

The Economics of IP Networks – Market, Technical and Public Policy Issues Relating to Internet Traffic Exchange

Executive Summary

Bad Honnef, May 2002

Disclaimer

The views expressed in this report are those of the authors and are not necessarily shared by the European Commission.

Authors:

Dieter Elixmann

Mark Scanlan

With contributions from:

Alberto E. García

Klaus Hackbarth

Annette Hillebrand

Gabriele Kulenkampff

Anette Metzler

Table of contents

Executive Summary	i
1 Introduction	i
<i>The main conclusions of the study:</i>	ii
2 Economic background to Internet public policy issues	v
2.1 Externalities	v
2.2 Market power	vii
2.3 Regulation	vii
3 Market failure and the Internet's addressing systems	viii
4 Quality of service	xi
5 Interconnection and partial substitutes	xiv
5.1 Peering	xiv
5.2 Transit	xv
5.3 Strategies involving interconnection	xvi
5.4 Interconnect price discrimination and fragmentation	xvii
6 Existing regulation	xviii

Executive Summary

1 Introduction

This study provides data, descriptions and analysis of issues concerned with commercial traffic exchange on the Internet, especially as it concerns backbone services, the providers of which are known as core ISPs or IBPs (Internet backbone providers). As such the study mainly looks back into the core of the Internet and only addresses what is happening between end-users and ISPs to the extent that this provides valuable insights for traffic exchange with the core.

We focus on those topics that relate to Internet traffic exchange that appear to be of particular interest to public policy makers and their advisers, and analyse in regard to each of the topic we discuss, whether there appears to be a level of market failure that might require the authorities to intervene in some form or another.

The causes of market failure can be grouped into three: (i) market power; (ii) externalities, and (iii) existing regulations. We look at all three causes in this study. Other rationale for the authorities to intervene in the market also exist, including: decency, confidentiality, security, and universal service issues. However, none of these issues feature to a significant degree in this study, primarily as it is not concerned *per se* with end-users and ISPs that serve them.

In order to fully examine the potential market failure issues, the study describes and analyses the following points:

- (i) the Internet addressing and routing system;
- (ii) arrangements that exist or are emerging in regard to Internet traffic exchange. The study includes a look at both the institutional, strategic and contractual side of traffic exchange arrangements;
- (iii) quality of service (QoS) issues, especially regarding the problems encountered when traffic is exchanged between different networks;
- (iv) the economics of congestion and why it is relevant to this study, and
- (v) the importance of network effects to understanding the strategic interest of the players and the risk that one (or a group) in the value chain might come to dominate the industry to the disadvantage of customers.

The main conclusions of the study:

Market structure and competition:

- In the complete absence of rules protecting competition, industries that display strong network effects like the Internet have a tendency to drift toward monopolisation, most probably through the aggressive takeover of rivals.
- The Internet has become less hierarchical in the last 5 or 6 years due to the introduction of new technologies which enable an increasing proportion of traffic to avoid the upper layer of the Internet hierarchy. This has reduced the market power of core ISPs.
 - The main technologies enabling this are:
 - those that enable content to be held closer to the edges of the internet (caching, mirroring, and content delivery networks), and
 - those that have made regional interconnection between smaller ISPs (secondary peering), and multi-homing (transit contracts with several core ISPs), economically attractive.
- Where market power is distributed among a sufficient number of interconnected networks, and services that function as partial substitutes for transit interconnection are widely purchased (those services noted in the above bullet), direct *ex ante* regulation of the core of the Internet is unnecessary in order to prevent monopolisation from occurring. Merger regulation should be relied upon instead.
- The loosening of the hierarchy has reduced the bargaining power of core ISPs when negotiating service contracts with 'downstream' ISPs and major content providers.
- Analysis of the strategic interests of the largest player(s) suggests that in many cases they will not gain by degrading interconnection unless they already have a share of the global market well in excess of anything observed presently.
- Research evidence regarding price discrimination suggests that in a wide variety of circumstances backbones have an incentive to set interconnection prices according to "the off-net pricing principle": that is, customers pay the same price for receiving traffic independently of whether the traffic was originated on the ISPs own network (on-net) or on another ISPs network (off-net). If ISPs have the power to price discriminate between on-net and off-net traffic, this situation may become unstable and the market tip in favour of the largest network.
- Where *seamless* interconnection can not be provided (e.g. where there are QoS problems at network borders as is the present case), new services that require high QoS attributes such as VoIP, may restore to a degree the importance of network size in the competitive process between core ISPs. If this occurred it would imply tipping in favour of the largest network.

- Where *seamless* interconnection can not be provided (e.g. where there are QoS problems at network borders as is the present case), new services that require high QoS attributes such as VoIP, may restore to a degree the importance of network size in the competitive process between core ISPs. If this occurred it would imply tipping in favour of the largest network.
- Part of our report provides an analysis of the relevance of theoretical models related to the topics discussed in this study, with several of these models motivated by recent merger cases. The strategies identified by these economic models provide valuable insight by focussing on key economic relationships. The models are, however, always simplifications of reality, and should be viewed more as a *failure to reject* the strategy predicted, and less as a confirmation that it will in practice occur.
- A level of market failure exists in the pricing of interconnected traffic as the indirect network benefits provided by content providers are not taken account of in interconnection prices. Given recent growth rates in subscriptions and content, the net cost to society of this failure appears to be relatively low, and not of the order that would warrant official intervention.
- The number of public Internet exchange points is increasing.
- The peering policies of core ISPs do not appear to be unfair and do not at this stage entail competition policy or regulatory concerns.

Addressing

- IPv4 addresses are likely to be exhausted before the end of this decade. The adoption of IPv6 is costly but appears unlikely to cause significant disruption, and thus to require official involvement to facilitate transition.
- In addition to a huge increase in address space provided by IPv6, it offers other advantages over IPv4, although in several areas the advantage has been significantly narrowed due to developments involving IPv4. Other aspects of the Internet such as those concerning measurement, billing, and grade of service pricing, need to develop before other advantages offered by IPv6 can be translated into benefits to end-users.
- IPv4 addresses are a scarce resource and failure to treat them as such has lead to pending address exhaustion and the need for the Internet Community to adopt a new addressing scheme before the end of this decade. The efficiency costs entailed in this could be considerable. However, intervention that would try to correct this inefficiency is not advised. The main reasons for this are as follows:
 - the Internet has no nationality, and is made up of over 100,000 networks world-wide. This makes regulation rather impractical;

- addressing is tied up with technology development, and in such cases intervention should only be considered where there are compelling reasons for doing so, and
- the internet community has shown that it is able to plan for the future and will likely avoid serious disruption such that existing and replacement addressing schemes will work with each other i.e. at a minimum IPv6 hosts will support IPv4 addresses.

Quality of service (QoS)

- QoS in packet-switched networks such as the Internet describes the statistical properties of the packet stream of network 'connections'
- Most parameters concerning Internet traffic performance are non deterministic and have to be defined probabilistically. Thus, strictly speaking packets do not receive equal treatment. However, as packets are randomised irrespective of the source or application to which that data is being put, then in this sense no packet can be said to receive preferred treatment under the current best effort transfer principle of the Internet.
- The key to convergence between the Internet and broadcasting and voice telephony services, is provided by a combination of:
 - solutions for QoS problems that exist between different networks (i.e. when traffic passes from being *on-net* to *off-net*), which arise for reasons including differences in propriety equipment and network management systems, and
 - the introduction of demand management techniques, such as through the provision of several grades of service (GoS), each with different end-to-end QoS statistics, and priced to customers accordingly.
- Greater bandwidth and processing power alone will not solve all congestion and QoS problems on the Internet. This is because the Internet involves the use of scarce resources, and when treated otherwise, theory and evidence suggests that congestion will become a problem, undermining convergence and the development of services that require superior QoS statistics.
- Technologies intended to provide for an improved QoS, such as *IntServ*, *DiffServ*, and Resource reSerVation Protocol (RSVP), must be implemented on all routers between origination and termination, and at present these technologies are mainly limited to intranets and some routes on a small number of ISP networks.

- The prospect of these technologies being widely deployed in the Internet seems low, especially as *IntServ* and RSVP have limited scalability, and increasing convergence between data and optical layers looks likely to lead to superior alternatives in the medium term.
- Available empirical evidence suggests that QoS on the public Internet increased significantly between 1998 and 2000. At the end of this period the *average* level of delay (not counting those experienced in LAINs and in general in obtaining access to the Internet), is getting closer to a level which if sustained could support voice.

2 Economic background to Internet public policy issues

In this study we investigate whether market failure is a significant problem in regard to commercial traffic exchange on the Internet. Our analysis is therefore primarily based on economics. However, a great deal of technological material has been referred to in order to support our analysis. Some of this material can be found in the main report, although a majority of the 60 pages of annexes concerns Internet engineering issues of relevance to this study.

In general, there are three courses of market failure: externalities, market power, and existing regulations, and we discuss these in turn. In the following we assume that end-users form a different group to content providers. In practice this is not absolutely true as end-users sometimes post content, and *visa versa*. However, the adoption of this simplification by us (and others) does not invalidate our (or their) analysis.

2.1 Externalities

Externalities are a benefit (or cost) caused by an event, that are not taken into account in the decision to go ahead with the event. In the case of the Internet, there are at least three types of network externality that appear to be significant, and these may effect industry structure and the development of the Internet: (i) direct network externality benefits and costs; (ii) indirect externality benefits, and (iii) congestion externality costs.

Perhaps the main reason network effects regarding the Internet are of interest in this study is that if left entirely unregulated they would likely see the Internet drift toward a national and possibly global monopoly. The reason is based on the fact that the value of a network to subscribers depends on the number of subscribers who are connected to it, i.e. subscribers will gravitate to the largest network (which may involve several interconnected networks). Absent merger regulations, the main way monopolisation would occur is through mergers and acquisitions, perhaps supported by refusals to interconnect with a network just prior to acquisition.

The existence of seamless interconnection for all network services effectively removes network effects from being an influence in the competitive process, thus allowing competition to focus on providing other benefits to users. A possibly negative implication of the introduction of seamless interconnection in a network made of multiple interconnected networks using different software and hardware solutions provided by different technology companies, is that it may undermine to a degree the incentives that drive technological progress. As we have noted in the main conclusions of the study, *off-net* interconnection is not presently seamless and thus network effects are more likely to be important to core ISP's competitive strategies. Network effects also have implications for the optimal pricing behaviour of ISPs.

Externalities and direct network effects

Direct network effects occur when the value of membership increases with the addition of new members. Such a situation often gives rise to network externalities which implies an under-consumption of subscriptions, i.e. that subscriber numbers are less at prevailing prices than is socially optimal. Networks that are owned or controlled by a single entity ought not suffer from this problem as the owner can adjust his prices so as to increase consumption (i.e. internalise the externality). This is not the case with the Internet which is made up of over 100,000 networks most of which are independently owned. Interconnection between these networks provides an opportunity for the largest network to prevent internalisation of network externalities through facilitating less than seamless interconnection, thus providing it with a competitive advantage over smaller rivals. Countering this incentive is the fact that an increase in direct network externality benefits that are lost due to this strategy, reduces demand for all ISP services, including those of the largest network.

Externalities and indirect network effects

Indirect network effects occur when the value of a network increases with the supply of complementary services. Because Internet ownership is so fragmented, indirect network externality benefits occur with the Internet due to the lack of co-ordination of pricing between both sides of the market (i.e. end-users and content providers). In particular, the sharing of costs between web-sites which provide indirect network benefits, and end-users, does not presently take account of the relative importance of web-sites in invigorating the *virtuous circle*, suggesting the existence of indirect network externalities.

Congestion externalities

A congestion externality occurs where each user fails to factor into her own decision to use the Internet at congested periods, the delay and packet loss her usage causes other users. This tends to occur when there is no congestion price, such as will occur when sending extra packets at congested periods is free. Similarly, for the interconnection of transit traffic or interconnection between peers, the lack of congestion pricing applied to firms purchasing interconnection means that these networks do not

normally face a charge for additional packets handed over to the other network during its most congested period. This means that networks owners themselves do not have the proper economic incentive to efficiently manage congestion on their networks. We discuss the economics of congestion in more detail below in the section on quality of service.

2.2 Market power

Firms have market power when they face too little competition or imminent threat of competition. Monopolies are an extreme example, although oligopolies are much more common and they too have market power even though they face some competition from rivals. Where firms have market power output tends to be lower and prices higher than would occur in a competitive market. This translates into lost trade and reduced economic welfare compared to what would have been obtained under competitive conditions. Moreover, firms that have market power tend to practise a range of strategies aimed at enhancing and protecting their market power, and some of these strategies may give rise to significant additional welfare costs. Typically, such strategies are not illegal and include such things as lobbying the authorities, branding and loyalty cards.

Where firms have substantial market power and it persists in the long term, there will be a level of market failure. In practice merger regulations would prevent mergers and takeovers being used to monopolise the industry. But in an unregulated network, such as the Internet, interconnection would remain a possible weapon for the largest network(s) to exert market power; not a refusal to interconnect as such since this would be too blatant and risk the attention of the authorities that might then seek to regulate interconnection, but possibly through deciding to provide lower quality (e.g. congested) interconnection.

2.3 Regulation

It is not uncommon for existing regulation to enhance or create market power. The way it does this may be very simple, such as in licensing only one or two firms, or it may be much more complex involving, for example, spill-over effects running from a regulated market to an unregulated one. While Internet backbones are not directly regulated and nor is Internet traffic exchange, regulations nevertheless surround this industry and potentially have far reaching implications for its development and the development of competition in the (tele)communications sector.

3 Market failure and the Internet's addressing systems

The Internet is arranged in a loose hierarchy of entities with IP communications devices like PCs, workstations and servers (generally denoted hosts) at the outer edges. A Local Area Network (LAN) is a collection of several of these devices connected with one another. Several LANs connect to the network of a single regionally focused ISP within a specific country (i.e. a local ISP). Several regionally focused ISPs connect to a network of a national ISP, with these in turn connected to the network of an international ISP which has, however, a regional focus, and these connect to a core ISP which has multi-regional or even a worldwide focus. Importantly, core ISPs are the only ones that have virtually complete addresses in their main routers. Traffic (datagrams) tends to make its way up this hierarchy searching for an ISP that can terminate the datagrams. For communication to take place between all these networks and the devices connected to them, traffic has to be addressed, routed and switched to its destination.

Routers exchange information with other routers concerning the optimal route to send packets, and this information is stored in routing tables. A collection of routers that is under the administrative control of a single organisation forms an *Autonomous System* (AS) – also known as a *routing domain*. Networks within an AS have a common addressing and routing policy. ASes are identified by numbers which are assigned by one of the three Registries ARIN, APNIC, and RIPE.

The basic technical elements of traffic exchange between hosts are “names” and “addresses”. Broadly speaking a name is used primarily to identify a host whereas the IP address contains the way to get there. The link between the two concepts is provided by the Domain Name System (DNS).

When the IP protocol was standardised in 1981, the specification required each host attached to an IP-based network to be assigned a unique 32 bit Internet address (the IPv4 addressing scheme). At this time the addressing system rested on three classes of addresses denoted by A, B and C. The IPv4 scheme in principle offers the possibility of connecting more than 4 billion (i.e. 2^{32}) devices.

In the early days of the Internet the seemingly unlimited address space resulted in blocks of IP addresses being assigned to any organisation asking for them. However, in a class scheme involving blocks of addresses there is inherently a great deal of inefficient use of addresses i.e. addressing space is allocated which remains largely unused today.

In the beginning of the 1990's the issue of address exhaustion especially in regard to Class B address space, and the rapid increase in the size of global Internet's routing table raised concerns among the Internet community. At this time several approaches were devised in order to deal with this problem, the most important of which were Classless Inter-Domain Routing (CIDR) and private and dynamic addressing, respectively.

CIDR led to much more efficient and flexible addressing by creating the capability for address aggregation, i.e. the possibility to aggregate a contiguous block of addresses into a single routing table entry. Private and dynamic addressing has also been very important in slowing the growth in allocation of IPv4 addresses as it has resulted in huge numbers of hosts not having to be permanently assigned an IPv4 address that is visible to the Internet. The same private address can be used contemporaneously by an unlimited number of organisations, with address translation used to provide IP addresses that are valid in the global Internet. Dynamic address allocation is usually applied if a private end-user is connecting to an ISP via a dial-up connection. In this case the ISP will not normally be assigned a permanent address (although dynamic addressing may limit the users use), but rather, an address that has been allocated to the ISP and which is not being used elsewhere at that moment, will be assigned to the end user for the duration of his session. This means that there is no need for an ISP to have as many addresses as customers, as had been the previous practice.

Despite the developments discussed above, IPv4 addresses are expected to be exhausted within the next 5 to 10 years. Wide differences in estimates of IPv4 addresses required by 3G networks is one reason why predicted IPv4 exhaustion dates vary so widely. A development that began approximately concurrently with CIDR and private addressing, was the design of a new version of the IP protocol called IPv6. It has a 128-bit long address space, enough to provide for several trillion IPv6 networks. and is being promoted as the successor to IPv4.

IPv6 has a number of other advantages over IPv4 in addition to a vast increase in address space, such as those concerning security and confidentiality, the possibility of providing a large number of service priorities or grades and flexible use options. There have, however, been several recent developments that have effectively resulted in IPv4 narrowing that gap with IPv6 in regard to several of these features.

While adoption of IPv6 is still in the planning stage, several (private) intranets have been using it for some years. The Internet community at large, however, does not appear to be in a rush to prepare for the introduction of IPv6, and there are a range of possible reasons for this. Perhaps the main ones are: (i) that moving to IPv6 will be costly, especially in terms of replacement equipment and the man-hours required to make adjustments to networks; (ii) IPv4 remains adequate given the present level of Internet development; (iii) IPv4 and IPv6 networks will be able to communicate with each other (i.e. a Y2K type of problem seems very unlikely), and (iv) because Internet related technology is evolving rapidly some IP networks may judge that it is in their interests to wait and see what else evolves in the meantime, i.e. it is feasible, if very unlikely, that IPv6 may be overtaken by events.

It is possible that there is a synchronisation problem that is causing a costly delay in the preparation for adoption of IPv6 on the Internet. This *synchronisation value* would exist if widespread adoption of IPv6 brought sufficiently large advantages compared to switching costs, but as these values are not captured by early adopters, there is a “you

first” delay. Due point (ii) in the above paragraph, it is arguable whether the benefits are likely to be large enough over the next few years to warrant official intervention to push through early adoption based on an the internalisation of an externality benefit. Moreover, even if the network externality or *synchronisation value* were shown to exist, further research would be needed to show that the private benefits for networks from switching to IPv6 were not large enough for the Internet to switch to IPv6 at a rate that did not imply a costly delay, without intervention. In other words, even if there were externality benefits present, the private benefits of adoption (i.e. exclusive of the network benefit each adopter would bring) may still be large enough to bring about a sufficiently timely adoption of IPv6. We also note that any such losses would be limited in time by IPv4 number exhaustion, which will ultimately drive networks to adopt IPv6.

A closely related issue concerns the possibility that Internet development as a whole is being held back by IPv4 compared to a situation where IPv6 was already widely deployed. It is doubtful, however, even assuming such lost benefits could be shown to exist, that they represent a genuine market failure. They might more accurately be described as spill-over benefits associated with *economic multipliers*. Profitable economic activities generally have positive spill-over benefits. The authorities do not mandate profitable investments as for one thing no one knows they are profitable until after the event. In the absence of a market failure there is no divergence between private and social cost and thus no market correction exists. The example of the authorities mandating that networks in a country switch to IPv6 by a certain date, as apparently has occurred in one OECD country, seems premature.

It has also been suggested that early adoption of IPv6 would give Europe an advantage over other world regions such as the USA. Without identifying the market failure that would prevent European Internet entities from taking this decision themselves, we could not recommend an official intervention. We are not aware of anyone having set out such arguments in detail.

Other address related issues are also of potential concern. These mainly involve the possible demand for worryingly high growth rates in processing power for routers that may not be able to be met by the computing industry. The specific issues include: routing table growth; the increasing length of route advertisements, and the rate of route announcements and withdrawals. All three remain legitimate concerns, although route table growth has declined in the last year or so, and at existing growth rates poses few concerns.

Internet IP-address management is an area where there appears to be significant market failure. Early IP-address allocation in particular failed to see that numbers were a depletable resource, and there remain huge amounts of allocated IPv4 numbers that are unused and will remain so. The degree to which this market failure imposes costs on society will depend on whether the adoption IPv6 provides a large enough increase in cumulated benefits net of conversion costs, compared to the situation where IPv4 continues to be used (assuming there is no IPv4 address shortage). We are not in

possession of the information that would give us the confidence needed to attempt to forecast an outcome of this cost benefit exercise. The Internet community is in the best position to make this judgement, although we note that few networks seem to be in any hurry to make the conversion. There appears to be a little more enthusiasm among firms that sell equipment to these companies or to end-users.

To a lesser degree, there appears to be a similar market failure in regard to AS numbers where the IETF is designing an increased space into the AS header. However, we suspect that it may not be practical or ultimately efficient to bring about the sort of fundamental change to the 'management' of the Internet needed to enable better IP-address management.

4 Quality of service

The Internet backbone sits on top of electrical and optical networks that sit on top of telecommunications networks. The internet is primarily a network based on TCP/IP protocols that connects over 100,000 computer networks, many of which operate different software and hardware solutions on top of TCP/IP protocols. This is both a strength and a weakness of the Internet. It allows non standardised equipment on which operates all manner of other software, to connect to each other over the Internet. Yet, the interoperability provided is not completely seamless, and not all functionality is retained between networks.

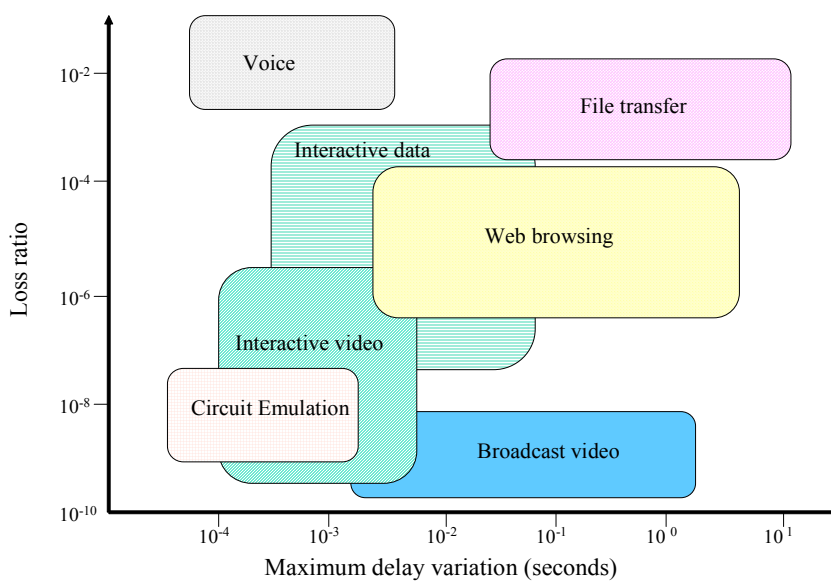
Unlike the PSTN which employs switched circuits, IP networks are based on packet switching technology, i.e. exclusive connections held open for the duration of a voice call do not exist in the case of the Internet. Rather, information is divided up into packets containing digitised information, and sent with other packets from different sources in a randomised fashion, ultimately to reach their destinations. As a rule, the reliability and quality of 'virtual' connections on the Internet falls well short of what can be provided over the PSTN. However, as well as the ability to interoperate with diverse systems, one reason for the attractiveness of packet networks compared with the PSTN, is that packet networks provide for a much greater level of flexibility for the bandwidth requirements for TCP/IP connections and greater utilisation efficiency of physical capacity. That means that the cost per bit are lower and the degree of service integration are higher compared with switched circuit networks.

QoS on the Internet is becoming more important than it has traditionally been as improved QoS is vital if convergence with other platforms (e.g. switched circuit and CATV networks) is to occur. There are a range of QoS measurements that are used to describe quality aspects of data streams on IP networks. There is a statistic that shows datagram loss (or arrival which is too late to be useful), a statistic for datagrams that have been incorrectly inserted and end up with datagrams addressed for some other 'connection'; a statistic that shows the time it takes for packets to go from sender to

receiver (known as latency) and a statistic showing the variation in latency (also known as jitter).

The range of services that can be provided over the Internet make different demands on these QoS parameters as well as on bandwidth. Figure I shows approximate ranges for datagram loss and time-to-arrive, for seven different types of end-use services. These end user services also differ according to their requirements with respect to packet loss and delay variation. Even putting aside the different demands for QoS of end-users, we can see from this figure that “one size” does *not* fit all.

Figure I: Application specific loss and delay variation QoS requirements



In recent years developments in technology have enabled IP networks to selectively manage traffic for QoS purposes. The main technologies are *IntServ*, *DiffServ*, and resource reSerVation Protocol (RSVP). While these QoS technologies are available on some intranets, and in parts of some ISP networks, they are not generally available on the Internet.

Where end-users experience QoS problems the most pervasive of them occur close to the edges of the Internet, such as bottlenecks in: the access network, LANs, end-users ISPs, and at interconnection points. However, there are also important QoS problems at the core of the Internet, the most important being caused by congestion and, importantly for the next generation Internet (NGI), a lack of seamless interoperability between networks.

One of the ways the Internet has tried to reduce quality of service (QoS) problems is through over-provisioning of capacity. For those advocating that this approach will

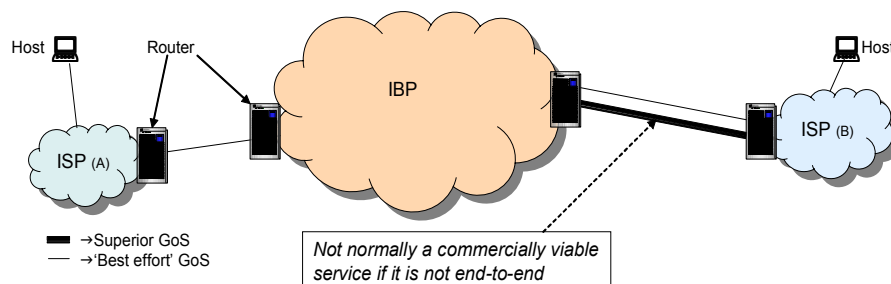
overcome QoS problems there is no need to change the structure of existing prices, as in the foreseeable future all services will receive a real-time QoS. It will not matter that the priority of treatment of packets on the Internet is according to the order of arrival, and that low priority e-mail gets the same QoS as do VoIP packets. All will be delivered in virtually real-time.

Economic theory and evidence suggests that where scarce resources are involved (i.e. investments in capacity), and users face a zero price for additional usage, congestion will occur and QoS will suffer accordingly. Supporting this prediction is the huge increase in internet traffic that is expected to occur as convergence progresses. Congestion occurs in large part because of the lack of economic signals (i.e. prices) to accurately match demand with incentives for investment in capacity. Usage based pricing can in principle be designed to shift some demand from peak periods to other times, and importantly can also signal to ISPs when demand is such as to make it economic for them to increase the capacity of their networks.

When the Internet is not congested, however, usage-based pricing is not helpful at all; it actually has a detrimental effect on economic welfare. At these times the cost of sending an additional number of packets is virtually zero. We say that the marginal cost of usage is zero, and it is a demonstrable economic axiom that under these circumstances a usage sensitive price is inefficient – it reduces economic welfare – flat rate pricing is optimal.

The end-customer is the principle source of value in the Internet and her willingness to pay depends on end-to-end QoS. If an end-user connected to ISP_(A) wants to communicate with another connected to ISP_(B) and the up-stream connectivity between the two ISPs is performed by an IBP_(C), then it is not normally commercially viable to deliver a superior QoS, say, between IBP_(C)'s network and ISP_(B) if this superior QoS is not supported by ISP_(A)'s network or the connection between ISP_(A)'s network and that of ISP_(B) (see Figure II).

Figure II: Links between ISPs and incentives to care for QoS



Some of the most important problems with end-to-end QoS on the Internet occur when traffic changes from being *on-net* to being *off-net*. The main off-net QoS problems appear to be explained by a lack of seamless interoperability between different vendor equipment. The problems concern both the network level and the management level. Network designs that are optimal for on-net traffic are often not suitable for maintaining QoS when receiving traffic from another network. Moreover, service level agreements (SLAs) offered by transit providers are all different, as are the statistical properties of ISP networks.

In summary, there are several factors presently holding back the development of the Internet into an integrated services network. These can be grouped into five overlapping categories: (i) Congestion management *on* ISP networks is not yet especially well developed, and often results in inadequate QoS for some types of service, e.g. VoIP; (ii) Although it is possible to provide superior QoS (e.g. consistent voice-grade QoS) on ISP networks, when achieved, it is often not be retained *between* ISPs due to technical reasons, such as software, and in some cases even hardware, incompatibility; (iii) There is a lack of accounting information systems able to provide the necessary measurement and billing between networks; (iv) There is no interface with end-users that enables them to select among connections with different QoS parameters, in a way that provides value to the user, and (v) The quality of access of many end-users is insufficient for congestion in the core to be presently noticed by them under most circumstances. Solutions to these QoS problems lie in technological developments informed by the economic ideas we have outlined above. We have not uncovered the sort of market failure that would be needed in order for us to recommend official involvement in fostering such technological developments. If market failure problems do materialise we think this is more likely to occur at a national / regional level, where there is weak regulation, and involve the incumbent telephone operator's control of local infrastructure.

5 Interconnection and partial substitutes

Interconnection between ISPs can be categorised into Peering and Transit.

5.1 Peering

Peering denotes a bilateral business and technical arrangement between two ISPs who agree to exchange traffic where the termination address is listed on the other ISPs network. In order for this to function they exchange (and periodically refresh) address tables and routing information. If two ISPs peer, they accept traffic from one another and from one another's customers (and thus from their customers' customers). ISPs do not carry (i.e. transit) traffic to third party ISPs under a peering contract.

When ISPs agree to peer with each other, no payment usually occurs between them. The ‘payment’ model is known as “sender-keeps-all”, or “bill-and-keep”. Of course the service is not free as each peer agrees to deliver the traffic it receives from the other. The implicit price is therefore the cost of delivering the other peer’s traffic. Peering contracts are conditional on each network perceiving that it is receiving a similar value to what it is providing. It is, therefore, not possible to put a price on interconnection between any two peers. Each such arrangement is largely about *impressions of value*. We do not think that this situation should *per se* be cause for concern by the authorities.

Until a few years ago peering was mainly associated with core ISPs (or IBPs), but it has become increasingly popular among lower level ISPs, such as between two local or two regional ISPs where it is known as *secondary peering*. Secondary peering enables regional and local ISPs to access end-users and content and application service providers’ services situated on the networks of neighbouring ISPs, without having to route traffic up through the Internet hierarchy.

Peering occurs at public or at private peering points. Public peering points are often called Network Access Points (NAPs), also known as Metropolitan Area Exchanges “MAEs” in the *United States*. NAPs are geographically dispersed and peering partners usually use “hot-potato routing” i.e. to pass traffic to each other at the earliest point of exchange. Private peering differs from public peering to the extent that it involves only two partners and occurs at a point agreed by the interconnecting network operators. Most Internet traffic is exchanged through private peering.

5.2 Transit

Transit is the means by which most ISPs obtain global connectivity. Unlike in a peering relationship, the ISP selling transit services will accept traffic that is not for termination on its network or a network of one of its customers, and will route this transit traffic to its peering partners, or will itself purchase transit from another ISP if the termination address for datagrams is not on the network of any of its peering partners. As such, a transit agreement offers connectivity to all end-users on the Internet. Datagrams with addresses not recognised by the ISP own address tables are handed up the hierarchy to a transit provider.

Prices for transit are commercially sensitive and thus confidential. However, even if we could see them we expect it would be very difficult to explain the price differences that we saw. One reason for this is that ISPs that purchase transit from an upstream ISP often also bring something of benefit to the transit provider, and to some degree the price the ISP pays for transit will reflect that benefit.

Our information suggests that there is no accepted industry model that governs the structure of transit prices. Some larger ISPs are able to negotiate a price structure with the transit provider, while others choose from a menu. There appear to be three basic

dimensions around which transit price offers are structured: (i) a fixed rate for a certain number of bits per month; (ii) a variable rate for bits in excess of this amount, and (iii) a rate based on peak throughput (which may include pipe size - representing the *option* for peak throughput), and some measure of *actual* peak throughput ('burstiness').

Transit providers offer service level agreement (SLA) regarding QoS, and they keep the statistical data necessary to verify their own QoS, and provide periodic reports to clients. Breaches of SLAs occur on occasions and so transit providers occasionally have to pay agreed compensation to their clients. It is common for ISPs to have transit contracts with several transit providers, and this is known as *multi-homing*. Analysis suggests that multi-homing, which became possible in the mid 1990s due to software and later hardware developments, has helped loosen the hierarchical structure of the Internet, and will have significantly reduced any market power that was held by core ISPs – the main transit providers.

5.3 Strategies involving interconnection

Several authors have attempted to analyse strategic incentives / opportunities that might be important in shaping the industry, and possibly giving rise to significant market failure. This research has used game theoretic models. The idea is that the key incentives that motivate actions by players should be captured in dynamic models where the players plan their own actions with some understanding of the strategies available to each other. We discuss this work in Chapter 8 of the main report.

Of the many issues that were discussed in the 1998 investigations by the European Commission into the proposed merger between MCI and WorldCom, and which re-emerged again in 2000 in the proposed Merger between WorldCom and Sprint, three stand out as having potentially the most significant implications in regard to the issues we are addressing in this section. The three are closely related, all of them concerning interconnection. The questions posed relate to whether the merged entity would control a sufficiently large proportion of the Internet that it had: (i) an incentive to degrade the quality of interconnection; (ii) an incentive to introduce proprietary standards (i.e. to differentiate on-net services from off-net services), and (iii) the ability to impose interconnection prices that are significantly above the relevant costs involved. Fundamental in finding answers to these questions will be the importance of network effects and strategies that will improve the competitive position of one or more networks vis-à-vis the others.

There are two main effects that need to be accounted for by the largest network when considering whether or not to employ seamless interconnection with rivals, or to interconnect with them at all: (i) without interconnection there will be a reduction in demand for all networks due to the loss of network benefits, and (ii) degradation will provide a competitive benefit for the largest network (or group of seamlessly connected networks).

For the largest network the effects of these two factors are mainly in conflict making the analysis of the outcome of a refusal to interconnect rather complicated to predict. Analysis suggests that a strategy of degrading or refusing interconnection by a dominant IBP *may* be profitable, leading to network tipping in favour of the dominant network. However, further analysis has suggested that it would only be profitable where the largest network already has a level market of market share well in excess of anything that has occurred to date, and even then the values for the model's other variables would often be implausible. This implies that core ISPs have incentives to overcome off-net QoS problems, and to co-operate in the area of standards, rather than trying to develop an on-net services strategy to differentiate themselves from their main competitors.

A strategy not yet examined by researchers is one that levers a competitive advantage in the provision of a particular service, in order to gain a competitive advantage in the provision of traditional Internet services. The scenario we have in mind is the following: while maintaining acceptable QoS for traditional Internet services, under what circumstances would a market leading ISP find it profitable to degrade the quality of interconnection (or not co-operate with rivals to overcome off-net problems) to prevent off-net VoIP from developing, given that traditional Internet service would continue to be provided according to a "best efforts" QoS? The largest network would then have an advantage in providing VoIP on-net, and could in the near future use the PSTN for off-net voice calls. Would the network advantage in providing voice service to its own subscribers be a telling factor in convincing existing subscribers to move from other ISPs to the market leader, and first time subscribers to choose the market leader ahead of other core ISPs? This is an issue for future research, but it is also one reason for viewing the results of theoretical research with caution. This research provides us with much insight, but the results should be viewed more as a failure to reject the behaviour predicted, rather than a confirmation that it will in practice occur.

Another branch of research has looked at the competitive effects of a looser hierarchy (as for example, has occurred through the rapid growth in secondary peering and multi-homing), on the market for backbone network services. This research suggests that the availability of secondary peering, multi-homing, and firms that move content closer to the edges of the Internet, reduces the ability of core ISPs to extract concessions from small networks during bargaining over interconnection. Indeed, increases in the proportion of ISP customers that have alternatives if one point of interconnection is degraded or cut, are shown to be similar in their effect on market power to reductions in the market share of a core ISP's network.

5.4 Interconnect price discrimination and fragmentation

Price discrimination is an area that might offer strategic benefits to the largest ISPs. In regard to traffic being passed between peers, research suggests that in a wide variety of circumstances ISPs' incentives are to price all traffic as if it originated off-net, i.e.

traffic that is originated by its own customers is charged as if it was originated on another ISP's network. In actuality, the details of price causation are buried behind peering arrangements when it comes to core ISPs, but research findings are no less valid for this. The reason behind the off-net pricing principle is that the price of interconnection governs the distribution of costs between both sides of the market – websites and end-users (senders and receivers). Higher or lower interconnection charges have a counter-balancing effect on the backbones' revenues in that they affect backbones' incoming and outgoing interconnection costs in regard to off-net traffic. If ISPs have the market power to price discriminate between on-net and off-net traffic, however, this situation is unstable and the market is likely to tip in favour of the largest network.

While the paragraph above addresses price discrimination between peers, a dominant core ISP could decide to price discriminate between transit traffic depending on whether it was terminated by the core ISP or its other transit customers, or was terminated by another core ISP or its transit customers. Were this to occur it would give rise to network externalities and network tipping. End-users and websites, and those that sign up to the Internet in the future, would be drawn to the network with which they can get the cheapest service, and where networks charge differently for on-net and off-net traffic this favours the largest network.

Perhaps counterintuitively, end-users prefer lower interconnection charges between networks and higher end-user charges than would result with a competitive outcome due to the indirect network benefits generated by websites who will be discouraged by higher interconnection charges, i.e. more content invigorates the virtuous circle of network benefits.

6 Existing regulation

While the Internet is virtually unregulated, it does not operate in a legal / regulatory vacuum. There are rules that govern the posting of information, such as copy-write and intellectual property law. Some firms that own ISPs, or firms that supply network services used by the Internet, are directly regulated, the most obvious example being incumbent telecommunications operators. More generally, firms that operate in markets close to and even overlapping Internet markets, are regulated. This raises the possibility, indeed likelihood, that firms that are competing for the same customers will not be equally treated under the law.

In the near future, changes in the process by which regulation is applied in the EU should significantly reduce the danger of competitors being covered by different laws and regulations. The approach that is making its way through the relevant European institutions at present will require any industry regulations to be targeted according to antitrust markets, and not to be placed on any firm that does not have dominance in that

market. This approach will not completely prevent regulatory non-neutrality, however, especially in industries where antitrust market boundaries are changing relatively rapidly, as appears to be happening with the convergence of different communications platforms, one of which is the Internet. One reason for this is that markets do not halt abruptly in either product or geographic space, and thus some regulatory non-neutrality is inevitable where different platforms are regulated differently.

An area where non-neutrality is already present concerns universal service subsidies, or more particular the way special universal service taxes (USO contributions) are raised from industry participants. As convergence occurs platforms other than the PSTN will be competing in the same market to provide voice services. We expect that voice services will be increasingly provided over the Internet in competition with fixed line switched circuit networks, and potentially also in competition with cellular networks. Clearly, when one type of firm pays extra taxes of the form that effect its marginal costs, and which its competitors do not pay, there would be a breach in neutrality principles.

We suggest that changes in the details of some aspects of regulation will be needed in the medium term. In addition to those that face universal service tax liabilities, regulations that govern interconnection pricing units appear to be another candidate for change. Where voice services can be provided over the Internet or the PSTN (or conversation could use parts of both networks), per minute PSTN interconnection charges may well prove to be distortionary as PSTN interconnection costs are capacity based, and while converting these to per minute charges has proved useful to date, per minute charges are an artificial construct, and an approximately similar unit should be used for both types of network to avoid regulatory arbitrage.
