Routers in Internetworks

How Data Travels through the Internet

Prashant Bharadwaj

Prashant Bharadwaj is a senior engineer-software in the Broadband Access unit of Telecom and Internetworking division, WIPRO Technologies, Bangalore. His current area of work includes developing datacom protocol stacks and providing software solutions for Network Infrastructure companies like Ericsson, Cisco Systems, Inc. among others.

Internetwork is a collection of individual networks, connected by intermediate networking devices, which function as a single large network. The advent of packet switching has revolutionized the manner in which an internetwork can be created. One of the most prominent packet switching devices is the router. This article provides a comprehensive overview of the necessity and working of a conventional router deployed in today's Internetwork.

Introduction to Networking

The Internet has changed the way humans work like never before. Millions of people use the Internet today to carry out a wide range of activities, from mundane checking of emails to conducting thousands of rupees worth online stock transactions. The Internet in simple terms is a network that connects millions of independent computing devices worldwide to create an illusion of a wired world. In Internet parlance, these computing devices are called hosts or end-systems. As these hosts may be heterogeneous, both architecturally and in the usage of software, there should be some rules that define how the data is to be transmitted from one host to another across a network. These rules are called protocols.

Managing millions of hosts at the same time is tedious to say the least. So a hierarchical structure of networks has been defined. Any network will tend to broadly fall into one of the following groups:

Local Area Networks (LAN) – A LAN is a high-speed data network that covers a relatively small geographic area (such as a building / small campus). It typically connects workstations,

Keywords
Internet, packet switching, route processor, packet routing.

PCs, printers and other devices. LANs offer computer users shared access to devices and applications, file exchange between connected users, and communication between users via electronic mail and other applications.

Wide Area Networks (WAN) – A WAN is a data network that covers a relatively broad geographic area and often uses the transmission facilities provided by common carriers such as telephone companies. The Internet is the interconnection of many LANs, and WANs. It is, therefore, appropriately called the 'network' of all networks.

The most prominent internetwork is the Internet. The hosts constitute the edge of an Internetwork and the mesh of devices that interconnect these hosts is called the core of an Internetwork.

Circuit Switching and Packet Switching

Internetwork can also be defined as a collection of hardware/software infrastructure that acts as a medium through which:

a) Geographically dispersed users communicate with each other.

b) Distributed services/applications are implemented.

The most prominent internetwork is the Internet. The hosts constitute the edge of an Internetwork and the mesh of devices that interconnect these hosts is called the core of an Internetwork.

There are fundamentally two approaches towards building a network core – *Circuit Switching* and *Packet Switching*. In a circuit switched network (*Figure* 1), the resources needed along a path to provide for communication between the end-systems are reserved in advance. Hence, when two remote hosts want to communicate, the network establishes a dedicated end-to-end circuit between two hosts. Each circuit is given a unique identifier. All traffic emanating from the source host will follow the same path to reach the destination. Data received on an input link is automatically switched onto the right output link and this process is called *switching*. The device that performs this function is called a *switch*. The telephone networks are the most well-known examples of circuit-switched networks. They were designed that way because a telephone conversation between two

In a circuit switched network the resources needed along a path to provide for communication between the end-systems are reserved in advance.

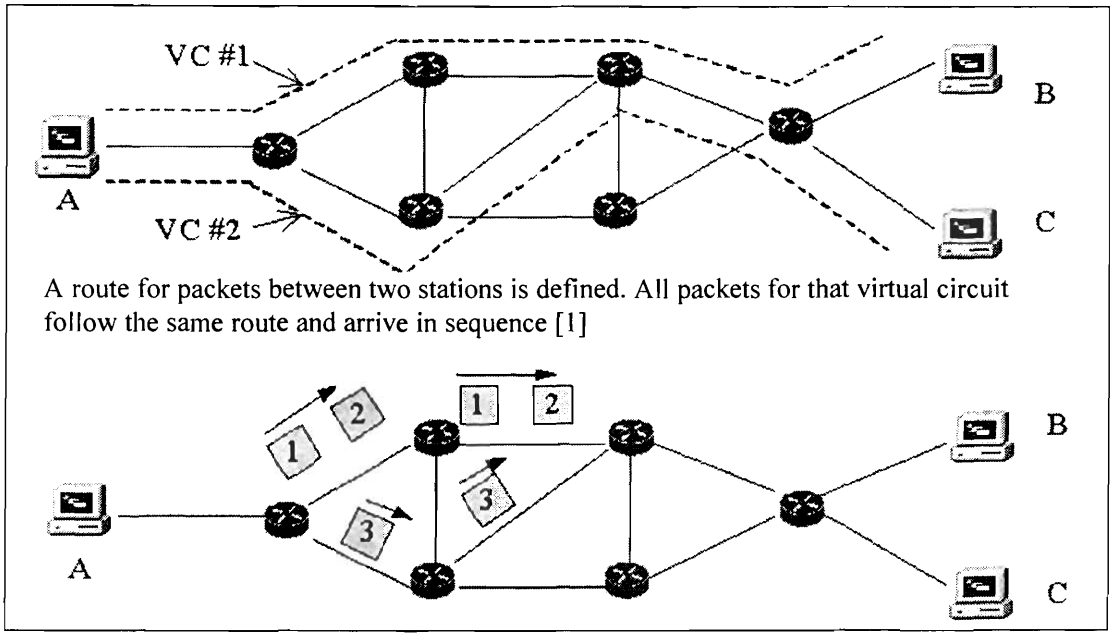

Figure 1. Virtual circuits and flow of packets in circuit switching.

individuals is a real time process. The process is inherently interactive and any delay/loss in reception of speech disturbs the smooth flow of conversation!

The most visible drawbacks of circuit switching are the overheads involved in establishing and reserving end-to-end paths. Consider an ophthalmic patient in a circuit-switched telemedicine facility; the surgeon will be remotely accessing a series of digital images of the infected eye to diagnose the problem. The surgeon sets up a connection, requests a digital image, contemplates the image and then requests for the next image. Observe that the network resources are being wasted for the period when the surgeon is musing! Also, applications such as messaging traffic, file transfers, web page downloads among others do not really need a circuit switched path. In order to provide a simple, less expensive yet effective alternate data transfer mechanism, researchers explored the concept of packet switching.

The most visible drawbacks of circuit switching are the overheads involved in establishing and reserving end-to-end paths.

In the case of a packet switched network, the source host divides the long message to be transferred into chunks called packets. The packets are processed individually, at each of the inter-

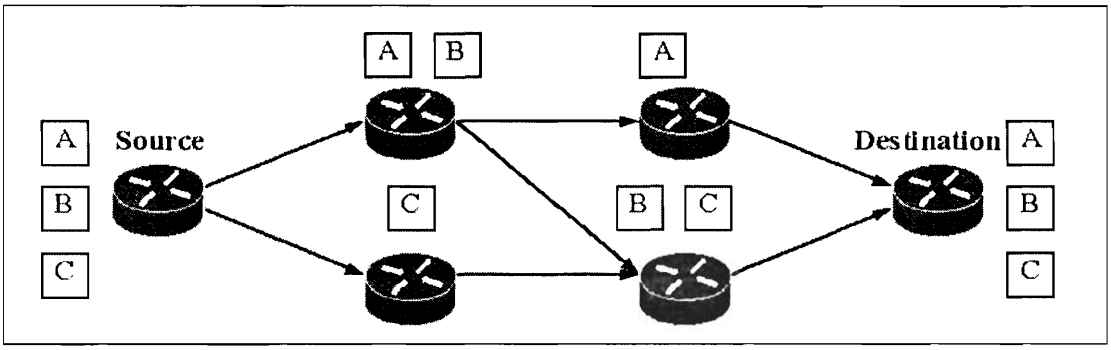

Figure 2. Store and forward mechanism of packets in packet switching.

mediate packet-switch devices, and typically share the network resources with other packets. Most packet switch networks use the concept of *Store and Forward* [2] (*Figure* 2) transmission for transmitting data. Store and forward transmission means that the packet switch must receive the entire packet before it can begin to transmit the first bit of the packet onto the outbound link. A packet need not follow the same path taken by its predecessor. Packet switching obviates the need to reserve resources between disparate hosts a-priori.

One of the pitfalls of the basic packet switch mechanism is its non-suitability for real-time services such as telephonic applications and streaming multimedia applications, due to variable and unpredictable delays that are inherent in packet switching. There are several approaches that have been developed which augment packet switching to support the smooth functioning of these real-time applications. However, the discussion of such topics is beyond the scope of this article.

One of the pitfalls of the basic packet switch mechanism is its non-suitability for real-time services such as telephonic applications and streaming multimedia applications

Basics of Routing

A complex internetwork consists of many remote hosts interconnected by communication links. It is however not feasible to have dedicated sets of links between every pair of hosts. So the data transmitted between a pair of end-hosts has to pass through intermediate devices. These inter-connecting devices may have many links emanating from them. The devices must be in a position to determine which link should be used so that the data is transmitted to the right destination. This process of decision-

The process of decision-making and transmission of packets from a source to a destination across an internetwork is called *routing*.

making and transmission of packets from a source to a destination across an internetwork is called *routing*. A computing device that is dedicated to perform this function is called a *router*. Routing involves two basic activities – determining optimal routing paths and transporting packets through such a path. Path determination is quite complex and is usually done at the software level. There are special programs, which run in a router to help make these decisions – these are called routing protocols. To aid the process of path determination, routing protocols maintain routing tables, containing the most feasible path to a particular destination.

A host is identified by a unique network address called the IP (Internet Protocol) address. Every packet has a header that contains the IP address of the destination along with details of the payload (actual data being carried). The routers use this IP address to route the packets. The Internet Protocol enables delivery of packets to the right destination host, but is incapable of delivering them to the specific application in that host. The Transmission Control Protocol (TCP) and the User Datagram Protocol (UDP) are two protocols that run on top of IP which facilitate the delivery of packets to the right application in the receiving host. Since TCP and UDP deal directly with the end hosts, they are called end-to-end protocols. In TCP, a connection is established between the sender and the receiver before the actual transfer of data. It is therefore called a *connection-oriented* protocol. It ensures that the packets of data are received in the same order in which they had been sent. So it can be used to reliably transfer packets of applications such as Video Conferencing. However, many commonly used applications do not require the overhead of connection establishment associated with TCP. A typical application is the email service. A connection need not be established between the two end hosts; emails have to be just delivered to the recipient's account. Also, it is not necessary that the packets arrive in an orderly fashion as long as the email can be reconstructed without errors. UDP is used for such applications. So UDP is called a *connectionless* protocol.

To aid the process of path determination, routing protocols maintain routing tables, containing the most feasible path to a particular destination.

Box 1. HUB

All the hosts of a LAN are connected via dedicated links to the Hub. A message received by a Hub is sent out on all links except the link from which it originated. This is called *Broadcasting*. A host that receives such a message determines whether it is the rightful recipient by looking at the IP address in the message. Hub is a device at the physical layer (layer 1) in the TCP/IP protocol stack.

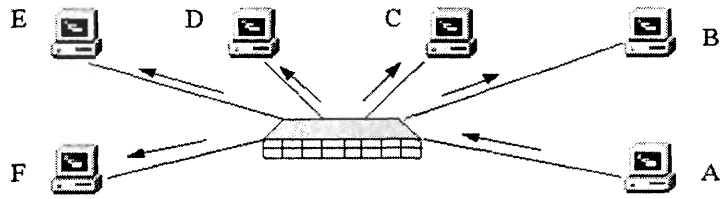

Figure 3. A Hub broadcasts data received from one host to all other hosts in the LAN.

The TCP/IP protocol stack consists of four layers. The topmost layer that interacts directly with the applications is called the *Application Layer*. The TCP and UDP protocols run in the *Transport Layer*. IP runs in the *Network Layer* and deals with the individual packets of a traffic flow. Then comes the *Physical Layer*, which deals with the actual transfer of the bits. Physical devices called *Network Interface Cards* act as the interface between a host and the network. Each network interface card has a unique address known as the MAC (Media Access Control) address. These MAC addresses are used to transmit data in a LAN environment. A separate layer that can handle the MAC addresses called the *Link Layer* has been defined in the OSI model. However, in the TCP/IP protocol stack model, the link layer is part of the Physical layer.

Physical devices called *Network Interface Cards* act as the interface between a host and the network. Each network interface card has a unique address known as the MAC (Media Access Control) address. These MAC addresses are used to transmit data in a LAN environment.

A large network combines many types of network elements. The most basic network element is a *Hub* (*Box* 1, *Figure* 3). It is used to interconnect hosts in a LAN. A *Bridge* (*Box* 2) is used to interconnect LAN segments. But the simple decision process adopted by a Bridge is not sufficient to successfully transfer data between two remote hosts. Hence, it is necessary to process the information in packets, determine which is the next router along the path to the destination (best 'next hop'), and prepare the packet for forwarding to the relevant next hop. This is where devices like Router come into the picture. A Router is capable of

Box 2. Bridge

Bridges 'learn' the location of hosts by looking at the sender's *Link Layer* address. Once they know where the hosts are located, they forward traffic across to another LAN as necessary. However they ensure that traffic destined for a host within the same LAN is not forwarded to other LANs. The decision process in a bridge is simple. If the destination address is not present within the LAN, the data is then broadcasted to the other LANs to which the bridge is connected. This simple decision process results in fast moving traffic. A bridge with 2 or more ports is referred to as a Layer 2 switch. Each port on the switch is capable of handling more than one MAC address (it can handle up to 1024 MAC addresses!!!)

making complex forwarding decisions. If the protocol used by the networks connected by a router are different, a router also performs protocol conversion.

Consider an example. The administrator of a network would typically like to exercise some control over data from some sources such as limiting the amount of bandwidth allocated to data from a specific source. This kind of control cannot be achieved by using the simple mechanism found in switches. This functionality is achieved by running complex network protocols on a router. So the network elements in the edge of a network are generally routers. The network core (network cloud in *Figure* 4) between two edge routers can contain either switches or routers. Traditionally, switches were used to enable faster transmission of data. However, rapid advances in routing technology has allowed contemporary routers to process packets between links as fast as conventional switching devices while at the same time retaining control over the data flows. Interestingly, switches are evolving to incorporate the features of routers. The current IP Switches are nothing but routers!

Generic Architecture of a router

The generic architecture of a router is as shown in *Figure* 5. The four components of a router are as follows:

Physical Interfaces: The incoming physical links are connected onto them. These interfaces are called Network Interfaces or Line Cards. They are also commonly referred to as ports.

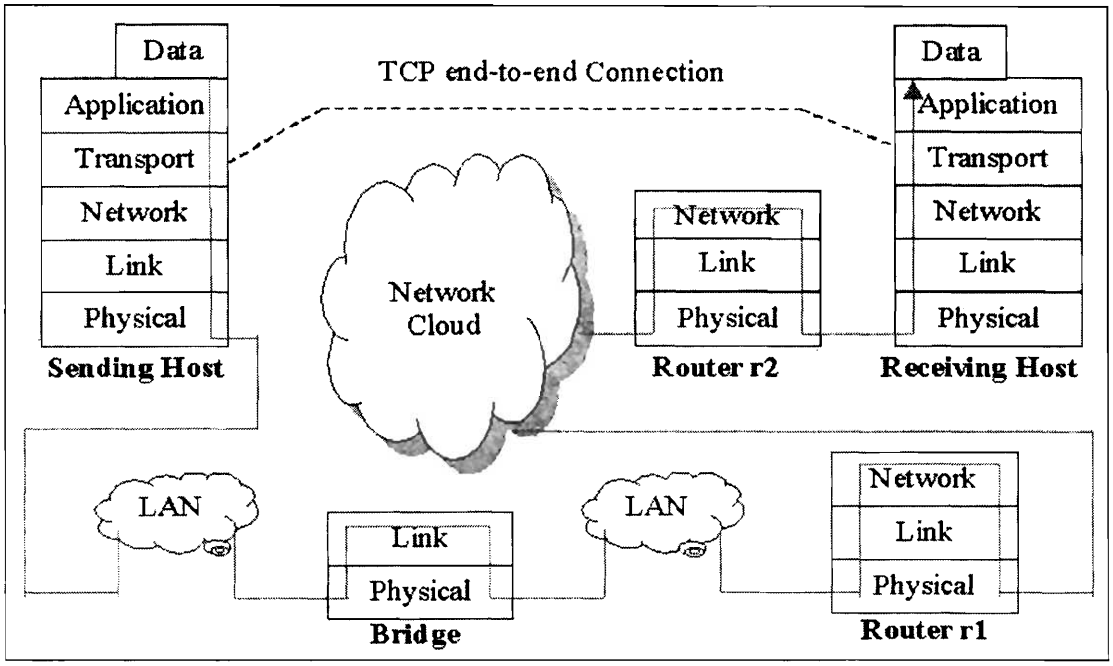

Figure 4. Flow of data between two Hosts via various Network Elements

Switching Fabric: The packets have to be switched between the incoming and outgoing interfaces. The switching fabric contains the Data/Address/Control buses through which packets are switched between two interfaces.

Memory: In case a packet is already being serviced, another packet wanting to use the same interface must not be discarded. So some queues are required. These queues are typically First In

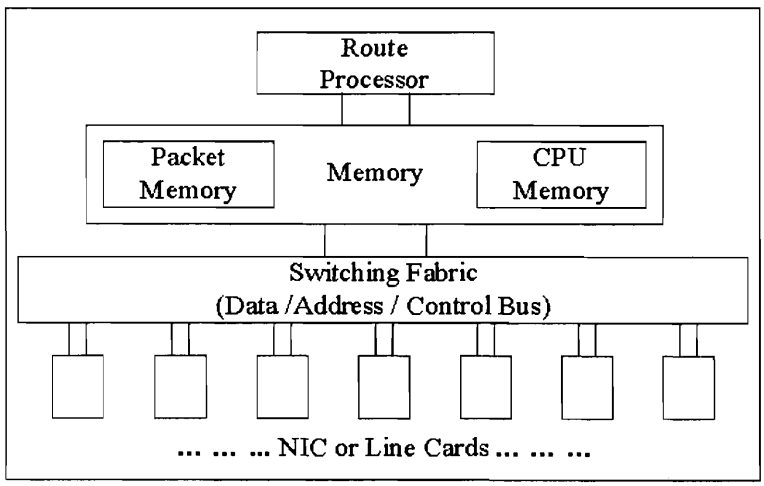

Figure 5. Generic architecture of a router.

The route processor executes the routing pro-tocols and computes the routing tables. It also looks up the rout-ing tables and forwards a packet to the right outgoing interface.

First Out (FIFO) queues. This part of memory is called Packet Memory. The protocols that have to be used to forward the data must also be stored in the memory. Also, the routing tables that are needed for making the forwarding decision must be stored in memory. This part of memory is called CPU Memory.

Route processor: The route processor executes the routing protocols and computes the routing tables. It also looks up the routing tables and forwards a packet to the right outgoing interface.

A router can be functionally divided into three parts:

Control Plane: A router would need to build and maintain tables that map the incoming and outgoing interfaces for reaching particular destinations. These tables are called forwarding tables. The control plane involves the building and maintenance of such tables.

Management Plane: Any router would need a configuration file that can be modified according to the needs of the management of an individual network. The management plane involves dealing with the configuration files, gathering and providing statistics like the rate of flow of traffic, link utilization, etc., and responding to messages from control protocols.

Data Plane: The packets have to be switched from an incoming interface to an outgoing interface. During this switching process sometimes the packet header and content is examined and manipulated. The packet delivery has to be scheduled in such a way so as to avoid starvation of packets on any interface. These functions are the prerogative of the data plane.

Evolution of Routers

Routers have steadily evolved over the years. Now they are capable of handling many Gigabits per second. The architecture of initial routers was similar to that of the conventional computers (*Figure* 6a) [4]. They had a shared central bus, a central route processor and memory and peripheral interfaces or line cards. When packets arrive at a link they are transferred from the line

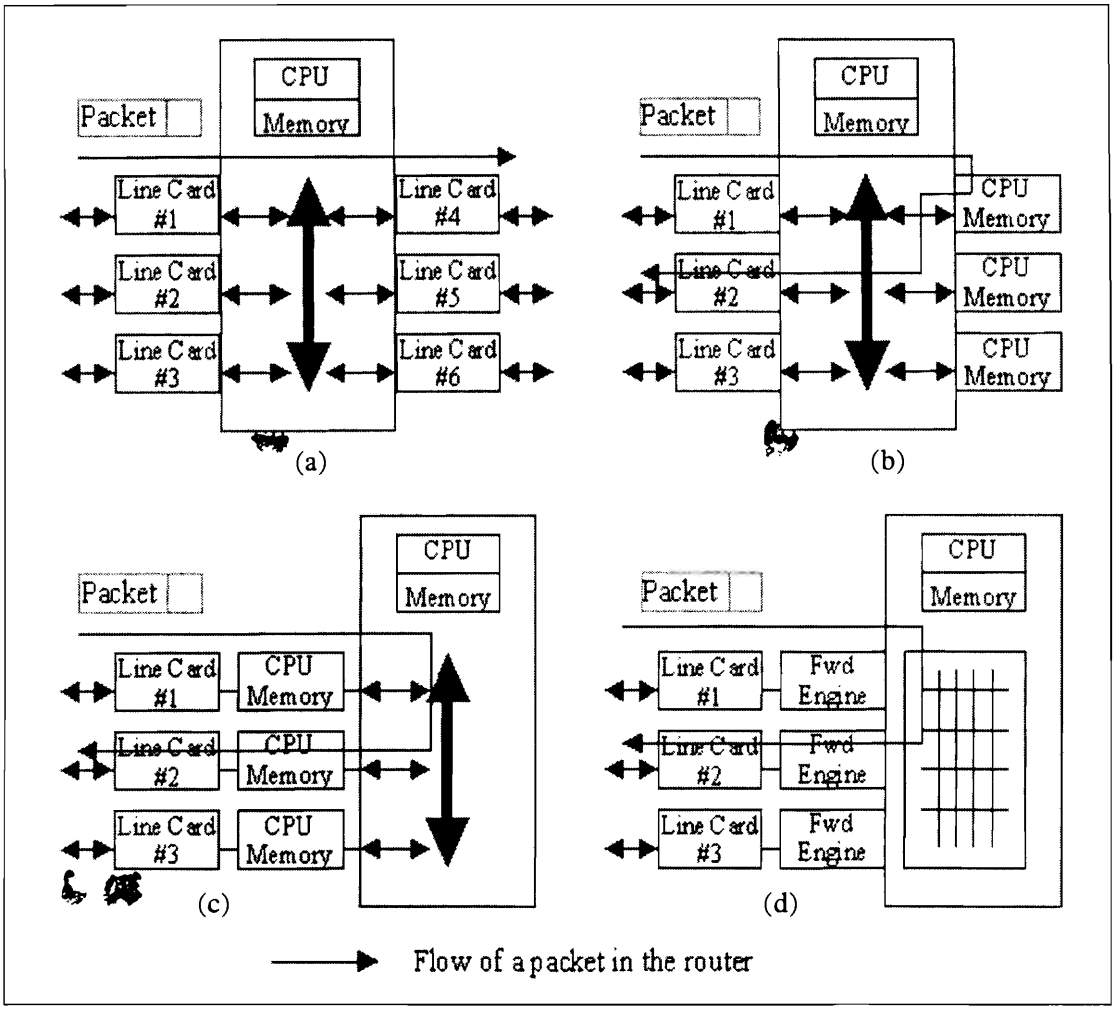

Figure 6. Evolution of routers.

cards to the CPU via the bus. The CPU makes the forwarding decision and then the packet is passed onto its external output link via the bus and a line card. This architecture had a major limitation. Every packet had to be processed by the central processor. This increased the average time for which a packet had to wait in a queue. This seriously decreased the throughput (number of bits transmitted per second).

To overcome this limitation, routers having multiple CPUs and Memories were introduced (*Figure* 6b). The incoming packets are handled in the same way, but a packet can be sent to a free

Today's high performance routers make use of both ASIC's (forwarding engines) and Crossbar Switches.

CPU or one CPU can be reserved for packets of a common destination. The potential parallelism in packet handling increased the throughput. But then, each packet had to travel across the bus twice, once from the Line card to the CPU and then again from the CPU to an output line card.

So in the next generation of routers, a CPU was placed at each interface (*Figure* 6c). The forwarding decision is made at the local dedicated CPU. This allowed the packets to be immediately forwarded to an outgoing interface. A central CPU is also needed. But this is limited to maintaining the forwarding tables in other CPUs and functions of the management plane.

General purpose CPUs are more suited for applications in which, the same data has to be examined many times by making efficient usage of a cache. But in the case of packet processing in a router the data to be examined is usually different in each case. Carefully designed special purpose Application-Specific Integrated Circuits (ASICs) can easily outperform a CPU in issues like managing queues, controlling access to the bus and making forwarding decisions. The older routers typically used a shared bus. This was the biggest bottleneck as only one packet can be in the bus at one time. Hence, a *Crossbar Switch* was used to allow multiple packets to traverse the bus simultaneously. Today's high performance routers make use of both ASICs (forwarding engines) and Crossbar Switches (*Figure* 6d).

Speedup of Routers

When a crossbar switch is used, packets from different source and destination interface pairs can be simultaneously transmitted. The packets are only delayed in case two packets from different inputs are competing for the same output. But even here a hidden problem arises. Many packets can arrive at the same time at an interface. They get queued up in the interface's FIFO queue. The forwarding engine can only service the packet at the head of the queue. In case the output interface that this packet has to traverse through is already busy, then all the

packets that are queued up behind have to wait until it is serviced, irrespective of whether their destination interface is free or not. This kind of blocking is called *Head of Line* (HOL) blocking [6]. HOL blocking degrades the router performance to just 60% of its potential bandwidth. This is a significant degradation issue and must be eliminated. Fortunately a solution does exist. A separate FIFO queue can be maintained for packets of each interface. When a packet arrives at an interface it moves into the queue reserved for its output interface. This method is called *Virtual Output Queuing* (VOQ). This enables the usage of the entire (100%) bandwidth of a router.

Input- and output-blocking do not reduce the throughput of a crossbar switch. They only make the delay that a packet needs to traverse the router unpredictable.

Each input and output line of a crossbar can transmit only one packet at a time. So if multiple packets from different VOQs in an interface try to access the crossbar, only one of them gets the access. This is called *Input-Blocking*. Also if packets from different input interfaces try to access the same output interface, then only one of them is given access. This is called *Output-Blocking*. Input- and output-blocking do not reduce the throughput of a crossbar switch. They only make the delay that a packet needs to traverse the router unpredictable. But some packets may need faster service than others, e.g., packets of multimedia applications. So priorities can be assigned to packets of different traffic flows. The higher priority packets are given preference. This is called *Prioritization*.

The fraction of multimedia applications traffic in the Internet is increasing day by day. So just prioritization may not suffice. What if, two packets of the same priority want to access the same crossbar at the same time? The amount of input- and output-blocking can be reduced by making the crossbar faster than the external line rate. For example, if the crossbar is made twice as fast as the external line, then two packets from each input and output interfaces can be transmitted in one packet cycle time. This technique is called *Speedup*. The ideal degree of speedup is currently an active research area.

Another key factor that can affect the efficacy of a router is the

End-systems access the Internet through a tiered hierarchy of Internet Service Providers (ISPs). Each ISP is essentially a network of routers and communication links.

Store and Forward method used to forward packets. In this method the entire packet needs to be read into an internal buffer thus necessitating the presence of a large buffer. This poses a significant overhead in busy interfaces. One way to eliminate it is by examining only that part of the packet header, which contains the destination address. This is called *Cut-Through* method. The inherent disadvantage of this method is that corrupted packets cannot be identified. A potential trade-off is to examine the first few bytes of a packet, thereby enabling the router to identify the destination and also verify the integrity of the packet. This is called the *Modified Cut-Through* method.

Deployment of Routers

End-systems access the Internet through a tiered hierarchy of Internet Service Providers (ISPs). Each ISP is essentially a network of routers and communication links. The ISPs may be residential ISPs (eg: AOL, Sify), university ISPs (eg: ERNET) and corporate ISPs (VSNL). Lower tier ISPs are interconnected with National and International ISPs (Tier1 ISPs) to ensure that users get access to worldwide Internet content. End-systems such as Desktop PCs, Unix-based workstations, Web servers and mobile computers running from home/small business environments connect into the Internet via an access network. Different ISPs provide a variety of different types of network access to the end-systems such as dial-up modem access, broadband access (Digital Subscriber line, cable modem), high-speed LAN access or wireless access. Enterprise/campus networks ensure interconnectivity among a large number of end-systems within a campus or enterprise. Enterprise routers apart from ensuring connectivity also need to make sure that the network performance does not degrade with increase in network size, support legacy LAN technologies, support multiple network protocols (IP, IPX, AppleTalk) and adheres to extensive administrative and security policies. Tier1 ISPs such as UUNet, Sprint, AT&T, Cable & Wireless are known as Internet backbone networks. The high-speed routers that interconnect these Tier1 ISPs and very large enterprise networks are called backbone routers.

Conclusion

Routers form an integral part of the Internet infrastructure. As we have seen in this article, without routers it is impossible to facilitate communication among Internet users. Rapid advances have led to the development of high-end routers, which are capable of transmitting many Gigabytes of data every second thereby satisfying the bandwidth needs of customers at competitive costs.

Suggested Reading

[1] William Stallings, *Data and Computer Communications*, Prentice-Hall, 2000.

[2] V Rajaraman, *Fundamentals of Computers*, Prentice-Hall, 2001.

[3] James F Kurose and Keith W Ross, *Computer Networking*, Addison Wesley, 2000.

[4] Nick McKeown, *Fast Switched Backplane for a Gigabit Switched Router*, Stanford University.

[5] Alberto Leon-Garcia and Indra Widjaja, *Communication Networks-Fundamental Concepts and Key Architectures*, McGraw-Hill, 2000.

[6] *The Evolution of High-End Router Architectures*, Cisco Systems.

[7] S Keshav and R Sharma, Issues and Trends in Router Design, *IEEE Communications Magazine*, May 1998.

Address for Correspondence
Prashant Bharadwaj
No. 273, 10th Cross
N.R.Colony
Bangalore 560019, India.
Email: prashant.bharadwaj@wipro.com

There is only one nature – the division into science and engineering is a human imposition, not a natural one. Indeed, the division is a human failure; it reflects our limited capacity to comprehend the whole.

– Sir William Cecil Dampier Whetham