Waldman, Helio; Bortoletto, Rodrigo Campos

Conference Paper

Game-theoretical models of the competitive dynamics in optical network service provision


Provided in Cooperation with:
International Telecommunications Society (ITS)


This Version is available at:
http://hdl.handle.net/10419/106860

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
I – Introduction

Modern society is organizing all its activities and interactions in the cyberspace, where it finds unprecedented opportunities to obtain information, connect with partners, collaborate, trade, disseminate information and ideas, and carry on so many interactions that had been the main attraction of the urban environment in the past. The enabling infrastructure that makes this feasible is a costly, underlying optical infrastructure based on optical fibers, lasers, advanced electronics and photonics that provide high-speed connections between optical client nodes, especially IP and/or MPLS routers. In order to promote optimal use of this asset by the users, regulating agencies must mitigate the dominance of significant market power among service providers, maintaining a competitive service provision environment and moderating the incumbent advantage through actions such as the reduction of sunk cost, network externalities and the cost of deployment of new technologies and services.

When the 21st century started, the optical infrastructure was quite overprovisioned, thus obviating the need to plan any further expansion for some time. More than a decade of strong growth of the Internet traffic, pulled by new services and new users, have now overcome this situation, leading to a somewhat unstable equilibrium between overprovisioning and congestion, always threatened by the unpredictability of growth and competitive outcomes. The multiservice nature of the Internet multiplies the number of stakeholders in its future, making it a global issue, in the geopolitical as in the social and corporate domains. Understanding the games played by the stakeholders is thus strategic for the formulation of public policies that make sure society will reap the full benefit of such emerging technologies.

In anticipation of this aim, technology itself is evolving in order to support full exploitation of the fiber bandwidth, allowing the supply of bandwidth to be shaped to match a heterogeneous demand in emerging flexible optical networks; developing advanced modulation schemes to increase spectral efficiency and support more bit/s per Hz of the fiber spectrum; allowing the coexistence of different line rates in the same fiber, etc. A concurrent effort is then due to provide efficient marketing of the available bandwidth. This requires the development of accurate models for user utility and behavior, service dimensioning and pricing strategies, regulator objectives, metrics and strategies.

We propose to model the dynamics of the interaction of the stakeholders and the resulting impact on the network. Flat-rate pricing breeds more congestion, which can then be overcome only through more bandwidth provisioning. The unsustainability of this scenario is highlighted by two conflicting and emerging imperatives: service universalization and energy conservation (“greenness”). The need to accommodate both targets may compel the future adoption of usage-based pricing, moderated by emerging regulation based on the network neutrality principle.

A model is always simpler than reality; otherwise, it would deny its purpose. Nevertheless, in order to be useful, a model must capture the essential features of reality. With this challenge in mind, we will lump the stakeholders into three broad categories: users, encompassing high-end users and bandwidth retailers that must share the available bandwidth; at least one, but possibly a few infrastructure providers that compete for market share and profit; and a regulator with the power to admit new entrants into the market, who strives to maximize social welfare.

We propose to apply standard game-theory concepts and techniques to discuss the emergence of equilibria in the deployment of infrastructure and the dynamics of allocation of its resources among competing customers in a regulated, competitive service provision environment. The provider’s game will comprise a CAPEX game that results in the network dimensioning; and the OPEX sub-game that is played by cost asymmetries introduced by new entrants. Each user strives to maximize her utility in a user’s game. Finally, the regulator controls the entrance to the market in order to enhance user satisfaction.

II – Related Work

The development of advanced lasers enabled by quaternary semiconductor materials in the 1980’s, followed by the advent of the erbium-doped fiber amplifier (EDFA) in the early 1990’s, allowed the deployment of the first wavelength-routed networks (WRN) based on the WDM technology. These networks provide optical connectivity between designated optical nodes through optical paths that may traverse intermediate nodes in a transparent fashion. Carrier optical frequencies, though, must be taken from a fixed grid made up of equally spaced frequencies. Spacings of 200 GHz were initially used, but continuous improvement of lasers and optical filters have allowed succeeding generations of WDM networks to reduce the frequency spacing to 100 GHz and currently to 50 GHz, thus enhancing the “bandwidth mining” of optical fibers.

The networking environment generated by optical paths evoked the circuit-switching networks that dominated the scene of communication networks in the old days of the telephone networks, which had been displaced by packet-
switching in the wake of digitalization. However, circuit-switching theory developed for telephone networks had to be revised in order to take into consideration new constraints to connection activation, i.e., new blocking conditions. The main new constraint was the need to keep the same wavelength over all links of a path, unless wavelength conversion is performed in a node either by a regenerator, which may not be rate-transparent; or by a wavelength converter, which may not be available or affordable, or may have only limited conversion capability.

The emergence of a wavelength (or “color”) continuity constraint, which was absent in old circuit-switching theory, motivated some authors to revisit the theory in the 1990’s, in order to evaluate the resulting performance degradation in terms of blocking probability [1, 2, 3]. This early work was successful in incorporating the wavelength continuity constraint into the classical blocking models, which assume an equilibrium condition between activation and deactivation of connections. This assumption was reasonable in old telephone networks because voice traffic used to grow only 10% a year while voice connections would last for only a few minutes. Current optical path dynamics is quite different, as Internet traffic may grow 50-100% a year while optical connections may last for months. In other words, traffic intensity may grow considerably during the duration of a connection, so the system can hardly approach an equilibrium. This issue has motivated some authors to propose new blocking models for non-stationary conditions under exponential traffic growth [4, 5]. Next generation optical networks, though, will deal with a much more dynamic traffic. If average connection duration is reduced to minutes or a few hours, so that traffic intensity may be seen as constant during this time, blocking models based on equilibrium will be useful again to support the network dimensioning and planning, so looking for equilibrium conditions is still valid.

Classical traffic theory was focused on the telephone network, which was dedicated to a single service, normally delivered by a single operator. In order to comply with regulation, this single operator had to provide enough channels to keep the call blocking probability below a specified level. The required number of channels could then be determined by the well-known Erlang-B formula [6]. In an unregulated environment, though, a single operator might be interested in dimensioning the network with the aim of maximizing its profit. For this reason, we have investigated the single-operator, profit-maximizing network dimensioning problem in [7]. We have found that profit maximization leads to decreasing levels of blocking probability with the growth of traffic intensity. Therefore, increasing levels of traffic intensity will eventually mitigate or eliminate any conflict between performance (as measured by blocking probability) and profit.

We have also discussed network dimensioning in competitive environments, in which more than one operator compete for customers in a single-service market [8]. For this purpose, we considered two interconnected games on a duopoly, assuming that both competitors charge the same price for the service, so competition is based only on their dimensioning and costs: a users game, where users look for a service provider that gives them the best performance in terms of blocking probability; and an operators game, where both competitors wish to maximize their profit. We have found that the users game always reaches a Nash equilibrium [9], with the largest operator (the one that provides more capacity) capturing more traffic than its share of capacity provision. The operators profit-maximizing game, though, does not necessarily reach equilibrium between pure strategies, leading to loops when the players alternate their moves. However, a player may overcome the loop by moderating his/her profit target so that the other player is encouraged by profit maximization to leave the game, leading to a monopoly. Under moderate cost asymmetries, both players can make this move without running a lossy operation. However, if cost asymmetries are large enough, only the operator with smaller costs will be able to force its competitor out while still keeping some profit.

III – Flexible Networks

Wavelength routing networks (WRN) provide a single service: they provide transparent optical paths between designated source and destination nodes. Usually, users of this service are multi-service providers, which may provide Internet or Intranet access to a community of end users. This situation generates a functionality mismatch between the optical layer and the application layers, which results in inefficient use of the optical spectrum. This inefficiency is exacerbated by constraining the carrier wavelengths to be taken from a fixed grid of equally spaced frequencies, as required by the WDM technology. Such uniform spacing means that users are forced to request a fixed amount of optical bandwidth, regardless of their real need for bandwidth at any time.

For this reason, it is expected that the next generation of optical networks will evolve to be more multi-service and flexible. The Elastic Optical Network (EON) is now regarded as the enabling system framework for these emerging functionalities [11]. Flexibility will be reached by allowing users to request any number of contiguous frequency slots to support a connection according to their real need for bandwidth, given by the required bit rate and distance to be reached without regeneration. Large noise margins may then be traded with spectral footprint through the use of advanced modulation schemes, thus leading to bandwidth conservation through enhanced spectral efficiency in metropolitan, short-reach environments. Frequency slots are sufficiently narrow to allow requests to be tight with respect to real bandwidth needs.

The supply of connections with \( n \) slots is a differentiated service for each \( n \). Thus, the EON’s are multi-service networks where all services are supported by a common resource, given by the physical links and their available spectral ranges. Thus, it is important to understand how the multiple demanded services can most efficiently share this common resource. This is the aim of the routing and spectrum assignment (RSA) problem, which succeeds the classic RWA (routing and wavelength assignment) problem of old WDM networks in the performance analysis of algorithmic solutions for the network operation. In multilink mesh networks, the coupling between the routing and the spectrum/wavelength assignment problems aggravates these problems, making them hard to approach analytically. For
this reason, it is instructive to consider the RSA problem in a single link environment, where only one routing is possible, so that the spectrum assignment (SA) problem is uncoupled from the routing problem.

In [10, 14], this is done for the case of incremental traffic (where all requested connections are permanent after acceptance) when the operation aims to minimize the mean time to exhaust the spectrum. A policy-based algorithm was derived that performs this optimization. In order to do this, the optimal algorithm must be aware of the traffic profile, which is defined by the rates of requests for connections with \( i \) slots, for all \( i \). By mapping the optimal policies on the space of traffic profiles, the following conclusions were reached:

- If the request rates increase with connection size (slot number), mean time to spectrum completion is minimized by the wholesale algorithm, which accepts only requests for all available slots.
- If request rates decrease with connection size, then mean time to spectrum exhaustion is minimized by the greedy algorithm, which accepts all requests that may be accommodated by the available resources.

These results assume a positive rate of requests for single slots. If there is no demand for single slots, there is a positive probability (which may reach 1 for certain values of initial number of available slots) that a remaining number of slots may not be filled by any of the connection sizes on demand, leading to deadlocked resources. The mean time to exhaust the spectrum would then run away to infinity, so no algorithm would be able to minimize it. However, the amount of deadlocked resources may be minimized by appropriate policies, as shown in [12].

We are now extending these results to the case of dynamic traffic, where connections are torn off after an exponentially distributed service time. For this purpose, the spectrum assignment process is modeled as a continuous-time Markov chain (CTMC). For each connection size on demand, the supply of new connections is treated as a differentiated service. With unlimited resources (spectral slots), the spectral assignment process for each service is independent of the other services and follows the well-known Erlang model, which is known to be reversible [6]. Since all services are independent of each other, the multi-service process is also reversible. When constrained by a finite amount of available resources, though, the services are not independent anymore. However, using the Kolmogorov criterion for reversibility, the multi-service process may be shown to keep its reversibility when constrained by a finite number of available slots that are either partitioned among the services or (fully or partially) shared by them [6, Theorem D.22]. Based on this reversibility property, we have devised a method to generate closed-form expressions for the steady-state probabilities of each state of the Markov chain, thus yielding easy calculations for the mean number of active slots, which allow the determination of the resource efficiency of different configurations for spectral partitioning or sharing among services. The flexibility gain for each configuration may then be obtained, thus revealing the most efficient one.

Notice that the implementation of flexible networks based on current generation of (non-flexible) equipment would imply added cost due the need to provide a larger inventory than needed for non-flexible operation. For this reason, it is very likely that a cost-efficient implementation of elastic optical networks will require the availability and affordability of bit-rate-variable transceivers (BVT’s), as well as flexible spectrum wavelength-selective switches (FS-WSS’s), and a whole new family of flexible gear [13]. Only then will system flexibility gains be fully reached, but technology is already evolving to provide the necessary functionalities for this purpose.

Notice that the implementation of flexible networks based on current generation of (non-flexible) equipment would imply added cost due the need to provide a larger inventory than needed for non-flexible operation. For this reason, it is very likely that a cost-efficient implementation of elastic optical networks will require the availability and affordability of bit-rate-variable transceivers (BVT’s), as well as flexible spectrum wavelength-selective switches (FS-WSS’s), and a whole new family of flexible gear [13]. Only then will system flexibility gains be fully reached, but technology is already evolving to provide the necessary functionalities for this purpose.

Client needs are specified not only in terms of bitrate, but also in terms of reach. With conventional modulation schemes, short reaches will normally imply large noise margins, which represent wasted power resources. In a flexible scheme, advanced modulation schemes may then use this extra power to increase spectral efficiency, i.e. to allow more bits per second to be supported by the same amount of bandwidth, thus reducing the need for bandwidth of short enough optical paths. Therefore, flexible equipment must be able to deal with a set of modulation schemes, switching from one to another as needed to optimize resource usage.

Notice that the flexible network will provide more efficiency through the exploitation of new degrees of freedom, thus requiring smarter management tools and heavier control planes. As evidenced by the results of [10], more degrees of freedom may lead to inefficient operation if management is unaware of the traffic environment or is not smart enough. For this reason, research must focus on smart RSA algorithms and their robust implementation.

Responding to this need, our research is planned to tackle increasing levels of complexity, in order to extend the results of [10] and [14] to contemplate:

- minimization of the total mean amount of resource idle time, under dynamic as well as incremental traffic, instead of just the mean time to exhaust the spectrum under incremental traffic, as already done;
- extension of the single link results to linear topologies;
- the RSA problem in mesh topologies;
- survivability in flexible networks, noticing that backup resources will then be used both to provide on-demand bandwidth enhancement for flexible service providers and protection capability to reliable service, so it is natural to exploit the possibility of sharing these backup resources as far as the corresponding SLA agreements will allow it without undue risk of non-conformity.

**IV – Competition in a flexible environment**

The additional degrees of freedom afforded by flexibility may lead competing flexible operators to specialize their service provision to different segments of the (differentiated) service space. This possibility motivates us to revisit the analysis made in [8] to reconsider the user and operator games in a flexible environment. The proposed research
may also generate new insights into appropriate methods to combine game theory and traffic theory in the Internet environment. Notice that traffic theory emerged during the first half of the 20th century in association with the public switched telephone network, i.e. with a circuit-switching network that services a Poissonian traffic in a fairly monopolistic setting. This was succeeded in the second half of the century by the packet-switched data network, i.e. by a best effort service network predicated on sufficient physical resources. So, neither of them provided a game setting. The current, - and foreseeable, - optical network of today and tomorrow is a circuit switching network that services a bursty, multi-service, escalating, flat-priced, unregulated traffic that is bound to generate a congested networking environment where all stakeholders must game in order to reach their desired payoffs. In this environment, we believe that traffic and game theories must cooperate in the production of useful modelling.

Acknowledgement. This work was supported by CNPq.

References