CERN-CN/96/18

Prospects for Internet Technology

Brian E. Carpenter
CERN, CH-1211 Geneva 23
Prof. Jon Crowcroft
University College London

March 1997

Abstract

This paper surveys the current developments in Internet technology, with a particular emphasis on performance, and the growing need for various guarantees of quality of service. It discusses hardware technologies for increased bandwidth, mechanisms for requesting and providing specific qualities of service, and various scaling issues. Finally it discusses mechanisms needed for (but not the economics of) the Internet in the mass market. To this end, we survey changes in the areas of addressing, and flow management.

1 Introduction

In the last few years the Internet has grown from being a useful tool for the international research and development community to its current status as a platform for national and international business of all types, including consumer access. While this development is exciting, and even gratifying to the pioneers and early adopters of Internet technology, it stretches the technology beyond its design limits and original principles [1].

The predecessor of the Internet, the ARPANET, began in 1969 as a research project with some military antecedents. At that time computers were slow (a few microseconds per instruction in many cases) and telecommunications lines were even slower (maybe 2400 baud) and lossy. It was desired that the ARPANET should be able to survive line outages of widely varying durations, i.e. robustness and automatic re-routing were vital. In a context of slow computers and slow, unreliable lines the survivability of network transactions, rather than their timeliness, became the main design goal. It was in this context that Baran [2] had invented packet switching and the connectionless datagram approach became the conventional wisdom [3].

As in any datagram network, end-to-end functions in the traditional Internet can best be realised by end-to-end protocols. The end-to-end argument is discussed in depth in [4]. The basic argument is that, as a first principle, certain required end-to-end functions can only be performed correctly by the end-systems themselves. A specific case is that any network, however carefully designed, will be subject to failures of transmission at some statistically determined rate. The best way to cope with this is to accept it, and give responsibility for the integrity of communication to the end systems. Another specific case is end-to-end security.

To quote from [4], “The function in question can completely and correctly be implemented only with the knowledge and help of the application standing at the end points of the communication system. Therefore, providing that questioned function as a feature of the communication system itself is not possible. (Sometimes an incomplete version of the function provided by the communication system may be useful as a performance enhancement.)”

From this starting point the Internet grew around its traditional applications. In their current state, these are known as ftp (file transfer protocol), smtp (simple mail transfer protocol), and nntp (network news transfer protocol). It will be noted that none of these applications are truly interactive in nature. Their poor relation was always telnet, the remote login application. This normally worked well on local-area networks but was, and is, notoriously unsuitable for wide-area network use over slow or congested lines with long response times.

Many other applications are of course in use, particularly on LANs. However, in recent years two related classes of application have appeared that seriously threaten the traditional Internet paradigm.

- The World-Wide Web [5] which not only needs a sophisticated window-based user interface, but also demands quick and reliable real-time response from arbitrary locations in the network. Indeed, because the Web was developed with little reference to constraints of the network infrastructure, its pattern of a multitude of small, urgent transport connections has placed a major strain on the Internet.
- Real time audio and video services evidently require adequate bandwidth, and bounded transit delay and delay jitter. Additionally, multi-party conferencing requires that these properties be delivered to dynamically configured sets of end points, each set being joined together as a multicast group.

Both the Web and audio/video have the characteristics of needing guaranteed quality of service (QoS) to satisfy their users, and of having enormous mass-market potential. Thus it is clear that continued expansion of the Internet, and perhaps even its survival, depend on its ability to deliver predictable QoS at a predictable and modest cost to the mass market.

2 Why QoS challenges the traditional paradigm

In traditional (legacy) telephony networks, signalling and call control are part of a monolithic system that provides resource control at a number of points on a path. IBM’s Systems Network Architecture was originally similar, with its relatively rigid path control mechanisms, and the need for cumbersome system generation procedures to manage the evolution of the network configuration. In more recent work in the telecommunications community, such monolithic systems have been divided into a number of separate modules, and at least call control (user to user) and QoS signalling (user to network) are distinct functions. Both signalling and call control have an implicit companion in the network, an admission control decision. This module or function allows new traffic into the network based on some knowledge of resources and current utilisation (whether local or global is a separate question).

In the Internet Architecture, call control is replaced by the presence of global source and destination addresses in every packet. This means that each packet "makes its own call" as it progresses.

However, in the Internet Protocol, there is no place for QoS. In fact, it is hard to envisage where in a datagram protocol to place a request for a facility which requires control over timescales longer than those of a single packet's transmission and propagation delay. Basically, the admission control decision must have a possible outcome of "no" to mean anything. Note that "no" means complete failure for a normal IP packet, whereas it may simply delay call completion in a circuit network. However, signalling for QoS can be either explicit or implicit. In the first case, a protocol can be used in band or out of band to inform the network of quality of service requirements for a sequence of one or more packets (known as a "flow"). In the second case, a well-defined set of standard fields are present, like the "call control" fields, in every packet, to indicate delay, error and throughput requirements of the flow to which it belongs. In both cases, the network uses these requirements to determine queueing (e.g. selection of priority queues according to varying load and routing topology conditions), and action in case of temporary congestion. The decision whether to delay or drop a packet from a given flow depends on its delay or loss tolerance compared with other packets. Under long term congestion, flows of packets must be sorted into those which meet their required performance goals and those which do not. Only some policy can decide this (e.g. monetary cost). Chosen flows can be discarded at source, or in intermediate nodes. A 100% discard of packets from some particular flow is this approach's counterpart to the "no" in an admission decision.[1]

All of this requires additional baggage for the Internet, although how much is very much open to question, and we look at this in later sections under separate headings.

3 Technical trends in response to the challenge

The Internet industry is nothing if not responsive to technical challenges. The responses under development fall into four major categories:

- technology to deliver higher bandwidth at lower cost
- technology to guarantee QoS over this bandwidth
- technology to help the Internet scale up by a further two or three orders of magnitude
- technology to make Internet products more suitable for the mass market

3.1 Higher bandwidth at lower cost

The simplest way to improve quality of service is to "throw bandwidth at the problem." In other words, by direct analogy with the road system, if the road is congested, widen it. It should be noted that this is not a panacea; as in the case of the road system, the effect of increased bandwidth in one part of the network may well only be to move the congestion elsewhere. Nevertheless, increased bandwidth is vital as the number of users and the demands of each user increase, especially as these two effects multiply together. It is also necessary that high bandwidth should cost little more money than low bandwidth, if a true mass market is to develop. Fortunately, there is no technical reason why this should not be the case, at least for urban areas and inter-urban trunks.

[1] One might ask why the packets should be sent at all in this case. Most applications are bi-directional, and in the absence of a response in these circumstances, they, or their user will "give up", just as a person gives up trying to complete a telephone call when there is no answer to the ringing tone.

3.1.1 Domestic and small business users

For Internet service to domestic and small business customers, the speeds offered by conventional analogue modems or by basic-rate ISDN are scarcely adequate today and will soon be left behind by increasingly fast PCs and more sophisticated applications. There are two obvious terrestrial media for delivering higher bandwidth bi-directional access to a huge percentage of homes and businesses: the telephone cable and the television cable.

Over the telephone system, normally providing a single twisted pair with rather poor transmission characteristics, it is already possible to deliver above 1 Mbit/s through many existing lines, using ADSL technology (asymmetric digital subscriber line) [6]. This is adequate for Internet use up to and including low frame rate video, even though the return path from the subscriber is much slower. VDSL (very high speed DSL) [7] will be several times faster but its scope of applicability will be less. Since ADSL and VDSL are both subject to quite short distance limits (5.5 and 1.5 km respectively) they are in any case limited to use in urban areas, without costly repeaters.

The television cable system normally provides a co-axial cable shared by several hundred households, with a single repeater connected to an upstream feeder cable or optical fibre. Such a system offers at least 30 Mbit/s of data capacity to be shared among the households, by using three or more TV channels to transmit data. Data modems conforming to the future IEEE 802.14 standard [8] will provide Ethernet service to paying households. Internet service can easily be provided over this infrastructure, by using conventional Ethernet-based routers. However, it is clear that if more than a few households simultaneously activate demanding Internet applications, then many more than three TV channels will have to be devoted to this usage.

For households which do not have telephone circuits good enough to run VDSL, and which do not have a television cable connection, i.e. typically households outside urban areas, it is not yet clear how high-speed Internet access can be provided. Wireless technology and radio spectrum allowing several Mbit/s access over long distances to many locations simultaneously is not in view at reasonable cost. Even the ambitious Teledesic low orbit satellite system is only expected to provide a global total of 20,000 circuits at 1.5 Mbit/s.

Digital satellite delivery might seem ideal for much existing Internet traffic, however there are currently some problems with the protocols over asymmetric links, particularly with implementation bugs in TCP, and with some routing systems. When these are overcome, we must include these as a third media for mass delivery of non time-critical (interactive) applications. The performance regime will otherwise be quite similar typically, to ADSL.

3.1.2 Large users and trunks

Large business users, universities, etc. already have high speed Internet access, albeit at excessive cost in many countries, using leased lines at up to 45 Mbit/s (or even 155 Mbit/s) to connect to the Internet service providers. Similarly, such users have internal Internet services running at any speed attainable with LAN technology (up to 800 Mbit/s in some cases, with 1 Gbit/s Ethernet under development [9]). Many different physical technologies are in use, with a strong trend towards increasing use of switched media rather than shared media for LANs. The advantage of a switched medium is that, in general, a transmission by a third party will never block the transmission between two systems. It should be noted, however, that the general use of client-server applications such as the Web, and the growing importance of multicast applications such as video-conferencing, both create traffic patterns that counteract the advantages of switching technology, and favour shared media networks. In the case of multiparty applications, the IP multicast service maps very well onto a shared transmission capacity; in the case of client server applications, the server is often a hot spot, and a symmetric deployment of i/o capacity between all clients and servers is wasted.

For local area networks, many different LAN switching technologies are available at a cost per port which is less than that of shared-medium Ethernet a few years ago. Unfortunately few of these technologies handle IP multicast groups effectively, with the result that (for example) all packets for a particular video-conference are liable to be delivered to every workstation on a LAN, although only one workstation wishes to receive them.

A particularly challenging switching technology, for both LAN and WAN use, is Asynchronous Transfer Mode (ATM) [10], offering flexible sharing of connections up to at least 622 Mbit/s. Since this is a virtual circuit technology, it is not well matched to the Internet paradigm. There is great complexity in adapting ATM for Internet use [11, 41]. A full discussion of this is outside the scope of the current paper, but we refer later to current work on integrating the Internet's provisions for QoS to those of ATM. There is an interesting interplay between routing

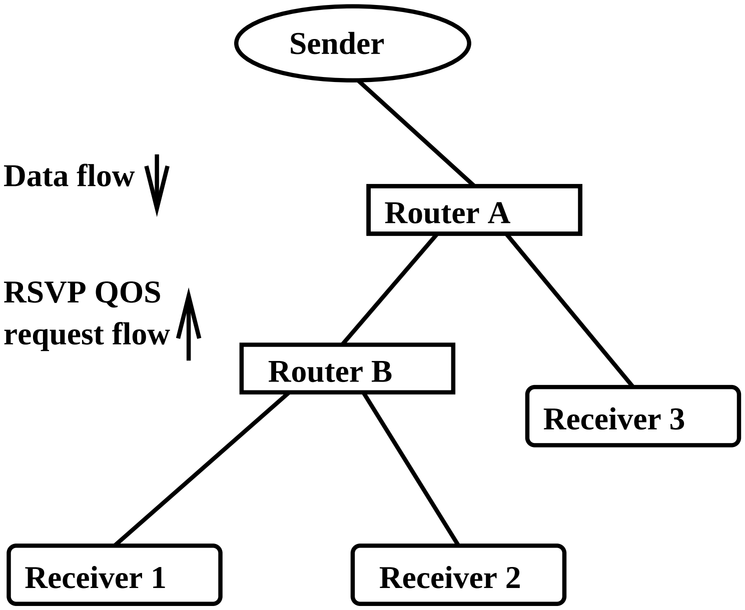

Figure 1: RSVP merge points

and ATM switching, since if a flow of IP packets can be identified, its component ATM cells could in theory be switched much more efficiently than the complete IP packets can be routed. Products already exist that attempt to identify packet flows automatically and switch them at the ATM cell level [12] and further work on cell-switching routers is to be expected. Note however that cell switches have even more difficulty handling multicast than do packet switches.

For inter-urban or international trunk connections, Internet service providers will increasingly use ATM or SDH/SONET connections at bandwidths ranging from 155 Mbit/s to 2.4 Gbit/s or more. There are significant challenges in routing IP packets at the highest bit rates [13] and in efficiently handling thousands of simultaneous transport connections (see Section 3.3.3).

However, it can be seen that for large users able to invest in advanced technology, there is no fundamental difficulty in scaling up the available bandwidth to any reasonable extent. For the remainder of this paper, we assume that bandwidth is not a technological limit.

3.2 Quality of Service

3.2.1 QoS Reservation Requests

As mentioned in Section 2, the Internet has met its enemy and its name is QoS. A large amount of effort has gone into adding an explicit signalling protocol to the suite of Internet management protocols (ICMP, SNMP, ARP, and Routing and so on), to address this perceived shortcoming, and the Resource reSerVation Protocol, RSVP [14] has been specified and largely implemented, although it has seen little deployment yet. RSVP allows receivers to send QoS requests upstream towards the sender, and RSVP-capable routers merge these requests appropriately (see Fig. 1).

RSVP is not enough to meet a QoS requirement: it is also necessary to specify a set of traffic classes (in the ATM community these are known as bearer service classes) which define the user settable traffic performance parameter sets and ranges. In the Internet community, these are known as the Integrated Services models; currently, Best-Effort (i.e. standard IP), Controlled-Load and Guaranteed-Service are the three specified service classes [15]. These define the parameters that are used to police traffic, in terms of a token bucket; they also define the treatment of traffic in a flow that exceeds its contracted (RSVP signalled) parameters. The way these service classes may be selected is by

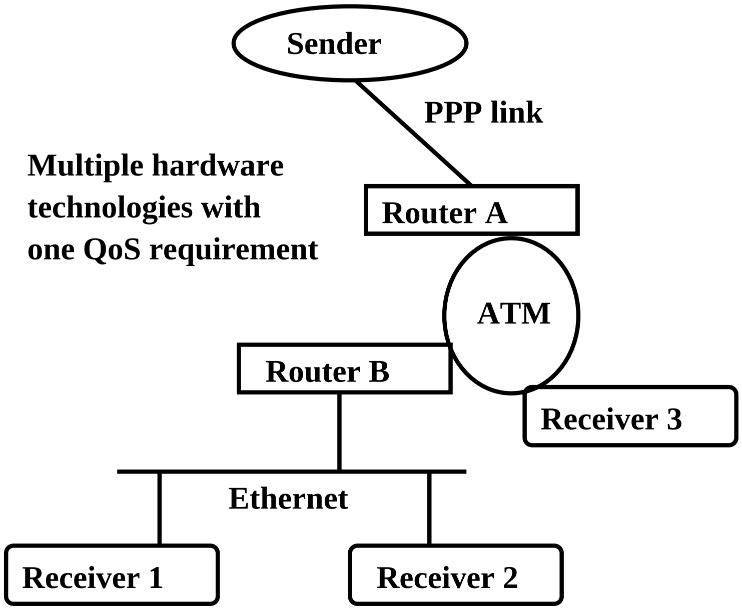

Figure 2: QoS over multiple link types

using RSVP to request a service class with a particular generalised source port (protocol, source address, ports etc), to be given a particular service with a particular flow specification (throughput, delay bounding and so on depending on service class). Note that the QoS request is issued by the receiver, which also issues the request to join a multicast group. This is important if multicast groups with thousands of members are to be possible, since the sender cannot possibly be aware of the QoS capabilities of all potential receivers.

Of course, QoS parameters could be set (even without RSVP) by the network manager of a collection of routers, so long as all routers along the path are under the control of that network manager, but this solution does not scale to massive networks.

For each traffic class in turn, there will be a defined mapping onto the link types that might join one router to another (see Fig. 2). In some cases, this definition may be simple- for example, the guaranteed service cannot be done with any degree of probability over an Ethernet; or the controlled load service can easily be done on a Token Ring subnet. There is a special working group of the IETF known as “Integrated Services over Specific Link Layers” which is defining these mappings. In some cases there are too many - for example, in recent work [16], at least eleven plausible combinations of Internet QoS classes with ATM QoS classes have been identified:

Guaranteed-Service could be provided over CBR or rtVBR.

Controlled-Load could be provided over CBR, rtVBR, nrtVBR or ABR

Best-Effort could be provided over CBR, rtVBR, nrtVBR, ABR or UBR

More recently, some of the Internet Providers have been looking at providing premium services in addition to standard (Best-Effort) IP service. Essentially, this would create a two level QoS system only. The model is like that of first and second class rail or air travel. Note that the “number of seats” is simply indicated by the amount you pay, and the “quality” by the price per seat. Taking this analogy back into the networking arena, there is something to be said for a simple QoS model where sources simply indicate delay tolerance, and send at the required rate - admission control is done by measurement, and feedback is done through pricing.

3.2.2 QoS Support

Requesting a specified QoS is not enough to enforce it. Two further components are needed:

1. Admission, and policing
2. QoS Queueing

Each of the traffic service classes that have been defined in the Integrated Services models has an associated admission test. At each stage (router hop) of a path/flow construction, requested by RSVP, a router carries out this test to ascertain, given the existing resources, whether sufficient remain to allow this additional utlization. There is a large body of theory behind the tests for admission, much stemming from Parekh's work[31].

Provided the traffic flow stays within its requested bounds (remembering that RSVP also allows re-negotiation), all is well. This implies that routers must monitor the traffic, and mark or discard excess traffic in times of overload (or, anytime, if the tariff structure calls for it!).

To provide deterministic delays, it is essential to use a more systematic approach to queueing in each router than has typically been used: in the past, many routers simply implemented FIFO queueing, which introduces highly variable delays. Several queueing disciplines serve traffic in a way that gives predictable, or even exact delay and delay variation performance, the simplest (at least to understand) being a round-robin, or Fair Queueing mechanism. Weighted Fair Queueing simply refers to a system that gives more than one go per round to important customers. This, and other equivalent schemes (at least equivalent in terms of being able to match request, admission, policing, and resource all together, if not in terms of utilisation), is being deployed by large ISPs today, even before they deploy RSVP. This is so that they can at least offer discernable QoS differences (traffic class differentation) through network management and configuration immediately. For example, it allows them to offer a ”premium service” to some customers, perhaps simply based on preferential queueing based on their IP network addresses.

One compelxity not yet touched on is the interaction between traffic management, and route management. We look at this next.

3.2.3 QoS routing

Traditional networks have been designed with very good knowledge of traffic distribution statistics. It is possible to do off-line link placement and dimensioning (albeit an expensive process) when *a priori* knowledge of the sources and sinks (the traffic matrix) is available.

The Internet has shown that in fact, enabling sources to “spring up” at will at new places in the network is good for the Information Economy. This has led to a far more richly (and some would say randomly) interconnected wide area mesh with seemingly random traffic patterns. In this mesh, hot spots come and go at quite a high rate, but link capacity deployment runs at a somewhat slower rate.

This means that there is often saturation on normal paths and spare capacity on alternate paths involving several or even many hops (see Fig. 3), not simply through dual redundant links between important nodes, as is often used in mission critical Intranets. At the time of writing, no Internet standard routing protocols support routing over these alternate paths on the basis of resource utilisation, although early research results show that it may be worth exploiting [17].

The key problem is that there is an inherent contradiction between the architecture of IP distributed dynamic routing which is based on statelessness (absence of prior knowledge) about flows in the network, and the requirement to allocate resources in the network based on signalled QoS requirements from those flows via, say, RSVP.

To get around this problem, routes associated with traffic flows with established QoS are currently “pinned” (locked in place, and removed from the set of routes that are dynamically updated). This has the obvious side effect that link outage causes reservations to be simply lost, but tends to reduce the risk of unstable routes.

Proposals for selecting initial routes based on current conditions rely on the ability to pin routes, but they go further by removing the flexibility of IP routing. Recent research [18] on “call” re-routing in ATM networks seems to point the way to possible future systems where flows can be re-routed around trouble spots or new hot spots, but there are few concrete results yet.

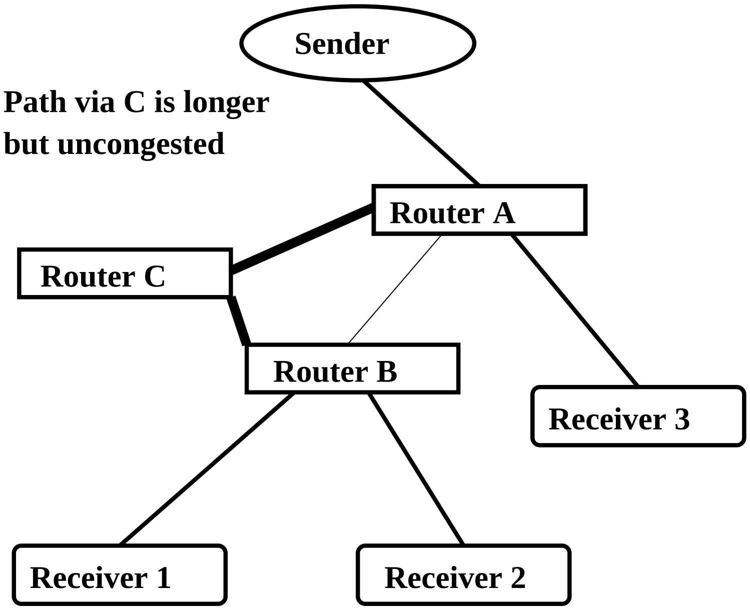

Figure 3: The need for QoS routing

3.2.4 QoS Scaling

One area of future research is in the scalability of the Internet QoS mechanisms. There has already been some work on layering flows - for example the virtual path concept in ATM represents one approach, and the wild-card filter mechanism for allocating a reservation to a group of senders in a many-to-many audio session in the integrated service Internet model represents another.

These are both *aggregation* techniques. As well as route and address aggregation, which we discuss below, there seems to be a need for flow aggregation, simply to reduce the amount of state (memory) required in routers to implement the mechanisms described above. How this shall come about is a matter for future work.

3.3 Scaling

The growth in the power of personal computers and workstations continues to demand more performance per user from the Internet. At the same time the number of users is growing rapidly. The product of these two growth rates means that various major Internet service providers have consistently reported traffic growth at a rate of about 100% every six months.

3.3.1 Scaling the current routing and addressing system

The first scaling problem to draw attention, in about 1992, was the growth of pressure on the addressing and routing system. It was noted [19] that if nothing changed, the Internet would start to run out of address space by 1995. The reason was that, from the beginning of the Internet, large blocks of address space had been allocated freely on a first-come, first-served basis. Essentially this meant that sites could be profligate with the use of addresses. It is unknown what percentage of allocated addresses are in actual use, but it is well under 10%. It also meant that, since address blocks were allocated chronologically, there was a random relationship between a site's address and its topological location in the Internet. Thus, if N different sites were all connected to the same Internet service provider's router, that router was obliged to announce N different routes to the rest of the Internet. Extrapolations showed that with continued growth, this would inevitably lead to routing tables and routing traffic throughout the Internet expanding beyond all reasonable bounds.

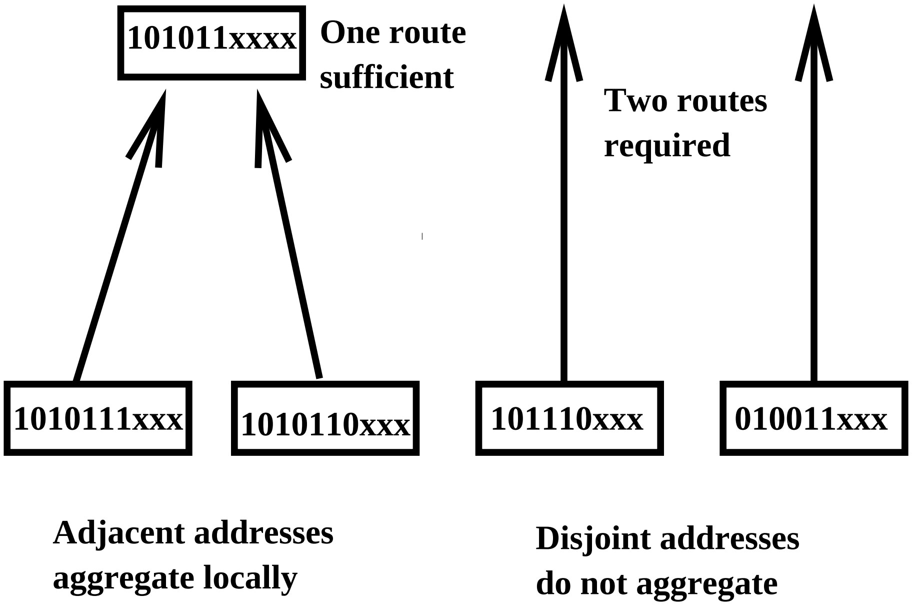

Figure 4: Schematic route aggregation

Three linked measures have been implemented to contain this problem. Firstly, the original policy of allocating address blocks only in units of 2^8, 2^{16} or 2^{24} addresses was changed to one of allocating address blocks of any size 2^n, while trying to keep n as small as possible [20]. Secondly, address blocks are now in general allocated to Internet service providers, not directly to user sites. This allows service providers to allocate contiguous address blocks to N users, such that one aggregated route can be announced for these N users [21]. Thirdly, a new routing protocol, Border Gateway Protocol 4 [22] was introduced that allows service providers and major sites to announce routes to address blocks of any reasonable size. Thus the three measures together allow for the introduction of topologically-based addressing and the resultant aggregation of routing announcements (see Fig. 4).

The combination of these three measures, and a firm policy from the Internet address registries, has succeeded in stabilising both the rate of address allocation and the rate of growth of routing tables, since 1993 when they were introduced. However, this is not enough. A defect of IP compared to some other network protocols is that it was not designed to allow for convenient automatic address allocation (often known as “plug and play”). Indeed, if the IP address of a computer has to be changed, this is normally an awkward, error-prone manual process. Yet to fully overcome the problem of scaling the addressing system, the original randomly allocated address blocks all need to be replaced by hopefully smaller topologically allocated blocks. Also, whenever a site changes from one service provider to another, it needs to change to an address block provided by its new provider.

Thus, automatic renumbering of TCP/IP systems is an important requirement for the future [23]. An alternative approach is the emergence of dynamic network address translators (NATs) and of application level gateways (ALGs). Corporate networks, also known as Intranets, may legitimately re-use a subset of the IPv4 address space, forming multiple routing realms. At the boundary between two (or more) routing realms, we may find devices in the spectrum of possibilities between ALGs and NATs.

At one end of the spectrum is a non-transparent Application Layer Gateway (ALG). Such a device acts as a termination point for the application layer data stream, and is visible to an end-user. A transparent ALG differs by making communication through the ALG transparent to an end user. Both non-transparent and transparent ALGs are required (by definition) to understand the syntax and semantics of the application data stream. ALGs are very simple from the viewpoint of network layer architecture, since they appear as Internet hosts in each realm, i.e. they act as origination and termination points for communication.

At the other end of the spectrum is a Network Address Translator (NAT) [40]. Since the NAT modifies address(es) in the IP header, the NAT also has to modify any transport layer information (such as checksums) that depend on addresses. By definition a NAT does not understand syntax and semantics of an application data stream. Therefore, a NAT cannot support applications that carry IP addresses at the application layer. Because of encryption, both ALGs and NATs are likely to force a boundary between two distinct IP Security domains, both for authentication and for confidentiality.

3.3.2 Replacing the routing and addressing system

The current version of the IP protocol is officially designated version 4 [24] and its addresses are rigidly limited to a length of 32 bits. Ultimately, this will not be enough. For this and many other reasons, a new version of IP, officially designated version 6, has been defined [25, 26] This expands the address space to 128 bits, adds a flow identifier to the packet header, and makes automatic renumbering a standard feature. Thus it is intended that IPv6 should both overcome the scaling problems of the IPv4 routing and addressing system, and assist the provision of specified quality of service for identified traffic flows.

One of the benefits of introducing a new addressing structure is that the new numbering scheme can allow for a more rational assignment of addresses. Two possible approaches that are popular are provider based addressing, and geographic based addressing. Both rely on hierarchy to achieve reduction of routing table sizes. Essentially, as described above, variable length prefixes are now used to look up a destination and find the output port from a router, using a “longest match” procedure.

Having said this, it is quite possible that there will continue to be large numbers of providers (and of complex policies concerning routes). This means that while local routing tables may often be much smaller in the future than they are now, they may not be in some parts of the network.

Another approach to dealing with routing hierarchies has been to derive the hierarchy from the topology, rather than from the provision or demography. This is done to some extent in two level routing hierarchies such as OSPF [27] and IS-IS [28] provide. However, a generalization of this through the exchange of topological maps is being worked on in the NIMROD working group of the IETF [29].

The same requirement exists for multicast routing - the Internet provides multicast delivery now in many places, but the cost in memory can be considerable. Work on next versions of PIM [37] and Inter-domain multicast routing is addressing these shortcomings.

3.3.3 Scaling transport capacity

Existing Internet applications typically use TCP [39], or perhaps RTP/UDP [38], to communicate. As the network gets larger and faster, these transport protocols are encountering a variety of problems of scale.

TCP adapts to conditions in the network, but does so without any explicit feedback from the network itself, just by monitoring its own perceived end-to-end behaviour. This works reasonably well whilst most TCP connections are relatively long lived. Unfortunately, as the network gets faster, the so called pipesize (bandwidth$*$delay product) of a typical path becomes such that it is often possible in principle to send all the data for an entire session before waiting for a single acknowledgement from the far end. In practice, TCP employs congestion avoidance schemes that do not permit this. To this end, three important scaling changes are being put in place in TCP (and its use)

- Window scaling - permits the use of large enough windows to exploit high bandwidth delay paths.
- Selective acknowledgement - permits a TCP to “keep going” rather than stall in the face of small percentage packet loss.
- Persistent HTTP - a client accessing the same server for a sequence of requests continues to use the same TCP connection, rather than closing and opening it (this has been done for some time by TCP based RPC and transaction processing systems, but is only now being introduced in the World-Wide Web).

These all go some way to mitigating the problems of millions of users fetching the same Web pages every day repeatedly from the 170 Gigabytes of HTML files estimated to be in the WWW at the time of writing, using a typical 11 packet exchange on average to do so!

Other changes in the pipeline include Transaction TCP, which conveys connection state between different TCP connections from a source to the same sink, and semi-reliable multicast transport protocols, such as the Scalable Reliable Multicast protocol employed by the Mbone whiteboard tool. [42]

TCP achieves a reasonable degree of fairness so long as all sources conform. However, in a mix of traffic containing non-TCP applications that adapt rapidly (e.g. receiver-driven multicast trees) and ones that adapt more slowly (e.g sender-driven adaptive multicast such as in IVS), fairness is impossible to ensure. Nevertheless, the network can give quite a bit of assistance, even without a QoS signalling system. Two approaches are:

- RED - Random Early Detection is a low cost, highly effective way to police traffic flows in the network. [30]
- WFQ - As discussed in 3.2.2 Weighted Fair Queueing may be used with RED to give different shares to different kinds of traffic. [31]

3.3.4 Multicast control

Multicast applications in the Internet have shown the most promise in terms of scalability. The receiver-initiated expansion and pruning of multicast trees effectively decentralise, and therefore resolve, scaling issues in multicast group management. In the late 1980s, the "killer application" was often seen to be multi-media conferencing; killer, in the sense of destroying the network by overloading it. In practice, the scaling of the Mbone (Multicast backbone) tools has meant quite the reverse. Today, while using around 11 percent of the Internet capacity at most, thousands of simultaneous participants watch a variety of sessions at conferences, supervise remote classes, watch satellite launches and so forth.

Two aspects of multicast scaling are illustrated well by the control protocols that are being standardised. The Session Directory is an application used to create sessions, for which the distributed directory itself makes use of multicast to carry out dynamic allocation of multicast addresses for a session, and to avoid clashes with other sessions' address allocations. The Session Directory then uses multicast to inform efficiently any interested party on the Internet of all multicast sessions in existence.

A session invitation protocol can make use of multicast to provide user location independent invitations (user-to-user call control protocol). Session control protocols can use multicast to carry out activities such as floor control (selection/suppression of speakers) and speaker activity in multicast conferences, by multicasting such information to all interested participants. This illustrates how the receiver-driven approach that is inherent in Internet multicast allows large new collaborative tools to be built.

3.3.5 Operations

All of the above developments tackle essentially technical aspects of scaling. Additionally, network operators are confronted with the need to scale up and make fully professional network operations that often started as a part-time hobby. Even so, it is still the case that even major Internet service providers depend on a very few brilliant individuals to keep their extremely complex networks running, often only by the installation of the very latest test versions of router software.

3.4 Mass Market

The Internet has now grown to the point where as many as 10% of the population have some form of access to it in certain developed countries. Several technical developments currently under way should open the door to a significantly larger fraction of the population.

Firstly, the "plug and play" aspect of Internet technology will improve. Today, most Internet access providers offer their customers software packages for the popular home computer environments, and the common mechanism for domestic access, PPP, requires little manual configuration. As mentioned above, it is a design goal of IPv6 that configuration should be as automatic as possible, and this is essential for the mass market. In IPv6, there are two options for auto-configuration. In "stateless" auto-configuration [32], a system that is starting up seeks to discover a local router which then informs it of the appropriate address prefix. The system then forms a complete 128-bit IPv6 address by concatenating that prefix and its local hardware address (typically a 48-bit Ethernet address). In “stateful” auto-configuration [33], the system subsequently seeks a configuration server that assigns it a definitive address. When IPv6 is fully deployed some years from now, the days of manual network configuration for Internet users will hopefully be over.

Secondly, there is active work on access server protocols. An access server is a device that subscribers dial, typically through a modem or ISDN today, and possibly through ADSL in future. The normal protocol between the subscriber and the server is PPP. However, a user may not always dial his or her “home” access server, for example while travelling with a portable PC. This raises multiple new requirements: authentication of the user (to prevent fraud or dissimulation); confidentiality of communication; remote billing procedures; and how the user’s traffic is tunnelled back to the “home” service provider. Work is in progress on standards for all these aspects, although only authentication [34] is at an advanced stage.

Another approach to this problem is Mobile IP [35]. In this model, there is no access server, but the user connects a mobile system to the Internet by any convenient means at any convenient point (by wire or by wireless means as the case may be). The mobile system then seeks an IP “foreign agent”, i.e. a nearby router willing to redirect traffic to or from the mobile. The “foreign agent” then identifies its peer at the mobile’s home site, naturally known as the “home agent”. The two agents collaborate to ensure that traffic addressed to the mobile reaches it, using an encapsulation technique.

It should be noted that these two approaches to the connection of portables have no particular performance limit and are independent of any particular type of hardware connection. In this respect they are superior to the currently available wireless solution, use of PPP over a modem attached to a mobile telephone. This solution is clumsy, slow (normally limited to 9600 baud) and relatively expensive, especially if it involves an inter-continental mobile telephone call.

3.4.1 Mass Market Security Concerns

Authentication and confidentiality were mentioned above for portable systems, but in fact they are fundamental requirements for the mass market Internet. If the network is to be used for private financial transactions, or any kind of private business, then the sender must be unambiguously authenticated to prevent fraud, and the message must be kept confidential, i.e. encrypted. Furthermore, since anybody who intercepts an Internet message must be assumed to have access to a powerful computer, the cryptographic mechanisms used must be strong ones. Standards for the authentication and encryption of IP packets have been defined [36] and are mandatory in IPv6 implementations. Unfortunately their deployment has been held up by technical difficulties with safe distribution of cryptographic keys, and more seriously by political restraints on the use and export of cryptographic technology - an unfortunate hangover from the Cold War.

4 Conclusion

In this paper we have concentrated on issues concerning the infrastructure of the Internet, up to the network and transport layer. Many new developments are to be expected in the applications built on this infrastructure, as it becomes more widespread, faster, more reliable, and cheaper. At the border between the infrastructure and the applications will lie mechanisms to measure the use of services, especially premium services delivering specified quality of service. Unless simple ways are found of making such measurements and using them to send bills to users are found, Internet service providers will not be able to match the quality of service they offer to the true needs of their customers. It is clear that the customers themselves may not know their needs. Recent promising work on measurement based signaling and admission control points the way to some part of the solution to the problem of getting users to specify their needs by neatly sidestepping it. In general, though, this is an open research topic.

The Internet has always evolved and will continue to evolve. The technology trends noted above suggest that within a few years, the Internet will offer the substrate for a wide range of applications, demanding a stable quality of service from the infrastructure, yet largely independent of the particular transmission mechanisms used. These applications will be limited only by the imagination of their creators.

5 Acknowledgements

Yakov Rekhter contributed to the text on NATs and ALGs. A few sentences are derived from [1]. Helpful comments were made by David O. Williams and Julian Bunn, and the anonymous referees.

6 References

Note: Internet "Request for Comment" (RFC) documents are available on-line, for example via http://ds.internic.net/ds/dspg1intdoc.html.

[1] B. Carpenter (ed.), Architectural Principles of the Internet, RFC 1958, 1996.

[2] P. Baran, On Distributed Communications Networks, IEEE Trans on Communications Systems, COM-12, 1-9, 1964.

[3] D.D. Clark, The Design Philosophy of the DARPA Internet Protocols, Proc SIGCOMM 88, ACM CCR Vol 18, Number 4, August 1988, pages 106-114 (reprinted in ACM CCR Vol 25, Number 1, January 1995, pages 102-111).

[4] J.H. Saltzer, D.P.Reed, D.D.Clark, End-To-End Arguments in System Design, ACM TOCS, Vol 2, Number 4, November 1984, pp 277-288.

[5] T.J. Berners-Lee, R. Cailliau, J-F. Groff, B. Pollermann, World-Wide Web: the Information Universe, Electronic Networking Vol. 2, 1992, pp. 52-58.

[6] Telecommunications - Network and Customer Installation Interfaces - Asymmetric Digital Subscriber Line (ADSL) Metallic Interface, ANSI T1.413-1995

[7] Tutorial: VDSL: Fiber-Copper Access to the Information Highway, ADSL Forum, 1996 (available at http://www.sbexpos.com/sbexpos

[8] Standard Protocol for Cable-TV Based Broadband Communication Network, IEEE Project 802.14, work in progress, 1996.

[9] Gigabit Ethernet, IEEE Project 802.3z, work in progress, 1996.

[10] M. de Prycker, Asynchronous Transfer Mode: Solution for Broadband ISDN, 3rd edition, Prentice Hall Communications, 1995

[11] R. Cole, D. Shur, C. Villamizar, IP over ATM: A Framework Document, RFC 1932, 1996

[12] P. Newman, W. Edwards, R. Hinden, E. Hoffman, F. Liaw, T. G. Minshall, Transmission of Flow Labelled IPv4 on ATM Data Links Ipsilon Version 1.0, RFC 1954, 1996

[13] A. Tantawy, M. Zitterbart, Multiprocessing in High-Performance IP Routers, Protocols for High-Speed Networks, III (Proc. IFIP 6.1/6.4 Workshop) Stockholm, 13-15 May 1992, Elsevier, B. Pehrson, P. Gunningberg, and S. Pink (ed.)

[14] Zhang, L., Braden, R., Estrin, D., Herzog, S., and S.Jamin, Resource ReSerVation Protocol (RSVP), IEEE Network, 1993.

[15] R. Braden, D. Clark, S. Shenker, Integrated Services in the Internet Architecture: an Overview., RFC 1633, 1994

[16] Mark Garrett and Marty Borden, Interoperation of Controlled-Load and Guaranteed-Service with ATM, proceedings of the IETF, Montreal, June 1996.

[17] QoS Routing BOF, proceedings of the IETF, Montreal, June 1996.

[18] Z.Wang, J.Crowcroft, Quality of Service Routing, to appear in JSAC.

[19] P. Gross, P. Almquist, IESG Deliberations on Routing and Addressing, RFC 1380, 1992.

[20] V. Fuller, T. Li, J. Yu, K. Varadhan, Classless Inter-Domain Routing (CIDR): an Address Assignment and

Aggregation Strategy, RFC 1519, 1993

[21] K. Hubbard, M. Kosters, D. Conrad, D. Karrenberg, J.Postel, Internet registry IP allocation guidelines, work in progress, July 1996

[22] Y. Rekhter, T. Li, A Border Gateway Protocol 4 (BGP-4), RFC 1771, 1995

[23] B. Carpenter, Y. Rekhter, Renumbering Needs Work, RFC 1900, 1996

[24] J. Postel, Internet Protocol, RFC 791, 1981

[25] S. Deering, R. Hinden, Internet Protocol, Version 6 (IPv6) Specification, RFC 1883, 1996

[26] Christian Huitema, IPv6: The New Internet Protocol, Prentice-Hall, ISBN: 0-13-241936-X, 1995

[27] J. Moy, OSPF Version 2, RFC 1583, 1994

[28] Intermediate system to intermediate system intra-domain-routeing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode Network Service (ISO 8473), ISO 10589, 1992.

[29] I. Castineyra, J. N. Chiappa, M. Steenstrup, The Nimrod Routing Architecture, work in progress, February 1996

[30] Floyd, S., and Jacobson, V. Random Early Detection gateways for Congestion Avoidance: Part 1, Part 2, Part 3, Part 4, Part 5. IEEE/ACM Transact Networking, V.1 N.4, August 1993, p. 397-413

[31] Abhey Parekh, Generalized Processor Sharing, PhD Thesis, MIT Laboratory for Decision Systems, MIT 1993.

[32] S. Thompson, T. Narten, IPv6 Stateless Address Autoconfiguration, work in progress, December 1995.

[33] J. Bound, Dynamic Host Configuration Protocol for IPv6 (DHCPv6), work in progress, February 1996.

[34] C. Rigney, A. Rubens, W. A. Simpson, S. Willens, Remote Authentication Dial In User Service (RADIUS), work in progress, July 1996

[35] C. Perkins, editor, IP Mobility Support, work in progress, May 1996

[36] R. Atkinson, Security Architecture for the Internet Protocol, RFC 1825, 1995

[37] S.Deering, D.Estrin, D.Farinacci, V.Jacobson, C-G.Liu, L.Wei, An Architecture for Wide Area Multicast Routing, ACM SIGCOMM 1994, London October 1994, ACM CCR Vol 24, No. 4, 126-135

[38] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, RTP: A Transport Protocol for Real-Time Applications, RFC 1889, 1996

[39] J. Postel, Transmission Control Protocol, RFC 793, 1981

[40] P. Francis, K. Egevang, The IP Network Address Translator (NAT), RFC 1631, 1994

[41] J.P.G. Sterbenz, Protocols for High Speed Networks: Life After ATM?, Protocols for High Speed Networks, G. Neufeld and M. Ito, editors, Chapman and Hall, London, pp. 3-18, 1995.

[42] S. Floyd, V. Jacobson, S. McCanne, C-G. Liu, L. Zhang, “A Reliable Multicast Framework for Light-weight Sessions and Application Level Framing”, Proc ACM SIGCOMM 1995, Cambridge, Mass.