

Interconnection Economics of All-IP Networks

KIHO YOON *

Department of Economics, Korea University

Abstract

This paper is an exploratory study on the interconnection economics of All-IP networks. To understand the economics of the yet-to-be-realized All-IP networks, this paper begins with discussing three important features of current IP networks: packet switching, receiver benefit, and two-sidedness. It then develops simple models of interconnection for basic fixed fee settlements and for session-based charges. The analysis shows that interconnection settlements depend on network size, market competitiveness, direct and cross network externalities, and costs. Some policy issues are also briefly discussed.

1 Introduction

It is generally expected that the current telecommunication infrastructure will evolve into NGN (next generation networks) over the next 10 years or so. Although there is no universal definition of NGN at this early stage of development, we may use the one in OECD (2004) that NGN are All-IP or packet-based integrated networks. In an NGN environment, applications and services will be separated from the transport network and all kinds of applications and services such as voice, data and video can be organized into packets and delivered on an integrated IP network. NGN are multiple service networks based on QoS-enabled transport technologies, and allow unrestricted access as well as generalized mobility.

The Internet Protocol (IP) has become the dominant design of networking after the breakthrough of HTML-based web and browser-based interfaces. Therefore, the transition towards All-IP networks is a natural evolutionary process. There are two main drivers. The first is cost savings: Legacy networks need replacements, and All-IP networks offer significant savings in both capital and operational expenses when compared to ATM (asynchronous transfer mode) networks. The second is value creation: Since the current best-effort IP networks do not provide significant value to customers nor sufficient revenues to operators, they may not be sustainable in the long-run.

* Contact author: Department of Economics, Korea University, Anam-dong, Sungbuk-gu, Seoul, Korea 136-701. Phone/Fax: +82-2-3290-2222/ +82-2-928-4948, E-mail kiho@korea.ac.kr The previous version of this paper was presented at the "All-IP Network Interconnection Forum", Seoul, Korea held on 18-19 October 2005. I thank the sponsors of this conference for support. Comments by Byoung Heon Jun, Nae-chan Lee, and an anonymous referee are acknowledged and appreciated.

All-IP networks will, if deployed as planned, accommodate traditional Internet services as well as communication services such as VoIP, MMS, and media services. In an All-IP network environment, head-to-head competition between all networks is expected. In addition, interconnection among different networks, possibly with differentiated services and customer bases, is required. Therefore, the issue of network interconnection is one of the subjects that should be addressed before All-IP networks are successfully promoted.

To understand the interconnection economics of upcoming All-IP networks, it is essential to detail the characteristics of current IP networks. This is so since the current technological, economic, and regulatory structures will persist for a good time and also affect the evolution path to the All-IP paradigm. Hence, we start by discussing what we believe are the most important features of IP networks. This is covered in Section 2. The following section develops the economics of All-IP networks and studies interconnection issues. Section 4 contains some concluding remarks.

2 Features of current IP networks

2.1 Packet switching

It is well-known that IP networks are packet-based. In particular, while calls in most telecommunication networks occupy physical circuits exclusively, calls in IP networks do not. Hence, while a telephone call without conversation still occupies 64Kbytes of physical resources and does not allow others to share, an Internet call without subsequent conversation does not occupy any physical resources.

Imagine that Amy makes a phone call to Bob. Bob talks incessantly while Amy listens carefully without speaking a word. In this case, while Amy initiated the call, most of the traffic flow is from Bob to Amy. That is, the initiator of the call is Amy, while the sender of (most) traffic is Bob. In a sense, the real user of the network resources is Bob. However, operators charge Amy for the call in most telecom networks since it is Amy who asked for the physical occupation of circuits.

The picture is remarkably different in IP networks. Imagine that Amy again makes an Internet phone call (through the current best-effort IP networks) to Bob the big talker. Now, all the words and sentences Bob speaks will be sliced into small packets and delivered to Amy independently through the complex web of routers. Contrary to the phone call above, Amy the initiator no longer asks exclusive occupation of the physical resources nor utilizes any network service during most of the conversation. On the other hand, Bob does make use of the network resources whenever he talks and sends packets to Amy.

At this point, it is very important to distinguish between the initiator and the sender, and likewise between the destination and the receiver. The *initiator* is the entity that initiates a call or a session, for example, a phone call or an Internet click to a particular website/file, and the *destination* is the entity that the call is destined to reach. In comparison, the *sender* (or the originator) is the entity that sends voice or data over the physical networks, and the *receiver* (or the termination) is the entity that receives such voice or data.

Regardless of actual traffic flow, traditional telecom networks treat the initiator as the sender (the originator) and charge the former for using of the physical network. Whether such practice is desirable and/or feasible in both technical and economic sense for IP networks is not completely settled yet. If we adopt the traditional telecom model, then each

TCP session can be regarded as a call through a virtual circuit and the initiator of the session can be charged for. If we are more concerned with the actual usage of network resources, then each packet flow at the IP level needs to be measured and the sender (the originator) of each packet can be charged for. See Huston (1999, 2005) for related discussion. Since the TCP protocol is at transport layer and the IP protocol is at network layer that corresponds to layers 4 and 3 of the OSI (Open System Interconnection) model respectively,¹ we will henceforth call the former charging scheme as *the layer 4 scheme*, and the latter as *the layer 3 scheme*.

This distinction can also be applied to the interconnection settlements between networks. Continue our example above and imagine that Amy subscribes to network A and Bob subscribes to network B. Under the layer 4 scheme, network A the initiating network pays interconnection charges to network B the destination network. In voice telecommunication, for example, Amy's network pays interconnection charges to Bob's network according to the settlement rates and the actual minutes of usage. On the other hand, each packet flow matters under the layer 3 scheme. For each packet sent from a network to another network during the conversation, *the originating network* that sends a particular packet pays interconnection charges to *the terminating network* that receives the packet according to the settlement rates. Networks will generally settle the payments based on net flows of packets, and for our example where Bob mostly talks network B pays interconnection charges to network A.

A typical Internet session, for example, when a consumer clicks on a website and the information at the website flows to the consumer closely resembles our second example. Therefore, depending on which scheme we use, the direction of interconnection charges is completely opposite. Moreover, since each packet finds its way to the termination independently, usually crossing a dozen routers and several network operators, the matter of interconnection settlements is by no means an easy task with the current best-effort IP networks.

We have so far assumed that networks charge variable fees according to actual usage. Alternatively, networks may settle interconnection charges according to some expected usage and thus may pay and receive fixed fees. In this case, networks may *bill and keep* (or *sender keeps all*), that is, networks do not charge each other, when each network's expected traffic inflow and outflow with respect to the other network is equal. In other words, when the net inflow of traffic is close to zero, networks need not actually pay money to each other.

2.2 Receiver benefit

Receivers in IP networks derive substantial benefits. Amy in our example may love listening to Bob. Consumers in the Internet get a lot of information from access to useful contents. Note that this also holds true for traditional telecom networks: We all enjoy receiving some phone calls, although not all phone calls, and what the senders give us over the network. Internet, however, is somewhat different from voice telephony in that receivers are active players. Consumers initiate a session to download contents from websites and receive useful information.

Simple economics principle suggests that it is efficient to charge receivers when they derive utilities from calls. That is, it is best for efficient consumption of calls to charge both

¹ The OSI model defines a networking framework for implementing protocols in seven layers: the physical layer (layer 1) – the data link layer (layer 2) – the network layer (layer 3) – the transport layer (layer 4) – the session layer (layer 5) – the presentation layer (layer 6) – the application layer (layer 7).

the sender and the receiver according to their relative benefits from the call. This way *the call externality* between the sender and the receiver can be internalized.

Traditional telecom networks essentially assume that receivers do not derive utilities from a call, and charge only the sender (who in turn is identified with the initiator as discussed above). Recently however, the so-called Receiver Pays Principle (RPP henceforth) in which receivers as well as senders share the call charges has been implemented in practice. For example, subscribers of mobile voice telephony in the United States, Canada, and Hong Kong currently pay for the calls they receive.² Current retail pricing structure in the Internet may also be viewed as incorporating the RPP: Consumers for instance subscribe and pay monthly fixed fees to Internet Service Providers even when they mostly receive traffic. If we postulate that networks can figure out the expected usage of consumers more or less accurately, then the current retail charges to Internet consumers hold non-negligible portion of the RPP.³

Similar reasoning applies to the structure of interconnection settlements. The originating network that sends the traffic no longer needs to pay as much amount of money as before since the receiver may contribute a certain part to the terminating network directly. For a more serious discussion on many important aspects of the RPP, such as the interplay between the interconnection charges and the retail charges, we refer the reader to the RPP literature that includes Doyle and Smith (1998), DeGraba (2001), Hermalin and Katz (2001), Kim and Lim (2001), and Jeon *et al.* (2004).

2.3 Two-sidedness

There are two distinct groups of subscribers in the Internet, namely consumers and websites. Consumers are those subscribers who get information and download files from websites, send and receive e-mails, or send and receive files via peer-to-peer (P2P) technologies, while websites are those subscribers who mostly post information to advertise their businesses or products, upload files for consumers' access, or engage in e-commerce.

In addition, significant part of traffic flows from websites to consumers. It is true that other traffic patterns exist such as traffic flows from consumers to websites, from consumers to consumers, and from websites to websites. P2P traffic that flows from consumers to consumers is rapidly increasing recently, and nothing is more certain than the fact the Internet traffic pattern will be ever-evolving. However, we can safely say that much of the Internet traffic is from website to consumers for the time being.

The Internet is thus two-sided. Moreover, direct payments from consumers to websites for useful contents are available. It is therefore necessary to know the intricate economic relationship among networks, consumers, and websites as well as other stakeholders. As a first step to this direction, we briefly review the recent literature on two-sided markets below.

Two-sided markets can be defined as markets in which one or several platforms enable interactions between end-users. Alternatively, Evans (2003) defines a market to be two-sided if there are (i) two distinct groups of customers, (ii) the value obtained by one kind of customers increases with the number of the other kind of customers, and (iii) an intermediary is necessary for internalizing the externalities created by one group for the other group. Examples of platforms or intermediaries in two-sided markets abound: (i)

² Collect calls and 800 numbers are other examples.

³ Since consumers send e-mails and files, the retail charges also hold a portion of the traditional "Sender Pays Principle".

software platforms such as Windows XP and videogame platforms such as Sony PlayStation or Microsoft X-Box, (ii) Shopping malls, (iii) Newspapers, TV stations, and Internet portals, (iv) Payment systems such as Visa and MasterCard, (v) Nightclubs and dating services, and so on.⁴

Cross network externality between groups is an important feature of two-sided markets. That is, the utility of each member of a group increases as the size of the other group gets larger. Since the transactions in these markets involve two sides, with the benefits and the costs arising jointly, the optimal pricing structure depends on the demand and cost conditions of both sides. In other words, as Evans (2003) put it, “there is no meaningful economic relationship between benefits and costs of either side of the market considered by itself. It takes two to tango”.

The optimal pricing structure in these markets is shaped by many other factors as well as the indirect network externalities, and different industries adopt diverse pricing strategies.⁵ There are several factors that determine the pricing structure. First of all, price elasticity matters. The price charged to a group goes down as its own demand elasticity goes up. This is a standard principle. Second, the externalities to the other side count. That is, if a group exerts strong externalities to the other group, then the price to that group will go down. For example, if women’s presence at a nightclub strongly influences men’s affiliation decision, women may be offered free admission. Third, as Hagiu (2004) convincingly demonstrates in his model, product diversity also affects the pricing structure: The price to consumers goes down as consumers’ preference for product diversity goes up. In addition, the market power of the sides is important. For example, if content providers in IP networks have strong market power, as in the case of Hollywood movie houses or the NFL (National Football League), then the price charged to this side will go down even below zero due to platform’s weak bargaining power.

Two-sided markets offer many interesting research topics, and the literature is rapidly expanding. See for example Armstrong (2006), Armstrong and Wright (2006), Rochet and Tirole (2004) as well as the papers cited above. In particular, the nature of competition among platforms and its implication on stakeholders’ strategies need to be studied in more depth.

Going back to the IP networks, the theory of two-sided markets implies that the retail prices charged to consumers and websites, whether in the form of fixed fees or variable fees, need not reflect the benefits or costs of either side by itself. Rather, the subscription fees or usage fees depend on various factors as outlined above: elasticities of demand, cross externalities, desire for variety, the pricing practice and the market power of content providers in the Internet, and so on.

The theory also implies that interconnection settlements between networks need not and should not be simply based on the benefits and costs of each side by itself. Especially when networks compete to attract subscribers in retail markets, the optimal level of interconnection charges may be quite different from the simple costs of access. This seems to be a fruitful research topic that has not been fully understood. In the next section, we will build a simple model that respects the two-sidedness of IP networks.

⁴ See Rochet and Tirole (2003) for more examples.

⁵ See Evans *et al.* (2005) for a survey of pricing strategies in computer-bases industries.

3 Economics of All-IP networks

In this section, we will discuss retail pricing and interconnection charges of All-IP networks. We want to state at the outset that the material is quite preliminary. This is partly due to the plain fact that All-IP networks are still a possibility, not a reality.

As defined in the Introduction, All-IP networks are multi-service networks or general QoS-enabled networks, in which all kinds of applications and services can be delivered. Therefore, it seems acceptable to postulate that each application session can be correctly identified and measured in All-IP networks. This can be technically achieved by encoding a classification code at the header of packets at the TCP level. The harder part is (i) how to classify application sessions and accommodate new services, and (ii) how to design the pricing scheme that is based upon the service parameters or the standardized charging factors including access bandwidth, peak delivery rate, delay, jitter, loss, call duration, rates, call priority class, call start time, distances, and agree on mutual interconnection charge rates for each specific application and/or service category. Moreover, the SLA (Service Level Agreement) between the network and the subscribers as well as between interconnecting networks is needed.

A natural structure of retail charges seems to be two-part tariffs in which connectivity and basic services are provided at fixed subscription fees, while usage fees are charged for application sessions. In determining the level of fixed fees, those factors discussed in the previous section may be considered. In particular, receiver benefit and cross externalities are the most important factors. As for usage fees, since each application session can be identified and measured, the layer 4 scheme can be applied and usage fees (either based on minutes or volumes) can be charged to the initiator, as in the traditional telecom charging scheme. We assume that this is feasible both technologically and economically.

Interconnection settlements may also mirror this structure. That is, fixed interconnection fee settlements between networks are based on some measure of subscription. Specific factors may include the number of consumers and websites, the importance of P2P, etc. Additionally, the initiating network pays variable interconnection fees to the destination network for each application session. Note that these interconnection payments are also similar to traditional telecom interconnection settlement scheme.

For the rest of this section, we first present a simple fixed interconnection fee settlements model for connectivity and basic services. We then briefly review and apply the existing literature on two-way interconnection to the session-based interconnection charges.

3.1 A basic fixed interconnection fee settlements model

Consider an All-IP network and let n be an appropriate measure of network size. Without giving an exact formula for the parameter n , we will interpret it to be a weighted sum of different customer classes. For example, each consumer may count as 1 while each website may count as 2 if websites exert more externalities to other subscribers of the network. As the interaction between consumers becomes more important as the P2P traffic grows, the relative weight to consumers may increase. Since the exact formula for the parameter n will depend on the actual features of All-IP network, we prefer to be rather agnostic of n in this paper.

The average utility of subscribers is given by $u(n)$. Thus, the total utility the network generates is $nu(n)$. We assume that $u'(\cdot) > 0$ so that positive network externalities are

present. The total cost of serving network size of n is given by $c(n)$. The revenue of the network is given by the formula

$$\frac{1}{1+\lambda} nu(n) + \frac{\lambda}{1+\lambda} c(n)$$

where λ is the measure of competitiveness that the network faces. The case when $\lambda = 0$ corresponds to a perfect price discriminating monopolist, the case when $\lambda = \infty$ corresponds to perfect competition, and the case when $\lambda \in (0, \infty)$ corresponds to the competitive situation in between. Note that this is a reduced-form model in that no explicit consideration of pricing is given. We will provide some justification below.

We will study the bargaining power in network interconnection using the Nash bargaining solution. This approach is taken previously in Besen *et al.* (2001). Let there exist two networks $i = 1, 2$ with respective installed subscriber bases. Network i 's pre-interconnection profits are given by

$$\pi_i = \frac{1}{1+\lambda_i} [n_i u(n_i) - c_i(n_i)]$$

and its post-interconnection profits without interconnection settlements, that is, under the “bill and keep” or “sender keeps all” arrangement are

$$\rho_i = \frac{1}{1+\lambda_i} [n_i u(n_1 + n_2) - c_i(n_1 + n_2)].$$

Network i 's Nash bargaining profits are

$$\phi_i = \frac{1}{2} [\rho_1 + \rho_2 - \pi_1 - \pi_2] + \pi_i = \frac{1}{2} [\rho_1 + \rho_2 + \pi_i - \pi_j]$$

where $i + j = 3$. Therefore, the difference

$$\phi_i - \rho_i = -\frac{1}{2} [\rho_i - \pi_i] + \frac{1}{2} [\rho_j - \pi_j]$$

is the net interconnection charge receipts network i gets from network j . (It becomes the net interconnection charge payments when negative.) Observe that

$$\rho_i - \pi_i = \frac{1}{1+\lambda_i} [n_i \{u(n_1 + n_2) - u(n_i)\} - \{c_i(n_1 + n_2) - c_i(n_i)\}] \quad (1)$$

is the incremental profits network i gets from the interconnection, and likewise $\rho_j - \pi_j$ is the incremental profits network j gets from the interconnection. Therefore, network i 's bargaining power increases as i 's incremental profits are smaller or j 's incremental profits are larger. The reason is that the network that obtains larger incremental profits from interconnection has a stronger desire to strike a deal, and so is willing to concede more.

We can determine the direction and the amount of monetary transfers by comparing $\rho_1 - \pi_1$ and $\rho_2 - \pi_2$. In particular, networks do not exchange monetary transfers when they are equal. On the other hand, when $\rho_1 - \pi_1 < \rho_2 - \pi_2$, network 1 receives interconnection charges from network 2. We list some important factors that determine the bargaining power of networks.

Proposition 1: If two networks face perfectly competitive environments, then they arrange interconnection without monetary transfers.

Proof: We observe from equation (1) above that $\rho_1 - \pi_1 = \rho_2 - \pi_2 = 0$ when $\lambda_1 = \lambda_2 = \infty$. Q.E.D.

The intuitive reason is that the incremental profits of networks are competed away under perfect competition, and no network is more willing to concede than the other.

Proposition 2: Other things being equal, the network under a more competitive environment has stronger bargaining power.

Proof: We observe from equation (1) above that $\rho_i - \pi_i$ decreases as λ_i increases. Therefore, network i 's bargaining power increases. Q.E.D.

The intuitive reason is that the network under a more competitive environment obtains a relatively smaller share of the increased network values from interconnection, and so is less keen to interconnect. Therefore, when networks in different market environments make interconnection deals, a network under a less competitive environment will pay interconnection charges to the other network *ceteris paribus*.

Assume $\lambda_1 = \lambda_2 = \lambda$, and $c_1(\cdot) = c_2(\cdot) = c(\cdot)$ with $c'(\cdot) \geq 0$. It is easy to see that $\rho_1 - \pi_1 = \rho_2 - \pi_2$ when $n_1 = n_2$. Therefore, networks with equal size and identical cost function that also face the same market competitiveness will interconnect without monetary transfers. To see the effect of network size, let $n_1 > n_2$ without loss of generality. Then,

$$\rho_i - \pi_i = \frac{1}{1+\lambda} [n_i \{u(n_1 + n_2) - u(n_i)\} - \{c(n_1 + n_2) - c(n_i)\}].$$

When $u''(\cdot) = 0$ so that network externalities exhibit constant returns to scale, say $u(n) = bn$, we have

$$[\rho_1 - \pi_1] - [\rho_2 - \pi_2] = \frac{1}{1+\lambda} [c(n_1) - c(n_2)] \geq 0,$$

implying that the larger network, network 1, needs to pay interconnection charges. The reason is that the incremental benefits from interconnection are the same across networks with linear network externalities, while the incremental costs are higher for the smaller network. Therefore, the larger network needs to subsidize the smaller network. The case when $u''(\cdot) > 0$ so that network externalities exhibit increasing returns to scale, or when $u''(\cdot) < 0$ so that network externalities exhibit decreasing returns to scale is not straightforward. It will depend on the relative magnitudes of the difference in the incremental benefits $n_1 \{u(n_1 + n_2) - u(n_1)\} - n_2 \{u(n_1 + n_2) - u(n_2)\}$ and the difference in the incremental costs $c(n_1) - c(n_2)$.

We see that when $u'(\cdot)$ decreases sufficiently fast so that the incremental benefits of the larger network are small enough, the larger network will receive interconnection charges from the smaller network. This is in fact the typical settlement pattern we observe in reality.⁶

Summarizing the above discussion, we have:

⁶ The transit arrangement between a larger network and a smaller network may also reflect the cost factor discussed shortly.

Proposition 3: Other things being equal, the likelihood that the larger network receives interconnection charges from the smaller network increases as the effect of network externalities decreases faster with respect to network size, that is, $u''(\cdot)$ gets smaller.

The cost structure also matters. To see this, let us now assume that $\lambda_1 = \lambda_2 = \lambda$, $n_1 = n_2 = n$, $c_1(n_1) = c_1 n_1$, $c_2(n_2) = c_2 n_2$ with $c_1 > c_2$. That is, the constant marginal cost of network 1 is higher than that of network 2. Then, $\rho_i - \pi_i = \frac{1}{1+\lambda} [n\{u(2n) - u(n)\} - c_i n]$ and we have

$$[\rho_1 - \pi_1] - [\rho_2 - \pi_2] = \frac{n}{1+\lambda} [c_2 - c_1] < 0.$$

Thus, the network with higher costs has stronger bargaining power.

Proposition 4: Other things being equal, the network with higher costs of providing services will receive interconnection charges.

This result has at least two implications. First, if a network provides transit service to another network instead of entering into a peering agreement, the costs will be higher so the former will receive interconnection charges. This is what we observe in practice: See Cremer *et al.* (2000) or Kende (2000) for the discussion of IP interconnection structure, in particular, peering and transit interconnection contracts. Second, as for the interconnection issue of a QoS-enabled IP network and a best-effort IP network, it is a reasonable assumption that a QoS network has higher marginal costs than a best-effort network. Hence, a QoS-enabled IP network will receive interconnection charges when it interconnects with a best-effort IP network.

We may enrich the model by incorporating the two-sidedness of IP networks quite straightforwardly. To do so, take the following simplistic picture of IP networks as depicted in some papers including Laffont *et al.* (2003) on Internet interconnection. Let there exist two networks as before, and network i initially has l_i consumers and m_i websites. Assume that the traffic flows only from websites to consumers, and every consumer visits every website once. Each consumer's utility is $f(m_i)$ and each website's utility is $g(l_i)$ with $f'(\cdot) > 0$ and $g'(\cdot) > 0$. Note that $f(\cdot)$ and $g(\cdot)$ measures the cross externalities. Thus, the total utility network i generates is $l_i f(m_i) + m_i g(l_i)$. Let the marginal cost of sending the traffic be c^o and the marginal cost of receiving the traffic be c^t . Then, network i 's total cost is $l_i m_i (c^o + c^t)$. Network i 's pre-interconnection profits are given by

$$\pi_i = \frac{1}{1+\lambda_i} [l_i f(m_i) + m_i g(l_i) - l_i m_i c^o - l_i m_i c^t]$$

and its post-interconnection profits without interconnection settlements, that is, under the "bill and keep" or "sender keeps all" arrangement are

$$\rho_i = \frac{1}{1+\lambda_i} [l_i f(m_1 + m_2) + m_i g(l_1 + l_2) - l_i m_i c^o - l_i m_i c^t - l_j m_i c^o - l_i m_j c^t].$$

As above, we need to compare, for $i = 1, 2$,

$$\rho_i - \pi_i = \frac{1}{1+\lambda_i} [l_i \{f(m_1 + m_2) - f(m_i)\} + m_i \{g(l_1 + l_2) - g(l_i)\} - l_j m_i c^o - l_i m_j c^t]$$

to see who pays whom how much. A moment's inspection establishes the following:

Proposition 5: Other things being equal, the more cross externalities websites (consumers, resp.) exert, the network with relatively more websites (consumers, resp.) will receive interconnection charges.

We have built a simple model to see what determines the relative bargaining power of networks when they interconnect. The important factors we have found are (i) the appropriate measure of network size, (ii) the market competition that each network faces, (iii) the magnitude and the change of network externalities with respect to network size, (iv) the costs, and (v) the cross network externalities. This simple model does have many limitations including the following. As noted above, this is a reduced -form model that does not account for endogenous pricing decisions. Neither does it consider the network competition *for* subscribers. Instead, it assumes that the subscriber bases are fixed for the purpose of bargaining. While this model certainly describes some interconnection arrangements, such as the international interconnection, the interconnection between networks with substantially differentiated customer bases, or the interconnection with locked-in or inertial customers, we may need to study the situation when customers actively choose and switch their network affiliation.

In what follows, we briefly discuss the case when networks compete for subscribers. We use the standard Hotelling framework. Customers of mass 1 are located uniformly on the unit interval. Network 1 is situated at the left-end (zero) and network 2 is situated at the right-end (one). Since we are concerned with fixed interconnection fee settlements in this subsection, assume that networks charge fixed subscription fees for connectivity and basic services. Assume also that every customer subscribes to one and only one network.

Network i 's market share is denoted by s_i and network i 's *network size* is denoted by z_i . Note that $z_i = s_i$ if networks do not interconnect, while $z_i = 1$ if they interconnect. Given network i 's fixed subscription fee r_i , customer x 's utility with $x \in [0,1]$ when joining network 1 is $u(z_1) - r_1 + t(1-x) + \theta t$ and that when joining network 2 is $u(z_2) - r_2 + tx$. The parameter t measures the degree of differentiation between networks, while the parameter θ captures the possible advantage of network 1 due to brand loyalty, installed base, and so forth.⁷ We assume zero costs for simplicity.⁸

Consider first the case when networks interconnect. Then, $z_1 = z_2 = 1$. Network 1's market share s_1 is determined by $u(1) - r_1 + t(1 - s_1) + \theta t = u(1) - r_2 + ts_1$, giving us

$$s_1 = \frac{1+\theta}{2} + \frac{r_2 - r_1}{2t} \text{ and } s_2 = 1 - s_1 = \frac{1-\theta}{2} + \frac{r_1 - r_2}{2t}.$$

Networks' post-interconnection profits are $\rho_i = r_i s_i$. Solving for the equilibrium outcome, we get

$$r_1^* = t + \frac{t\theta}{3}, r_2^* = t - \frac{t\theta}{3}, \rho_1^* = \frac{t}{18}(3+\theta)^2, \rho_2^* = \frac{t}{18}(3-\theta)^2.$$

⁷ Assume that $t > 1$ and $\theta < 1$ to ensure both networks attract customers in equilibrium.

⁸ Note that cost differentials between networks may be equivalently represented in the model by the parameter θ .

Network 1's market share in equilibrium is $1/2 + \theta/6$.

Consider now the case when networks do not interconnect. Then, $z_i = s_i$. To analytically solve the model, let us assume that $u(z) = z$ so that constant returns to scale hold. Network 1's market share s_1 in this case is determined by $s_1 - r_1 + t(1 - s_1) + \theta = 1 - s_1 - r_2 + ts_1$, giving us

$$s_1 = \frac{t(1 + \theta) - 1}{2(t - 1)} + \frac{r_2 - r_1}{2(t - 1)} \text{ and } s_2 = 1 - s_1 = \frac{t(1 - \theta) - 1}{2(t - 1)} + \frac{r_1 - r_2}{2(t - 1)}.$$

Networks' pre-interconnection profits are $\pi_i = r_i s_i$. Solving for the equilibrium outcome, we get

$$r_1^* = t - 1 + \frac{t\theta}{3}, r_2^* = t - 1 - \frac{t\theta}{3}, \pi_1^* = \frac{[3(t - 1) + t\theta]^2}{18(t - 1)}, \pi_2^* = \frac{[3(t - 1) - t\theta]^2}{18(t - 1)}.$$

Network 1' market share in equilibrium is $1/2 + t\theta/6(t - 1)$.⁹

We observe several facts. First, subscription fees are lower when networks do not interconnect. The reason is that each network's own subscriber base is important from the customers' perspective, which leads to fiercer competition. Second, profits are lower when networks do not interconnect. Therefore, networks have an incentive to interconnect. Finally and most importantly, we observe that $(\rho_1 - \pi_1) - (\rho_2 - \pi_2) = 0$, hence networks interconnect without monetary transfers. That is, their bargaining power is the same. This is interesting since networks are asymmetric. Note that this neutrality result is in contrast with proposition 3 when networks do not compete for subscribers.

One may wonder whether these results, particularly the last one, are due to constant returns to network scale, that is, $u(z) = z$. Calculations show, however, that the same results obtain under increasing returns $u(z) = z^2$ or decreasing returns $u(z) = (2 - z)z$. Moreover, it can be straightforwardly shown that, in all cases, consumer surplus decreases after interconnection.

We mention that Carter and Wright (1999) derive similar results in a competition model where networks charge usage-based fees to subscribers. In particular, they show that usage fees increase after interconnection and that interconnection is not generally in the interest of subscribers. These facts are reaffirmed in this simple competition model where networks charge fixed subscription fees.

3.2 The session-based interconnection charges

We assume that an agreement is established among relevant parties as for the classification and service levels of application sessions or services such as VoIP, VOD, or IPTV. In addition, we assume that it is technically possible to measure the usage of each application or service. Given these assumptions, it is rather a straightforward matter to agree on the interconnection charges.¹⁰ We may apply the traditional telecom interconnection charge scheme, or the layer 4 scheme, to each application or service category. Of course, it may be necessary to calculate or agree on the physical costs of providing each service category in an NGN environment, which is an issue of technical or accounting nature.

⁹ Note that $t\theta/6(t - 1) < 1/2$ to have $s_2 > 0$.

¹⁰ If these assumptions are not satisfied, we may apply the settlements model of the previous subsection.

There exists voluminous literature on the economics of two-way interconnection. This literature studies the situations where each network needs the interconnection to the other network for the provision of end-to-end services to its subscribers. Its main emphases include the relationship between the interconnection charges and the retail charges, the determination of market-mediated and socially optimal levels of interconnection charges, and some guiding principles for the regulators. Rather than summarizing the literature in this paper, we only refer the reader to Laffont and Tirole (2000) and Armstrong (2002) for authoritative surveys.¹¹ As for Internet interconnection, Laffont *et al.* (2003) present a simplistic but neat viewpoint. In particular, they explicitly consider the two-sidedness of IP networks as well as receiver benefit, and establish the *off-net-cost principle* which states that each network's retail charges to its subscribers are set to the actual cost level as if the subscriber belongs to the other network. Recall that the discussion in the previous section on two-sided market implies that this conclusion may not hold when networks do have market power, since in that case each network will set retail charges to reflect the cross externalities as well as other factors.

The two-way interconnection literature in fact informs us that a lot depends on specific market environments, such as the competitiveness of backbones, the structure of retail charges, the nature of subscriber demand, and so forth. The IP networks further pose many difficult but fascinating research topics for network interconnection. First of all, direct payments between subscribers are possible in IP networks. It is yet to be known how it will affect the interconnection agreement between networks. One may argue that the interconnection charges are *neutral* for networks' profits or social welfare since networks pass the costs along to subscribers. However, this conclusion may hold only in special circumstances.

Moreover, contents are important in All-IP networks and the competitiveness of content will matter a lot for retail and interconnection charges. If a content provider has significant market power, as in the case of NFL games or Hollywood movies, it may exert strong bargaining power and a network or several networks will compete for the access to the content. The economic structure in this area of All-IP networks seems similar to that in the CATV industries or broadcasting industries. In particular, networks will not charge for the network service to the content provider. On the contrary, they will pay a handsome amount of money to secure the provision of such content. On the other hand, if a content provider faces tough competition from other content providers, her bargaining power against the network is weak and networks will generally charge for the network service in the form of revenue sharing of the content provider's sales to consumers or in the form of a fixed admission fee. Observe that the actual charges are dependent on the factors discussed earlier in two-sided markets. It seems hard to say anything concrete about the level of interconnection charges at this early stage of development. If we have to conjecture, however, we would guess that the interconnection charges will be mainly dependent on pure network costs since, for the session-based charges, the network functions as the provider of network services. Note that this feature is in contrast to the previous case in subsection 3.1 where networks function as providers of basic information services as well as network services.

There are many interesting questions concerning content in All-IP networks, which remain to be answered satisfactorily. For example, is it desirable to vertically integrate contents and network services or not, from the viewpoint of firms and of society as a whole?

¹¹ See, however, the next section for some discussion in relation to policy issues.

On the one hand, vertical integration can eliminate the well-known double marginalization problem. On the other hand, the market power of the integrated firm may increase so as to be detrimental to economic welfare. One may also wish to understand the effects of exclusive contracts between a content provider and a network, and the effects of multi-homing of content providers. Another interesting phenomenon is the network admission policy of content providers when some content categories are competitively supplied by many differentiated providers: Networks may have an incentive to limit the number of content providers to control the quality of contents and to generate profits. This is an issue related to the nature of information goods and the welfare effects are yet to be understood. These contractual arrangements will affect the interconnection agreements between networks in a nontrivial way. These and many other interconnection issues of All-IP networks await meaningful research efforts.

4 Concluding remarks

Network interconnection has always been a concern for regulators. Historically, interconnection regulation has become necessary to introduce competition into the monopolistic telecom industries. The regulation has aimed to ensure the open and fair interconnection of new entrants to the facilities of incumbents. The objectives of interconnection regulation are (i) to prevent inefficient foreclosure by dominant firms, (ii) to prevent excessive entry, and (iii) to encourage innovation and long-term investment. These objectives are achieved mainly by imposing mandatory interconnection obligation on the firms with essential facilities, and by controlling interconnection charges.

These objectives are mainly for one-way interconnection in which a dominant firm holds essential facilities to which other firms need to access. For two-way interconnection, the objective to prevent collusion may be added. In fact, one of a few general principles that the two-way interconnection literature provides is that networks generally have an incentive to collude on the interconnection charges to mitigate the competition for subscribers.

The traditional regulatory goal of promoting competition may be less important in All-IP network environments, while the competition policy concerns gain more importance. As a specific example, consider the strategic refusal to interconnect. Unless there is an essential facility that was acquired by historical government intervention, it seems not wise to impose mandatory interconnection because it may inefficiently intervene in the natural process of firms' activities, especially investment and innovation motives. And, in an All-IP network environment, there may not exist essential facilities. This is obviously true for the backbone level, and also for the access level since alternative technologies including traditional fixed lines, cable lines, FTTH, wireless technologies such as IEEE802.11, high-speed mobile technology exist. However, there may emerge market power at different levels such as the service level, and competition concerns may well apply.

One of the upcoming issues in All-IP networks is the issue of quality or capacity degradation of interconnection between QoS-enabled and best-effort networks. As Crémer *et al.* (2000) and Foros *et al.* (2005) show in related papers, it is plausible that telecom operators have an incentive to lower the interconnection quality or capacity between QoS-enabled networks and current best-effort networks. The reason is simple: By refusing to interconnect fully, the operator can attract more customers to the QoS-enabled networks because customers value network externalities positively.

Finally, while traditionally interconnection charges are regulated at costs, the economic analysis generally indicates that a flexible approach is needed. This is all the more so in an All-IP networks environment since it is extremely hard to calculate proper costs when multiple services are provided within the common network. Moreover, two-sidedness implies that it is simply inappropriate to base any charges on the benefits or costs of one side considered by itself.

5 References

Armstrong, M. (2002) “The Theory of Access Pricing and Interconnection,” in M. Cave, S. Majumdar and I. Vogelsang (ed.), *Handbook of Telecommunications Economics*, North-Holland: Amsterdam.

Armstrong, M. (2006) “Competition in Two-sided Markets,” *Rand Journal of Economics*, forthcoming.

Armstrong, M. and J. Wright (2006) “Two-sided Markets, Competitive Bottlenecks and Exclusive Contracts,” *Economic Theory*, forthcoming.

Besen, S., P. Milgrom, B. Mitchell and P. Srinagesh (2001) “Advances in Routing Technologies and Internet Peering Agreements,” *American Economic Review*, 91: 292-296.

Carter, M. and J. Wright (1999) “Interconnection in Network Industries,” *Review of Industrial Organization*, 14: 1-25.

Cr  mer, J., P. Rey and J. Tirole (2000) “Connectivity in the Commercial Internet,” *Journal of Industrial Economics*, 48: 433-472.

DeGraba, P. (2001) “Efficient Interconnection for Competing Networks when Customers Share the Value of a Call,” mimeo, Federal Trade Commission, U.S.A.

Doyle, C. and J. Smith (1998) “Market Structure in Mobile Telecoms: Qualified Indirect Access and the Receiver Pays Principle,” *Information Economics and Policy*, 10: 471-488.

Evans, D. (2003) “The Antitrust Economics of Multi-sided Platform Markets,” *Yale Journal on Regulation*, 20, 325-382.

Evans, D., A. Hagiu and R. Schmalensee (2005) “A Survey of the Economic Role of Software Platforms in Computer-based Industries,” *CESifo Economic Studies*, 51: 189-224.

Foros, O., H. Kind and J. Sand (2005) “Do Internet Incumbents Choose Low Interconnection Quality,” *Information Economics and Policy*, 17: 149-164.

Hagiu, A. (2004) “Two-sided Platforms: Pricing and Social Efficiency,” mimeo, RIETI and Harvard Business School.

Hermalin, B. and M. Katz (2001) “Network Interconnection with Two-sided User Benefit,” mimeo, University of California at Berkeley.

Huston, G. (1999) "Interconnection, Peering, and Settlements," mimeo, <http://www.potaroo.net>.

Huston, G. (2005) "Where's the Money? – Internet Interconnection and Financial Settlements," mimeo, <http://www.potaroo.net>.

Kende, M. (2000) "The Digital Handshake: Connecting Internet Backbones," OPP Working Paper No. 32, Federal Communications Commission, U.S.A.

Kim, J. and Y. Lim (2001) "An Economic Analysis of the Receiver Pays Principle," *Information Economics and Policy*, 13: 231-260.

Jeon, D., J. Laffont and J. Tirole (2004) "On the 'Receiver-pays' Principle," *Rand Journal of Economics*, 35: 85-110.

Laffont, J., S. Marcus, P. Rey and J. Tirole (2003) "Internet Interconnection and the Off-net-cost Pricing Principle," *Rand Journal of Economics*, 34: 370-390.

Laffont, J. and J. Tirole (2000) *Competition in Telecommunications*. MIT Press: Cambridge.

OECD (2004) "Next Generation Network Development in OECD Countries," Final Report, Directorate for Science, Technology and Industry/Committee for Information, Computer and Communications Policy.

Rochet, J. and J. Tirole (2003) "Platform Competition in Two-sided Markets," *Journal of the European Economic Association*, 1: 990-1029.

Rochet, J. and J. Tirole (2004) "Two-sided Markets: An Overview," mimeo, University of Toulouse.