Information Asymmetry Models in the Internet Connectivity Market

Ioanna D.Constantiou
Costas A. Courcoubetis

May 13, 2001

Athens University of Economics and Business
47a Evelpidon and 33 Lefkados str.
11362 Athens,Greece
Phone: +301 8203663, Fax: +301 8203664
Email: {`ioanna,courcou`}`@aueb.gr`

Abstract

This paper discusses the structure of the Internet connectivity market by focusing on the business relations of stakeholders involved in network services provision. We believe that the role of information asymmetry is critical when considering interconnection agreements, and should be taken into account in the structure of the contract. Information asymmetry due to incomplete information concerning important operating parameters such as network load, capacity, cost, gives rise to adverse selection during negotiation and contract preparation. The current flat structure of interconnection agreements does not address such information asymmetries. In many cases, the difficulty of observing the actual effort allocated by the contracted network for providing quality of service, and in particular, the absence of appropriate incentives in the contract, allows for the possibility of opportunistic behaviour in the form of moral hazard. We formulate two simple analytical models which demonstrate the effects of moral hazard in the market for network transport services. The first deals with the case where the network is contracted for short duration transport where the customer can not use the statistical information obtained during the life time of the service to determine with certainty the actual effort allocated by the network. The second model deals with transit contracts of longer duration, where the actual cost for provisioning the service at various quality levels is only statistically known at the time the contract is set up. Although these models are too simple for capturing the complexity of interconnection agreements between ISPs, they demonstrate the bad effects of information asymmetry and motivate the importance of incentive contracts for improving efficiency.

1 Introduction

The Internet connectivity market is structured hierarchically, comprising three main levels of participants, namely, end users, Internet Service Providers (ISP) and Internet Backbone Providers (IBP). End users are at the bottom of the hierarchy and access the Internet via ISPs. End users include individual and business customers. At the top of the hierarchy, IBPs own high speed and high capacity networks to provide global access and interconnectivity. They sell primarily wholesale Internet connectivity services to the ISPs, (Kende and Oxman 1999). ISPs then resell connectivity services or add value and sell new services to their customers. However, IBPs may also get involved in ISP business activities by selling retail Internet connectivity services to end-users. Both IBPs and ISPs provide complementary inputs to the bundled network services that end-users consume, (Foros and Kind 2000).

In the Internet connectivity value chain two markets are identified, the wholesale and the retail for global access and connectivity to end-users respectively, (Huston 1999a). There are two main types of pricing contracts: pricing between end-user and ISP for primary Internet access and pricing between ISP and IBP for interconnection. In the early days when the Internet was serving exclusively the public sector, mainly for research and education purposes, interconnection was a public good and its provision was organized outside competitive markets. Today interconnection is primarily commercial, yet its basic architectures remain unchanged. Internet connectivity in itself possesses public good properties, the most pervasive being network externalities. Externalities generate powerful incentives for interconnection while setting the stage for potential opportunistic exploitation of shared network resources.

The simple market mechanisms governing Internet connectivity are cracking under current pressures and may not be able to sustain future growth. The problem is more complicated once we realize that even if we devise an efficient and practical mechanism for resource allocation and price discrimination for primary access, actual performance depends on the conditions and behavior of several networks that mediate data transfers throughout the world. In the past, ISPs and IBPs have been agreeing to service each other's traffic without charge, for their obvious mutual benefit. However, competitive market dynamics have tilted the balance in such peering agreements when, for example, one partner makes heavier use of another's resources. Commercial wholesale contracts, on the other hand, cannot always verify or enforce the agreed performance levels. For example, a wholesale network provider may disguise his low effort (e.g. neglect to upgrade bottleneck network components) as adverse system-wide demand conditions. An even more elementary problem is to agree upon what constitutes performance, effort and cost and how that is built into an effective pricing scheme. Participants in wholesale markets for Internet connectivity lack perfect information regarding each other's capacity, demand, resource allocation, effort and cost (Cukier 1998). As a result, they cannot enforce any contracts based on performance and they have an incentive to act opportunistically against each other (to take advantage of the other party's poor information and deviate from agreed performance).

This paper explores the current types of interconnection agreements and outlines future research directions for taking information asymmetry explicitly into account. Such directions are motivated through two simple but precise models capturing the information asymmetry issues that arise when a transport service contract is to be established between two ISPs (or an ISP and an end user). The first deals with the case where the network is contracted for a transport service of short duration, and hence the customer can not use the statistical information obtained during the life time of the service to determine with certainty the actual effort allocated by the network. The second model deals with transit contracts of longer duration, where the actual cost for provisioning the service at various quality levels is only statistically known at the time the contract is set up. In both cases the optimal contract must include the appropriate incentives in order to motivate the contractor to allocate the necessary effort, in contrast to the current practice where such contracts are flat and do not include a quality component.

The paper is structured as follows. Section 2 presents current types of interconnection agreements. Current practices are discussed and their weaknesses are identified. Section 3 explores in detail the nature, manifestations and implications of asymmetric information in interconnection agreements. This section also sets out the requirements for sustainable service quality expectations from such agreements. Section 4 outlines the modeling framework for incentive compatible mechanisms for interconnection, satisfying such requirements. Section 5 presents the two interesting cases of moral hazard in transit agreements. Finally, we conclude with some directions for further research.

2 Interconnection Agreements

The dominant economic driving force for interconnection between network providers is positive network externalities. Externalities result from connectivity, the possibility of every party connected to the Internet to be able to communicate with any other party, and from universal access, the possibility to have access to all network resources independently of the user's physical location. In many cases two network service providers may negotiate special interconnection conditions in order to provide a given service or application posing special performance demands (e.g. streaming media). Furthermore, dense interconnection (i.e. having agreements with many network service providers) facilitates packet routing though shorter and less congested paths that decreases the possibility of packet loss, thus supporting the provision of better quality of service.

Network services providers have been implementing two types of interconnection agreements, namely peering and transit. These agreements involve provision and delivery of network services between them. As a result the exchange of Internet traffic operates within two parallel systems of contracts. Peering agreements involve the exchange of traffic and routing information between the two networks with no interconnection charge. This exchange occurs at public and/or private Internet exchange points, which are points of traffic exchange and provide access to backbone networks.

There are four distinct characteristics of a peering system:

1. Peering partners only exchange traffic that originates from the customer of one partner and terminates to the customer of the other partner, on a bilateral basis. A partner will not act as an intermediary and accept the traffic of one peering partner and transit this traffic to another peering partner.
2. Peering partners exchange traffic on a settlement-free basis also known as sender-keeps-all. The only costs involved in peering are the purchase of equipment and the provision of transmission capacity needed for each partner to meet the requirements deriving from peering.
3. Peering partners exchange their customers' traffic at the geographically nearest possible exchange point.
4. The recipient of traffic on a peering agreement provides best effort services when terminating traffic to its customer.

Transit agreements are the alternative to peering. There are two main differences between peering and transit. First, one partner pays another partner for interconnection and therefore becomes a customer. The partner selling transit services will route traffic from the transit customer to its own peering partners as well as to other customers (on Telecommunications and Policies 1998). Second, transit does not involve the same service as peering and, therefore, refusing peering in favour of transit is not a means of charging for a service that was otherwise provided free of charge. When regional ISPs pay for transit they benefit from the infrastructure investments of national or global backbones without themselves having to make the same investments. Transit gives an ISP customer access to the entire Internet, not just the customers of the peering partner, thus the transit provider must either maintain peering arrangements with a number of other backbones or must pay for transit from another backbone.

During negotiations for peering several factors will be considered from both parties, such as the prospective peer's customer base (e.g. number and type of customers), as well as the reach and size of their network. Peering between ISPs of equal size and shape is relatively straightforward because they recognise the mutual benefits. However negotiations for peering do not just occur horizontally but also vertically between local ISPs and national ISPs. The national ISPs are in a stronger bargaining position because they not only provide access to their own customer base, but sometimes also act as a gateway to the rest of the Internet. Some have customers providing content and services of great demand to the customers of other (possibly smaller) ISPs. At the same time, some ISPs have a customer base that the customers of other ISPs want to access. Peering in such an environment is best expressed as the balance of perceptions, in which each party perceives an acceptable approximation of equal benefit in the interconnection relationship. Uneven parties may fail to negotiate peering successfully or a peering agreement may not be sustainable.

Interconnection partners face conflicting incentives. On one hand, they have an incentive to cooperate with one another in order to provide their customers with access to the full range of Internet users and content. On the other hand, they have an incentive to compete with one another for both retail and wholesale customers. The strategies and growth of individual networks vary significantly and this has led to some breakdowns in the peering system (Frieden 1998). Peering is based on the implicit agreement that either party can terminate the interconnection relationship and that the other party would not consider it a competitively hostile act. If one party has a high reliance on the interconnection arrangement and the other does not, then the most stable business outcome is that this reliance is expressed in terms of a service contract as a customer-supplier relationship (Frieden 2000). The exponential growth of the Internet has put enormous pressure on the backbones and their interconnection points connecting the backbones (Cremer, Rey and Tirole 1999). As a result, performance is hampered at these points and peering often turns out to be inferior in terms of service quality.

In many cases a hybrid approach to interconnection has been adopted, peering with a number of ISPs and paying for transit from one or more backbones in order to have full access (Kende and Oxman 1999). Interconnection agreements are also influenced by the dynamic nature of the Internet, which often leads to a form of arbitrage that is played behind the scenes by the different ISPs negotiating new interconnection agreements. For example an IBP that provides connectivity to smaller ISPs must also interconnect with other IBPs and act similarly to foreign exchange arbiters, as he seeks to extract revenue in both directions. The resultant business environment is one characterised by a degree of fluidity. Many network providers operate both as a client and as a provider (Huston 1999b).

3 Information Asymmetry in Internet Connectivity Markets

The bilateral transactions in the Internet connectivity market are characterised by severe information asymmetries (Macho-Stadler and Castrillo 1997). Network service providers control all the information pertaining to the characteristics of their networks (e.g. capacity, usage etc.) and may or may not disclose it to potential interconnection partners. From an economic perspective, such information is critical for the structure and efficiency of interconnection agreements. Current practices are often based on the subjective perceptions of the parties involved and may not be optimal or sustainable because of asymmetrically available information.

Asymmetric information in current types of interconnection agreements gives rise to opportunistic behaviour in different guises. The first is called backbone free riding. A national ISP has to build and maintain a nation-wide network, connecting different regions, whereas a local ISP, concentrating on a single region does not. If both ISPs agree to interconnect, the local ISP may use national ISP capacity to service traffic between customers in distant regions. For example, when a customer of the regional ISP requests a web page from a customer of the national backbone whose server is far away, the request will be carried through the national ISP, from one region to the other and the response back. The national ISP may thus refuse to peer on the grounds that it is bearing the expense for a national infrastructure that the regional ISP can exploit at no cost. As a result, a number of ISPs include in their publicly stated peering policies that potential peer partners should be willing and able to peer at a number of geographically dispersed locations (Ergas 2000).

The second manifestation of opportunism is called business stealing effect. Interconnection naturally lowers end user switching costs. End customers may switch network providers seeking better price/performance ratios without losing connectivity or access to shared network resources. Lower switching costs increase competition and, as a result, weaken ISP incentives to interconnect (Shapiro and Varian 1998). An alternative strategy is to raise switching cost by differentiation. An ISP may bundle exclusive services or content to its main Internet access offering in order to achieve customer lock-in (e.g. AOL).

Another example of perceived free riding that may arise in a peering relationship derives from the business strategy of an ISP. One ISP may choose for a variety of reasons to focus on providing service to users that generate high traffic volumes and use extensively the web servers of the peer ISP. In such cases the second ISP will carry extra traffic volume that will negatively affect its network performance, and decrease the quality of services provided to its own customers. If usage patterns are not reciprocal, peering is not sustainable.

Opportunistic behaviour may also arise in transit agreements. When an ISP signs a transit agreement he is expecting to have global access to the Internet. It is, however, difficult to know the network coverage of his provider and the performance levels of its network. As ISPs are trying to increase their revenues through higher utilisation of their network, they often overbook it. This behaviour in combination with best effort service provision may end up to increased delays and packet losses for client traffic. Thus, ISPs entering transit agreements do not always receive their expected benefits.

When ISP A is not able to identify the type of ISP B, with respect to certain characteristics that will affect the outcome of an interconnection agreement, there is an adverse selection problem. The result might be that desirable interconnections may not be agreed or that agreements may be settled under unfair or inefficient conditions. The main information components that may be asymmetrically available to candidate interconnection partners include, among others, the following:

- The types of customers. Customer demand is notoriously unsystematic and difficult to predict. However, an ISP can obtain demographic and usage characteristics (as indicators of demand patterns) of its client base. Such information is not available to third parties.
- The net loading of traffic carried by the network. This information is directly related to customer demand, which is not predictable.
- The existing interconnection agreements. Such information concerns the business strategy of the ISP and its core competence. An ISP has no incentive to reveal this type of information that will directly reveal the cost of managing its network
- The available capacity and the resource allocation policies. This information includes decisions on statistical multiplexing, overbooking, attracting new customers. Resource allocation has strong implications for network performance.

In the stage of negotiations for peering or transit agreements such information is critical. However, it is not readily available and ISPs have little incentive to reveal it or report it truthfully. Current market practices address this problem only in part. Large ISPs exert their bargaining power to extract such information from smaller potential partners. The requirements and terms of such agreements are privately communicated and undisclosed.

Information asymmetries are also manifest in the form of moral hazard after an interconnection agreement is signed. When ISP A is not able to fully observe or monitor the effort of ISP B after an interconnection agreement, ISP B may alter its effort opportunistically for its own private benefit and to the detriment of ISP A (or vice versa). Moral hazard arises as a result of actions such as the following:

- An ISP may not keep upgrading his network capacity after an interconnection agreement. This will result in poorer servicing of the partner's traffic. As peering agreements currently are based on best effort services, such behaviour cannot be verified.
- An ISP may actively discriminate against IP packets that enter into his network from the interconnected partner when its network has large amount of local traffic.

- An ISP may overbook its network in order to maximise economies of scale. To avoid congestion the ISP may delay or not admit interconnected traffic. This not the predictable outcome under 'naturally' arising congestion but the result of intentional unilateral overbooking.

Moral hazard appears because one ISP's profit maximisation strategy may not be aligned with the interests of its interconnection partners and because he can hide or disguise his effort. The result is inefficient and unstable agreements. Incentive compatible contracts can be devised so as to safeguard interconnection agreements from opportunism and sustain the undeniable benefits of network externalities.

4 Incentive Contract Issues

Current interconnection agreements do not always provide sufficient incentives for partners to collaborate on exploiting positive network externalities. New Internet applications appear to be increasingly demanding in terms of specific network performance guarantees. We argue that new types of interconnection agreements based on contracts with incentive mechanisms, will mitigate the adverse implications of asymmetric information and will provide a sound basis for sustaining quality of service requirements.

There are many issues open to future research on interconnection agreements. Our ultimate goal is more specifically

- To investigate problems of asymmetric information in existing interconnection agreements and pricing schemes

- To obtain optimal pricing schemes that will be consistent with the Internet practice (i.e. based on traffic measurements).

in order to model the asymmetric information problems in the Internet connectivity market, certain basic parameters have to be defined. These are the effort, the outcome and the cost of providing such effort.

The *effort* of a network service provider (e.g., an IBP) is defined as his decision on how to treat client (e.g., ISP) traffic and is described in terms of service classes and/or multiplexing strategies applied. When multiplexing traffic from different sources and applications, the network manager can assign different priorities to different kinds of packets according to subjective criteria. Such criteria may include, among others, the type of application being serviced (e.g. email vs. videoconferencing), the identity of the sender (or recipient), or the revenue generated by the traffic transferred.

The inability to verify such a level of effort can be alleviated by devising pricing mechanisms that provide suitable incentives to the IBP to exert the appropriate effort as to ensure the required performance. In effect, such mechanisms make the IBP responsible for the effort he exerts by tying his profit to the outcome after accounting for uncertain conditions. Performance indicators such as average delay or packet loss may measure the observable *outcome* in an interconnection agreement.

The choice by the IBP of a particular priority class or of the scheduling algorithms for serving the traffic resulting from a transport contract with an ISP, has a *cost*. An candidate definition of such a cost is the opportunity cost for not serving (or reducing the quality of service for) other client ISPs of the same network. An alternative but equivalent definition of this cost is in terms of negative externality (congestion) imposed on the network and its other users. It is quite difficult to estimate this cost as it depends on parameters that an IBP may not reveal. In many cases, a key parameter that affects this cost is the traffic load that the network is contracted to cary due to its own customers (large end-user organizations and ISPs). This "local traffic" information may be available to the network provider before deciding how to treat transit traffic from other ISPs he is peering with. In this setting, the cost of high effort (extra delay or packet losses experienced by the local traffic due to the traffic of the specific contract when this is serviced with high priority instead of best-effort) is small under some threshold local traffic level and increases fast above that threshold. In turn, this threshold may depend on the total available capacity, on the multiplexing algorithms, and on the burstiness of the traffic. In principle, the more effective bandwidth is allocated to the specific contract, the less such bandwidth is available for the rest of the traffic, resulting in some opportunity or congestion cost.

In order for an incentive contract to be successful, one has to be able to quantify reasonably well the expected cost of the required effort and the value of the resulting quality. These issues are made precise through the examples presented in the following section. We describe two cases of moral hazard where there is information asymmetry at the time the contract is established. In both cases, unless provided with the appropriate incentives, a rational service provider will provide under all circumstances the minimum possible effort.

5 Moral hazard in interconnection agreements

We describe two cases of moral hazard where there is information asymmetry at the time the contract is established. In both cases, unless provided with the appropriate incentives, a rational service provider will provide under all circumstances the minimum possible effort. We first focus on the classical problem of moral hazard with unobservable effort in a real-time transport service model. Then we present a case of observable effort but uncertain network conditions, which is a reasonable model of a transit service between ISPs. For simplicity, we focus on the modelling issues and the resulting optimal incentive schemes, omitting the complete analysis.

5.1 A real-time service model with unobservable effort

Consider the case of a network offering end-to-end services with some quality of service guarantee, such as a lower bound on the maximum delay a packet may encounter during its sojourn in the network. In many practical cases such as the ATM technology or the diffserv architecture of the Internet, such a guarantee is of a statistical nature, namely it specifies the probability for the delay of a packet to exceed a certain level. It is not surprising that, in such a case, it is hard for a customer to prove that the network did not keep its side of the contract when excessive delays were observed to an unreasonably high percentage of packets on single short lived connection. The reason is that statistical guarantees can not preclude “bad luck”.

It is common to expect that the network uses different service classes for the packets that transit its links. For, instance, there may be a service class offering a small probability for the queuing delay of a packet to exceed a given level d, and an other class that offers a substantially higher such probability. Suppose now that a connection of fixed duration T for which low packet delays are valuable makes a contract with the network for the high quality service. The network incurs a cost for supporting such a contract, which may be the added delay of the packets of its internal traffic due to the packets of new connection. If this delay cost is higher for the higher quality service, then we have a conflict of interest: since the customer can not prove which internal service class was actually used by the network, the later will be tempted to use the least cost service even if in the contract the customer paid for the higher quality one. This is a typical case of a moral hazard problem that occurs due to the asymmetric information between the network and the customer.

Note that if the customer could observe the service class in some provable way, then he could insist on a clause in the contract which would severely penalize the network in the case of cheating.

What can the customer do in this case? the only possibility is to offer incentives to the network instead of a flat fee. Such incentives may be through a mechanism which specifies different payment amounts that will take place after the service is completed, and which depend on some observable results which can not be refuted by the provider or the customer. Hence the provider incurs a risk of a low payment which may be a good enough reason to offer the service desired by the customer. We will illustrate the above concepts through a concrete example.

The provider's network consists of a single link that implements two FCFS packet queues, the high-quality queue Q_H, and the low-quality queue Q_L. For simplicity of the analysis, suppose that both queues are M/M/1, served with the same service rate μ, and that the arrival rates of packets due to internal traffic are λ_H and λ_L respectively, where $\lambda_H < \lambda_L$. If $\rho_i = \lambda_i/\mu$ denotes the utilization of Q_i, then the probability of an arriving packet to find more than K packets in the queue is $q_i = (1-\rho_i)\rho_i^K$. Note that $q_H < q_L$. Observe that this is equivalent with an arriving packet to be delayed by more than $K \times 1/\mu$, since $1/\mu$ is the average service time of a packet.

The connection needs to transport N packets in total time T. These quantities can be part of the contract and can be verified by the network, i.e., the customer can not cheat. The reason for customer demanding such a data connection originates from some higher level application which needs some performance guarantees in order to function properly. Typical example is a video conference or an IP telephony call. For simplicity assume that such an application can distinguish between two performance levels: performance is bad if a percentage of packets larger than π is delayed excessively, where excessive delay is say K average packet service times.

Lets first define the cost of the network for supplying the service to the above contract. If queue i is selected, then the extra cost for serving the N packets in time T is due to the increased delay encountered by internal traffic packets, and is equal to

$$v_i = \gamma_i T \lambda_i \frac{\partial D_i(\lambda_i)}{\partial \lambda_i} \frac{N}{T} = \gamma_i \lambda_i N \frac{\partial D_i(\lambda_i)}{\partial \lambda_i},\ i = L, H\,, \tag{1}$$

where $D_i = 1/(\mu - \lambda_i)$ is the average delay per packet in queue i, γ_i in \$/s is the monetary equivalent of delay seconds for the packet streams using queue i, and the rate increase N/T

is small compared to λ_i. Note that although $\partial D_i(\lambda_i)/\partial\lambda_i$ is higher in Q_L due the higher utilization, substantially higher values of γ_H than γ_L may well justify that $v_H > v_L$, i.e., allocating high effort is more costly than low effort. This may occur in practice if the flows in Q_H are due to other customers which have contracts that pay the network in proportion to the average delay experienced during different time periods.

We have already mentioned that there are two possible outcomes regarding the performance of the service. Outcome 1 (bad) corresponds to the event in which more than πN packets have been delayed by more than $K\mu^{-1}$ seconds, and outcome 2 (good) corresponds to the complementary event. Assume that the value of outcome j to the customer is r_j where $r_H > r_L$. If p_j^i denotes the probability of outcome j given an effort level i, then by Sanov's Theorem, we can approximate p_j^i by the expression

$$p_1^i = e^{-NI(\pi,q_i)} \quad \text{when} \quad \pi > q_i \tag{2}$$

$$p_2^i = e^{-NI(\pi,q_i)} \quad \text{when} \quad \pi < q_i\,. \tag{3}$$

where $I(\pi, q_i) = \pi \log \pi/q_i + (1-\pi)\log(1-\pi)/(1-q_i)$. Observe that $p_j^i > 0$, hence the customer can not be certain about the effort deployed by the network after observing the outcome.

Asymmetric information contract theory suggest the use of an incentive payment mechanism, where the customer pays a different amount w_j depending on the outcome j. Such an incentive mechanism may turn useful to the customer for maximizing his net profit for such a contract.

Let $B(r-w)$ denote the profit function of the customer in the case of a service of value r and payment w, which is assumed to be concave. We will construct the optimal payment w_1, w_2 which induces the network to use the high quality service. Note that this may not be always the desired strategy for the customer. If r_1 is close to r_2, there may be no justification to pay the extra amount for higher quality. In such a case, the contract will include a flat fee enough to pay for the extra cost v_L of the low quality service.

In the general case, using the Principal-Agent model formulation where the principal is the customer and the agent is the network, the mathematical programming problem becomes

$$\begin{aligned} \max_{\{w_j\}} \quad & \sum_j p_j^H B(r_j - w_j) && (4) \\ \text{s.t.} \quad & \sum_j p_j^H w_j - v_H \geq 0 && (5) \\ & \sum_j p_j^H w_j - v_H \geq \sum_j p_j^L w_j - v_L \,. && (6) \end{aligned}$$

Note that the first inequality is the participation condition, whereas the second inequality guarantees the incentive compatibility (choosing high effort).

The Lagrangian of this problem is

$$L = \sum_j p_j^H B(r_j - w_j) + \theta_1[\sum_j p_j^H w_j - v_H] + \theta_2[\sum_j (p_j^H - p_j^L) w_j - v_H + v_L], \qquad (7)$$

which gives the first order conditions

$$B'(r_j - w_j) = \theta_1 + \theta_2[1 - \frac{p_j^L}{p_j^H}] \,. \qquad (8)$$

Note that since (5) must be tight and increasing the right hand side reduces the optimal value of the problem, we must have $\theta_1 > 0$. Also, $\theta_2 = 0$ implies that $B'(r_1 - w_1) = B'(r_2 - w_2)$, and hence $r_1 - w_1 = r_2 - w_2 = d$. The latter is equivalent with the complete insurance policy for the customer (in all cases keeps d for himself, transfers the rest of his value back to the network). If this is true, we need (from the incentive compatibility condition) that

$$\sum_j p_j^H r_H - \sum_j p_j^L r_L \geq v_H - v_L \,. \qquad (9)$$

Hence there are two cases to consider. If (9) holds, then the customer can deploy the optimal strategy of the complete information case (the best he could ever hope for), which is full insurance. If (9) does not hold, then $\theta_2 > 0$ (due to positive shadow cost of incentive compatibility), and (8) implies that when the signal of high effort is strong, the customer rewards the network by an even larger payment keeping a smaller amount of r_j for himself.

Note that one could analytically solve the above problem using the explicite expressions in (1), (2) and (3).

5.2 A model with observable effort but uncertain network conditions

In this section we provide a simple example of moral hazard that may appear in interconnection agreements. In particular, we focus on a transit agreement between two network service providers. We again use the formulation of the Principal-Agent model. The transit customer is the principal **P** who contracts the agent **A** (the transit service provider) for the transport of a packet flow of rate x through the agent's network.

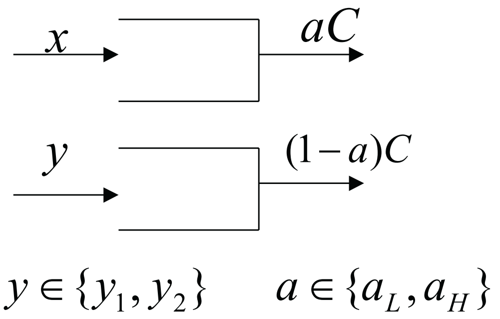

Figure 1: The network consists of a link serving two queues, one for the traffic resulting from the transit agreement, and one for the internal traffic. The level of effort offered to the transit contract corresponds to the fraction of the capacity dedicated to the first queue. The rate of internal traffic at the time the contract will be instantiated is random taking the values y_1, y_2 with probabilities p_1, p_2.

The network of the provider consists of two queues, see Figure 1, one dedicated to the traffic of P, and the other dedicated to the traffic of the rest of the customers of A (the internal traffic). The capacity of the network is C, and a fraction α of the above capacity, where $\alpha \in \{\alpha_L, \alpha_H\}$, $\alpha_L < \alpha_H$, must be allocated to the flow x of the principal. Here α corresponds to the effort that is provided by the agent in the context of the contract with the principal. The average rate of the flow of the internal traffic of A is denoted by y and takes the values y_1, y_2, where $y_1 < y_2$, with probabilities p_1 and $p_2 = 1 - p_1$ respectively. In our model, the network has no control on the value of the rate of its internal traffic at the time of the activation of the contract. It can only control the fraction of its capacity that it will allocate. On the other hand, the distribution of the value of the above internal rate is known at the time of the establishment of the contract. This is common in many practical situations where the contract defines a service to be provided at some later point in time, for which only statistical information is available on the state of the network.

There is an associated cost for allocating capacity to the flow of the principal in terms of the extra delay that this implies to the packets of its internal flow y. A simple way to calculate this cost is by using the simple M/M/1 formula where the average delay per packet for a flow of rate y packets/s served by a server of rate C packets/s is equal to $1/(C-y)$. This implies a rate of delay cost (in money per seconds) equal to $\gamma y/(C-y)$, where γ denotes the monetary value of the cost of one second of delay. In this framework, the cost of allocating an amount of effort α to the contract with P is

$$v(y,\alpha) = \gamma y[\frac{1}{(1-\alpha)C-y} - \frac{1}{C-y}]\,. \tag{10}$$

Let $v(ij)$, $i \in \{1,2\}$, $j \in \{L,H\}$, denote $v(y_i,\alpha_j)$. It can be proved that

$$v(\mathit{2H}) - v(\mathit{2L}) > v(\mathit{1H}) - v(\mathit{1L}) > \mathit{0} \tag{11}$$

In other words, shifting from low effort to high is more costly when the system has a higher load. On the other hand, allocating a higher effort has a positive effect for P since this reduces the average delay of his packets. Let r_L, r_H denote that monetary value of the service received by P when the effort levels are low and high respectively.

Our task is to design an incentive contract where P pays A an amount $w(\alpha)$ determined after the completion of the service, where the payment depends on the level of effort allocated by A. We assume that P can estimate in an un-contestable manner the level of effort provided by A as follows: he measures the average delay of his traffic and then uses the M/M/1 delay formula to compute the amount of capacity offered by the network.

Let w_L, w_H denote the payments when the effort was low and high respectively. These are included as part of the contract. Once these are known, the network operator first needs to decide whether to accept or to reject the contract. Such a decision is based on the statistical information about the future network state (the rate of its internal traffic), when the service will be instantiated. At that later time, he determines the actual network state and dynamically decides on the level of effort to be provided. This decision is rational and is based on the available information so far. The goal is to maximize net benefit. Note that simply computing the net benefit that will result from each of the (two) possible actions can do this.

The problem that the principal needs to solve is $\max_{\{w_L,w_H\}} \mathbb{E}B(r-w)$, where B is the profit function of the principal (assumed concave), and r, w are the random variables that define the value obtained by the principal and the value of the payment to the agent.

These are well defined for each fixed pair of w_L, w_H. (If the agent rejects the contract, then the principal gets nothing).

In order to solve this problem, we need to better understand the strategy of the agent. Such a strategy is easy to define for a known pair w_L, w_H. First, observe that if the value of the state is i, the rational action is $j = \arg\max_l\{w_l - v(il)\}$, and the payoff is $w_j - v(ij)$. Equivalently, the sign of $w_L - w_H - [v(iL) - v(iH)]$ determines the most profitable action for the agent. The participation condition can be written as

$$p_1 \max\{w_L - v(1L), w_H - v(1H)\} + p_2 \max\{w_L - v(2L), w_H - v(2H)\} \geq 0\,. \qquad (12)$$

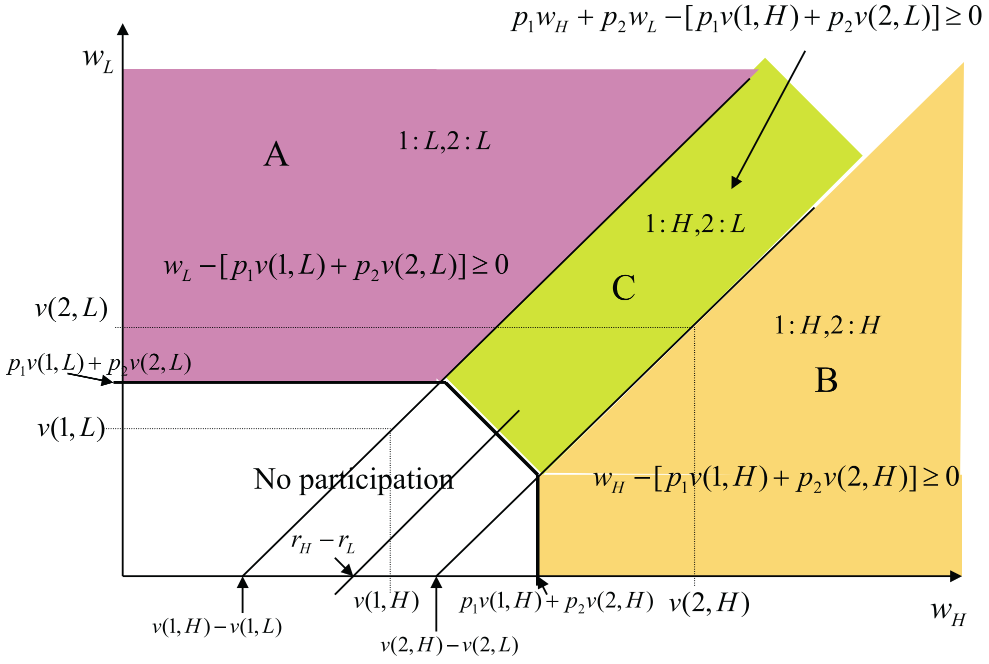

Figure 2: We show the optimal decisions of the agent (network) for all points on the w_L, w_H plane. The notation $k : f$ refers to the decision of offering effort level f when the load state is k.

In Figure 2 we show the optimal decisions of A for all points on the w_L, w_H plane. A participates only in regions A, B, C. In A the decision is always for low effort, in B always for high, and in C it depends on the network state. When the state is 1 (i.e., y_1), it decides for high effort, and when the state is 2, it decides for low.

Now, the principal must choose the optimal pair w_L, w_H , which maximizes his net profit. We can analyze the properties of each region separately. In region A, the maximum profit is $B_A = B(r_L - p_1 v(1L) - p_2 v(2L))$, and is achieved for the minimum payment $w_L = p_1 v(1L) + p_2 v(2L)$ and small w_H. In region B the maximum profit is $B_B = B(r_H - p_1 v(1H) - p_2 v(2H))$, and is achieved for the minimum payment $w_H = p_1 v(1H) + p_2 v(2H)$ and small w_L. In region C the optimum B_C is achieved on the boundary line $p_1 v(1H) + p_2 v(2L)$ since from any interior point we can decrease both w_L, w_H while keeping the agent participating.

One can show that if the overall optimum is achieved in B_C, then it must hold that

$$v(1H) - v(1L) < r_H - r_L < v(2H) - v(2L)\,. \tag{13}$$

Furthermore, $r_H - w_H = r_L - w_L = d$, which implies that the optimal policy is for the principal to be fully insured. He keeps d for himself and offers the value he obtains reduced by d back to the agent.

To prove this note that $B_C > B_A, B_B$ implies for some w_L, w_H on the boundary that

$$\begin{aligned} p_1 B(r_H - w_H) + p_2 B(r_L - w_L) \quad &> \\ B(r_L - p_1 v(1L) - p_2 v(2L)) \quad &, \quad B(r_H - p_1 v(1H) - p_2 v(2H))\,. \end{aligned} \tag{14}$$

By concavity, $p_1 B(r_H - w_H) + p_2 B(r_L - w_L) < B(p_1(r_H - w_H) + p_2(r_L - w_L))$, and since $B(p_1(r_H - w_H) + p_2(r_L - w_L)) = B(p_1 r_H + p_2 r_L - p_1 v(1H) - p_2 v(2L)$, we obtain that

$$p_1 B(r_H - w_H) + p_2 B(r_L - w_L) < B(p_1 r_H + p_2 r_L - p_1 v(1H) - p_2 v(2L))\,. \tag{15}$$

Now, combining (15) with (14), we obtain

$$\begin{aligned} p_1 r_H + p_2 r_L - p_1 v(1H) + p_2 v(2L) \quad &> \\ r_L - p_1 v(1L) - p_2 v(2L) \quad &, \quad r_H - p_1 v(1H) - p_2 v(2H)\,. \end{aligned}$$

From this, simple algebra proves that (13) holds. Now, since the optimum of

$$\begin{aligned} \max_{\{w_L, w_H\}} \quad & p_1 B(r_H - w_H) + p_2 B(r_L - w_L) \\ \text{s.t.} \quad & p_1 w_H + p_2 w_L = p_1 v(1H) + p_2 v(2L) \end{aligned}$$

is achieved for $r_H - w_H = r_L - w_L$, this point corresponds to the intersection of the above line with the boundary line of region C.

The reverse also holds. If (13), then $B_C > B_A, B_B$. This holds since in this case the maximum of B_C is achieved at the point

$$\begin{aligned} w_L &= p_1 v(1H) + p_2 v(2L) - p_1(r_H - r_L)\,, \\ w_H &= p_1 v(1H) + p_2 v(2L) - p_2(r_H - r_L)\,. \end{aligned}$$

where $r_H - w_H = r_L - w_L$, and hence

$$\begin{aligned} B_C &= B(r_L - [p_1 v(1H) + p_2 v(2L) - p_1(r_H - r_L)]) \\ &= B(r_H - [p_1 v(1H) + p_2 v(2L) - p_2(r_H - r_L)])\,. \end{aligned}$$

Substituting (13) and since B is increasing, simple algebra implies $B_C > B_A, B_B$.

The above analysis provides for a methodology to compute the optimal values B_A, B_B, B_C, and hence determine the optimal contract.

In summarizing the above analysis, the interesting case where an incentive contract must be designed corresponds to the case where (13) holds. In this case, if the internal load is low, appropriate incentives will motivate the network provider to offer a high quality service to the traffic of the contract. As already discussed, the form of the optimal policy is for the customer to promise to give back to the network all the value obtained by the service (independently on the amount of effort provided by the network), reduced by some fixed amount d determined by the previous analysis.

6 Conclusions

This paper presents a brief overview of interconnection agreements and the effects of information asymmetry. We believe that the role of information is critical when considering the Internet connectivity market. Asymmetric information in current types of interconnection agreements gives rise to opportunistic behaviour with negative implications. Interconnection agreements exhibit great instability because of the difficulties in enforcing them, and the effects of adverse selection. Both peering and transit agreements do not provide sufficient incentives for partners to collaborate on exploiting network externalities. Our research approach is aimed at addressing adverse selection and moral hazard in interconnection agreements directly, with a view to devising incentive mechanisms suitable for stability and quality of service.

We our currently extending the work presented in this paper to include network architectures such as diffserv, and transport services such as Frame Relay and ABR. The above services have in common the fact that quality of service is rather implicit than explicit, and there are important aspects of information asymmetry regarding the parties that are involved in the contract. We believe that the theory of asymmetric information could play an important role in improving network contracts in terms of the value the resulting services. Clearly, in order for such contracts to be realistic, we need a better understanding of the cost models of network service providers and of the factors that influence network management decisions.

References

Cremer, J., Rey, P. and Tirole, J. (1999) Connectivity in the Commercial Internet. *Institut dEconomie Industrielle (IDEI)* http://www.idei.asso.fr/English/EPresent/index.html (Access on 14/11/00).

Cukier, K. (1998) Peering and Fearing: ISP Interconnection and Regulatory Issues. *Harvard Information Infrastructure Project* http://www.ksg.harvard.edu/iip/iicompol/Papers/Cukier.html (Access on 14/11/00).

Ergas, H. (2000) Internet Peering: A Case Study of the ACCCs Use of its Powers Under Part XIB of the Trade Practices Act, 1974. *In Proceedings of the 3rd Berlin Internet Economics Workshop* .

Foros, O. and Kind, H. (2000) National and Global Regulation on the Market of Internet Connectivity. *In Proceedings of the 3rd Berlin Internet Economics Workshop* .

Frieden, R. (1998) Without Public Peer: The Potential Regulatory and Universal Service Consequences of Internet Balkanization. *Virginia Journal of Law and Technology* **3**, Fall.

Frieden, R. (2000) When Internet Peers Become Customers: The Consequences of Settlement-based Interconnection. *In Proceedings of the 3rd Berlin Internet Economics Workshop* .

Huston, G. (1999a) Interconnection, Peering and Settlements. *In Proceedings of the INET99 Internet Society Conference* .

Huston, G. (1999b) Interconnection, Peering and Settlements-Part II. *The Internet Protocol Journal* **2**, 2, 2–23.

Kende, M. and Oxman, J. (1999) The Information Interchange: Interconnection on the Internet. *In Proceedings of the Common Telecommunications Research Policy*

Conference .

Macho-Stadler, I. and Castrillo, D. P. (1997) *An Introduction to the Economics of Information: Incentives and Contracts.* Oxford University Press, Oxford, first edition.

on Telecommunications, W. P. and Policies, I. S. (1998) Internet Traffic Exchange: Developments and Policy. *OECD* .

Shapiro, C. and Varian, H. (1998) *Information Rules: A strategic guide to the network economy.* Harvard Business School Press.