Routing in a Bridged Network

by
Joachim Martillo
Chief Scientist
Clearpoint Research Corporation
35 Parkwood Drive
Hopkinton, MA 01748

Voice: (800) 800-7898x7466
(508) 435-7466
(617) 298-4107
FAX: (508) 435-7504
E-MAIL: martillo@azea.clearpoint.com

I. Introduction

Stand-alone networking devices which support both routing and intelligent bridging among LANs are increasing their presence and importance in corporate, research and manufacturing networks. In fact, just as more hackers are acquiring their own LANs, hackers are likely to start acquiring, hacking on or building their own stand-alone networking devices which both route and bridge data packets. Using, hacking on or building such devices requires understanding how routing and bridging functionalities can or should interact.

In many cases like DEC Network Architecture (DNA) and in the case of DOD Internetwork Architecture in particular, the coexistence of routing[1] and bridging functionality in a single hybrid bridge router device is conceptually problematic. Implementing DOD IP router (IP gateway)[2] functionality in a device which performs P802.1d Spanning Tree Protocol (STP) MAC bridging requires especial care.

The IEEE document, *P802.1d MAC Bridges*, makes the following important observation on p. 13.

> IEEE 802 Local Area Networks of all types may be connected together with MAC Bridges. Each individual LAN has its own independent MAC. The Bridged Local Area Network created allows the interconnection of stations attached to separate LANS as if they were attached to a single LAN. A MAC Bridge operates below the MAC Service Boundary, and is transparent to protocols operating above this boundary, in the Logical Link Control Sublayer or Network Layer [1].

Essentially and for almost all intents and purposes, a collection of Ethernets connected by such a MAC bridge is a single Ethernet.

---

[1] DECNET Digital Network Architecture (DNA) terminology is different from DOD Internet Architecture terminology. A DOD IP router corresponds to a DECNET Level 2 Router.

[2] DARPA Internet literature typically calls IP routers gateways while OSI literature typically reserves the term gateway for protocol converter.

Douglas E. Comer, who is a well-respected implementer of computer networks which follow DARPA Internet architectural principles, elucidates the issue of bridged Ethernets in *Internetworking with TCP/IP, Volume I, Principles, Protocols, and Architecture, Second Edition*. On p. 28, he gives the view of bridged networks from the standpoint of DARPA Internet-style networking.

> Because the connection among physical cables provided by bridges and repeaters is transparent to machines using the Ethernet, we think of bridged Ethernets as a single physical network system.

More precisely, a collection or mesh of Ethernet or other LAN technology segments interconnected via MAC bridges comprise an intelligent communications subnet to which end hosts may be attached via the usual Ethernet interfaces. The MAC bridges as special purpose communications subnet processors, perform the same function for their communications subnet as ARPANET Packet Switch Nodes (PSNs in the general case, formerly called IMPs in ARPANET-specific terminology) perform for the ARPANET communications subnet.

Terminology

Address Resolution Protocol (ARP). A DOD Internet protocol used on some communications technologies to find a physical address which corresponds to a given IP address. This determination of corresponding physical address is called "address resolution." Usually, ARP procedures are implemented on physical technologies which support broadcast. RFC 826 defines generalized ARP procedures for Ethernet. ARP is not strictly necessary for Ethernet. Rather than using ARP to resolve an IP address to a physical address, the Ethernet driver could simply send IP packets to the broadcast address when the destination physical address is unknown. The Ethernet driver could learn the IP address/physical address mapping from the IP source address and Ethernet source address of received IP packets. After broadcasting an IP packet when the destination physical address is unknown, the transmitting host might receive ICMP redirects from IP routers attached to the same IP subnetwork as the transmitting host, but the ICMP redirects are harmless and could serve as a means for the transmitting host to learn the IP and physical addresses of IP routers attached to the same IP subnetwork to which the transmitting host attaches. Of course, there may be security problems with broadcasting IP packets, and if the destination host never sends a reply, the source host would never learn a physical address to use in association with the destination IP address. In such a case, the source host would always broadcast IP packets to that destination. Such broadcasting, especially if frequent and if the IP packets are large would be rather unneighborly. Even if these specific IP broadcasting problems did not exist, separating address resolution from IP packet transmission would still represent a reasonable partitioning of the problems in using the Ethernet medium for the carriage of IP traffic.

ARPANET. A packet switching computer network developed under auspices of the US Defense Advanced Research Projects Administration (DARPA). The ARPANET first went on line in December, 1969. In current jargon, this type of network is a "switching-fabric" network. The ARPANET IP subnetwork was a central backbone communications subnet within the DARPA Internet until 1983 when it was split into two backbone communications subnets, the ARPANET and MILNET. The ARPANET served the research community while MILNET has served the military community. Since 1990 the ARPANET has been supplanted by faster backbone communications subnets.

Bridgelet. A logical subbridge in Clearpoint parlance.

Catenet. A concatenation of separate networks. The individuals networks need not have the same type.

Communications Subnet. A network technology concept in contrast to IP subnetwork or DECNET area which are network architecture concepts. A communications subnet transports data units among hosts. Finding paths between hosts is the responsibility of the communications subnet. A communications subnet consisting of a single Ethernet coaxial cable provides an extremely trivial form of path selection. All hosts hear all packets. Communications subnets built with packet switches like IEEE P802.1d MAC bridges or ARPANET IMPs provide somewhat more sophisticated path selection. An internet like the DARPA Internet is a second order communications subnet because it provides host to host path selection by routing packets among IP subnetworks, which themselves correspond to communications subnetworks which select host-to-host paths. IP routers are, in a sense, second order packet switches. A DECNET area corresponds to a single communications subnet. A DECNET Level 1 router cooperatively routes packets among hosts within the communications subnet using a distributed version of the ARPANET *decentralized* routing algorithm which is based on Ford and Fulkerson's network routing algorithm. *Flows in Networks* by L. R. Ford, Jr. and by D. R. Fulkerson describes this algorithm. A DECNET which comprises multiple areas constitutes a second order communications subnet, and the DECNET Level 2 routers use the same routing algorithm to route between areas as Level 1 Routers use to route among hosts. DECNET Level 2 Routers like IP routers act as second order packet switches.

DECNET. A computer network which has been implemented according to the Digital Network Architecture and which uses the communications protocols of this network architecture.

Digital Network Architecture (DNA). An evolving layered computer network architecture specified by Digital Equipment Corporation (DEC). DNA provides end-to-end network virtual circuit service to applications via routing datagrams at lower protocol layers. The DEC publication, *The DECnet DIGITAL Network Architecture (Phase IV) General Description* (Order No. AA-K759B-TK), contains a DNA overview.

DOD Internet Architecture. The ex post facto articulated architecture of the DARPA Internet. The DOD Internet Architecture provides end-to-end virtual circuit and datagram transport service within a large catenet. The core of the DOD Internet Architecture lies in the IP layer and the IP router. The IP layer conceals the details of the individual physical networks. The IP encapsulation provides to end hosts the appearance of a single global virtualized network, which carries packets between the end hosts. The IP routers perform the actual routing of packets within the Internet by routing packets among IP subnetworks. The DOD Internet architecture says little about the selection by a communications subnet of paths between hosts within a single communications subnet which corresponds to an IP subnetwork. The DOD Internet architecture leaves the application interface mostly unspecified, but many implementations use a Berkeley socket interface or an AT&T streams interface. *The Elements of Networking Style* by Michael Padlipsky contains a flippant introduction to the DOD Internet Architecture.

Internet. A computer network composed of multiple physical networks which may be a different technology types and which appears to end hosts as a single virtualized global network. The DARPA Internet is the most prominent example of an internet.

Internetwork Control Message Protocol (ICMP). A protocol, which is a necessary, integral and obligatory part of IP. ICMP carries control and error messages.

Interface Message Processor (IMP). A packet switch designed the ARPANET communications subnet. IMPs connect either to end hosts or other IMPs. Current jargon designates IMPs as packet switch nodes (PSNs).

IP Address. A 32 bit number used to address hosts within the DARPA Internet. Hosts may be either end hosts or routers. Hosts have at least one unique IP address. The DARPA Internet also supports broadcast and logical group addresses. Host addresses contain a network ID and host ID. The leading bits of the network ID specify the network class which in turn determines the number of bits in the network and host IDs. In the early days of DARPA Internet, IP networks typically corresponded to a single communications subnetwork. Net 10 was the class A network ID of the ARPANET.

IP Subnet Mask. An IP subnet mask defines IP subnetworks within an IP subnetwork, where an IP subnetwork is either an IP network of one of the defined classes or an IP subnetwork. The bitwise "and" of host IP address and IP subnet mask identifies the IP subnetwork to which an IP host attaches.

IP Subnetwork. A hack which enables the division of a single IP network among several communications subnets. The number of possible host IDs on a Class A or a Class B network is much larger than the number of hosts which a single LAN-based communications subnet typically interconnects. The number of possible host IDs on a Class C network may be smaller than the number of hosts to be connected to single LAN based communications subnet. Also using a new Class C network ID for each LAN-based communications subnet can be administratively inconvenient. Dividing an IP network (usually Class B) into IP subnetworks represents one possible solution. IP routers which support IP subnetworks route between IP subnetworks and determine the IP subnetwork to which a host address corresponds by "anding" bitwise the host IP address with configured IP subnet masks. If an IP router knows a path to the destination IP subnetwork, the IP router forwards the packet appropriately. IP routers external to a given IP network need not know the internal IP subnetwork structure to route packets to a host within an IP network which has been divided into subnetworks. An external IP router need only know a route to a border router which can route the packets to subnetworks within the destination IP network. An IP subnetwork itself may also be subdivided into smaller subnetworks about which IP routers external to this subdivided subnetwork need not know.

LAT. A DEC local area asynchronous terminal concentration protocol for Ethernet.

Logical Link Control (LLC). A sublayer within the OSI link layer of IEEE 802 LANs. This sublayer is the medium independent part of the OSI link layer. LLC provides connectionless and connection-oriented services to the network layer. The ANSI/IEEE 802.2 (ISO/DIS 8802/2) standard, *Logical Link Control,* defines LLC protocol and services for the 802 LAN technologies.

Logical Subbridge (LSB). A Clearpoint Constellation Series hybrid bridge router concept. The network administrator allocates bridge ports to different logical subbridges. The hybrid bridge router bridges frames between ports allocated to the same logical subbridge. The router functionality routes packets between logical subbridges to which the router functionality has a logical interface. Alternatively, a Constellation hybrid bridge router network-layer interface to a communications subnet built from LAN segments can be described as a software-defined embedded bridge.

MAC Bridge. A device for LAN segment interconnection which operates beneath the MAC/LLC interface. The MAC layer of each segment is independent, but the operation of the MAC bridge is completely transparent to end hosts. The IEEE *P802.1d MAC Bridges* draft proposes a transparent bridge architecture for IEEE 802.3 (Ethernet), 802.4 (Token Bus) and 802.5 (Token Ring) LANs. This document uses MAC bridge to refer to P802.1d MAC bridges. P802.1d MAC bridges use the Spanning Tree Protocol (STP) to find unique paths between hosts. Source Routing (SR) and Source Routing Transparent (SRT) are alternative techniques for routing frames between end hosts over LAN segments. From the standpoint of routing between bridged networks, the same analysis applies to SR bridges and SRT bridges as applies to P802.1d STP MAC bridges.

Medium Access Control (MAC). LAN link layer sublayer which arbitrates access to the medium, which defines medium addressing and which accepts frames from the physical layer and passes them up to the logical link layer. The IEEE 802.3, 802.4, and 802.5 standards define Ethernet, Token Bus and Token Ring MAC layers respectively.

Multiply Connected. A concept from graph theory. A graph is multiply connected if there exist nodes within the graph between which there is no path (viz diagram 1).

Packet Switch Node (PSN). Currently, the official jargon for a packet switching computer within the ARPANET. This article uses PSN generically for any packet switching device within a communications subnet. Communication computer, node, Intermediate System (IS), packet switch and data switching exchange are equivalent terms.

Packet Switching. A mode of operation for a communications network. A packet switching network routes data packets on the basis of an identifier (e.g. destination address or virtual circuit ID) in the data packet. In most cases, an end host receives only packets which it was meant to receive. Circuit switching is an alternative mode of operation for a communications network. In the context of telephony, a distinction can be made between Packet Switching, Packet Relay, Frame Switching and Frame Relay. From the standpoint of computer networking, these techniques are simply variations of one another.

Packet Filtering. A reception mode for end stations connected to a broadcast medium like most LANs. The receiver discards (e.g. Ethernet) or forwards unchanged (Token Ring) packets not meant for the receiver. Promiscuous reception is another end station mode. Monitors and bridges generally receive promiscuously.

Port. Physical interface from the standpoint of the MAC layer according to P802.1d. Several MAC layer physical interfaces may as a group correspond to a single interface at the network layer.

Request for Comments (RFC). Primary documentation for the DARPA Internet. RFCs comprise reports of work and research, various proposals, current DARPA Internet protocol standards and some unclassifiable essays.

Root Bridge. Highest priority bridge in a simply connected network of P802.1d MAC bridges. According to the P802.1d path costs calculated relative to this bridge are used to determine which ports are to be *blocked*.

Simply Connected. A concept from graph theory. A graph is simply connected if there exist no nodes within the graph between which there is no path (viz diagram 2).

Spanning Tree Protocol (STP). A protocol specified in the IEEE *P802.1d MAC Bridges* document. A physical network of bridges connect by LAN segments can be mapped to a mathematically equivalent graph to which concepts of graph theory apply. In mapping a physical bridged network to a graph, busses (either internal bridge busses or LAN segments) correspond to graph nodes while bridge ports correspond to arcs (line segments) between nodes (viz diagram 3). Subgraphs of a simply connected graph which consist of all graph nodes and a subset of arcs such that the subgraph has one unique path between any two graph nodes is a spanning tree for the original graph. If arcs have assigned weights, a spanning tree subgraph with minimum sum of arc weights is a minimum spanning tree subgraph. According to the P802.1d spanning tree procedure, after the determination of the root bridge in the network, a distributed procedure finds a unique minimum spanning tree for the corresponding graph on the basis of minimum path costs from the root bridge. Those ports whose associated arcs are absent in the minimum spanning tree are *blocked* and traffic is not forwarded either to or from these ports. The Spanning Tree Protocol specifies the Protocol Data Units (PDUs) which carry the spanning tree information between bridges. Algorithms based on various forms of spanning tree procedures most commonly serve as tools for solving the subnet design problem when unique paths between hosts provide sufficient network reliability. The application of spanning tree procedures (in conjunction with a source location learning procedure and a frame filtering procedure) provides a clever solution to the problem of selecting paths between hosts for frames to transit within a communications subnet which consists of MAC bridges acting as PSNs and LAN segments acting as connecting links between PSNs.

An intelligent communications subnet with communications subnet processors providing the subnet's intelligence is a typical characteristic of Wide Area Networks (WANs) while LANs typically have extremely simple communications subnets like passive coaxial cables. Intelligent bridging technology blurs the distinction between WANs and LANs by creating an intelligent communications subnet from component LANS and MAC bridges acting like PSNs. In the DOD Internet Architecture, a single communications subnet typically corresponds to a single IP subnetwork.

A communications subnet in most networking architectures counts as a single physical network. Unfortunately, in the case of bridged LANs where individual component pieces of the intelligent communications subnet are themselves individual physical (local area) networks, this terminology, albeit unavoidable in the literature, can become rather confusing.

Considering bridged LANs as a single physical network (or -- with less ambiguity -- as a single communications subnet) is an important point because of the internet interconnection principle, which Comer states on p. 55.

> In a TCP/IP internet, computers called gateways provide all interconnections among physical networks.

Comer elaborates this idea of interconnection via routers on p. 110.

> To understand IP routing completely, we must go back and look at the architecture of a TCP/IP internet. First, recall that an internet is composed of multiple physical networks interconnected by computers called *gateways*. Each gateway has direct connections to two or more networks. Unlike a gateway, a host usually connects directly to one physical network.

Minimally, an IP router should have interfaces on at least two physical networks (i.e. communications subnets) while a MAC bridge connects separate physical networks so that from the standpoint of the DARPA internet architecture the bridged LANs constitute a single physical network. In some sense, IP routers and P802.1d MAC bridges work at cross purposes. If a single device bridges among a set of interfaces and simultaneously performs IP routing among the same set of interfaces, that device is actually routing packets from a single physical network back onto that single network. Even though as a bridge this device has multiple physical network interfaces, the physical network interfaces are all interfaces to the same communications subnet, and as a router this device has but one logical network interface onto a single IP subnetwork.

To be fair, nothing in DARPA specifications forbids the degenerate case where the physical networks, to which the IP router connects, could be the same, but implementing an hybrid bridge router which can only route from a communications subnet back onto the same communications subnet is conceptually unsatisfactory and defeats the purpose of implementing a router. These considerations imply that implementing a useful hybrid bridge router reasonably and correctly requires serious thought and careful consideration of the essential natures of routers and bridges.

II. IP Routers

In DARPA internet-style computer networking, routers serve the purpose of finding paths between IP subnetworks. When a computer host needs to transmit an IP packet to another computer host over the internet, the source host determines from the source and destination IP addresses and according to the source host IP subnet mask whether the source and destination IP interfaces attach to the same IP subnetwork. If the IP subnetworks are the same, the source host can presumably transmit the IP packet to the destination host directly without the mediation of a router.

If the IP subnetworks are not the same, the source host transmits the IP packet to an appropriate router which is attached to the same IP subnetwork as the source host interface. Routing the IP packet to the IP subnetwork of the destination host becomes the responsibility of the router. The router strips off source IP subnetwork link layer headers and trailer, modifies the IP headers and forwards the IP packet out the appropriate IP interface to the next IP subnetwork in the IP packet's route. In the forwarding of this IP packet, the router must reencapsulate the IP packet with a new link layer header and trailer. Reencapsulation generally requires resolving the destination IP address to a new physical address on the communications subnet to which the router is forwarding the IP packet.

If the physical network technology is Ethernet, transmitting to the destination host or router attached to the same Ethernet communications subnet as the source host, means resolving the IP address to an Ethernet address via ARP (Address Resolution Protocol) procedures, encapsulating the IP packet within an Ethernet packet and then transmitting the Ethernet packet onto the Ethernet medium.

If the destination host and source host interfaces attach to different IP subnetworks according to their IP addresses and IP subnet masks but really attach to the same communications subnet, these hosts will from the standpoint of the IP protocol software attach to separate IP subnetworks and will not be able to exchange data unless there is a router attached to the same communications subnet which has two IP addresses corresponding to both subnetworks.[3] One of the router IP addresses must correspond to the same IP subnetwork to which the IP address of the source host corresponds, while the other router IP address must correspond to the same IP subnetwork to which the IP address of the destination host corresponds. (Actually, to be completely accurate, there must exist an interface on the router to a subnetwork from which there is a route to the subnetwork to which the destination IP address corresponds.)

Suppose a single communications subnet supports multiple IP subnetworks and where the IP software does not support resolving of destination IP addresses on any but the source subnetwork. If a source host is to send a packet to the destination on separate subnetwork on the same communications subnet, the source determines that it must transmit the packet to the IP router. The router receives the packet and then resolves the IP address of the destination and retransmits the IP packet to the destination after the IP packet has been reencapsulated within a new link layer frame.

If the single communications subnet consists of a single LAN segment, this procedure will work but is less than sensible

1. because the router acts as unnecessary potential single point of failure for pairs of hosts on different IP subnetworks but on a single communications subnet,
2. because the router could restrict bandwidth if the router becomes too busy to route packets at full bandwidth and
3. because the number of packets on the LAN segment increase as the router forwards packets received from the LAN segment back onto the LAN segment.

---

[3] In some cases a mistake in configuration or a poor IP implementation will cause a host to attempt to resolve destination IP addresses via ARP procedures even when these IP addresses correspond to separate subnetworks. Such implementations permit hosts on the same communications subnet to communicate without the mediation of a router even if their IP addresses correspond to separate subnetworks. Such configurations and implementations violate the spirit of IP. Depending on incorrect IP layer behavior for correct functioning of a network is wrong for the following reasons.

a) An incorrect implementation will probably behave incorrectly in undesired ways.

b) Correcting or even making minor changes to an incorrect configuration could make communications connectivity between hosts suddenly vanish. And

c) incorrect IP layer behavior might have an unspecified impact relative to the behavior of other protocols.

Routing from a communications subnet back onto itself makes more sense when the communications subnet consists of bridged LAN segments, but in any less than trivial topology, the problem of configuring bridging and routing functionalities compatibly quickly becomes so complex as to be completely intractable.[4]

Routers are most useful in networks which consist of multiple physical networks of different technologies where they actually perform a protocol conversion at the link layer via deencapsulation and reencapsulation.[5] Yet, if the physical networks are of the same technology type, deencapsulation and reencapsulation represent wasted computer cycles because the router is performing a protocol conversion from a link layer protocol back to the same link layer protocol. Such wasted overhead increases the cost of routers because in order to obtain high throughput, a router might need an expensive microprocessor and associated hardware to provide the extra computation power which the router would be free to waste in unnecessary overhead. Still, network designers have found that routers can in some cases serve usefully as an interconnection technique even when all physical networks are the same type.

Routers can reduce contention between transmitters in some multidrop technologies like Ethernet. If a single Ethernet segment connects hosts, one host's transmission may block or collide with the transmission of other hosts. If a router connects several Ethernet segments, the transmissions of a host on one Ethernet segment will neither block nor collide with the transmissions of hosts on other Ethernet segments. Routers may also improve performance in an alternative case. If a router interconnects multiple LAN segments, if each LAN segment connected by the router corresponds to a single IP subnetwork and if a host on one LAN segment is only communicating with peers on its own LAN segment, the router does not forward the packets from that host's LAN segment to the LAN segments which correspond to other IP subnetworks. In this way, routers can reduce contention and increase the effective bandwidth of some technologies (viz diagram 4).

---

[4] This observation is explained below in the discussion of the parallel router filtering bridge architecture for an hybrid bridge router.

[5] For this reason, even within the OSI formalism, calling an IP router a gateway is not a misdesignation.

Unfortunately, traffic lossage can accompany contention reduction.

1. If a router has M + N identical interfaces,
2. if the router routes traffic from M incoming interfaces to N outgoing interfaces,
3. if the total incoming traffic for the M incoming interfaces is greater than the smaller of the maximum possible aggregate bandwidth on the N outgoing interfaces or the maximum bandwidth through the router and
4. if the data rates are sustained long enough (beyond the internal buffering capability of the router),

eventually the router will drop packets no matter how much buffer capacity the router has (viz diagram 5). Even on a two interface router such traffic lossage can occur because wire-speed routing is rarely a feature of IP routers (viz diagram 6) unless the router quite costly hardware. As a consequence, even with contention reduction the maximum data rates for communication between hosts can drop tremendously when pairs of communicating hosts attach to different LAN segments joined by a IP router.

Despite their expense, using high-performance wire-speed routers to interconnect IP subnetworks may be economically reasonable because their use in a judiciously designed and interconnected network could obviate the need to recable for more expensive transmissions media. In most cases, a need for higher bandwidth technologies like FDDI appears when aggregate bandwidth exceeds the carrying capabilities of lower speed technologies currently in use. Except perhaps for massive file servers, data base servers, supercomputers collecting or generating data for some real time applications, some image processing systems and some other special devices, a single host rarely has need for the total bandwidth of technologies like FDDI. Judicious network design, which properly divides traffic flow, could avoid the need for aggregate bandwidths comparable to FDDI bandwidths on any link or might confine the need for such bandwidth to internal router busses where providing even larger bandwidths than FDDI provides is much less costly.

If budgetary limitations argue against the purchase of high performance wire-speed routers and if traffic between networks requires less than maximum data rates, the potential communications bottleneck, which a router may cause, may be a less important consideration of network design than the following benefits which all IP routers provide.

1. Highly redundant interconnection of IP subnetworks via multiple IP routers can provide many alternate routes within a network so that if a router or link becomes inoperative, alternate routes through the network may still be available (viz diagram 7).[6]
2. IP routers act as firewalls between IP subnetworks by isolating some sorts of network anomalies like broadcast storms (viz diagram 8).

---

[6] Unless each end host has multiple links to different routers within the catenet, the redundancy described does not provide true fault tolerance in the sense of survivability of the failure of individual links or routers. Without such multiple links from the end hosts to the routers, an internet with redundant path only provides higher network availability probabilistically.

IP routers can act as firewalls against link layer and physical layer anomalies because packets are only forwarded by a router from one physical network to another physical network when fairly restrictive conditions are met and because interconnection via routers maintains the electrical isolation of separate communications subnets.

III. P802.1d MAC Bridges

P802.1d MAC bridges serve the purpose of finding paths between the hosts attached to the LAN segments which the MAC bridges connect. Because a bridged collection of LAN segments comprise a single communications subnet, the MAC bridges actually address the problem of providing paths between hosts attached to different communications links in such a way that a single communications subnet is created. MAC bridges address precisely the same problem which packet switches address in an intelligent switching fabric communications subnet. MAC bridges and the LAN segments which connect them comprise an updated and advanced version of switching fabric communications subnets. In a sense, MAC bridges are new and improved versions of old-fashioned PSNs like ARPANET IMPs or DECNET Level 1 routers while LAN segments are advanced versions of the serial links which interconnected PSNs or connected PSNs to end hosts.

The MAC bridging approach of building switching fabric communications subnets from MAC bridges and LAN segments is superior to the original ARPANET or DECNET approaches to construction of communications subnets for the following reasons.[7]

1. Ethernet is a much simpler and much faster access technology than the original ARPANET or DECNET packet switch access technologies.

2. P802.1d bridges are much more transparent to network end hosts than ARPANET IMPs or DECNET Level 1 Routers in that end hosts have no awareness whatsoever of their interaction with the MAC bridges and in that a MAC bridge within the communications subnet can be temporarily replaced by a passive LAN segment at any time if there is a need and if degraded performance can be tolerated.[8]

3. Ethernet is a much more flexible access technology than the original ARPANET or DECNET packet switch access technologies in that the Ethernet can provide either multidrop or point-to-point access to the MAC bridge packet switch according to networking needs.

---

[7] If MAC bridges are advanced packet switches, Ethernet concentrators are advanced access lines. Concentrators, as bus extenders, far surpass the capabilities or ordinary Ethernet cable and often provide all sorts of sophisticated monitoring, test and management capabilities. Those few learning bridges on the market which forward packets as they are received and which may be considered a sort of learning repeater are also a sort of advanced access line, but the smaller forwarding delay which learning repeaters can claim hardly justifies their costliness in comparison with smart hubs.

[8] Usually, multiport repeaters replace MAC bridges or vice versa. In the normal network growth path, a MAC bridge replaces a multiport repeater to increase performance by reducing contention and providing higher effective bandwidth. If the MAC bridge breaks, a multiport repeater can easily replace it temporarily with degraded performance.

Switching fabric communications subnets off-load onto PSNs the work of moving and routing packets between hosts. Such off-loading is a desirable feature of a communications technology because computer cycles lost to the overhead of communications processing are unavailable for other applications which run on the end host. Because early computers often barely had enough computational capacity for their main work, most early communications technologies and almost all current wide area technologies are designed to remove as much of the computer networking overhead from end hosts as possible.

Unfortunately, in the early days of computer networking, both processors and memories were expensive, software engineering techniques were much less developed and implementers did not have as good an understanding of software and hardware issues as we do today. As a consequence, early communications subnets tended to be quite expensive, were prone to flakiness and provided less performance than desired especially in relationship to the cost of the technology.

In the next attempt at providing good computer networking technologies, designers switched in many cases to LAN technologies consisting of passive broadcast communications subnets consisting of serial busses like coaxial cables or similar technologies and replaced packet switches with host-resident network controllers. These host-resident network controllers arbitrate access to the serial bus and act as packet filters which used simple hardware filtering strategies rather than complex software packet switching strategies.

A LAN system consisting of a serial bus, some number of attached computers and LAN controllers, which perform packet filtering and which arbitrate access to the bus, is a NORMA (NO Remote Memory Access) message passing multiprocessor parallel computer system.[9] Potentially, many problems which can be solved on a parallel processing supercomputer are amenable to solution at much less cost through the use of ordinary computers connected to a LAN. If solving such a problem requires more computer power executing in parallel than is possible on a single LAN segment, concentrators are available for Ethernet LANs which are bus extenders and which act as sophisticated repeater devices. Concentrators obviate the problems of signal degradation and late collision from which overly long Ethernets suffer.

LANs have proven themselves to be excellent price-performers on a wide range of parallel processing and computer networking tasks in comparison with more tightly coupled parallel computation systems and with earlier attempts at packet switching communications subnets. Yet, compared with parallel multiprocessor computer systems, LANs suffer a disadvantage in using a serial bus rather than a parallel bus. Also, LANs lack the hardware optimizations commonly found in message-passing parallel multiprocessor computers. Moreover, the usual LAN procedure of broadcasting all packets to all end hosts and filtering at the end hosts is an inherently low performance architecture especially because of the relatively slow serial nature of typical LAN busses.[10]

---

[9] Other common multiprocessor architectures are UMA (Uniform Memory Access) and NUMA (Non-Uniform Memory Access) architectures.

[10] Such a broadcasting procedure is also a security nightmare.

With the goal of improving LAN performance against the low-performance characteristics of LAN technology, network system designers eventually ended up revisiting (perhaps unconsciously) via bridging techniques the earlier intelligent switching fabric communications subnet architectures. The use of LAN technologies as components of a new type of intelligent communications subnet has become reasonable for the following reasons.

1. LAN technologies have matured, increased in reliability and have decreased so much in expense that it makes sense to use LAN technologies as access technologies to packet switches.
2. The software problem of packet switching and large software system development in general is better understood nowadays.
3. Software and hardware tools have improved immensely since the early days of packet switched network development.
4. Memories and microprocessors have steadily decreased in price as their packages have shrunken and as the power of microprocessors has increased.
5. Interconnectivity requirements have grown, but many protocol suites do not provide internetworking capabilities.

Early attempts at constructing such communications subnets with Ethernet segments substituted devices like simple bridges, simple hubs, smart bridges or smart hubs in place of repeaters. Two-interface simple bridges and multiple-interface simple hubs (henceforth denominated collectively as simple hubs -- note that hub when applied to token ring networks has a slightly different meaning) receive frames into memory and forward them out the other interfaces without the overhead of deencapsulation and reencapsulation. Two-interface smart bridges and multiple-interface smart hubs (henceforth denominated collectively as smart hubs) receive frames into memory, learn the source addresses to build an associative address map between interface and physical address, forward frames with unknown destination, broadcast and multicast addresses out all interfaces and forward frames with destination physical addresses known in the address map out the associated physical interface.[11]

---

[11] A bridge or smart hub might also transmit frames as they are being received but a bridge or smart hub which operates in this fashion may still need to receive the frame into memory because two frames, which are to be forward out the same Ethernet interface, may be received simultaneously. Alternatively, a host connected to this Ethernet interface may be transmitting a frame while a frame destined for this interface is being received at the bridge or smart hub. There may be some benefit to a smart bridging or smart hubbing device which can forward packets as they are being received (a "learning repeater"), but the benefit of sometimes reducing the tiny latencies involved in receiving the packet completely into memory before transmission can hardly justify the cost of building the special purpose hardware to carry out this procedure.

Simple hubs behave like repeaters except for the following differences.

1. Simple hubs provide more electrical isolation of attached LAN segments than repeaters often provide and can filter out some aberrant frames (e.g. frames with bad CRC) which repeaters generally repeat.

2. A simple hub can reduce contention because hosts attached to different LAN segments can transmit packets at the same time.

If the simple hub forwards at wire speed, the internal buffering of simple hubs increase the effective bandwidth of a network because at any instant the total number of frames in the process of being transceived is the total number of frames on the transmission line plus the total number of frames stored in the smart hubs memory. This effect is sometimes called "the equivalence of memory and bandwidth."

Smart hubs like simple hubs often forward packets at wire-speeds. But smart hubs can increase effective bandwidth even more than simple hubs. If the smart hub has learned the location of all hosts actively transceiving on the attached LAN segments, if the smart hub has N physical interfaces, $I_{n=1..N}$, and N physical interfaces, $O_{n=1..N}$, if there is a single host attached to each such interface, if host $H_{I(n)}$ is transmitting to $H_{O(n)}$ at maximum bandwidth and if the smart hub has sufficient computational power to sustain wire-speed forwarding, the total aggregate bandwidth of this configuration is N * 10 Mbps or 10 Mbps per receiver.

If a multiport repeater were to replace the smart hub, the total *aggregate* bandwidth of the alternative repeater-based configuration would be approximately 3 Mbps[12] or 3 Mbps/N per receiver (viz diagram 9). A simple hub because of internal buffering would have a slightly higher effective bandwidth than a multiport transceiver but would achieve nothing near the type of bandwidth increase which a smart hub can supply. Connecting LAN segments with smart hubs capable of wire-speed bridging can reduce contention and increase effective bandwidth at a relatively low cost because smart hubs lack the overhead of deencapsulation and reencapsulation. Smart hubs have the same lossage problem as routers have when aggregate incoming bandwidth on M interfaces directed towards N outgoing interfaces exceeds the maximum possible aggregate bandwidth on the outgoing interfaces.

While connecting hosts to smart hubs in most LAN environments has many benefits over stringing hosts together via a long cable snaked through a building, smart hubs cannot provide the ability to build complex network topologies which can provide fault tolerance through redundant links and paths between hosts. A redundant link or alternate path in a hubbed network would create a routing loop within which frames would circulate forever. Some topologies could cause frames to replicate at geometric rates (viz diagram 10) until network "meltdown" occurs.

---

[12] For sufficiently large N, if N hosts attach to an Ethernet and attempt to transmit at maximum data rate simultaneously, because of collision and transmitter back-off, the total possible aggregate bandwidth obtained on the Ethernet will be approximately 3 Mbps.

P802.1d STP MAC bridging procedures provide a solution to this problem[13] of providing redundant links without creating routing loops. MAC bridges provide all the capabilities of smart hubs and much more. If MAC bridges and LAN links connect to form a highly redundant topology such that between any two MAC bridges there are many possible paths which pass through many different MAC bridges and LAN links, the end hosts may still be able to reach one another in the event of LAN link or MAC bridge failure.[14] While P802.1d STP MAC bridging theoretically enables transparent bridging between all the IEEE 802.x technologies, different bit ordering and data rates make applying transparent P802.1d MAC bridging between different types of LAN technologies a far from non-trivial problem.

P802.1d applies graph theory to the problem of finding transit paths for frames in a communications subnet. P802.1d applies a spanning tree algorithm to the graph which is topologically equivalent to the actual physical network of MAC bridges and LAN segments. The nodes of this graph are the bridges and the LAN segments (viz diagram 3) i.e. each actual physical bus, either an internal MAC bridge bus or a LAN segment bus is a node on the graph. The weighted arcs of the graph correspond to the ports[15] which connect bridges to LAN segments.

---

[13] Alternative MAC bridging designs use Source Routing (SR) and Source Routing Transparent (SRT) procedures. The analysis which this article applies to STP MAC bridges applies to SR and SRT MAC bridges as well.

[14] As with the case of highly redundant topologies of LAN segments and routers, highly redundant topologies of LAN segments and MAC bridges do not provide guaranteed connectivity if a link or MAC bridge fails unless end hosts have multiple links which connect to different MAC bridges. The articles discusses below a straightforward way to provide end hosts themselves with such redundant connectivity.

[15] *P802.1d MAC Bridges* uses "port" as a term for a LAN interface at the MAC layer. Keeping the MAC and network layer terminology straight can be tricky. MAC bridges have MAC ports while routers have network interfaces. A MAC port connects to a physical segment to which hosts and other MAC bridges may connect. A router connects to an IP subnetwork via a network link. The Clearpoint Constellation Series Pyxis bridge (not bridge-router) has four Ethernet MAC ports but has only one IP network interface, which supports TELNET and SNMP. Each MAC port may connect to a physical segment, but the Pyxis bridge can have only one network link to an IP subnetwork. While there are other interpretations of the SNMP LinkUp/LinkDown trap, the Pyxis should probably send LinkUp messages to an interested Network Management Station (NMS) sometime after at least one MAC port has entered forwarding state. Of course, the network administrator would probably be interested to learn when a MAC port has disconnected (an indication that an Ethernet cable has been kicked or has otherwise become unterminated). Probably, a new PortConnected/PortDisconnected trap should be added to the set of standard SNMP traps. Currently, many SNMP bridge implementations use LinkUp/LinkDown traps in lieue of PortConnected/PortDisconnected traps, but this usage may prove confusing and misleading.

The P802.1d spanning tree procedure enables bridges to *block* the bridging of LAN frames on algorithmically determined bridge ports. This blocking has the effect of removing redundant paths in the simply but redundantly connected graph which has connections topologically equivalent to the actual physical network of MAC bridges and LAN segments (viz diagram 11). The active network topology provided by the spanning tree procedure possesses no routing loops, and this topology provides exactly one path between any two end hosts no matter to which LAN segment either of the end hosts is connected.

Once an active spanning tree topology has established itself on an actual physical topology, the MAC bridges just like smart hubs learn which bridge ports lead to which end hosts by building address tables which associate a received frame's source MAC address with the port in *forwarding* state on which the bridge received the frame.[16] When a bridge forwards a frame, the bridge looks up the destination MAC address in this table. If the bridge finds the destination address in the table, the bridge forwards the frame out the port associated with this address. Otherwise, the bridge forwards the packet out all *forwarding* ports whose filtering/forwarding restrictions, if any, the frame meets.

When filtering/forwarding restrictions are absent and the destination MAC address is an unknown or broadcast MAC address, the frame will be forwarded along all spanning tree paths to all end hosts. While consuming bandwidth, this broadcast is harmless because packet filtering at end hosts can permit all but the destination end host to discard the received frame. If the destination MAC address is a known MAC address, the frame travels along the unique spanning tree path from the source to the destination without any transmission onto any LAN segments outside the path. In this way, LANs connected by MAC bridges appear to end hosts as single LANs, and the MAC bridges act as a sort of "glue" which sticks LAN segments together.

---

[16] A P802.1d MAC bridge may also receive frames on *blocked* ports, but except for STP frames these ports are treated as inactive from, and the address table does not learn an association between source physical address of a received frame and the source port on which the MAC bridge receives the frame. A bridge may receive a frame on many ports in *blocked* state, but a bridge will receive a frame on at most one port in *forwarding* state.

IV. Contrasting Bridges and Routers

Dogmatic IP partisans may consider such "glue" architecturally inelegant, assert that IP routers are an architecturally superior alternative to MAC bridges and that sophisticated computer networks should be built only with routers. This assertion relies more an ignorance of networking principles and basic economics than on any true insight.[17] The categories of routers and bridges in the computer network hierarchy are exactly analogous to the categories of toll switches and local office switches in the telephone network hierarchy.

Just as toll switches and local office switches carry out exactly the same circuit switching function albeit at different hierarchical levels, routers and bridges perform exactly the same packet switching function albeit at different hierarchical levels. Thus, the distinction between routers and bridges is purely terminological.[18] In fact, arguing the superiority of bridging versus routing indicates a complete lack of understanding and is about as awesomely silly as arguing the relative merits of toll switching versus local office switching.

Just as building a telephone network from toll and local office switches requires attention to the appropriate position of each device in the hierarchy, building a computer network from routers and bridges requires exactly the same sort of attention. And just as toll and local office switching functions can reside in a single device, a single computer network device can host routing and bridging functions simultaneously.

---

[17] This assertion -- in my experience -- most commonly appears in the IETF (Internet Engineering Task Force) working group context. Since these groups handle protocol standardization and requirements specifications, the frequency and general acceptance of such irrational beliefs among working group members suggests some serious problems with the IETF working group mechanism and bodes ill for the continued economic and technical preeminence of the USA in computer networking. To be fair, such levels of ignorance, dogmatism, misunderstanding and general silliness are hardly unusual in other standards groups.

[18] Bridges and routers both perform path selection. STP, SR and SRT bridges select paths from one host to another host. IP routers generally select paths from one network to another, but host routes are definable as well at the IP layer. Either a bridge or a router can act as a firewall against "bad" packets. Because IP routers require that forwarded packets might more stringent requirements, there is an argument that IP routers provide better firewalls, but the more network protocols a router supports, the more ways a network packet can meet forwarding criteria. Consequently, claims that routers provide stronger firewalls are less obviously true in the case of multiprotocol routers. Bridge detractors often identify as particularly vile the need for link layer protocol conversion in bridging of a frame from one medium (e.g. Ethernet) to another medium (i.e. token ring). But in routing a network packet from one medium (e.g. token ring) to another (e.g. Ethernet) the IP router may actually need to perform network layer protocol conversion along with packet deencapsulation and reencapsulation. Somehow, dogmatic router partisans consider protocol conversion acceptable when it is called fragmentation. In short, there are no unique router-specific or bridge-specific functionalities.

Of course, just as blindly mixing toll and local switching office functions relative to local loops and trunk lines would make no sense in a telephone network, combining routing and bridging functions without due attention to the nature of computer network architectural principles might have the potential to completely mess up a computer network.

In ISO OSI Reference Model terminology, a router is a network layer or layer 3 device while a bridge is link layer or layer 2 device. In DOD Internet Architecture terminology, a router is an IP layer device while a bridge is a communications subnet device. As devices of different protocol layers, bridges and routers interoperate by being transparent to one another. To be specific, a bridge forwards a frame to a router just as it would forward a frame to any other device attached to the same communications subnet, and a router will always transmit a frame in the same way even if the frame is to be forwarded by a bridge.[19]

Router and bridges are transparent to one another because routing and bridging functionalities are "orthogonal" to one another. Routers mediate inter-network path selection within an internet while bridges mediate inter-host path selection within a communications subnet. Despite the orthogonality of routing and bridging functionality, if routers and end hosts are correctly configured and if all network traffic is routable, stand-alone routers could replace all stand-alone bridges in a network to create a routed network with exactly the same connectivity as the bridged network.

If the routers were sufficiently powerful, the routers could provide the same contention reduction, higher effective bandwidth and increased network availability via redundant connectivity which bridges provide. In fact, since routable packets have to meet much stricter criteria of correctness for routing than bridgeable frames have to meet for bridging, single-protocol routers can potentially provide much stronger firewalls against network anomalies than bridges ever can.

---

[19] The meaning of bridge-router interoperability is perfectly straightforward, but understanding has completely eluded the IETF working groups in charge of PPP (Point-to-Point Protocol) and Multiprotocol Interconnect over Frame Relay. Not only do these protocols ruin the ability of bridges and routers to interoperate correctly and reasonably over WAN links, but by treating bridging and routing layer protocols as if they belonged to the same level in the protocol hierarchy, these protocols violate the fundamental architectural principles of both the ISO OSI Reference Model and the DOD Internet Architecture.

Of course, routing might not be available for all the traffic supported in the bridged network, and routers which can route at wire-speed will use tremendously more expensive hardware and cost tremendously more than a wire-speed bridge. Even if these two drawbacks happen to be of lesser importance than the need to replace a single WAN with multiple LANs, reconfiguring a single subnetwork to be multiple subnetworks can easily be a nightmare. By comparison with routers, bridges are essentially "shrink-wrap" interconnection devices. A P802.1d MAC bridge right out of the packing box can generally provide such connectivity between separate LAN segments of the same IEEE 802 technology without configuration.[20] For simplicity and price performance in basic connectivity, nothing can beat bridges. In interconnecting physical networks of different technologies or where there is a need for extremely strong firewalls, routers will generally be the device of choice. To handle all possible needs, the optimum dream device for building complex networks is a hybrid bridge router which supports multiple technologies and which support flexible configuration so that the device can route and bridge between interface according to the network exigencies.

Because bridges as communications subnet devices function at an architectural layer invisible to end hosts, MAC bridges unlike IP routers need not have network addresses and interbridge traffic is conceptually "out of band" from the data traffic which MAC bridges switch. Yet, because MAC bridges can be fairly powerful computers and because remote management of MAC bridges in complex networks is desirable or even necessary, many modern MAC bridges support end host functionality, are assigned network addresses, support management functionality as provided by protocols like SNMP and sometimes even support remote terminal functionality as provided by protocols like TELNET.

Such network end host functionality is logically independent of MAC bridging, but when present MAC bridges can provide the remote access functionality which is almost always found built into IP routers because IP routers always have network addresses. The Constellation products implement this functionality through a logical MAC bridging port to which the internal network end host functionality attaches. The host functionality has its own IP and physical addresses on this port, and Ethernet frames with this physical address as a destination address are forwarded to this port.

In this case, the MAC bridge circuity is performing interworking between the external LAN busses and the internal MAC bridge bus, and the LAN addressing logic has been implemented in software on top of the actual physical internal bus. Alternatively, LAN-style bus addressing could have been implemented as part of the internal MAC bridge bus, and the software logic could have migrated to the hardware. In either case, this sort external-to-internal bus interworking suggests perhaps the most creative use of MAC bridging technology.

---

[20] According to the theory, MAC bridging between different IEEE 802 technologies should work, but there are sufficient number of "gotchas" in bit-ordering, frame structure, frame size and data rate that routing (link layer protocol conversion) makes a lot more sense than attempting to bridge transparently.

Constellation communications products use external-to-internal bus interworking to deliver bandwidth to an end host functionality far in excess of the Ethernet 10 Mbps maximum. The built-in end host functionality actually interfaces directly to the internal Constellation 120 Mbps bus. Because MAC bridging is invisible to the network layer, the MAC bridging functionality provides completely transparent internal-to-external LAN bus interworking for the use of the internal network end host functionality. In some network topologies, this interworking can provide the end host functionality of the Constellation Series Little Dipper (a high performance intelligent WAN LAN bridge router supporting 8 Ethernet interfaces and three WAN interfaces) with an effective network interface of a full 80 Mbps aggregate bandwidth.

In alternative physical topology configurations, some of this bandwidth can be used for redundant links to increase the probability of network connectivity even if an individual link or MAC bridge within the communications subnet should fail. The high bandwidths and redundancy capabilities which MAC bridges provide become much more interesting when the MAC bridge is implemented in the Ethernet controller which is resident in an end host like a VME bus SUN or a Novel file server.[21] While an Ethernet controller which contains a multiport MAC bridge (an Etherbridge controller or -- in the general case -- a LANbridge controller), cannot provide precisely all the functionality which an FDDI interface might provide, in most cases such an controller provides all the functionality which is important at less cost for the controller and with no cost to recable for FDDI.

The redundant connectivity capabilities of an Etherbridge controller are superior to the redundancy capabilities of FDDI. Etherbridge controller capabilities simplify the problem of creating fault tolerant network topologies which guarantee fault tolerance in the case of the failure of any single link, stand-alone bridge or stand-alone router. To provide an end host network connectivity despite the failure of any single link through the use of ordinary Ethernet controllers on end hosts which support no MAC bridging functionality, the host must be multihomed.[22] Transparent recovery from link or bridge failure could require extensive rewriting of network and application level software because the rethreading of sessions from one network interwork interface to another is a non-trivial operation in most protocol architectures. An Etherbridge controller could automatically switch over to a redundant link in the case of link failure. Since the switchover takes place at the MAC layer, the network layer would never even notice.[23]

---

[21] In the context of IBM PCs, controller cards are often called adaptor cards.

[22] A multihomed network end host in a purely bridged environment has LAN interfaces which connect to LAN segments which connect to different MAC bridges. A MAC bridge, which supports network end host functionality, is a multihomed host, but not all multihomed hosts are MAC bridges. A multihomed network end host in a purely routed environment has network interfaces which connect to multiple subnetworks which connect to different routers. A router is a multihomed host, but not all multihomed hosts are routers. Various connectivity configurations can characterize a multihomed host in a heterogeneously bridged and routed environment.

[23] If the Etherbridge controller is considered one of the network bridges which might fail, the end host would have to support multiple Etherbridge controllers to guard against the possibility of controller failure. But this notion of fault tolerance belongs more to the architecting of a fault tolerant computer than to the architecting of a fault tolerant network.

The Constellation solution provides a capability which no simple FDDI approach to high bandwidth connectivity provides.[24] An Ethernet controller which supports hybrid bridging routing would contain an outboard router functionality.[25]

A host using such a controller would communicate with this outboard router at the full bandwidth of the controller card's internal bus. The hybrid bridge router on the controller card has the potential to improve the performance of well designed networks in terms of either bandwidth or availability via redundant fault tolerance. Performance increases of at least an order of magnitude should be obtainable without a comparable increase in cost.

---

[24] Of course, the next step would be to put an FDDI bridge on a single controller card.

[25] Such an outboard router functionality is not comparable to a controller based TCP/IP implementation like those found on the Prime 50 Series LHC (LAN Host Controller) and some other Ethernet controller cards. Such controller-based TCP/IP implementations act simply as part of the hosts normal networking end-host functionality. Outboard router functionality is not part of the end-host functionality of the computer which hosts the controller card. Instead the controller card contains its own LAN constructed around the internal bus of the LAN controller and the host computer and the outboard router functionality both attach to this internal bus network. The internal bus network may -- of course -- interwork with external LAN networks. An outboard router functionality functions just like an stand-alone router attached to a LAN segment. In both cases, the overhead of network routing processing has been moved from an ordinary host computer to a special purpose computer optimized for this purpose. A controller card with outboard router functionality has the benefit of providing a communications path between end host and router at a typical computer bus bandwidth which can be orders of magnitude larger than typical LAN bandwidths.

Controller-based implementations of TCP/IP are somewhat problematic in contrast to an outboard router functionality. The type of intelligent network controllers which contain implementations of TCP/IP tend to be somewhat pricey for low-end hosts, but on a high end hosts, which probably had computer cycles available for network protocol processing, controller-based implementations tend to become a networking bottleneck. In any case, controller-based implementations of TCP/IP are in a sense a solution in search of a problem because in common host-based implementations of TCP/IP for virtual memory operating systems, the overhead of network processing pales by comparison with the overhead of context switches between user and privileged mode. Such context switches are generally obligatory in the case of host-based and controller-based implementations of TCP/IP. When the host operating system is DOS or a similar operating system, which makes no use of context switching between user and privileged mode, controller-based implementations of TCP/IP may make some sense.

V. Disjoint Functionality Hybrid Bridge Routers

The logically simplest form of multiport hybrid bridge router supports both bridging and routing functionality disjointly. Physical ports are allocated either to the internal router(s) or to the internal bridge(s) exclusively. Single-unit disjoint functionality hybrid bridge routers offer little in comparison with separate pure routers and pure bridges except that

1) they may save money and space by having a single box for bridging and routing and

2) the ability to configure which ports perform bridging and which ports perform routing is a useful sort of flexibility which pure routers and pure bridges by definition cannot provide.

VI. Parallel Router Filtering Bridge Architecture for Hybrid Bridge Routers

More sophisticated hybridization of bridging and routing would involve providing a method whereby a single port can provide access to both the bridging and routing functionalities. Routing the routable network traffic and leaving the rest of the traffic for bridging is perhaps the most obvious way to implement a hybrid bridge router which combines routing and bridging functionality more intimately.

Such an hybrid architecture makes sense if an unroutable protocol like LAT must be bridged across several LAN segments while all other protocols carried within the link layer are to be routed across the LAN segments. From the standpoint of LAT the LAN segments would have to constitute a single communications subnet while from the standpoint of the other protocols the LAN segments must correspond to separate communications subnets.

This methodology of routing and bridging in parallel across the same set of LAN segments corresponds to a parallel router filtering bridge architecture. An hybrid bridge router with this architecture is equivalent to connecting in parallel a router which routes the routed traffic and a bridge which filters out the routed traffic and bridges the remainder of the traffic (viz diagram 12).

Filtering is a useful technique at all protocol layers in the constructions of sophisticated computer networks when there is a need to prevent certain traffic from traversing some portion of a network, to stop some traffic from reaching some portion of a network, to block some traffic from reaching some hosts or to block communication among some set of hosts. Unfortunately, there are insurmountable configuration and reachability problems for this type of hybrid bridge router with respect to any but the simplest topologies when such a device tries to use filtering to provide sufficient separation between bridged LAN segments so that the device can route some traffic between subnetworks which correspond to apparently separate communications subnets while the device bridges some traffic among the very same LAN segments which compose the apparently separate communications subnets.

The following example elucidates the problem (viz diagram 13).

1. Suppose three LAN segments (L1, L2 and L3) connect three MAC bridges (B1, B2 and B3) in a triangular configuration to provide redundancy.

2. Suppose that B1 routes IP traffic between LAN L1 and LAN L2 while bridging LAT traffic.

3. Suppose IP subnetwork 1 (S1) corresponds to LAN L1 while IP subnetwork 2 (S2) corresponds to the bridged network consisting of LAN L2, MAC bridge B2 and LAN L3.

If the port from B3 to L3 blocks, the configuration works correctly. If the port from MAC bridge B2 to LAN L2 blocks, the configuration is broken and subnetwork S2 IP hosts on LAN L2 will be unable to communicate with subnetwork S2 IP hosts.

In this particularly simple topology, keeping the network layer configuration compatible with the MAC layer bridge priority and port cost configurations is an annoying but less than insuperable management problem. Unfortunately, for almost any topology of greater complexity, the problem of maintaining compatible configurations is daunting if not impossible. Even if a network administrator has established a compatible MAC layer/network layer configuration for a complex topology, minor purposeful topology changes might break the configuration, or -- even worse -- someone might for some perfectly legitimate reason connect a MAC bridge of unspecified configuration across a set of LAN segments and break the configuration beyond repair. Determining the root cause of such connectivity breakdowns could be extremely difficult especially

1. because connectivity breakdowns can occur for many reasons besides incompatible configuration,

2. because the problem might not appear immediately and

3. because the relationship between the topology change and the connectivity breakdown might be non-obvious.

Parallel router filtering bridges have an even more serious drawback than the difficulty or complexity of establishing compatible MAC layer/network layer configurations. Routers are often used precisely because they can provide a sort of firewall against network anomalies. In the simple example of a hybrid bridge router which bridges LAT traffic and which routes all other traffic, an insane LAT box would wipe out the whole network.

An even trickier problem in the case of P802.1d MAC bridges lies in handling spanning tree PDUs (Protocol Data Units). Filtering out spanning tree PDUs could reasonably be part of creating apparently separate communications subnets because if such PDUs were accepted and processed by the hybrid bridge router and if somewhere in the bridge network bogus spanning tree PDUs were appearing, the hybrid bridge router would propagate rather than isolate the network anomaly which such bogus PDUs would cause. Unfortunately, if the P802.1d MAC bridge discards the spanning tree PDUs at the receive port before processing. The logic of P802.1d MAC bridging forces that port into the *forwarding* state. If ports, which should *block*, instead enter *forwarding*, the bridged network will suffer routing loops and will eventually "melt down" or fail miserably.

The parallel router filtering bridge architecture is simply an incorrect architecture for an hybrid bridge router and will not function properly except in the most trivial of topologies. This architecture is incorrect for the following reasons.

1. Filtering is simply a useful technique to control local access and only improperly a method to establish global topologies, and

2. the parallel application of routing and bridging to the same LAN segments is equivalent to treating routing and bridging as operations at the same protocol layer.

The parallel router filtering bridge architecture violates the DOD Internet Architecture, the ISO OSI Reference Model, DNA and practically any other reasonable computer networking architecture. There is no way for such a device to avoid dysfunctional behavior.

VII. Logical Subbridge Architecture for Hybrid Bridge Routers

The key to architecting a non-disjoint functionality hybrid bridge router correctly lies in the principle that routers must connect separate communications subnets. In a MAC-bridged environment, a separate spanning tree topology defines a separate communications communications subnet. This mapping of spanning tree topology to communications subnet implies that an hybrid bridge router device must take part in at least two spanning trees. This requirement means that the hybrid device from the standpoint of P802.1d must contain at least two logical P802.1d transparent MAC subbridges. The router functionality would have an interface to each logical subbridge (LSB or bridgelet in Clearpoint parlance). Each bridgelet would take part in its own separate active spanning tree topology which has its own root bridge (viz diagram 14). The LSB bridge router architecture, which gives the network layer routing functionality subordinate interfaces to link layer MAC bridging functionalities, is completely consistent with the protocol layering of typical network architectures.

Of course, either by accident or conscious decision, different bridgelets on a single Constellation Series LSB bridge router could take part in the same active spanning tree topology. Such a condition is readily apparent in the information, which Constellation bridge routers provide, because the root bridges for the active spanning tree topology in which such bridgelets take part will all be the same.

If all the LAN interfaces belonged to the same bridgelet, the LSB bridge router really should not be configured to route but should act as a pure bridge. Of course, there is no reason to forbid routing from one communications subnet back to that communications subnet if the network administrator really wants such a configuration. If each LAN interface belongs to a different bridgelet, and if each bridgelet has an IP address which corresponds to a distinct IP subnetwork, the device cannot perform any bridging but can act as a pure router.

The Constellation LSB bridge routers can support as many bridgelets as they have interfaces so that pure routing is possible. The Constellation bridge routers support multiple logical IP interfaces which can map to different bridgelets. The IP router functionality routes packets between separate IP subnetworks corresponding to the different communications subnets in which the bridgelets take part, by routing packets between the logical IP interfaces which the bridge router supports. IP Interfaces are assigned to different bridgelets depending on whether traffic between different groups of interfaces should be bridged or routed.

Assignment of multiple IP interfaces to the same bridgelet is allowed so that multiple IP subnetworks can reside within one communications subnet. Or alternatively, there may be bridgelets to which no IP interfaces are assigned so that the communications subnet, which corresponds to the active spanning tree topology in which this bridgelet takes part, is isolated from its sibling bridgelets and their associated communications subnets. Of course, the communications subnet of an isolated bridgelet my be connected to other communications subnets by some other router or router functionality elsewhere in the network.

LSB hybrid bridge routers provide firewalls between communications subnets for even the most complex technologies (viz diagram 15) just as routers are supposed to provide. Anomalies within one communications subnet will not propagate to other communications subnets simply because some traffic is being bridged. Likewise, the effects of bogus spanning tree packets will confine themselves to one communications subnets. Yet because the LSB formalism does not interfere with the normal operation of STP, there is no possibility of creating routing loops within the communications subnet. LSB bridge routers have none of the configuration and reachability problems associated with parallel router filtering bridge architectures.

LSB hybrid bridge routers are superior to ordinary pure routers because such bridge routers provide flexibility in the allocation of bandwidth for throughput or redundancy. An N port bridgelet can provide redundant bandwidth of M * (the bandwidth of an individual link), where M < N, and throughput bandwidth of (N - M) * (the bandwidth of an individual link) to a given IP interface. No pure router can provide such a capability to craft a network in this way. In fact, the LSB hybrid bridge router architecture is flexible enough that if a need should occur for pathological configuration, an LSB hybrid bridge router, which supports filtering capabilities, can act like a parallel router filtering bridge by configuring the bridge router with one or more LSBs

1. which bridge all non-routed traffic,

2. which partake in only one spanning tree via an internal virtual LAN segment connection,[26]

3. which support multiple IP interfaces corresponding to different IP subnetworks and

4. which discard all routable traffic across the internal virtual connections while such traffic is forwarded to the router functionality.

Because the LSB architecture is a correct architecture for hybrid bridge routing, LSB bridge routers do not limit network design as parallel router filtering bridge architectures would, but rather LSB bridge routers provide hitherto unavailable techniques and potentialities in the construction of sophisticated networks.

VIII. Conclusion

If a networking technology were available which provided tremendous bandwidth at low cost, which could span large distances and which could easily provide network security and redundancy, there would be less need to use complex and sophisticated networking techniques to provide needed functionality. But in the absence of such technology, hubs, bridges, routers and high bandwidth media like FDDI all have their place as useful components in the design and construction of computer networks. A disjoint functionality hybrid bridge router can probably provide some savings, but the LSB hybrid bridge router for its price, flexibility and capabilities is perhaps one of the most useful network components yet introduced.

---

[26] There are alternative methods of creating parallel router filtering bridge functionality within an LSB hybrid bridge router architecture. The proposed procedure handles the specific problem where the router functionality must stop advertising routes to an IP subnetwork if all physical interfaces to the corresponding communications subnet are inoperative. Tony Bono pointed out that treating bridgelets (LSBs) joined by internal virtual LAN segments as if they corresponded to separate well-defined communications subnets makes it possible to inform the router functionality when all physical interfaces to an apparent communications subnet are inoperative. When this condition occurs, the router functionality can refrain from advertising connectivity to the corresponding IP subnetwork.

In fact, in a period of economic retrenchment during which recabling for FDDI might be dauntingly expensive, using LSB hybrid bridge routers as a lower cost substitute for FDDI is quite reasonable because LSB hybrid bridge routers can in many cases satisfy those needs which FDDI addresses. LSB hybrid bridge routers even address issues of fault tolerance and redundancy which FDDI does not address very well. While LSB hybrid bridge routers will prove extremely useful network components, a parallel router filtering bridge is in practically all cases a disaster as a networking component when the complexity of the network topology is anything other than trivial. Purchasing or developing a parallel router filtering bridge would be a mistake.