Network Turbochargers: 100M Technology for the Brunel Network

Andrew Findlay

Head of Networking and Systems
Computer Centre
Brunel University

January 1996

Andrew.Findlay@brunel.ac.uk

ABSTRACT

The current topology and state of the Brunel Network is reviewed, performance limits are identified, and actions are proposed to progressively increase the performance of the network. Attention is focused on the Uxbridge site, where the heaviest network use occurs.

1. Background

The Brunel Network serves around 3000 end-systems, spread across four main sites. It is primarily an IP network, with heavy use of routing (there are currently 47 active subnets). Most of the network is based on Ethernet, with some inter-site serial links and one FDDI ring connecting the main routers at Uxbridge.

2. Terminology

This paper makes reference to the bottom three layers of the ISO reference model for networking. These are:

Layer 1:
: The physical layer, concerned with cables, fibres, signal levels, data rates, and line encodings.

Layer 2:
: The datalink layer, concerned with low-level packet framing, checksums, and local addressing (e.g. Ethernet hardware addresses). This layer also deals with medium access control (MAC) in shared-medium networks like ethernet, so it is often referred to as the MAC layer.

Layer 3:
: The network layer, concerned with global addressing and routing of packets across diverse media from source to destination.

Layers 4 to 7 build on the services of the lower layers, and are implemented entirely in the end-systems. The network infrastructure is not normally aware of these higher layers.

3. Current Topology

The network at Uxbridge is logically two stars, each with a Cisco router at the centre. The spokes radiating from each router to the client buildings are fibres carrying Ethernet to multi-port repeaters or hubs. Resilience is provided by Sun fileservers with multiple ethernet interfaces, which can route IP packets if the main routers should be unreachable.

The topology at the other sites is a simpler version of the same plan: see the network diagrams for an overview of the current network. Router and server capacity at Osterley and Twickenham has been substantially upgraded as part of the Common Computing Project. Runnymede needs a similar upgrade in the near future.

Within each client building, distribution to end-systems is mainly by thin-wire ethernet (10-Base-2). There are usually several segments, joined by the multi-port repeaters. In some buildings there are as many as three eight-port repeaters serving a single network.

In a few recently-wired areas, distribution is by UTP structured cabling. In these cases the multi-port repeaters are replaced by 10-Base-T hubs, usually of the 'stackable' variety.

It is common to have 100 or more end-systems and several fileservers connected to each subnet. In teaching areas the number of end-systems is usually lower due to the high peak loads imposed by classroom machines. In office areas the number may be rather higher. Our subnetting scheme does not allow more than 254 nodes per subnet, but we aim to divide a subnet long before it reaches that figure.

There are many buildings with very few computers. These buildings are grouped onto a few 'catch-all' subnets by linking the fibres to multi-port fibre repeaters at the centre of the star.

4. Limitations

The thinwire ethernet wiring is a major problem. It looks simple, so people take liberties with it. Because a single cable is shared by several machines (often a whole row of offices), a single fault or deliberate disconnection can inconvenience many people. Further, it is impossible to run this wiring above 10Mbit/s.

Many subnets become very busy during peak hours. The result is poor performance, often affecting users on other subnets because they need data from a server on the overloaded net.

The problems can get a lot worse when filesystem backups are running. There are now over 40 machines that are backed up by Computer Centre, with a total of 140GB of data at the time of writing. The backup traffic is usually several megabits per second from each server to its tapedrive host. There have been cases of three such datastreams passing across a single ethernet segment at once. Such a load almost always brings other applications to a halt. Filesystem use (and hence backup requirement) approximately doubles each year.

Almost all the available router ports have been used at Uxbridge, and Runnymede has a severe shortage as well. This prevents us from splitting up overloaded nets and also means that several buildings must often share one subnet, which makes fault location difficult.

Brunel's SuperJANET connection is capable of delivering data at 25Mbit/s yet a fully-loaded ethernet handles less than 10Mbit/s. In practice this has not been a limitation so far because few applications need high speeds to single machines, but the requirements are growing.

5. Technologies

When the current network was designed, ethernet was the prime high-speed networking technology. Traffic isolation by subnetting was the only way to increase aggregate bandwidth. FDDI was expensive so it was only appropriate for inter-router links.

The situation today has changed in several important ways:

Ethernet is still the prime technology, but there is now a 100Mbit/s version with hardware available at less than double the cost of 10Mbit/s hardware. UTP Category 5 structured cabling systems are available that will carry well over 100Mbit/s on relatively low-cost cable.

Several types of multi-port bridge are now available. These are usually referred to as 'switches' by analogy with telephone switches (telephone exchanges are usually called switches in the trade). Most switches are layer-2 devices so they do not give such good traffic isolation as routers, but they are often very much cheaper so it is more appropriate to think of switches as replacements for hubs or repeaters. The advantage of ethernet switching is that it allows the *aggregate* bandwidth of a subnet to be increased without changing any of the end-systems in any way. Traffic flows do not impinge on systems that are not part of them so the shared-bandwidth ethernet is translated into something more like a dedicated-bandwidth star network. Some types of 10Mbit/s ether switch have additional 100Mbit/s ports for linking to fast servers or routers.

FDDI is cheaper than it was, and there is now a copper-based version, but it is still much more expensive than 100-Base-T ethernet.

A new technology called ATM (Asynchronous Transfer Mode) has appeared and is catching most of the popular attention at present. It has great potential for the future, but current products are relatively expensive and not all the technical issues have been resolved yet. Any ATM equipment bought in the near future should be regarded as having a relatively short service life.

Routers are still a vital part of any large network. Price/Performance has improved, and there are also 'level 3 switches' which provide most router functions using lower-cost technology.

6. Proposed Topology

The stars-with-crosslinks topology has served us well and remains a good model for the future. The main improvements required are:

(1) Prevent heavy traffic flows (particularly from backups) from impacting normal user performance.

(2) Make 100Mbit/s connections available where they could be useful.

(3) Provide more router ports to allow more subnets to be used.

To achieve these aims I propose to start installing small ethernet switches beside (or in place of) the multi-port repeaters and hubs in the client buildings. This will be of most benefit where there are local fileservers and where the building wiring is structured UTP. A pilot installation of this type on the Computer Centre staff network has been very successful, eliminating the frequent network overloads caused by backup traffic.

The ethernet switches will all have 100Mbit/s fibre connections in addition to their large number of 10Mbit/s local connections. These 100Mbit/s links will be connected through the inter-building fibre to Computer Centre.

A new switch/router will be installed in Computer Centre. It will have both 10Mbit/s and 100Mbit/s ethernet ports, which can be grouped in any combination to provide both switching and routing for a number of subnets. Servers located in Computer Centre will connect directly to this box in many cases. A similar unit will be installed in Tower A to serve the second star.

Traffic will pass freely between the existing network and the new overlay because they will be joined at many points. Attention will have to be paid to routing policies to make sure that traffic does not pass through more client networks than necessary.

Refer to the attached diagrams showing a portion of the existing network and a stage in its future evolution. It can be seen that equipment can be introduced gradually without disrupting the service, and that some benefit will be gained from each new item of equipment even if the building wiring will not support its full capacity.

A new site-boundary router will be needed within the next two years, to cope with increasing inter-site traffic and the coming transition to IPng. These issues must also be taken into account when selecting new equipment for Runnymede. The Osterley router is a current model so it should be upgradable when the time comes.

7. Choice of equipment

7.1. Switches

A wide range of switches is available, with port-counts from six to a potential of over 100 per unit. Cost per port waries widely.

Most current ether switches are capable of doing a useful job in our network, but very few manufacturers have considered the impact of multi-media traffic. The special features of this traffic are that it is usually constant-bit-rate (often over 1Mbit/s) and is commonly multicast to serve several users at once. The simpler switches treat multicast packets in the same way as broadcast: every packet is copied to every port. This means that one or two multi-media sessions can have a serious impact on all users of a given switch. Smarter devices exist that take part in the group membership protocol so they need only forward the multicast traffic to those end-systems that need it. Unfortunately, this feature usually comes as part of a large and expensive package. I have made detailed enquiries of several manufacturers, but have not yet located any who plan to put intelligent multicast handling on the cheaper boxes in the near future.

As the need for progress is urgent, I propose to start by using simple low-cost ether-switches. When smarter devices become affordable at a later date, we can redeploy the simpler units to parts of the network with lower multi-media demands. Even a simple ether-switch is much better than the existing equipment.

The most suitable device at present is the 3Com LinkSwitch 1000, which provides 24 10Mbit/s ether ports and two 100Mbit/s ether ports for a cost of about 2000 pounds per box.

Possible alternatives include several from the Cisco Catalyst range. The 1700 is low-cost (about 2500 pounds for 25 10-base-T and two 100-base-T) but is only suitable where an entire subnet can be converted to switching, as it only supports a single MAC address per port. The 2100 costs more (about 3500 for the same configuration) but can handle multiple addresses per port, so its ability to handle up to four separate groups of hosts would allow it to support a small amount of switching on several nearby subnets. For large installations, the Catalyst 5000 chassis-based switch may be more appropriate, as it can accept up to five cards each supporting 24 10-base-T or 12 100-base-T or 12 100-base-FX interfaces.

7.2. Routers

At Uxbridge in particular we need routers with very many ports. The two existing star-point routers have 18 and 12 ethernet ports respectively. The new units need to be capable of at least three times this port capacity at a reasonable price.

The most likely way to achieve the high port densities required is to use level-3 switches. The Cisco Catalyst 5000 is an attractive option, but the routing module is not yet available and the form it will take is still uncertain. Of the existing options, the most attractive is the Alantec Powerhub. An appropriate initial configuration would provide six 100Mbit/s ports and 16 10Mbit/s ports, leaving room to expand in the same chassis. This would cost around 22000 pounds, with further groups of six 100Mbit/s ports costing around 11000 pounds.

8. Wiring

As mentioned above, the old distribution wiring is a significant barrier to improved performance. In 1993 I submitted estimates for the progressive rewiring of all buildings to a modern standard. It turns out that these estimates were far too low, but even so the work was not funded. I believe the University simply reaffirmed its policy that Computer Centre's responsibility for network wiring ended at the 'front door' of each building. Departments are still expected to install and maintain the in-building wiring. The result of this strategy has been a continuation of the piecemeal approach to wiring, and (worse) no planning for development or replacement.

At its meeting on 7th February 1996, the Computer Services User Group resolved to press the University to make network wiring a central responsibility, on a par with electricity and plumbing. This will require a policy decision at a high level, with adequate funding to implement a wiring programme across all four sites.

Because of the central importance of the wiring infrastructure, I propose to write to all Heads of Departments explaining the limitations of the current system and advising them to budget for re-wiring programmes over the next few years.

9. Priorities, and how we get there from here

Easing the bottlenecks on the 10Mbit/s network must come first. There are at least six subnets that could benefit immediately from the installation of ether switches. They would benefit even more if we added some UTP-to-thinwire media converters so that the switches could replace the multiport repeaters.

Installing these switches would release some hubs and repeaters which could be usefully redeployed into lower-traffic subnets and used in expansion projects. Indeed, a policy of buying switches and redeploying hubs from now on would cost very little more than simply buying new hubs as the network expands.

Once there are a few subnets with switches in place, the new star router will be able to make a useful contribution by switching more traffic at 100Mbit/s.

I currently believe that the purchase of a new site boundary router can be delayed, but that depends on a number of factors which are not all within our control.

Document version: 1.5 96/03/12
/usr/src/local/DOC/NETWORK/plan-for-100m/plan-for-100m.ms