}}, \forall l \\ f_{l,\text{flex},A}^{\text{comp}} + f_{l,\text{flex},B}^{\text{comp}} &= D_{\text{flex}} = f_{l,\text{flex},A}^{\text{eff}} + f_{l,\text{flex},B}^{\text{eff}}, \forall l \\ f_{l,\text{cap},A}^{\text{comp}} = f_{l,\text{cap},B}^{\text{comp}} &= D_{\text{cap}} = f_{l,\text{cap},A}^{\text{eff}} = f_{l,\text{cap},B}^{\text{eff}}, \forall l. \end{aligned}$$

Case 2, $D_{\text{flex}} \geq \sqrt{D_{\text{cap}}(\frac{C^T - C^O - C^S}{(n+1)C^G} - D_{\text{cap}})} - D_{\text{cap}}$:

Case 2(i1), meeting flexible demand threshold, $D_{\text{flex}} \leq \frac{C^T - 2C^O - 2C^S}{3(n+1)C^G} - D_{\text{cap}}$:

$$\begin{aligned} \tau_A^{\text{comp}} = \tau_B^{\text{comp}} &= C^S + 2\frac{n+1}{2n}C^G n(D_{\text{cap}} + D_{\text{flex}}) \\ f_{l,\text{flex},A}^{\text{comp}} &= \frac{1}{2}D_{\text{flex}} + \frac{\tau_B - \tau_A}{2(n+1)C^G} = \frac{1}{2}D_{\text{flex}} = f_{l,\text{flex},A}^{\text{eff}}, \forall l \\ f_{l,\text{flex},B}^{\text{comp}} &= D_{\text{flex}} - f_{l,\text{flex},A}^{\text{comp}} = \frac{1}{2}D_{\text{flex}} = f_{l,\text{flex},B}^{\text{eff}}, \forall l \\ f_{l,\text{cap},A}^{\text{comp}} = f_{l,\text{cap},B}^{\text{comp}} &= D_{\text{cap}} = f_{l,\text{cap},A}^{\text{eff}} = f_{l,\text{cap},B}^{\text{eff}}, \forall l. \end{aligned}$$

Case 2(i2), increasing flexible demand, $\frac{C^T - 2C^O - 2C^S}{3(n+1)C^G} - D_{\text{cap}} < D_{\text{flex}} < \frac{C^T - 2C^O - 2C^S}{2(n+1)C^G} - D_{\text{cap}}$:

$$\begin{aligned} \tau_A^{\text{comp}} = \tau_B^{\text{comp}} &= C^S + \frac{1}{2}[C^T - 2C^O - 2C^S - 2\frac{n+1}{2n}C^G n(D_{\text{cap}} + D_{\text{flex}})] \\ f_{l,\text{flex},A}^{\text{comp}} &= \frac{1}{2}D_{\text{flex}} + \frac{\tau_B - \tau_A}{2(n+1)C^G} = \frac{1}{2}D_{\text{flex}} = f_{l,\text{flex},A}^{\text{eff}}, \forall l \\ f_{l,\text{flex},B}^{\text{comp}} &= D_{\text{flex}} - f_{l,\text{flex},A}^{\text{comp}} = \frac{1}{2}D_{\text{flex}} = f_{l,\text{flex},B}^{\text{eff}}, \forall l \\ f_{l,\text{cap},A}^{\text{comp}} = f_{l,\text{cap},B}^{\text{comp}} &= D_{\text{cap}} = f_{l,\text{cap},A}^{\text{eff}} = f_{l,\text{cap},B}^{\text{eff}}, \forall l. \end{aligned}$$

Case 2(ii), excessive flexible demand, $D_{\text{flex}} \geq \frac{C^T - 2C^O - 2C^S}{2(n+1)C^G} - D_{\text{cap}} \geq 0$:

$$\begin{aligned}\tau_A^{\text{comp}} &= \tau_B^{\text{comp}} = C^S + \frac{1}{4}(C^T - 2C^O - 2C^S) \\ f_{l,\text{flex},A}^{\text{comp}} &= \frac{C^T - 2C^O - 2\tau_A}{2(n+1)C^G} - \frac{1}{2}D_{\text{cap}} = \frac{C^T - 2C^O - 2C^S}{4(n+1)C^G} - \frac{1}{2}D_{\text{cap}} < f_{l,\text{flex},A}^{\text{eff}}, \forall l \\ f_{l,\text{flex},B}^{\text{comp}} &= \frac{C^T - 2C^O - 2\tau_B}{2(n+1)C^G} - \frac{1}{2}D_{\text{cap}} = \frac{C^T - 2C^O - 2C^S}{4(n+1)C^G} - \frac{1}{2}D_{\text{cap}} < f_{l,\text{flex},B}^{\text{eff}}, \forall l \\ f_{l,\text{cap},A}^{\text{comp}} &= f_{l,\text{cap},B}^{\text{comp}} = D_{\text{cap}} = f_{l,\text{cap},A}^{\text{eff}} = f_{l,\text{cap},B}^{\text{eff}}, \forall l.\end{aligned}$$

Proof. The equilibrium is constructed using a folding back procedure. Under the simplified network, in the second stage of the game customer l takes as given the choices of other customers and solves the problem

$$\begin{aligned}\min \quad & \left(C^O + C^G[f_{l,\text{cap},A} + f_{l,\text{flex},A} + \sum_{l' \neq l} (f_{l',\text{cap},A} + f_{l',\text{flex},A})] + \tau_A \right) (f_{l,\text{cap},A} + f_{l,\text{flex},A}) \\ & + \left(C^O + C^G[f_{l,\text{flex},A} + \sum_{l' \neq l} f_{l',\text{flex},A}] + \tau_A \right) f_{l,\text{flex},A} \\ & + \left(C^O + C^G[f_{l,\text{cap},B} + f_{l,\text{flex},B} + \sum_{l' \neq l} (f_{l',\text{cap},B} + f_{l',\text{flex},B})] + \tau_B \right) (f_{l,\text{cap},B} + f_{l,\text{flex},B}) \\ & + \left(C^O + C^G[f_{l,\text{flex},B} + \sum_{l' \neq l} f_{l',\text{flex},B}] + \tau_B \right) f_{l,\text{flex},B} \\ & + C^T[(D_{\text{cap}} - f_{l,\text{cap},A}) + (D_{\text{cap}} - f_{l,\text{cap},B}) + (D_{\text{flex}} - f_{l,\text{flex},A} - f_{l,\text{flex},B})] \\ \text{s.t.} \quad & f_{l,\text{cap},A} \leq D_{\text{cap}} \\ & f_{l,\text{cap},B} \leq D_{\text{cap}} \\ & f_{l,\text{flex},A} + f_{l,\text{flex},B} \leq D_{\text{flex}} \\ & f_{l,\text{cap},A}, f_{l,\text{cap},B}, f_{l,\text{flex},A}, f_{l,\text{flex},B} \geq 0.\end{aligned}$$

Since there exists a symmetric equilibrium by the assumed symmetry, any index l may be omitted. As the objective function is convex and the constraints are linear, the solution is obtained using the first order conditions

$$\begin{aligned}0 &= C^O + (n+1)C^G(f_{\text{cap},A} + f_{\text{flex},A}) + \tau_A - C^T - \lambda_{\text{cap},A} + \mu_{\text{cap},A} \\ &= C^O + (n+1)C^G(f_{\text{cap},B} + f_{\text{flex},B}) + \tau_B - C^T - \lambda_{\text{cap},B} + \mu_{\text{cap},B} \\ &= 2C^O + (n+1)C^G(f_{\text{cap},A} + 2f_{\text{flex},A}) + 2\tau_A - C^T - \lambda_{\text{flex},A} + \mu_{\text{flex}} \\ &= 2C^O + (n+1)C^G(f_{\text{cap},B} + 2f_{\text{flex},B}) + 2\tau_B - C^T - \lambda_{\text{flex},B} + \mu_{\text{flex}},\end{aligned}$$

where the Lagrange multipliers are $\mu_{\text{cap},A}$, $\mu_{\text{cap},B}$, μ_{flex} for the first three constraints and $\lambda_{\text{cap},A}$, $\lambda_{\text{cap},B}$, $\lambda_{\text{flex},A}$, $\lambda_{\text{flex},B}$ for the non-negativity constraints. Considering all possible cases for the

KKT conditions, we find the solution

$$f_{\text{cap},B}^{\text{comp}} = \begin{cases} D_{\text{cap}} & \text{if } 0 \leq \tau_B \leq C^T - C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}}), \\ \frac{C^T - C^O - \tau_B}{(n+1)C^G} & \text{if } C^T - C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}}) \leq \tau_B \leq C^T - C^O, \\ 0 & \text{if } \tau_B \geq C^T - C^O \end{cases}$$

and

$$f_{\text{flex},B}^{\text{comp}} = \begin{cases} D_{\text{flex}} & \text{if } 0 \leq \tau_B \leq \tau_A - 2\frac{n+1}{2n}C^G(nD_{\text{flex}}) \text{ and } \\ & \tau_B \leq \frac{1}{2}[C^T - 2C^O - 2\frac{n+1}{2n}C^G n(D_{\text{cap}} + 2D_{\text{flex}})], \\ \frac{1}{2}D_{\text{flex}} + \frac{\tau_A - \tau_B}{2(n+1)C^G} & \text{if } \tau_A - 2\frac{n+1}{2n}C^G(nD_{\text{flex}}) \leq \tau_B \leq \tau_A + 2\frac{n+1}{2n}C^G(nD_{\text{flex}}) \text{ and } \\ & \tau_B \leq C^T - 2C^O - 2\frac{n+1}{2n}C^G n(D_{\text{cap}} + D_{\text{flex}}) - \tau_A, \\ \frac{C^T - 2C^O - 2\tau_B}{2(n+1)C^G} - \frac{1}{2}D_{\text{cap}} & \text{if } \tau_B \leq \frac{1}{2}[C^T - 2C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})] \text{ and } \\ & \tau_B \geq \frac{1}{2}[C^T - 2C^O - 2\frac{n+1}{2n}C^G n(D_{\text{cap}} + 2D_{\text{flex}})] \text{ and } \\ & \tau_B \geq C^T - 2C^O - 2\frac{n+1}{2n}C^G n(D_{\text{cap}} + D_{\text{flex}}) - \tau_A, \\ 0 & \text{if } \tau_B \geq \tau_A + 2\frac{n+1}{2n}C^G(nD_{\text{flex}}) \text{ or } \\ & \tau_B \geq \frac{1}{2}[C^T - 2C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})], \end{cases}$$

where $f_{\text{cap},A}^{\text{comp}}$ and $f_{\text{flex},A}^{\text{comp}}$ have the same expressions except for swapping everywhere the indices A and B .

Next we analyze the first stage of the game. Each service provider s , taking the charge of the other service provider as given, sets the charge τ_s to maximize the profit

$$(\tau_s - C^S)n(f_{\text{cap},s}^{\text{comp}} + 2f_{\text{flex},s}^{\text{comp}}).$$

We analyze the best response maximal profit depending on the charge regions for $f_{\text{cap},s}^{\text{comp}}$ and $f_{\text{flex},s}^{\text{comp}}$. For $\tau_s \geq C^T - C^O$, $\tau_s \geq \frac{1}{2}[C^T - 2C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})]$ and $f_{\text{cap},s}^{\text{comp}} = f_{\text{flex},s}^{\text{comp}} = 0$, so the profit is 0. For $C^T - C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}}) \leq \tau_s \leq C^T - C^O$, since $D_{\text{cap}} \leq \frac{C^T}{(n+1)C^G}$ by (A.1), $C^T - C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}}) \geq \frac{1}{2}[C^T - 2C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})]$. Thus $f_{\text{flex},s}^{\text{comp}} = 0$ because $\tau_s \geq \frac{1}{2}[C^T - 2C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})]$. Therefore maximizing the concave profit function $(\tau_s - C^S)n\frac{C^T - C^O - \tau_s}{(n+1)C^G}$, first order condition implies the optimal charge $\tau_s^{\text{comp}} = \frac{C^T - C^O + C^S}{2}$ and the optimal frequency $f_{\text{cap},s}^{\text{comp}} = \frac{C^T - C^O - C^S}{2(n+1)C^G} \geq D_{\text{cap}}$. The constraint $f_{\text{cap},s}^{\text{comp}} \leq D_{\text{cap}}$ then implies $\tau_s^{\text{comp}} = C^T - C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})$ and $f_{\text{cap},s}^{\text{comp}} = D_{\text{cap}}$ with positive profit. For any $\frac{1}{2}[C^T - 2C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})] \leq \tau_s \leq C^T - C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})$, we still have $f_{\text{flex},s}^{\text{comp}} = 0$, thus the profit is maximized again at the upper bound of this interval. We conclude that in equilibrium, for both s , $f_{\text{cap},s}^{\text{comp}} = D_{\text{cap}}$. Moreover, either $\tau_s^{\text{comp}} \leq \frac{1}{2}[C^T - 2C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})]$, or we have the following solution:

(a) $\tau_s^{\text{comp}} = C^S + [C^T - C^O - C^S - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})]$, $f_{\text{flex},s}^{\text{comp}} = 0$ and $f_{\text{cap},s}^{\text{comp}} = D_{\text{cap}}$ for both s , with profit $[C^T - C^O - C^S - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})]nD_{\text{cap}}$.

For the flexible flow, when it is positive there are two potential solutions:

(b) Given that for each service provider s , τ_s is in the range for which $f_{\text{flex},s}^{\text{comp}} = \frac{C^T - 2C^O - 2\tau_s}{2(n+1)C^G} -$

$\frac{1}{2}D_{\text{cap}}$ and $f_{\text{cap},s}^{\text{comp}} = D_{\text{cap}}$, maximizing the concave profit function $(\tau_s - C^S)n(\frac{C^T - 2C^O - 2\tau_s}{(n+1)C^G})$, the first order condition implies $\tau_s^{\text{comp}} = \frac{C^T - 2C^O + 2C^S}{4}$, consequently $f_{\text{flex},s}^{\text{comp}} = \frac{1}{2}[\frac{C^T - 2C^O - 2C^S}{2(n+1)C^G} - D_{\text{cap}}]$ and $f_{\text{cap},s}^{\text{comp}} = D_{\text{cap}}$ for both s . This is possible if and only if $D_{\text{flex}} \geq \frac{C^T - 2C^O - 2C^S}{2(n+1)C^G} - D_{\text{cap}} \geq 0$, in order to have $0 \leq f_{\text{flex},A}^{\text{comp}} + f_{\text{flex},B}^{\text{comp}} \leq D_{\text{flex}}$.

(c) Given that for both s , τ_s is in the range for which $f_{\text{flex},A}^{\text{comp}} = \frac{1}{2}D_{\text{flex}} + \frac{\tau_B - \tau_A}{2(n+1)C^G}$, $f_{\text{flex},B}^{\text{comp}} = \frac{1}{2}D_{\text{flex}} + \frac{\tau_A - \tau_B}{2(n+1)C^G}$ and $f_{\text{cap},s}^{\text{comp}} = D_{\text{cap}}$ for both s , maximizing simultaneously the service providers' concave profit functions $(\tau_A - C^S)n(D_{\text{cap}} + D_{\text{flex}} + \frac{\tau_B - \tau_A}{(n+1)C^G})$ and $(\tau_B - C^S)n(D_{\text{cap}} + D_{\text{flex}} + \frac{\tau_A - \tau_B}{(n+1)C^G})$, the first order conditions under symmetry imply $\tau_s^{\text{comp}} = C^S + 2\frac{n+1}{2n}C^G n(D_{\text{cap}} + D_{\text{flex}})$, consequently $f_{\text{flex},s}^{\text{comp}} = \frac{1}{2}D_{\text{flex}}$ and $f_{\text{cap},s}^{\text{comp}} = D_{\text{cap}}$ for both s . This is possible if and only if $0 \leq D_{\text{flex}} \leq \frac{C^T - 2C^O - 2C^S}{3(n+1)C^G} - D_{\text{cap}}$ in order to have $\tau_A^{\text{comp}} + \tau_B^{\text{comp}} \leq C^T - 2C^O - 2\frac{n+1}{2n}C^G n(D_{\text{cap}} + D_{\text{flex}})$.

Thus we consider several cases depending on the values of D_{cap} and D_{flex} :

(Case 2(ii)) $D_{\text{flex}} \geq \frac{C^T - 2C^O - 2C^S}{2(n+1)C^G} - D_{\text{cap}} \geq 0$ and $0 \leq D_{\text{cap}} \leq \check{D}_{\text{cap}}$, where

$$\check{D}_{\text{cap}} \equiv \frac{C^T - C^O - C^S - \sqrt{\frac{1}{2}(C^T)^2 - (C^O + C^S)^2}}{2(n+1)C^G},$$

and we note that $\check{D}_{\text{cap}} \leq \frac{C^T - 2C^O - 2C^S}{3(n+1)C^G} \leq \frac{C^T - 2C^O - 2C^S}{2(n+1)C^G}$. In this case, by concavity, the only relevant solution is solution (b). Both service provider's profit is $\frac{n}{n+1} \frac{(C^T - 2C^O - 2C^S)^2}{8C^G}$. For this profit to be weakly higher than the profit of solution (a), it is necessary and sufficient that $0 \leq D_{\text{cap}} \leq \check{D}_{\text{cap}}$.

(Case 2(i2)) $\frac{C^T - 2C^O - 2C^S}{3(n+1)C^G} - D_{\text{cap}} < D_{\text{flex}} < \frac{C^T - 2C^O - 2C^S}{2(n+1)C^G} - D_{\text{cap}}$ and $0 \leq D_{\text{cap}} \leq \check{D}_{\text{cap}}$. In this case, solution (a) is not relevant because $\frac{C^T - 2C^O - 2C^S}{2(n+1)C^G} - D_{\text{cap}} > 0$ implies that each service provider prefers to serve some flexible demand. Moreover, solutions (b) and (c) are outside their respective regions, and by concavity, the only relevant solution is the one on the boundary of both regions, i.e. $\tau_s^{\text{comp}} = C^S + \frac{C^T - 2C^O - 2C^S - 2\frac{n+1}{2n}C^G n(D_{\text{cap}} + D_{\text{flex}})}{2}$. By symmetry $f_{\text{flex},s}^{\text{comp}} = \frac{1}{2}D_{\text{flex}}$ and $f_{\text{cap},s}^{\text{comp}} = D_{\text{cap}}$ for both s . Each service provider's profit is $(\frac{C^T - 2C^O - 2C^S - 2\frac{n+1}{2n}C^G n(D_{\text{cap}} + D_{\text{flex}})}{2})n(D_{\text{cap}} + D_{\text{flex}})$. As in Case 2(ii), this profit is weakly higher than the profit of solution (a) if and only if $0 \leq D_{\text{cap}} \leq \check{D}_{\text{cap}}$.

(Case 2(i1)) $\bar{D}_{\text{flex}} \leq D_{\text{flex}} \leq \frac{C^T - 2C^O - 2C^S}{3(n+1)C^G} - D_{\text{cap}}$ and $0 \leq D_{\text{cap}} \leq \bar{D}_{\text{cap}}$, where

$$\bar{D}_{\text{flex}} \equiv \sqrt{D_{\text{cap}}[\frac{C^T - C^O - C^S}{(n+1)C^G} - D_{\text{cap}}]} - D_{\text{cap}}$$

and

$$\bar{D}_{\text{cap}} \equiv \frac{C^T - C^O - C^S - \frac{1}{3}\sqrt{(C^T + C^O + C^S)(5C^T - 7C^O - 7C^S)}}{2(n+1)C^G},$$

and we note that $\bar{D}_{\text{cap}} \leq \check{D}_{\text{cap}}$, and that $\bar{D}_{\text{flex}} \leq \frac{C^T - 2C^O - 2C^S}{3(n+1)C^G} - D_{\text{cap}}$ if and only if $D_{\text{cap}} \leq \bar{D}_{\text{cap}}$. When $D_{\text{flex}} \leq \frac{C^T - 2C^O - 2C^S}{3(n+1)C^G} - D_{\text{cap}}$, the relevant solution is solution (c). Each service provider's profit is $2\frac{n+1}{2n}C^G[n(D_{\text{cap}} + D_{\text{flex}})]^2$. This profit is higher than the profit of solution (a) if and only if $\bar{D}_{\text{flex}} \leq D_{\text{flex}}$ and $D_{\text{cap}} \leq \bar{D}_{\text{cap}}$.

(Case 1(ii)) $\hat{D}_{\text{flex}} < D_{\text{flex}} < \min\{\bar{D}_{\text{flex}}, \frac{C^T - 2C^O - 2C^S}{3(n+1)C^G} - D_{\text{cap}}\}$ and $\hat{D}_{\text{cap}} \leq D_{\text{cap}} \leq \check{D}_{\text{cap}}$, where

$$\begin{aligned} \hat{D}_{\text{flex}} &\equiv \frac{C^T - 2C^O - 2C^S}{4(n+1)C^G} - \frac{D_{\text{cap}}}{2} \\ &\quad - \sqrt{\left(\frac{C^T - 2C^O - 2C^S}{4(n+1)C^G}\right)^2 - \frac{D_{\text{cap}}}{2}\left(\frac{C^T - C^O - C^S}{(n+1)C^G} - D_{\text{cap}}\right)} \end{aligned}$$

and

$$\hat{D}_{\text{cap}} \equiv \frac{7C^T - 6C^O - 6C^S - 4\sqrt{2(C^T)^2 - 2(C^O + C^S)^2 - C^T(C^O + C^S)}}{17(n+1)C^G}.$$

In this case, solution (b) is not relevant because $D_{\text{flex}} < \frac{C^T - 2C^O - 2C^S}{2(n+1)C^G} - D_{\text{cap}}$. Solution (c) is not relevant because each service provider's profit is strictly lower than the profit of solution (a) (as explained in Case 2(i1), $\bar{D}_{\text{flex}}$ is exactly the cut-off point for this comparison). Moreover, solution (a) is not relevant because each service provider's profit when $f_{\text{flex},s}^{\text{comp}} = D_{\text{flex}}$ and $f_{\text{cap},s}^{\text{comp}} = D_{\text{cap}}$ is strictly higher (the lower bound of the interval, $\hat{D}_{\text{flex}}$, is exactly the cut-off point for this comparison). Thus we have an asymmetric equilibrium in which one service provider, say A , sets the charge $\tau_A^{\text{comp}} = C^S + [C^T - C^O - C^S - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})]$ consequently $f_{\text{flex},A}^{\text{comp}} = 0$, while the other service provider best responds by setting the charge $\tau_s^{\text{comp}} = C^S + \frac{1}{2}[C^T - 2C^O - 2C^S + -2\frac{n+1}{2n}C^G n(D_{\text{cap}} + 2D_{\text{flex}})]$ consequently $f_{\text{flex},B}^{\text{comp}} = D_{\text{flex}}$ (or the analogous equilibrium where A and B interchange).

(Case 1(i)) $[0 \leq D_{\text{flex}} \leq \hat{D}_{\text{flex}} \text{ and } 0 \leq D_{\text{cap}} \leq \bar{D}_{\text{cap}}]$ or $D_{\text{cap}} > \check{D}_{\text{cap}}$. In this case, the only relevant solution is solution (a).

Finally, the system optimal solution flows for all cases are obtained by considering the charge dependent user equilibrium flow expressions in $f_{l,\text{flex},s}^{\text{comp}}$ for case 2(i) and 2(ii) (where case 2(i) is applicable also below the threshold $\sqrt{D_{\text{cap}}(\frac{C^T - C^O - C^S}{(n+1)C^G} - D_{\text{cap}})} - D_{\text{cap}}$ for the flexible demand D_{flex}), and by replacing the term $n+1$ everywhere by the term $2n$ in order to fully internalize congestion costs. The efficient flows that minimize total social costs are then derived when setting $\tau_A = \tau_B = C^S$. ■

Theorem A.5 (formal version of Theorem V) *There exists a unique user equilibrium, with a potential flexible demand D_{flex} threshold, whereby compared to no integration (as in Theorem A.4):*

Case 1, lowest flexible demand, $[0 \leq D_{\text{flex}} \leq 2\hat{D}_{\text{flex}} \text{ and } 0 \leq D_{\text{cap}} \leq \check{D}_{\text{cap}}]$ or $D_{\text{cap}} > \check{D}_{\text{cap}}$ (where $\hat{D}_{\text{flex}}, \check{D}_{\text{cap}}$ are defined in the proof of Theorem A.4): the outcome is as in case 1(i) of Theorem A.4.

Case 2(i), intermediate flexible demand, $2\hat{D}_{\text{flex}} < D_{\text{flex}} < \frac{C^T - 2C^O - 2C^S}{2(n+1)C^G} - D_{\text{cap}}$ and $0 \leq D_{\text{cap}} \leq \check{D}_{\text{cap}}$: the outcome is as in case 2(i2) of Theorem A.4

Case 2(ii), excessive flexible demand, $D_{\text{flex}} \geq \frac{C^T - 2C^O - 2C^S}{2(n+1)C^G} - D_{\text{cap}} \geq 0$ and $0 \leq D_{\text{cap}} \leq \check{D}_{\text{cap}}$: the outcome is as in case 2(ii) of Theorem A.4.

Proof. The equilibrium is constructed using a folding back procedure. Under the simplified network, the analysis of the second stage of the game is the same as in Theorem A.4, and even simplifies further because the single service provider must set a single charge $\tau = \tau_A = \tau_B$.

Thus $f_{\text{cap}}^{\text{int } A,B} = f_{\text{cap},A}^{\text{int } A,B} = f_{\text{cap},B}^{\text{int } A,B}$ and $f_{\text{flex}}^{\text{int } A,B} = f_{\text{flex},A}^{\text{int } A,B} = f_{\text{flex},B}^{\text{int } A,B}$ are:

$$f_{\text{cap}}^{\text{int } A,B} = \begin{cases} D_{\text{cap}} & \text{if } 0 \leq \tau \leq C^T - C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}}), \\ \frac{C^T - C^O - \tau}{(n+1)C^G} & \text{if } C^T - C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}}) \leq \tau \leq C^T - C^O, \\ 0 & \text{if } \tau \geq C^T - C^O \end{cases}$$

and

$$f_{\text{flex}}^{\text{int } A,B} = \begin{cases} \frac{1}{2}D_{\text{flex}} & \text{if } \tau \leq \frac{1}{2}[C^T - 2C^O - 2\frac{n+1}{2n}C^Gn(D_{\text{cap}} + D_{\text{flex}})], \\ \frac{C^T - 2C^O - 2\tau}{2(n+1)C^G} - \frac{1}{2}D_{\text{cap}} & \text{if } \frac{1}{2}[C^T - 2C^O - 2\frac{n+1}{2n}C^Gn(D_{\text{cap}} + D_{\text{flex}})] \leq \tau \leq \frac{1}{2}[C^T - 2C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})], \\ 0 & \text{if } \tau \geq \frac{1}{2}[C^T - 2C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})]. \end{cases}$$

In the first stage, the single service provider sets τ to maximize the per route profit

$$(\tau - C^S)n(f_{\text{cap}}^{\text{int } A,B} + 2f_{\text{flex}}^{\text{int } A,B}).$$

As in Theorem A.2, (A.1) implies that $f_{\text{cap}}^{\text{int } A,B} = D_{\text{cap}}$. Moreover, either $\tau^{\text{int } A,B} \leq \frac{1}{2}[C^T - 2C^O - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})]$, or we have the following solution:

(a') $\tau^{\text{int } A,B} = C^S + [C^T - C^O - C^S - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})]$, $f_{\text{flex}}^{\text{int } A,B} = 0$ and $f_{\text{cap}}^{\text{int } A,B} = D_{\text{cap}}$, with per route profit $[C^T - C^O - C^S - 2\frac{n+1}{2n}C^G(nD_{\text{cap}})]nD_{\text{cap}}$.

For the flexible flow, when it is positive there are two potential solutions:

(b') Given that τ is in the range for which $f_{\text{flex}}^{\text{int } A,B} = \frac{C^T - 2C^O - 2\tau}{2(n+1)C^G} - \frac{1}{2}D_{\text{cap}}$ and $f_{\text{cap}}^{\text{int } A,B} = D_{\text{cap}}$, maximizing the concave profit function $(\tau - C^S)n(\frac{C^T - 2C^O - 2\tau}{(n+1)C^G})$, we have $\tau^{\text{int } A,B} = C^S + \frac{C^T - 2C^O - 2C^S}{4}$, $f_{\text{flex}}^{\text{int } A,B} = \frac{1}{2}[\frac{C^T - 2C^O - 2C^S}{2(n+1)C^G} - D_{\text{cap}}]$ and $f_{\text{cap}}^{\text{int } A,B} = D_{\text{cap}}$.

(c') Given that τ is in the range for which $f_{\text{flex}}^{\text{int } A,B} = \frac{1}{2}D_{\text{flex}}$ and $f_{\text{cap}}^{\text{int } A,B} = D_{\text{cap}}$, the profit is maximized at the upper bound of this range with $\tau^{\text{int } A,B} = C^S + \frac{1}{2}[C^T - 2C^O - 2C^S - 2\frac{n+1}{2n}C^Gn(D_{\text{cap}} + D_{\text{flex}})]$.

Thus we consider several cases depending on the values of D_{cap} and D_{flex} :

(Case 2(ii)) $D_{\text{flex}} \geq \frac{C^T - 2C^O - 2C^S}{2(n+1)C^G} - D_{\text{cap}} \geq 0$ and $0 \leq D_{\text{cap}} \leq \check{D}_{\text{cap}}$, where $\check{D}_{\text{cap}}$ is defined in the proof of Theorem A.4. In this case, by concavity, the only relevant solution is solution (b'), with per route profit $\frac{n}{n+1} \frac{(C^T - 2C^O - 2C^S)^2}{8C^G}$. For this profit to be weakly higher than the profit of solution (a'), it is necessary and sufficient that $0 \leq D_{\text{cap}} \leq \check{D}_{\text{cap}}$.

(Case 2(1), which replaces cases 1(ii) and 2(i1), 2(i2) in Theorem A.4) $2\hat{D}_{\text{flex}} < D_{\text{flex}} < \frac{C^T - 2C^O - 2C^S}{2(n+1)C^G} - D_{\text{cap}}$ and $0 \leq D_{\text{cap}} \leq \check{D}_{\text{cap}}$, where $\hat{D}_{\text{flex}}$ is defined in the proof of Theorem A.4. In this case, the only relevant solution is solution (c'), with per route profit $\frac{1}{2}[C^T - 2C^O - 2C^S - 2\frac{n+1}{2n}C^Gn(D_{\text{cap}} + D_{\text{flex}})] \cdot n(D_{\text{cap}} + D_{\text{flex}})$. As in case 2(ii), this profit is weakly higher than the profit of solution (a') if and only if $0 \leq D_{\text{cap}} \leq \check{D}_{\text{cap}}$.

(Case 1(i)) $[0 \leq D_{\text{flex}} \leq 2\hat{D}_{\text{flex}} \text{ and } 0 \leq D_{\text{cap}} \leq \check{D}_{\text{cap}}] \text{ or } D_{\text{cap}} > \check{D}_{\text{cap}}$. In this case, the only relevant solution is the above solution (a'). ■

Theorem A.6 (formal version of Theorem VI) *There exists a unique user equilibrium,*

whereby compared to parallel competition without technology adoption (as in Theorem A.4):

$$\begin{aligned}
\tau_A^{\text{int } A,1} &= C^S - 2\frac{C^S - C_A^S}{3} + \frac{1}{3n}(C^O - C_{1A}^O) + 2\frac{n+1}{2n}C^G n(D_{\text{cap}} + D_{\text{flex}}) \\
\tau_B^{\text{int } A,1} &= C^S - \frac{C^S - C_A^S}{3} - \frac{1}{3n}(C^O - C_{1A}^O) + 2\frac{n+1}{2n}C^G n(D_{\text{cap}} + D_{\text{flex}}) \\
f_{1,\text{flex},A}^{\text{int } A,1} &= \frac{D_{\text{flex}}}{2} + \frac{C^S - C_A^S}{6(n+1)C^G} + \left(\frac{n}{2} - \frac{1}{3n}\right)\frac{C^O - C_{1A}^O}{(n+1)C^G} \\
f_{o,\text{flex},A}^{\text{int } A,1} &= \frac{D_{\text{flex}}}{2} + \frac{C^S - C_A^S}{6(n+1)C^G} - \left(\frac{1}{2} + \frac{1}{3n}\right)\frac{C^O - C_{1A}^O}{(n+1)C^G} \\
f_{1,\text{flex},B}^{\text{int } A,1} &= \frac{D_{\text{flex}}}{2} - \frac{C^S - C_A^S}{6(n+1)C^G} - \left(\frac{n}{2} - \frac{1}{3n}\right)\frac{C^O - C_{1A}^O}{(n+1)C^G} \\
f_{o,\text{flex},B}^{\text{int } A,1} &= \frac{D_{\text{flex}}}{2} - \frac{C^S - C_A^S}{6(n+1)C^G} + \left(\frac{1}{2} + \frac{1}{3n}\right)\frac{C^O - C_{1A}^O}{(n+1)C^G}.
\end{aligned}$$

Proof. We concentrate on the case of competitive pricing (as in Case 3 of Theorem A.4) in which all demand is served, i.e. $f_{l,\text{cap},A} = f_{l,\text{cap},B} = D_{\text{cap}}$ and $f_{l,\text{flex},A} + f_{l,\text{flex},B} = D_{\text{flex}}$ for each customer l under the simplified network. The equilibrium is constructed using a folding back procedure. Customer l takes as given the choices of other customers and sets $f_{l,\text{flex},A}$ to solve the problem

$$\begin{aligned}
\min \quad & \left(C_{l,A}^O + C^G[nD_{\text{cap}} + f_{l,\text{flex},A} + \sum_{l' \neq l} f_{l',\text{flex},A}] + \tau_A \right) (D_{\text{cap}} + f_{l,\text{flex},A}) \\
& + \left(C_{l,A}^O + C^G[f_{l,\text{flex},A} + \sum_{l' \neq l} f_{l',\text{flex},A}] + \tau_A \right) f_{l,\text{flex},A} \\
& + \left(C^O + C^G[nD_{\text{cap}} + nD_{\text{flex}} - f_{l,\text{flex},A} - \sum_{l' \neq l} f_{l',\text{flex},A}] + \tau_B \right) (D_{\text{cap}} + D_{\text{flex}} - f_{l,\text{flex},A}) \\
& + \left(C^O + C^G[nD_{\text{flex}} - f_{l,\text{flex},A} - \sum_{l' \neq l} f_{l',\text{flex},A}] + \tau_B \right) (D_{\text{flex}} - f_{l,\text{flex},A}),
\end{aligned}$$

where we assume that $C_{1,A}^O \leq C^O$ and $C_{l',A}^O = C^O$ for all $l' \neq 1$. By the assumed symmetry, there exists a symmetric equilibrium such that for each service provider s , $f_{l',\text{flex},s}$ is equal for all $l' \neq 1$, and is denoted by $f_{o,\text{flex},s}$. The first order conditions are

$$\begin{aligned}
& 2C_{1,A}^O + C^G[(n+1)D_{\text{cap}} + 4f_{1,\text{flex},A} + 2(n-1)f_{o,\text{flex},A}] + 2\tau_A \\
& = 2C^O + C^G[(n+1)D_{\text{cap}} + 2(n+1)D_{\text{flex}} - 4f_{1,\text{flex},A} - 2(n-1)f_{o,\text{flex},A}] + 2\tau_B
\end{aligned}$$

for customer 1, and

$$\begin{aligned}
& 2C^O + C^G[(n+1)D_{\text{cap}} + 2f_{1,\text{flex},A} + 2nf_{o,\text{flex},A}] + 2\tau_A \\
& = 2C^O + C^G[(n+1)D_{\text{cap}} + 2(n+1)D_{\text{flex}} - 2f_{1,\text{flex},A} - 2nf_{o,\text{flex},A}] + 2\tau_B
\end{aligned}$$

for each customer $l' \neq 1$. Thus the second stage solution is

$$\begin{aligned} f_{1,\text{flex},A}^{\text{int } A,1} &= \frac{D_{\text{flex}}}{2} + n \frac{C^O - C_{1A}^O}{2(n+1)C^G} + \frac{\tau_B - \tau_A}{2(n+1)C^G} \\ f_{o,\text{flex},A}^{\text{int } A,1} &= \frac{D_{\text{flex}}}{2} - \frac{C^O - C_{1A}^O}{2(n+1)C^G} + \frac{\tau_B - \tau_A}{2(n+1)C^G} \\ f_{1,\text{flex},B}^{\text{int } A,1} &= \frac{D_{\text{flex}}}{2} - n \frac{C^O - C_{1A}^O}{2(n+1)C^G} + \frac{\tau_A - \tau_B}{2(n+1)C^G} \\ f_{o,\text{flex},B}^{\text{int } A,1} &= \frac{D_{\text{flex}}}{2} + \frac{C^O - C_{1A}^O}{2(n+1)C^G} + \frac{\tau_A - \tau_B}{2(n+1)C^G}. \end{aligned}$$

In the first stage each service provider s , taking the charge of the other service provider as given, sets the charge τ_s to maximize the profit

$$(\tau_s - C_s^S)[nD_{\text{cap}} + 2(f_{1,\text{flex},s}^{\text{int } A,1} + (n-1)f_{o,\text{flex},s}^{\text{int } A,1})],$$

and we denote the service cost of service providers B by C^S . Solving the first order conditions

$$\begin{aligned} 2\tau_A - \tau_B &= C_A^S + (n+1)C^G(D_{\text{cap}} + D_{\text{flex}}) + \frac{C^O - C_{1A}^O}{n} \text{ and} \\ -\tau_A + 2\tau_B &= C^S + (n+1)C^G(D_{\text{cap}} + D_{\text{flex}}) - \frac{C^O - C_{1A}^O}{n}, \end{aligned}$$

the outcome is as stated in the theorem. ■