Congestion, Private Peering and Capacity Investment on the Internet*

Fabio M. Manenti[†]

March 2003

PRELIMINARY - COMMENTS ARE WELCOME

Abstract

This paper presents a model of private bilateral and multilateral peering arrangements between Internet backbone providers when the network is congested. We study how different forms of interconnection and the competitive conditions of the market affect backbones' investments in network and peering point infrastructures. We show that networks and peering point capacities are equilibrium complements; increasing competition reduces capacity investments (under-investment), thus worsening the quality of service both with multilateral and bilateral peering. Under bilateral peering the inefficiency is less severe. Because of under-investment, welfare may be lower when the market is more competitive. We also show that asymmetries between backbones, which can take the form of unequal content distribution or product differentiation, may reduce under-investment and improve the quality of service. The introduction of an "inverse capacity interconnection fee" where providers pay each other a fee which is negatively correlated with their installed capacity may play the role of a coordinating mechanism towards a Pareto superior outcome.

Keywords: Internet, peering, congestion, QoS, capacity investment, interconnection.
J.E.L. codes: D4, L13, L96.

---

[*] This paper has been completed while the author was visiting the Haas Business School at the University of California, Berkeley. The author would like to thank Glenn Woroch for the hospitality provided. I would also like to thank Hal Varian, Tommaso Valetti, Ernesto Somma, Benedetto Gui and Stefano Comino for useful comments. The usual disclaimer applies.

[†] Address for correspondence: Dipartimento di Scienze Economiche "M. Fanno", Università di Padova, Via del Santo 33, 35123 PADOVA (Italy), Tel. (39) 49 8274238, email: `manenti@decon.unipd.it`.

1 Introduction

The Federal Communications Commission reports that on the Internet the volume of traffic is doubling every 100 days, which is remarkable given that telephone traffic has typically increased only by about 5 percent a year. Estimates on the number of U.S. households that are connected to the Internet show an increase up to 60 million in 2003 (Carmel et al., 1999); Europe and the other continents are actually lagging behind but as long as populations will gain access, it is likely that in the next few years these countries will experience an intense growth rate in the number of Internet access. At the same time, due to a proliferation of data intensive real-time applications such as Internet telephony, instant messaging and video conferencing, this substantial growth in general demand for access is accompanied by an intense increase in demand for broadband, high-quality access to the Internet.

Although the transmission capacity of the international Internet backbone has increased at a fast rate in the last few years, the rapid growth in demand for access combined with the increase in the demand for broadband applications are actually creating congestion of network's infrastructures, usually measured in terms of either excessive delay or loss of service altogether.[1] Furthermore, the transmission protocol adopted for the Internet and its infrastructures was not designed for delivering continuous speech or videos; for this reason the control of congestion and the related issue of "quality of service" (usually referred as QoS) have become two major concerns on the Internet both at the theoretical and practical level.[2]

The issue of congestion is tightly linked with that of interconnection. Providers compete against each other in the market for individual "surfers" but, at the same time, they are in various way interconnected in order to offer global connectivity. Clearly, the way in which providers interconnect and/or how they share common infrastructures affects the degree of congestion of their networks and the QoS that they are able to offer.

Congestion can be roughly defined as the amount of traffic per unit of transmission infrastructures; the most direct way that providers have to mitigate network congestion is through improving their transmission capacity. The aim of the paper is to analyse the relationship between networks interconnection, congestion and providers' incentives to expand capacity in order to improve the quality of their

---

[1] There are many web sites offering real time measures of the reliability of Internet connections. See, for example, the Internet Traffic Report (ITR) web site where the flow of data around the world is constantly monitored. In order to provide a measure of the current level of congestion of the world wide web, ITR offers measures of packet loss at each router and an index of the response time of transmissions (see `www.internettrafficreport.com`).

[2] Although there is no general definition, the International Telecommunication Union (ITU), defines a number of QoS parameters to describe the speed and reliability of data transmission, e.g., throughput, transit delay and error rate.

services.

Before proceeding with the model, it is useful to discuss briefly the issues at stake. These are described in the next section where we also offer a short synthesis of our results. Section 3 presents the basics of the model and formally defines the concepts of traffic and congestion; section 4 and 5 discuss and compare congestion under different forms of interconnection. Section 6 extends further the model to the case of product differentiation and discusses interconnection charges while section 7 concludes the paper.

2 Background and main results

Together with several technological innovations, a growing attention has been directed towards the definition of distribution mechanisms, primarily pricing mechanisms, in order to provide the efficient use of network's scarce resources. The literature on how to price congestible network resources is by now quite well developed; the proposed solutions range from the simple use of flat pricing to very complicated approaches based on auctions over packets,[3] to the use of multi-class service pricing[4] or on static and dynamic priority pricing.[5]

A full survey of the various proposals of congestion pricing is beyond the scope of this paper.[6] At a theoretical level, these pricing schemes may alleviate congestion and also induce an efficient allocation of resources; nevertheless, the use of pricing mechanisms to control congestion requires the individuals' usage of resources to be constantly monitored and billed at the various nodes where packets' transmission occurs. Internet traffic "accountability" is both extremely costly and difficult to realise with the current network management technology and this often makes congestion prices more a theoretical than a practical issue.

---

[3] This is the well known approach proposed by MacKie-Mason and Varian (1995a,b); they introduce a smart market pricing where, prior to transmission, users inform the network on how much they are willing to pay for the transmission of a packet: the network runs an auction which takes account of the marginal congestion cost of an extra packet and accepts only those packets with a bid that exceeds this marginal cost.

[4] This scheme was originally proposed in Odlyzko (1997 and 1999) and then refined by Gibbens *et al.* (1998 and 2000); Odlyzko's proposal is based on the pricing of the Paris Metro: the total network capacity is divided into several sub-networks. Each network is priced differently; the service level is implicitly defined by users that select one of the sub-networks according to their willingness to pay.

[5] Priority pricing schemes are schemes that force users to indicate the value of their traffic by selecting a priority level and to pay according to the priority. These pricing policies may be dynamic if prices fluctuate as a result of different network conditions. See Cocchi *et al.* (1993) and Gupta, Sthal and Winston (1995, 1996 and 1997).

[6] See Wiseman (2000), DaSilva (2000) and Falkner (2000) *et al.* for non technical discussions of current proposals.

Pricing mechanisms act on the demand side of the market; the alternative, and more direct way to ameliorate congestion and to improve the quality of service, is to act on the supply side. This may be done indirectly, implementing new technologies such us "caching", which can effectively reduce congestion through traffic aggregation,[7] or directly, by expanding network infrastructures. In this paper we focus on this latter alternative.

Clearly, a provider endowed with more capacity, is generally able to offer a better QoS to its customers. Very few papers have analysed the incentive to expand capacity by Internet backbone providers; MacKie-Mason and Varian (1995a) present a model of individual demand for access to the Internet with congestion which takes the form of a negative network externality: individuals' willingness to pay for access decreases with network congestion, with this latter positively related to the number of individuals having access to the Internet and negatively to providers' capacity. In a very simple framework, the authors show that with perfect competition between providers, the socially efficient level of capacity is achieved. As a consequence, in order to induce both an efficient utilization of resources and an adequate level of investment in transmission capacity it is sufficient to promote competition between backbone providers.

Nevertheless, this result heavily depends on the assumption that the degree of congestion suffered by a provider depends only on the amount of its customers and capacity and not on those of the rivals. This is a very strong limitation which also drives the model away from what we observe in the reality where backbone providers do not generally operate in isolation: providers not only compete against each other but they need also to find a way to collaborate in order to interconnect their infrastructures and to provide global connectivity to their respective customers.

Backbone providers are private firms that own and operate large capacity links, routers and switching equipments. These firms connect at the so called *public peering points*, also known as Network Access Points (NAP), where the traffic between customers of different providers is processed, routed and redirected to its destination.

The term "peering" indicates a settlement free interconnection between providers. Following Weiss and Shin (2001), peering arrangements can be divided into three categories: 1) according to the openness of the agreement, peering can be *public* when anyone can join it,[8] or *private* otherwise, 2) according to the number of the peering partners it can be *bilateral* or *multilateral* and 3) according to the market in which it occurs, it can be *primary* peering between backbone providers

---

[7] Caches come in many types, but they all work the same way: they store information where it can be accessed quickly. A web browser cache stores the pages, graphics, sounds, and URL's of online places visited by an individual on his/her hard drive. When the customer goes back to the page, everything doesn't have to be downloaded all over again. This speeds things up.

[8] The only requirement is that each partner pays for the equipment needed to interconnect at the peering point.

or *secondary* peering if it occurs in the downstream market.[9]

In this paper we concentrate only on the second category and focus on the characteristics of bilateral *vs* multilateral private peering. Private peering has emerged in recent years as a way to avoid the very congested public peering points; originally, at the beginning of the commercial Internet, there were only four public peering points[10] and networks exchanged traffic bound for each other's destinations at these sites only; although these points have so far guaranteed global connectivity, given the rapid growth in Internet traffic, the NAPs became soon congested, leading to delayed and dropped packets.

In order to avoid transmission delays and failures, backbones started to get around the public peering points and to interconnect directly with one another signing bilateral/multilateral private peering arrangements.[11] Thanks to these private arrangements, providers are now able to exchange directly reciprocal off-net traffic. It is estimated that 80% of the traffic is now exchanged via private peering.[12]

Usually, these private peering points are run on a quid pro quo basis and frequently they permit traffic exchange free of charge.[13] There isn't a clear rule that governs private peering; the term "peer" suggests equality of treatment and, usually, backbones decide to set a private peering point when they are "similar in size". This is a very broad definition; more precisely, providers may peer when they have a similar geographic coverage, similar traffic volume or similar number of customers and also when they have similar capacity. Clearly, it is very unlikely that two backbones are similar along so many dimensions, and the question then is how backbones weight one dimension against the other.

When two providers interconnect, they agree to share their infrastructures and to exchange reciprocal traffic; having observed this, the main effect of an interconnection agreement is to make the degree of congestion incurred by a provider no longer independent from the level of congestion of the interconnected rival. An example may help to understand this crucial point: when an individual downloads a document from a certain web site which is hosted on a different provider, the "sending"

---

[9] The Internet has a hierarchical structure and it is usually broken down into different layers or tiers: tier 1 providers' are those that have access to the global Internet Routing Table and that provide international connectivity. Examples include UUNet, Cable & Wireless, Sprint, Qwest, Genuity and AT&T. The second tier includes smaller providers, usually those with national presence while tier 3 providers are players with only a regional presence and no national backbone. Secondary peering occurs between providers of tiers 2 and 3.

[10] They were set up in San Francisco, Chicago, New York and Washington D.C.

[11] For a complete non technical discussion of the evolution of peering arrangements, see Kende (2000) and Bailey (1995).

[12] See Kende and Oxman (1999).

[13] When the agreement between providers involves a settlement payment, thus configuring a provider-customer relationship between them, then the agreement is known as traffic interconnection.

and the "receiving" providers are different;[14] in this case, the time needed to deliver the document depends on the degree of congestion of both providers and possibly on other factors if the transmission is routed through transit providers.

More generally, the level of congestion suffered by customers of a certain provider, and the QoS they are offered, is the result of a complex mix of factors which includes the amount of capacity installed by all the providers involved in the transmission, the structure of the interconnection agreements between these providers and, finally, also the degree of competition between them.

It is therefore evident that the two issues of interconnection and congestion are tightly linked together; the aim of this paper is to analyse these relationship and to describe how the incentives for backbone providers to invest in transmission capacity in order to offer better QoS change with the type of interconnection agreement between firms and with the competitive conditions of the market.

The analysis is conducted building on a recent contribution proposed by Dewan *et al.* (2000). These authors present a simple and tractable model of private peering arrangement in a duopoly setting. The originality of their approach is the way in which they model interconnection: backbones interconnect at a common private peering point where they exchange traffic bounded for each other destination. Firms compete after having invested in their network capacity and in the capacity of the common peering point; according to this framework, the degree of congestion on the network of each provider depends also on the congestion on rivals' network. The authors focus mainly on the characteristics of the peering point and show that, due to the negative congestion externality, the larger network (the one with more customers and content) prefers to have a lower capacity for the private interconnection point.

We go further on this analysis by extending the model in two directions: we solve for different peering arrangements (bilateral *vs* multilateral) and for different competitive conditions of the market. We show that once the relationship between interconnection and congestion is taken into account, the efficiency result of MacKie-Mason and Varian does non longer apply. Transmission capacity is a public good and as consequence there is a serious problem of under-investment, with competition that reduces incentives for upgrading network capacities; this has a negative impact on the quality of services and, therefore, on social welfare. We also show that under-investment is less severe with bilateral peering and, generally, when providers are somewhat differentiated either in their QoS or in the product they offer to their customers.

Although the focus is on pricing strategies, an interesting paper which spirit is close to ours' is Gupta *et. al* (1999); they run an experiment to analyse, in a monopolistic framework, the effects of different congestion based pricing on incentives for

---

[14] A transmission which is entirely controlled by the same provider is said "end-to-end" transmission; this is a situation which is rarely observed.

network capacity expansion as the demand for services increases. Their results show that the ability of a pricing scheme to provide incentives for capacity expansion is strictly related to the cost of capacity and to the individuals' willingness to pay for network services. Nevertheless, since the model is restricted to monopoly, the issues of interconnection and competition, which are our main focus, are not taken into account.

3 The model

3.1 Demand for access and social surplus

Throughout the paper, we refer to backbones or providers and we will use these two terms interchangeably; the market that we have in mind is the one between Internet backbone providers, i.e. those operators providing international connectivity. Never the less, with small changes, our model can be applied also to national Internet service providers or to all those providers with backbone infrastructures.

There is an continuum population of consumers; each consumer is indexed by his/her basic willingness to pay for having access to the Internet. Let the utility of joining the Internet through provider i ($i = 1, \ldots, n$) enjoyed by consumer indexed by r be

$$U = r - D_i - p_i \qquad r \in [0, A], \quad A > 0$$

where $r > 0$ is a positive benefit of joining the Internet which is uniformly distributed in $[0, A]$, p_i is the per unit time price charged by backbone i and D_i is the disutility (or dis-benefit) of delay incurred by customers. While r is independent from which backbone is providing access, the disutility of delay depends on the degree of congestion on network i. Consumers are assumed heterogeneous in their basic willingness to pay and homogeneous in the disutility of congestion. This very simplified utility function assumes away service differentiation: for instance, real-time voice and video applications are very sensitive to delay and jitter while traditional data applications are more sensitive to losses.[15] The only differentiation between backbones comes from the degree of congestion, and it is not intrinsically related to the type of services offered; we will reintroduce product differentiation in the last part of the paper

This functional form for utility implies two further assumptions. First, users do not *multihome*, i.e. they join one and only one provider; secondly, users pay a price per unit time of connection which does not depend on the volume of data

---

[15] See Cocchi et al. (1993) and Jiang and Jordan (1996) for a theoretical discussion. Recently the project "Internet Demand Experiment" (INDEX) conducted at the University of California at Berkeley has attempted to experimentally characterise users behavior when confronted to differentiated services; preliminary results and implications are in Varian (2002).

transmitted. These assumptions are taken to maintain model tractability, and we believe that they do not imply a major loss of generality. In particular, the second assumption, which is common in the literature,[16] is reasonable when users generate the same amount of traffic. In this simplified framework, subscription and usage decisions coincide. This fact, combined with the assumption of linear prices, is also consistent with the scope of our analysis that is to model the relationship between congestion and interconnection policies when providers charge flat prices.

There are n providers on the market. At the equilibrium, consumers must be indifferent between joining provider i and provider j; from the indifferent condition:

$$p_i + D_i = p_j + D_j \qquad \forall j \neq j \tag{1}$$

Let us call ϕ this "hedonic" price. Only customers with $r \geq \phi$ buy the connection, therefore at the equilibrium there is a total amount of consumers $x_t = A - \phi$ connecting to the Internet. Hence, using (1), the demand function for ISP i is:

$$p_i = A - D_i - x_t \tag{2}$$

From the demand function we compute the consumers surplus. Total surplus enjoyed by consumer of type r when she/he is a customer of ISP i is:

$$CS(r) = r - D_i - p_i \tag{3}$$

where p_i is given in (2). At the equilibrium, only those consumers with a basic willingness to pay r bigger than $\phi = A - x_t$ join the network. Integrating (3) over all consumers who do join the network we derive the consumers' surplus:

$$CS = \int_{A-x_t}^{A} [\gamma - A + x_t] d\gamma = \frac{x_t^2}{2} \tag{4}$$

Social welfare is simply defined as the sum of consumers' surplus and firms' profits: $W = CS + \sum_i \pi_i$; π_i represents the profit of firm i and it will be defined shortly, once the concepts of network traffic and congestion have been introduced.

3.2 On-net and off-net traffic

In order to characterise congestion, it is crucial to differentiate between *on-net* and *off-net* traffic flows.

There are mainly two categories of users on the Internet: sites/portals that place their information content on the web (and for this reason they are often referred as "content providers") and individuals who "consume" the information. Nowadays

---

[16] See Gibbens *et al.* (2000) among others.

85% and more of all Internet traffic is generated by receiver requests, such as web page retrieval or downloading files;[17] this implies that the request traffic generated by individuals is generally very small compared to the traffic of web pages or file downloads. Accordingly, without loss of generality we assume away to zero the traffic flow from individuals to content providers and we concentrate only on the traffic from web sites to individuals.[18]

Consider an individual who downloads the content from a certain web site. The web site and the individual may either be customers of the same provider or customers belonging to different operators. In the first case, the flow of information from the web to the consumer goes entirely on the infrastructures of the same provider; for this reason, this traffic is said *on-net* traffic. If, conversely, the consumer downloading the content and the web site offering the content belong to different providers, then the traffic is said *off-net* since it is goes from a provider, the one hosting the web site, to a different one, the one where the consumer is located.

3.3 Private peering arrangements, traffic and congestion

Although very interesting, the aim of this paper is not to analyse when and why providers decide to peer; we do not analyse the bargaining process that yields backbones to set up a point of interconnection but we assume that providers actually agree to peer.[19] Our aim is to analyse how different interconnection arrangements and market structures affect providers' incentives in mitigating congestion through capacity investment. Each provider has its own infrastructures, routers and switches; let k_i be the capacity of provider i's network. In addition to their own capacity, providers invest in infrastructure and capacity to build the interconnection point.

We restrict the attention to the analysis of two types of arrangements which are commonly observed: *bilateral* and *multilateral* peering. In the first scenario, each provider peers with any one on the market in a one-to-one relationship. Conversely, peering is said multilateral when more than two backbones peer. In the reality, these two regimes often coexist; furthermore, other and more complex forms of interconnection can be observed. For example, many backbones have adopted a hybrid approach to interconnection, peering with some backbones and signing

---

[17] See Weiss and Shin (2001).

[18] This simplifying assumption is common in the literature on the Internet, see Laffont *et al.*(2001a, 2001b).

[19] The characteristics of the bargaining process between Internet service providers have been recently studied in Besen *et al.* (2001). These authors provide analytical support against the conventional wisdom that large backbones may refuse to peer with smaller ones. They claim that as the Internet develops a richer sets of interconnection arrangements such as secondary peering, transit and multi-homing, the incentives of large backbones to refuse to enter into or degrade peering arrangements will be reduced.

"transit" agreements with others.[20] Although our model does not encompass these forms of interconnection arrangements, we believe that the approach we provide in this paper represents a valid starting point to analyse these controversial issues.

The regimes that we present share three characteristics: *i)* providers who interconnect build a private peering point where reciprocal off-net traffic is exchanged, *ii)* traffic is exchanged on a settlement-free basis and *iii)* the cost of building the peering point is equally divided among providers: each backbone pays for the equipment and the transmission capacity needed to build the peering point.

It is useful to provide a graphical representation of the two regimes analysed in the paper; suppose, for simplicity, that there are only 3 providers on the market: figure 1 presents multilateral peering where backbones set up a common peering point and figure 2 shows the bilateral peering scheme where each provider peers with all the others. Let k_p be the capacity of the multilateral peering point and $k_{p,i}^{j}$ the capacity of the peering point between provider j and provider i in the case of bilateral arrangement. All through the paper, we indicate with $K = \sum_i k_i$ the total (own) capacity installed by providers.

We can now define formally the traffic flows and the degree of congestion in the two interconnection regimes; we generalise the model presented in Dewan *et al.* (2000). All through the paper we make the following assumption:

Assumption 1. *Let $\alpha_i \in (0,1)$, with $i = 1, \ldots, n$, be the share of the total content available on line hosted on provider i's infrastructures, with $\sum_i \alpha_i = 1$.*

In other words we assume exogenous the amount of content hosted by each provider, which therefore compete only for individual users. Assumption 1 is taken on practical grounds: while individuals can change provider very easily, businesses and web sites incur in very high switching costs when changing provider. These costs are mainly due to the need of adapting the technology and the architecture of their sites to different requirements when changing provider.

This assumption, which makes the approach a short period analysis, helps to simplify the model considerably; nevertheless, in front of a little loss of generality, we are able to characterise how the equilibrium of the model changes when content distribution changes. This is a very interesting issue that we address in the paper.

Note that since individuals can download any of the content irrespective of the provider they use to access the Net, asymmetric content distribution among providers is not "per-se" a form of providers' product differentiation; nevertheless, as it will become clear in the coming section, α_i affects the traffic transmitted on provider i's infrastructures and therefore, through its impact on congestion, it is indirectly a form of quality differentiation.

---

[20] In a transit relationship, if provider A is refused to peer with B, then it can take transit instead. In this case, A pays B for access to B's customers and also to its peering relationships.

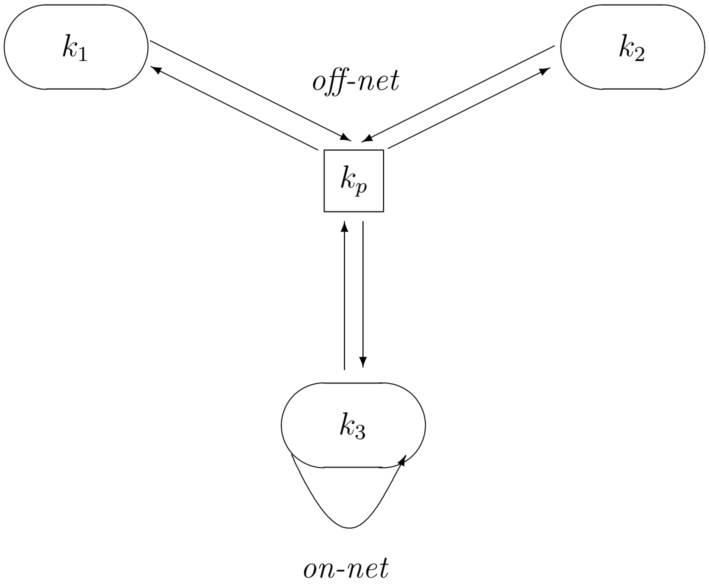

Figure 1: multilateral peering and traffic flows

3.3.1 Traffic flows and the level of congestion with multilateral peering

Consider now the multilateral peering framework. The total traffic on provider i's infrastructure positively depends on the amount of content available on the web sites hosted by the same provider: the more content is available on provider i web sites, the more users contact these sites and the higher the traffic carried on provider i's network.

Under the assumption that users download a fixed amount of content (i.e. they generate the same amount of traffic), then the total traffic carried on a given network, t_i, is proportional to the content hosted on the same network, α_i, and to the total number of customers on-line x_t. Therefore the total traffic on provider i's network is:

$$t_i = \alpha_i x_t \qquad i = 1, \ldots, n \tag{5}$$

t_i may be higher either because there are more customers on-line or because provider i hosts more content. Part of this traffic stays on-net and part goes off-net; the latter depends on the amount of content hosted by the other providers and is given by $(1 - \alpha_i)\, x_i$, where x_i represents provider's i amount of customers. The peering point gets all the cross traffic between networks; adding up this last expression for all the n backbones, the total amount of off-net traffic is given by

$$t_p = \sum_{i=1}^{n} (1 - \alpha_i)\, x_i \tag{6}$$

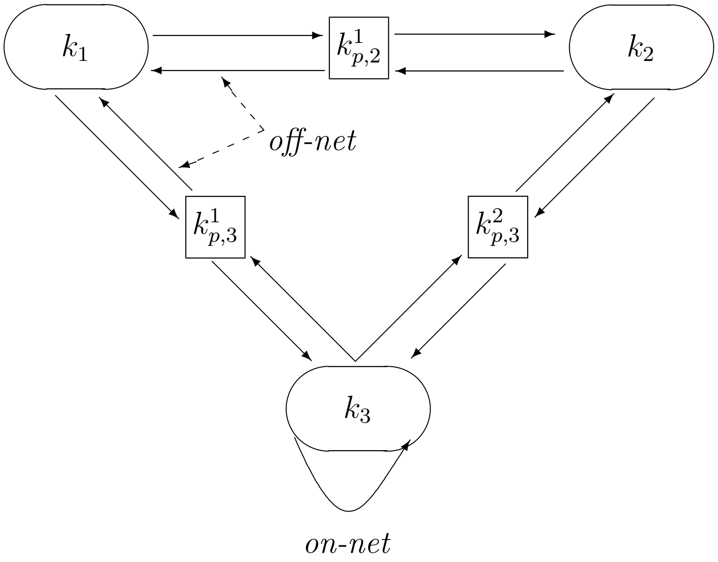

Figure 2: bilateral peering and traffic flows

We use these expressions to determine the delay suffered by customers on the network of each provider. The delay, and therefore the QoS, that provider i is able to offer depends on the degree of congestion of its infrastructures. Broadly speaking, the congestion of a network is proportional to the ratio between traffic t and capacity k: for a given traffic, congestion is higher the less capacity is in place.

In this framework, the average (expected) delay suffered by an individual when surfing the net is a combined measure of the congestion of his provider's network and the congestion at the peering point. Given that k_i and k_p are the capacities of provider i's network and of the peering point respectively, the average delay suffered by a customer on backbone i is:

$$D_i^m = \alpha_i \frac{t_i}{k_i} + \sum_{j \neq i}^{n} \alpha_j \left(\frac{t_j}{k_j} + \frac{t_p}{k_p} \right) \tag{7}$$

where the superscript m stays for "multilateral". This expression simply states that the delay/QoS offered by provider i is an average between the delay incurred for on-net transmissions and for off-net transmissions, the former that depends only on provider i's own traffic and capacity, while the latter related to both rivals' and the peering point traffic and capacities. It is now clear how both the provider j's own capacity, k_j, and the amount of traffic it generates, t_j, affect i's congestion and quality of service, thus creating an externality.

3.3.2 Traffic flows and the level of congestion with bilateral peering

Suppose now that backbones peer only bilaterally as presented in figure 2; while expression (5) still represents the total traffic on provider i's infrastructures, the expressions for the traffic at the peering point and the delay must be slightly modified. With n providers there are $\frac{n(n-1)}{2}$ private peering points; the average total cross traffic transiting through the peering point between backbone i and backbone j is given by the sum of the traffic from j to i and the traffic in the reverse direction; formally:

$$t^i_{p,j} = \alpha_i x_j + \alpha_j x_i \quad \text{with} \quad i \neq j, \quad i,j = 1,\ldots,n. \tag{8}$$

As a consequence, the average delay suffered by the representative customer of provider i is given by:

$$D^b_i = \alpha_i \frac{t_i}{k_i} + \sum_{j\neq i}^{n} \alpha_j \left(\frac{t_j}{k_j} + \frac{t^i_{p,j}}{k^i_{p,j}} \right) \tag{9}$$

where the superscript b stays for "bilateral".

Clearly, when there are only two providers on the market, the multilateral and the bilateral frameworks coincide. The formal simplicity of this case makes it possible to analyse the equilibrium in detail; for these reasons it is useful to start the analysis assuming $n = 2$. In the last part of the paper we move to the case of $n = 3$; three is the smaller number of providers to distinguish between bilateral and multilateral peering but that at the same time allows to maintain model's tractability.

4 The duopoly benchmark

4.1 Timing and profit

Investment in capacity is a long run decision; therefore it is reasonable to assume that providers choose their capacity at the beginning of the game and compete afterwards. Furthermore, without loss of generality, we assume that providers decide simultaneously how much to invest in capacity for their own network and how much for the peering point.[21] Finally, we assume that providers compete á la Cournot, by setting quantities.

With only two providers, bilateral and multilateral peering coincide. The demand function faced by each provider is simply obtained by plugging (7) into (2). For the sake of simplicity, we normalise away to zero the marginal cost of providing access to an additional customer. This assumption is taken also on practical grounds: residential users usually connect to the Internet via the existing telecom infrastructure (through "dial-up" connection) and the provider does not incur any cost of connecting the user. Business users may connect through dedicated leased lines; although in this case the cost of connecting one more customer may be positive, this feature can be easily incorporated into the model without changing its qualitative nature. On the contrary, providers face the cost of upgrading transmission capacity in order to improve network efficiency and the QoS; for simplicity we assume the following:

Assumption 2. *The cost of installing k_i amount of capacity is given by*

$$c(k_i) = \theta k_i \qquad\qquad i = 1, 2, p$$

where $\theta > 0$ is the marginal cost of installing capacity.[22]

Finally, according to the collaborative nature of a peering agreement, it is natural to assume that

Assumption 3. *The capacity of the peering point k_p is chosen to maximise backbones joint profits and the cost is equally shared between the interconnecting parties.*

Therefore, provider i's profit is:

$$\pi_i = p_i x_i - \theta\left(k_i + \frac{k_p}{2}\right) \qquad i = 1, 2$$

where the second term represents provider's total cost of installing capacity and it is made of two components: the first one is the cost of building its own capacity, k_i, while the second one is the share of the cost of the private interconnection infrastructure which is equally divided between the two providers. Using the demand function given in (2) and the delay function as in (7), provider i's profit is therefore:

$$\pi_i = \left(A - \frac{{\alpha_i}^2 x_t}{k_i} - \alpha_j\left(\frac{\alpha_j x_t}{k_j} + \frac{\alpha_j x_i + \alpha_i\, x_j}{k_p}\right) - x_t\right) x_i - \theta\left(k_i + \frac{k_p}{2}\right)$$

---

[21] As it will become clear later, the solution does not change if k_p is chosen once network capacities are in place.

[22] Capacity is often provided by public telecommunications operators (PTOs) through leased lines; θ can therefore be interpreted as the price charged by PTOs for a leased line of a certain capacity.

We solve the model by backward induction. In the second stage, competition occurs given capacity and in the first stage, backbones anticipate second stage equilibrium and choose capacities.

4.2 The symmetric solution

Each provider chooses x_i to maximise profits; firms' equilibrium outputs given capacities are obtained by solving the usual system of first order conditions and are given by:

$$x_1(k_1, k_2, k_p) =$$

$$= \frac{A k_1 k_2 k_p \left[\left((k_1 + \alpha^2)k_2 + (1-\alpha)^2 k_1\right) k_p + (3\alpha - 1)\alpha k_1 k_2\right]}{\left((k_1 + \alpha^2)k_2 + (1-\alpha)^2 k_1\right) k_p \left[3\left((k_1 + \alpha^2)k_2 + (1-\alpha)^2 k_1\right) k_p + 2(5\alpha^2 - 5\alpha + 2)k_1 k_2\right] + 3(\alpha - 1)\alpha^2 k_1^2 k_2^2}$$

$$x_2(k_1, k_2, k_p) =$$

$$\frac{A k_1 k_2 k_p \left[\left((k_2 + \alpha^2)k_1 + (1-\alpha)^2 k_2\right) k_p + (3\alpha - 2)(\alpha - 1) k_1 k_2\right]}{\left((k_2 + \alpha^2)k_1 + (1-\alpha)^2 k_2\right) k_p \left[3\left((k_2 + \alpha^2)k_1 + (1-\alpha)^2 k_2\right) k_p + 2(5\alpha^2 - 5\alpha + 2)k_1 k_2\right] + 3(\alpha - 1)\alpha^2 k_1^2 k_2^2}$$

where we have replaced α_1 with α and α_2 with $1-\alpha$ to avoid cumbersome notation.[23]

In the first stage, each provider decides its level of capacity k_i, given rivals' choice k_j; furthermore, the two providers decide cooperatively the capacity of the private peering point k_p. Indicating joint profits with $\Pi = \pi_1 + \pi_2$, then the solution vector $[k_1, k_2, k_p]$ of firms' choice of capacities, is the simultaneous solution of the following set of first order conditions:

$$\frac{d\pi_1\left(x_1(k_1, k_2, k_p), x_2(k_1, k_2, k_p), k_1, k_p\right)}{dk_1} = 0 \tag{10}$$

$$\frac{d\pi_2\left(x_1(k_1, k_2, k_p), x_2(k_1, k_2, k_p), k_2, k_p\right)}{dk_2} = 0 \tag{11}$$

$$\frac{d\Pi\left(x_1(k_1, k_2, k_p), x_2(k_1, k_2, k_p), k_1, k_2, k_p\right)}{dk_p} = 0 \tag{12}$$

While the general solution to these equations is quite complex and can be analysed only numerically, the symmetric case can be easily derived. Symmetry occurs at the equilibrium when the two providers host the same amount of content, formally when $\alpha = 1/2$. In the symmetric case, we prove the following:[24]

---

[23] Second order conditions are

$$\frac{d^2\pi_i}{dx_i^2} = -2\,\frac{(1-\alpha)^2}{k_i} - 2\,\alpha\left(\frac{\alpha}{k_j} + \frac{\alpha}{k_p}\right) - 2 < 0 \qquad i, j = 1, 2$$

which are clearly satisfied.

[24] The proof of this and of all the mathematical results are in the Appendix.

Proposition 1. *In a bilateral peering relationship, providers' and peering point capacities are complements:*

$$\frac{dk_i}{dk_j} > 0, \quad \text{and} \quad \frac{dk_p}{dk_i} > 0 \quad i, j = 1, 2.$$

At the symmetric equilibrium, the total amount of network capacity and the capacity of the peering point are respectively

$$K = \frac{4\,A^2 - 3A\,\sqrt{\theta}(8+\sqrt{2}) + 9\theta(4+\sqrt{2})}{12(3\theta + A\sqrt{\theta})} \qquad k_p = \frac{2A\,\sqrt{2\theta} - 3\,\theta(1+2\sqrt{2\theta})}{12\theta} \tag{13}$$

Interestingly, network capacities are strategic complements: the more a peering partner invests in capacity, the more capacity is installed also by the other party. This is due to a standard "business stealing" effect. By improving its transmission infrastructures, provider i increases its QoS; as a consequence, some consumers of provider j now prefers to demand access to provider i. In order to protect its profits, provider j reacts to i's investment by also improving its capacity (and QoS). This competitive process stimulates capacity investments up to the point where the marginal benefit of additional capacity is perfectly off set by the additional cost of installing it.

Furthermore, network and peering point capacities are also complements; they too move together although, in this case, is not entirely correct to define them as strategic complements given that k_p is chosen in collaboration between the two firms. This last result is interesting especially if we think to the widespread concern among practitioners and academics about the possible trade-off facing backbones when upgrading their networks: in terms of choosing between improving peering capacity or network capacity, backbone may concentrate their resources on the latter at the expense of the former (see Kende and Oxman, 1999). We have shown that the two objectives are not necessarily one against the other.

The complementarity between k_i and k_p can be intuitively explained by observing that the QoS that a single provider is able to offer to its customers does not depend only on the capacity of its network but also on both the capacity installed by the rival and on the transmission efficiency of the peering point. When a backbone decides how much transmission capacity to install, it knows that its customers will enjoy full benefits of it only if the peering point is sufficiently efficient to exchange traffic without delays: the more efficient the transmission through the peering point, the stronger the incentive for each provider to expand its own network; or, equivalently, the marginal productivity of provider i's investment in its own capacity k_i, increases with k_p.

A consequence of this observation is that any regime or regulatory policy that provides incentives for the peering parties to invest in peering capacity may have the positive effect of stimulating a general network expansion.

4.3 The asymmetric solution

Suppose now that the two providers host a different amount of content: $\alpha \neq 1/2$. In this case the solution can no longer be obtained formally and we need to resort to numerical simulations. In the following table, we present the solution for different degrees of asymmetry when $A = 10$ and $\theta = 1$.[25] The table also reports providers profits, welfare and, as a proxy for the QoS offered, the amount of delay experienced by customers of each provider.

α	0.5	0.45	0.4	0.3	0.2
k_1	0.99	0.83	0.69	0.41	0.07
k_2	0.99	1.16	1.34	1.76	1.75
K	1.98	1.99	2.03	2.16	1.82
k_p	1.4	1.38	1.33	1.1	0.37
π_1	4.9	4.66	4.4	3.77	1.56
π_2	4.9	5.18	5.48	6.28	6.31
W	17.66	17.67	17.69	17.62	11.5
D_1	2.7	2.78	2.86	3.08	4.7
D_2	2.7	2.63	2.55	2.43	3.16

Table 1: bilateral peering with asymmetric content

The asymmetry between backbones very much affects the equilibrium of the model. In particular, as α gets smaller up to a certain level of asymmetry:

1. the lower k_1 and the larger k_2; total own capacity K increases with the asymmetry;

2. the lower the private peering capacity;

3. the lower provider 1 profits and the larger provider 2 profits;

4. the higher the welfare;

5. the worst provider 1 QoS (higher delay) and the better the QoS guaranteed by provider 2.

These results hold true up to a certain degree of asymmetry, while for sufficiently small values of α (high asymmetry), a further increase in α induces a decrease in all

---

[25] Different values of the parameters can be considered but the qualitative properties of the solution are unaffected. Second order conditions have been checked.

the investments in capacities, a general deterioration in the QoS for both providers and also a fall in social welfare.

Nevertheless, the table shows that a certain degree of asymmetric content distribution may play a positive role since it induces an increases in the installed total network capacity and despite the lower capacity of the private peering point, also an increase in the level of welfare, at least for $\alpha \in (0.4, 0.5)$ in the table.

Interestingly, the delay suffered on the two backbones' infrastructures moves opposite than expected: the provider hosting more content, which should be therefore more congested, at the equilibrium suffers less delay/congestion than the smaller backbone. As α moves away from $1/2$, the provider hosting more content, here provider 2, has incentives in installing more capacity while provider 1 does the opposite. Furthermore, as α gets smaller, provider 2 increases its capacity more than the reduction of the rival, thus increasing total network capacity. As a consequence, as content is more asymmetrically distributed, the Quality of Service offered by provider 2 improves while that of provider 1 is worsened even if provider 2 handles more traffic.

Consumers are willing to pay more for a better service and this translates into higher profits for provider 2; finally, the overall impact on welfare is also positive.

This analysis does not contemplate the possibility that providers do not actually reach a private agreement; this would require to model a "disagreement point" due to lack of interconnection. As previously discussed, backbones usually decide to set a private peering point when they are "similar in size". Nevertheless, as shown in Besen *et al.* (2001), private peering might be the equilibrium outcome of the bargaining process between Internet providers also when they have different size and capacity. A peering agreement is reached when parties gain equally from interconnection. This may happen also when providers are not symmetric: lack of interconnection means service degradation for both providers since their reciprocal traffic must transit through the congested public peering point. A lower QoS implies less per customer revenues, due to lower individual willingness to pay, and therefore lower profits for both backbones. Under some circumstances, these losses are identical for both parties: while without agreement the smaller provider loses more per customer, the larger provider suffers a smaller loss but over a larger customer base.

Besen *al.* (2001) show that this happens when the individual benefit is linear in the QoS, as it is in our case; in this scenario, private interconnection is an equilibrium outcome also between asymmetric providers.

5 The model with three providers

To distinguish between bilateral and multilateral peering we need at least three providers; things get very complicated, nevertheless we are still able to solve the model at least numerically and to compare the different scenarios. Let us start with multilateral peering.

5.1 Multilateral peering

With three providers, traffic and delay for provider i are given in (5), (6) and (7). The cost of installing the peering infrastructure is equally shared between the three rivals; therefore backbone i's profit is:

$$\pi_i = p_i x_i - \theta \left(k_i + \frac{k_p}{3} \right)$$

where p_i is the demand function given in (2).

Proposition 2. *At the symmetric equilibrium ($\alpha_i = 1/3$) with multilateral peering between 3 backbones, the total amount of network capacity and the capacity of the peering point are respectively*

$$K = \frac{9\,A^2 - 8\,A\sqrt{\theta}(\sqrt{3}+9) + 16\theta(9+2\,\sqrt{3})}{36(A\sqrt{\theta} - 4\,\theta)} \qquad k_p = \frac{3\,A\sqrt{3\theta} - 4\theta(2+3\sqrt{3})}{18\theta} \tag{14}$$

Comparing Proposition 1 with Proposition 2, the following corollary easily follows:

Corollary 1 (Under-investment). *With multilateral peering, total network capacity decreases with the degree of market competition:*

$$K|_{n=2} > K|_{n=3}$$

The effect of enhanced competition on the capacity of the peering point is ambiguous:

$$k_p|_{n=3} > k_p|_{n=2} \qquad if \qquad \theta < \frac{68}{1000}A^2$$

Contrary to Varian and MacKie-Mason (1995), total transmission capacity decreases with competition. This result is not surprising in this set up and it is due to a standard "public good" argument. The negative externality related to congestion induces firms to under-invest in network capacity and this is true although capacities are strategic complements. On the one side, strategic interactions push investments up; on the other side firms tend to behave opportunistically decreasing their investment in capacity as the market becomes more competitive. Corollary 1 shows that this second effect dominates.

The impact of higher competition on the capacity of the common interconnection point is different; in this case, if the cost of capacity is not too big, then three backbones tend to invest in the multilateral peering point more than two. Clearly, under-investment is less severe at the peering given that k_p is chosen in collaboration between providers.

Never the less, the under-investment in K dominates and the overall impact of higher competition on the delay/QoS of transmission is clear; to evaluate this, we have computed the values of the delay suffered by customers under the two scenarios:

$$D|_{n=2} = \frac{1}{2}\frac{3\,A\theta^{3/2}\left(17\sqrt{2}+12\right) - 9\,\theta^{2}\left(9\sqrt{2}+8\right) - 4\,A^{2}\theta\,\left(1+2\sqrt{2}\right)}{3\,A\theta\,\left(1+4\sqrt{2}\right) - 2\,A^{2}\sqrt{2\theta} - 9\,\theta^{3/2}\left(1+2\sqrt{2}\right)}$$

$$D|_{n=3} = \frac{1}{3}\frac{9\,A^{2}\theta\,\left(3\sqrt{3}+2\right) + 16\,\theta^{2}\left(31\sqrt{3}+36\right) - 8\,\theta^{3/2}A\left(29\sqrt{3}+27\right)}{3\,A^{2}\sqrt{3\theta} - 8\,A\theta\,\left(1+3\sqrt{3}\right) + 16\,\theta^{3/2}\left(3\sqrt{3}+2\right)}$$

It is easy to verify that when there are only two providers the quality of service is higher than when there are three active firms: $D_i|_{n=2} < D_i|_{n=3}$.

The most relevant consequence of under-investment is that competition may harm welfare; competition reduces prices and increases output and, therefore, welfare; nevertheless this positive effect may be more than offset by the negative welfare effect of higher congestion due to the lower amount of transmission capacity.

5.2 Bilateral peering

When peering occurs only bilaterally, we are in the framework presented in figure 2 in which each backbone peers with all the others in the market. With three providers, there are three peering points. Traffic and delay are given in expressions (8) and (9). Provider i's profit is:

$$\pi_i = p_i x_i - \theta\left(k_i + \sum_{j\neq i}\frac{k^i_{p,j}}{2}\right)$$

Competition occurs in the second stage, once capacities are in place; this stage is identical to the previous case. The difference here lyes in the first stage when each provider chooses its network capacity, k_i, to maximise its own profits and also sets the capacity of the peering point, $k^i_{p,j}$, together with each peering partner j.

Unfortunately, in this case we are not able to solve the model formally not even in the symmetric case; the maximisation problem is still well behaved but the solution can be obtained only numerically. The following table presents the symmetric solution for $A = 10$ and $\theta = 1$:[26]

	bilateral	multilateral	duopoly
k_i	0.44	0.37	0.99
K	1.32	1.11	1.98
k_p	2.49	1.29	1.58
π_i	2.03	1.99	4.9
W	13.42	11.56	17.67
D_i	3.58	4.15	2.7

Table 2: bilateral *vs* multilateral peering.

For easiness of comparison, table 2 also reports the solution values in the multilateral peering case and in the duopoly case.

Comparing the first two columns, it is apparent that the solution with bilateral peering is better than the multilateral one in all respects: providers install more capacity, this reduces the average delay suffered on their networks and both profits and welfare are higher than with multilateral peering. Finally, comparing the bilateral solution with the duopoly case, it is clear that the negative congestion externality induces under-investment also when providers peer bilaterally, although the overall impact on capacity investment and QoS is less severe than with multilateral peering.

6 Two further extensions

We end the paper presenting two further extensions of the basic duopoly model. In the first one we study the impact of product differentiation on backbones incentives to invest in capacity while in the last section we propose a simple rule to ameliorate the under-investment problem based on a settlement payment between providers.

6.1 Horizontal product differentiation

So far we have assumed intrinsic homogeneity between providers' products and services; eventually, in the asymmetric equilibrium, (quality) differentiation occurs due to different degrees of congestion and QoS. In the reality, providers can differentiate

---

[26] Second order conditions have been checked. Note that in the table, $k_p = k^2_{p,1} + k^3_{p,1} + k^3_{p,2}$.

their products in many respects, for example through the supply of different customer services, or through different offerings in terms of enhanced emails services or fidelity programs.

Consider again the duopoly benchmark; following Bowley (1924), when backbones offer horizontally differentiated services, then we can generalise the inverse demand faced by backbone i as follows:

$$p_i = A - D_i - (x_i + \gamma x_j) \qquad i, j = 1, 2, \qquad i \neq j \tag{15}$$

where $0 \leq \gamma \leq 1$ measures the degree of product differentiation between providers: the lower γ, the more consumers find the products to be different; if $\gamma = 0$, the two products are independent in demand while as γ approaches to 1, the varieties become closer and closer substitutes.

We have solved the model for different degrees of product differentiation; in order to isolate the impact of intrinsic product differentiation from that due to (vertical) quality differentiation and congestion, we have solved for the symmetric equilibrium by assuming identical content shares across the two providers ($\alpha = 1/2$). The following table presents the numerical solutions of the model for A=10 and θ=1.

γ	1	0.9	0.8	0.7	0.5
k_i	0.99	1.06	1.14	1.22	1.43
K	1.98	2.12	2.27	2.44	2.86
k_p	1.4	1.45	1.61	1.73	2.02
π_i	4.9	5.26	5.64	6.08	7.15
W	17.66	19.13	20.8	22.69	27.3
D_i	2.7	2.65	2.6	2.54	2.42

Table 3: the solution with product differentiation

As expected, provider profits increase the more the products are differentiated; furthermore differentiation stimulates capacity investment, both at the peering level and at the level of each single provider. Interestingly, product differentiation induces a decrease in providers' congestion, with a better QoS the more customers perceive products as different.[27] The more products are differentiated, the more the

---

[27] It must be noted that in a model of product differentiation "á la Bowley", the overall size of the market increases with the number of varieties and with the degree of product substitutability. If $\gamma = 0$, varieties are independent in demand and adding an additional variety amounts to adding a completely new market. Therefore, the results on capacity, profits and welfare of a decrease of γ must be considered with caution. On the contrary, the positive effect of differentiation on networks' delay can be fully accepted given that D_i measures the Quality of Service relatively to the total aggregate demand.

two markets can be viewed as separated entities, although providers are still fully interconnected, and therefore the weaker the firms incentives to behave opportunistically.

From this numerical simulation, product differentiation seems to play a role very much similar to that played by the asymmetry in content shares; indeed, a difference in the amount of content/web sites hosted by networks determines, at the equilibrium, different degrees of congestion across firms and as a consequence a differentiation in the quality of access services. Therefore the similarity between the equilibrium impact of these two forms of differentiation is not surprising.

6.2 A simple proposal: a reciprocal investment capacity fee

So far, according to the commonly observed practice between backbones, we have assumed that providers exchange traffic at the peering point for free, in a "bill-and-keep" fashion. The main message resulting from the previous pages is that, in the presence of a negative congestion externality, this form of interconnection induces a sub-optimal level of investment in capacity.

As clearly discussed, this form of interconnection has become very popular among providers for its simplicity and easiness of implementation; nevertheless this simplicity may also result in strong welfare losses and quality degradation. From microeconomic theory, we know that a way to get out of this Pareto inefficiency, can be obtained by allowing side payments between parties.

In this framework, a side payment is simply an interconnection fee payed by each provider for being interconnected to rival's network through the peering point. One possible fee may be the following

$$a(k_i) = \frac{1}{k_i} \qquad i = 1, 2$$

where each backbone pays to the rival a fee for interconnecting at the peering point which is decreasing with its own installed capacity.[28] This fee is simple and it is based on backbones' capacities which, contrary to traffic, can be easily observed, measured and verified by the interconnecting parties.

If such a payment is introduced in the duopoly benchmark, firm i's profit becomes:

$$\pi_i = p_i x_i - \theta \left(k_i + \frac{k_p}{2} \right) - \frac{1}{k_i} + \frac{1}{k_j}$$

The game remains symmetric; this implies that at the equilibrium what a provider pays for interconnection is identical to what it receives from the rival:

---

[28] Note that the term "interconnection fee" may be inappropriate since it is not a fee for handling traffic but simply a payment to enter the peering arrangement.

settlement payments cancel each other out and equilibrium profits are not affected by them. Nevertheless, although backbones do not make or loose money on interconnection, the presence of a fee alters their investment strategies and consequently the equilibrium.

Also in this case to solve the first stage capacity game, we need to resort numerical simulations; the following table compares the solution of the duopoly game with and without the interconnection fee.

	no fee	fee
k_i	0.99	1.5
K	1.98	3
k_p	1.4	1.58
π_i	4.9	5.16
W	17.66	20.3
D_i	2.7	2.2

Table 4: the solution with interconnection fee

From this table clearly emerges that the introduction of a reciprocal payment which is negatively related to the installed capacity may have a positive effect on capacities, profits, welfare and the quality of service. The fee plays the role of a coordinating mechanism that induces backbones to invest more in capacity. In other words, an interconnection fee which is negatively related to providers' capacity moves the solution of the game towards a Pareto superior outcome where both firms and consumers are better off.

7 Conclusion

We have developed a model to study the relationship between interconnection and congestion on the Internet. This is a particularly interesting issue which importance is growing as the demand for bandwidth on the Internet increases. We have extended a recent framework proposed by Dewan *et al.* (2000) in order to analyse firms' incentives to upgrade their capacities and the capacity of peering points under different interconnection structures and when the degree of market competition changes.

The framework takes full account of the impact of negative network externalities due to congestion by explicitly modeling on-net and off-net traffic such that the degrees of congestion across backbones are correlated. Our main message is that under-investment occurs at the equilibrium: congestion induces firms to invest less than the optimal level and this sub-optimality gets stronger as market competition increases; this is true both when firms agree to peer multilaterally and when they peer only bilaterally, although in this case under-investment appears to be less severe.

We also study firms' investment decisions when they offer differentiated products. Product differentiation may occur in two ways: through asymmetric content distribution which affects equilibrium quality of service or, more simply, because firms offer horizontally differentiated products. In both cases, differentiation may stimulate firms' capacity investments thus reducing congestion.

A way to ameliorate congestion due to under-investment, is by allowing side payments between firms which can take the form of reciprocal interconnection charges; we show that when each provider pays the rival for interconnection a fee negatively related to its capacity, then firms invest more in capacity and the model converges towards a Pareto superior outcome.

The model is very simple and it is only a starting point for further developments. A first extension would be to take full account of the characteristics of the bargaining process, considering the impact on the final outcome of uneven bargaining power between providers; this would imply, for example, the possibility that the parties sign a transit rather than a peering agreement. This will be discussed in future research.

Appendix

Proof. of Proposition 1. To solve for the symmetric equilibrium, rewrite the system of first order conditions (10), (11) and (12) imposing $\alpha_1 = \alpha_2 = 1/2$; in this case, expressions (10) and (11) are identical and the system boils down to

$$\frac{d\pi_i}{dk_i} = \frac{9\,\theta\,\left(k_j + 4\,k_j k_p + k_p\right)^2 {k_i}^2 + 18\,k_p k_j \theta\,\left(k_j + 4\,k_j k_p + k_p\right) k_i - {k_j}^2 {k_p}^2 \left(4\,A^2 - 9\,\theta\right)}{-9\left(\left(k_j + 4\,k_j k_p + k_p\right) k_i + k_j k_p\right)^2} \tag{16}$$

$$\frac{d\Pi}{dk_p} = \frac{\left(9k_p\theta\left(1 + 4k_j\right)\left(4k_jk_p + 2k_j + k_p\right) - {k_j}^2\left(8A^2 - 9\theta\right)\right){k_i}^2 + 18k_pk_j\theta\left(k_j + 4k_jk_p + k_p\right)k_i + 9{k_j}^2{k_p}^2\theta}{-9\left(\left(k_i + k_p + 4k_ik_p\right)k_j + k_ik_p\right)^2} \tag{17}$$

where $i, j = 1, 2$. Rearranging these conditions and solving for k_1, k_2 and k_p, it is easy to obtain the following:

$$k_i\left(k_j, k_p\right) = \frac{2\,A\sqrt{\theta} - 3\,\theta}{3\theta\,\left(k_j + 4\,k_j k_p + k_p\right)} k_p k_j \qquad k_p(k_i, k_j) = \frac{2A\,\sqrt{2\theta} - 3\,\theta}{3\theta\left(4\,k_j k_i + k_i + k_j\right)} k_i k_j \tag{18}$$

Solving for k_i and k_p, the solution given in the proposition follows.
The above expressions (18) describe the strategic interactions in networks and peering point capacities. By simple differentiation it is easy to verify that capacities are strategic complements:

$$\frac{dk_i}{dk_j} = \frac{\left(2\,A\sqrt{\theta} - 3\,\theta\right){k_p}^2}{3\theta\,\left(k_j + 4\,k_j k_p + k_p\right)^2} > 0 \qquad \frac{dk_i}{dk_p} = \frac{\left(2\,A\sqrt{\theta} - 3\,\theta\right){k_j}^2}{3\theta\,\left(k_j + 4\,k_j k_p + k_p\right)^2} > 0$$

and

$$\frac{dk_p}{dk_i} = \frac{\left(2A\,\sqrt{2\theta} - 3\,\theta\right){k_j}^2}{3\theta\,\left(4\,k_i k_j + k_j + k_i\right)^2} > 0$$

Finally, to verify that the solution is a maximum, take the second order derivatives of (16) and (17):

$$\frac{d^2\pi_i}{dk_i^2} = -\frac{8}{9}\,\frac{\left(k_j + 4\,k_j k_p + k_p\right){k_p}^2{k_j}^2A^2}{\left(k_i k_p + k_i k_j + k_j k_p + 4\,k_j k_p k_i\right)^3} < 0$$

$$\frac{d^2\Pi}{dk_p^2} = -\frac{16}{9}\,\frac{\left(4\,k_i k_j + k_j + k_i\right){k_i}^2{k_j}^2A^2}{\left(k_i k_p + k_i k_j + k_j k_p + 4\,k_j k_p k_i\right)^3} < 0$$

which are always satisfied for positive values of k_i, k_j and k_p.

□

Proof. of Proposition 2. As before, the solution is obtained by backward induction: we first derive the cournot equilibrium outputs in the second stage and then, plugging the obtained outputs into the profit functions, we solve for capacities.[29] In the symmetric case, formally when $\alpha_i = 1/3$, the optimal level of capacities are the solution to the following first order conditions:

$$\frac{d\pi_i}{dk_i} = \frac{16\theta \left(k_j k_p + 4k_j k_s + k_p k_s + 9k_j k_s k_p\right)^2 {k_i}^2 + 32\theta k_j k_p k_s \left(k_j k_p + 4k_j k_s + k_p k_s + 9k_j k_s k_p\right) k_i - {k_j}^2 {k_p}^2 {k_s}^2 \left(9A^2 - 16\theta\right)}{-16 \left(k_j k_p k_i + 4k_j k_s k_i + k_j k_s k_p + k_i k_p k_s + 9k_j k_p k_i k_s\right)^2} = 0$$

$$\frac{d\Pi}{dk_p} = \frac{4\theta \left(k_j k_i + k_j k_s + k_s k_i + 9k_j k_s k_i\right)^2 {k_p}^2 + 32k_j k_i \theta k_s \left(k_j k_i + k_j k_s + k_s k_i + 9k_j k_s k_i\right) k_p - {k_j}^2 {k_i}^2 {k_s}^2 \left(27A^2 - 64\theta\right)}{-4 \left(k_j k_p k_i + 4k_j k_s k_i + k_j k_s k_p + k_i k_p k_s + 9k_j k_p k_i k_s\right)^2} = 0$$

with $i, j, s = 1, 2, 3$ and where $\Pi = \sum_{i=1}^{3} \pi_i$. Rearranging these conditions and solving for capacities, the proposition follows.
Second order conditions are

$$\frac{d^2\pi_i}{dk_i^2} = -\frac{9}{8} \frac{\left(k_s k_p + 4\,k_s k_j + k_j k_p + 9\,k_j k_p k_s\right) {k_s}^2 {k_p}^2 {k_j}^2 A^2}{\left(4\,k_j k_s k_i + k_j k_p k_s + k_p k_s k_i + k_i k_p k_j + 9\,k_j k_p k_i k_s\right)^3} < 0$$

$$\frac{d^2\Pi}{dk_p^2} = -\frac{27}{2} \frac{\left(k_j k_i + k_s k_i + k_s k_j + 9\,k_j k_s k_i\right) {k_s}^2 {k_j}^2 {k_i}^2 A^2}{\left(4\,k_j k_s k_i + k_j k_p k_s + k_p k_s k_i + k_i k_p k_j + 9\,k_j k_p k_i k_s\right)^3} < 0$$

which clearly are always satisfied.

□

---

[29] For the sake of brevity, we omit to present the second stage Cournot outputs given capacities; these results are available on request.

References

Bailey, J. P. (1995). *Economics and Internet Interconnection Agreements.* in Bailey and McKnight (1997).

Bailey, J. P. and McKnight, L. (1997). *Internet Economics.* MIT Press, Cambridge MA.

Besen, S., Milgrom, P., Mitchell, B., and Srinagesh, P. (2001). Advancing in Routing Technologies and Internet Peering Arrangements. *American Economic Review*, 91:292–296.

Bowley, A. L. (1924). *Mathematical Groundwork of Economics.* Oxford University Press, Oxford.

Carmel, E., Eisenach, J., and Lenard, T. (1999). *Digital Economy Fact Book.* Progress and Freedom Foundation, Washington D.C.

Cocchi, R., Shenker, S., Estrin, D., and Zhang, L. (1993). Pricing Computer Networks: Motivation, Formulation, and Example. *IEEE/ACM Trans. Net.*, 1, vol. 6:614–627.

DaSilva, L. A. (2000). Pricing for QoS-Enabled Networks: A Survey. *IEEE Communications Surveys*, 2nd Quarter:2–8.

Dewan, R., Freimer, M., and Gundepudi, P. (1999). Evolution of Internet Infrustructure in the 21^st Century: The Role of Private Interconnection Agreements. Proceedings of the ICIS.

Dewan, R., Freimer, M., and Gundepudi, P. (2000). Interconnection Agreements Between Competing Internet Service Providers. IEEE - Proceedings of the 33th Hawaii International Conference on System Sciences.

Falkner, M., Devetsikiotis, M., and Lambardis, I. (2000). An Overview of Pricing Concepts for Broadband IP Networks. *IEEE Communications Surveys*, 2nd Quarter.

Gibbens, R., Mason, R., and Steinberg, R. Internet Service Classes Under Competition. *IEEE Journal on Selected Areas in Communications*, 18/12:2490–2498.

Gibbens, R., Mason, R., and Steinberg, R. (1998). Multiproduct Competition Between Congestible Networks. University of Southampton Discussion Paper in Economics and Econometrics, No. 9816of Cambridge.

Gupta, A., Jukic, B., Stahl, D. O., and Whinston, A. B. (1999). Impact of Congestion Pricing on Network Infrastructures Investment. paper presented at ???

Gupta, A., Stahl, D. O., and Whinston, A. B. (1995). *A Priority Pricing Approach to Manage Multi-Service Class Networks in Real-Time*. in Bailey and McKnight (1997).

Gupta, A., Stahl, D. O., and Whinston, A. B. (1996). An Economic Approach to Network Computing with Priority Classes. *Journal of Organizational Computing and Electronic Commerce*, 6:71–95.

Gupta, A., Stahl, D. O., and Whinston, A. B. (1997). A Stochastic Equilibrium Model of Internet Pricing. *Journal of Economic Dynamics and Control*, 21:697–722.

Jiang, H. and Jordan, S. (1996). A Pricing Model for High Speed Networks and Guaranteed QoS. *Proc. IEEE INFOCOM*, 1:888–895.

Kahin, B. and Keller, J. (1995). *Public Access to the Internet*. MIT Press, Cambridge MA.

Kende, M. (2000). The Digital Handshake: Connecting Internet Backbones. FCC - OPP Working Paper No. 32.

Kende, M. and Oxman, J. (1999). The Information Interchange: Interconnection on the Internet. Paper presented at the Workshop on Internet Service Quality Economics, MIT, Cambridge MA.

Laffont, J. J., Marcus, S., Rey, P., and Tirole, J. (2001a). Internet Interconnection and the Off-Net Costing Principle. mimeo, Institute d'Economie Industrielle (IDEI).

Laffont, J. J., Marcus, S., Rey, P., and Tirole, J. (2001b). Internet Peering. *American Economic Review*, 91:287–291.

MacKie-Mason, J. K. and Varian, H. (1995a). Pricing Congestible Resources. *IEEE Journal of Selected Areas in Communications*, 13:1141–1149.

MacKie-Mason, J. K. and Varian, H. (1995b). *Pricing the Internet*. in Kahin and Keller (1995).

Odlyzko, A. (1997). A Modest Proposal for Preventing Internet Congestion. AT&T Labs - Research, mimeo.

Odlyzko, A. (1999). Paris Metro Pricing for the Internet. *Proc. ACM Conference on Electronic Commerce*, pages 140–147.

Varian, H. (2002). The Demand for Bandwidth. Evidence from the INDEX Project. mimeo, University of California at Berkeley.

Weiss, M. B. and Shin, S. (2001). Internet Interconnection Economic Models and its Analysis: Peering and Settlement. School of Information Science - University of Pittsburgh, mimeo.

Wiseman, A. E. (2000). Economic Perspectives on the Internet. Working Paper, Bureau of Economics - Federal Trade Commission.