
Lesson 1: Quality of Service

At a Glance



Anyone working with networks is concerned with the issues behind the quality of the service provided by the network. Quality of Service (QoS) is a measurement of a network's performance, reliability, and availability. The merging of existing voice, data and video networks and the development of new voice applications, such as call centers and IP telephony, require improvements in QoS.

QoS consists of many confusing technologies, applications, and issues. This lesson is designed to provide only an overview of QoS basics, including the key concepts behind QoS and some of the emerging QoS standards.

What You Will Learn

After completing this lesson, you will be able to do the following:

- Define QoS and describe why it is important
- Identify key concepts in QoS
- Identify congestion services and management tools
- Compare RSVP, QoS, and MPLS
- Create a network architecture and a network diagram

Student Notes:

Tech Talk

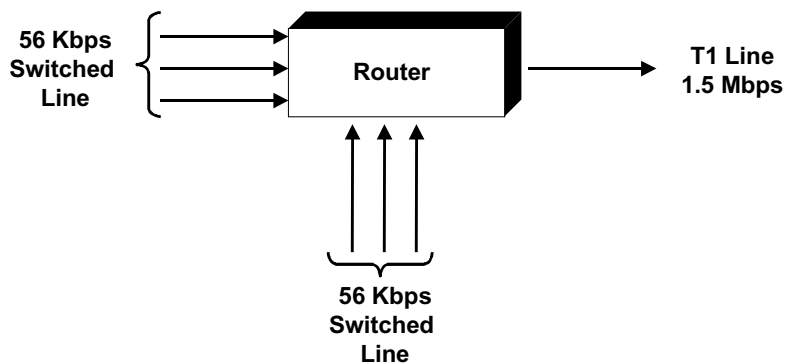


- **Forward Equivalence Class**—Priority packet-forwarding class within MPLS based on different criteria, such as destination or specified service requirements.
- **Forward Information Base**—MPLS routing table that contains information matching the best route for a specific forward equivalence class.
- **Label Edge Router (LER)**—MPLS router that segments IP packets and attaches MPLS label at the entrance of the Label Switching network. At the exit of the network, the LER removes the labels of the cells and reassembles them into the IP packet.
- **Label Switched Path**—A routing path created by Label Swapping.
- **Label Switching**—A MPLS process where the incoming label encapsulated within the IP header of a packet arriving to a router is with an outgoing label.
- **Label Switching Network**—An MPLS network domain consisting of label switching and edge routers.
- **Label Switching Router (LSR)**—MPLS router that forwards cells within a Label Switching network based on the cells' label.
- **Mission Critical Applications**—The most important applications that an organization relies upon for its success, e.g., payroll processing or inventory control. The requirements of these applications define those services for which a network must provide acceptable performance.
- **Priority**—Having importance over another. In networking, some packets may be assigned a higher priority than others, thus allowing them to travel faster across the network.
- **Queue Depth**—The maximum number of packets that can reside in a queue.
- **Queuing**—The process of storing packets for later transmission across a network.
- **Service**—A specific function that provides support or a specific network condition.
- **Traffic Engineering**—The ability to efficiently direct traffic onto an existing network topology in such a way as to optimize the use of network resources.

What is Quality of Service?

In the past, applications were designed to be self-reliant. Most applications could sense packet loss and delays, and then recover automatically. Additionally, networks operated on a best effort basis, either the packet was transmitted and received successfully or it wasn't. Network managers engineered their networks to maintain an expected level of service by segmenting the network or adding additional bandwidth at points of line convergence, so that there was never more demand than the network could accommodate. The goal was to simply keep the traffic flowing. This practice wastes bandwidth.

Overengineering



Prior to the rise of the Internet, network traffic was usually a local concern. Networks were typically self-contained and methods for handling data collisions, traffic congestion, processing priority, and security were developed for each site. Sudden changes in traffic volume or patterns were not as common prior to the commercial Internet.

The effective management of network traffic beyond simple success or failure is a recent concept. With the advent of the Internet and the move toward the unification of voice, data, and video networks, the nature of network traffic is changing. Today network managers are implementing methods to ensure high quality of service, thus providing the network and application performance expected by the end user.

Quality of Service (QoS) refers to the ability of a network to provide better and more consistent service (support or network provision) to selected packets and network traffic based on pre-established criteria. These criteria may be related to the speed with which transmission occurs, reserving a portion of the network for exclusive use, or the setting of priorities that cause some packets to be processed before others.

QoS provides better and more predictable network services by:

- Supporting dedicated bandwidth for a specific traffic flow.
- Reducing the occurrence of lost packets.
- Managing network congestion.
- Setting traffic priorities across a network. Packets with a higher priority are considered more important and typically passed across the network before lower priority packets.

Ultimately, the goal of QoS centers on the KISS principle, Keep It Simple, Stupid. It is important to create the best possible and most reliable network, while implementing the simplest QoS approach that supports the network.

Check Your Understanding

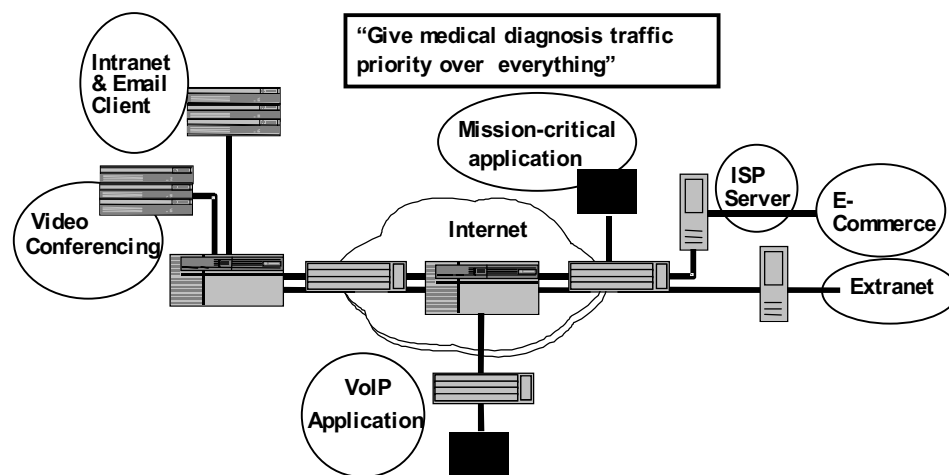
- ◆ Overengineering a network wastes bandwidth because the line serving the point of convergence (the router) has far more bandwidth than needed, even with all the lines operating at full capacity. Network managers practice another method called "oversubscribing" the line. In this case, the network manager assumes that no lines will ever operate at full capacity. In fact, much of the time each line may only operate at 10% capacity. Therefore, the manager configures the network so that several lines converge to one line with a lower transmission rate than would be needed to accommodate the incoming lines if they were transmitting at full capacity. Diagram oversubscribing using the same model presented in the lesson for overengineering.
- ◆ Speculate as to what problems using oversubscribing may cause.

Why is QoS Important?

Quality of Service is already offered in some traditional network environments, such as voice, ATM, and Frame Relay. ATM offers the ability to classify network traffic into service categories that match specific quality of service features with the performance requirements of the data traffic. For example, Constant Bit Rate (CBR) is often implemented for voice connections. ATM is typically a WAN solution and it is rarely implemented over LANs. The QoS provided by ATM is typically not extended to the desktop. Consequently, network managers must implement other QoS technologies in the LAN and then make sure the different technologies can interoperate. Frame Relay allows a network manager to change the allocation of bandwidth based on network traffic conditions, but it does not provide any guarantees against packet loss or delay.

Despite the current QoS available for ATM and Frame Relay, there is a need to increase and standardize quality of service to address the issues underlying the diverse nature of the Internet and the complexities of unifying voice, data and video networks. The number of users is growing exponentially due to the Internet. Progressive applications are under rapid development that have special bandwidth requirements and create new traffic dynamics, e.g., VoIP. Network managers must be able to provide high quality of service to maintain the performance expectations of network users.

Forces Driving QoS



The principal benefits that are gained from using a Quality of Service approach to network configuration are:

- **Control over resources**—The network manager may give priority to specific processes; for example, payroll processing may be given higher priority, over FTP or other file transfer protocols.
- **Efficient use of network resources**—QoS tools help the network manager to learn how the network is being used and if it is processing transmissions in the most efficient manner.
- **Tailored services**—Quality of Service allows the network manager to tailor specific types of processes to particular needs, such as providing a high-speed path for an Internet Service Provider (ISP) or pre-processing packets for faster flow across the network.
- **Reduction of interference with mission-critical applications**—The most important applications for an organization are guaranteed resources (e.g., specific bandwidth allocation), while others can share resources without being completely brought to a halt.
- **An integrated network**—The fully integrated network of the future is recognized as one that can be tailored to all of its users needs, and QoS is a first step in that direction.

Key Concepts in QoS

To understand the implementation of QoS, there are a few key concepts that must be understood. Flows, delay or latency, jitter, bandwidth, and reliability are important factors that must be monitored to effectively measure QoS.

Flows

A flow is defined as a stream or set of packets that have similar characteristics. Packets can belong to a flow based on an observable quality. The packets within a flow may share the same source network, the same source and/or destination address, or they were generated by the same application.

Each device that moves data within a network keeps track of the data flow passing through it. In networking, the word "states" is used to describe the way the network keeps track of the flow. The critical issue centered on states is the maintenance of information about the data packets flowing through a network device, whether it is a router or a switch. States are divided into two categories: stateful and stateless.

- **Stateless**—In stateless mode, the device is keeping no record of the history of the data flow. Routers maintain routing tables that inform the router the next step a packet should take toward its destination. The creation and maintenance of the routing table is independent of the actual data flows that pass through the router. In terms of QoS, as long as the packets are not entering the router at a greater rate than they are leaving, there is no need for the router to examine the packet headers for QoS information. The need to prioritize is only necessary when there is competition for network resources (e.g., bandwidth).
- **Stateful**—Stateful mode can be divided into two types: hard and soft. A device in hard state is maintaining information about a preset path for a particular data flow. For example, voice calls require signaling that sets up a specific path from one telephone to another across multiple switches. Each switch along the path must track the status of the call. This status remains for the length of the call. In terms of QoS, maintaining a hard state taxes the resources of the switch or router and may result in delaying the flow. A soft state also sets up a path, but the path information will age out and requires periodic refreshing. In other words, the information becomes "old" and is dropped by the router or switch. The device must then re-gather the information. This process allows the device to recover its resources periodically and allows the network to recover faster during a network outage.

Latency

Latency or delay is the elapsed time for a packet to travel from the sender to the receiver. There are many factors that can slow this packet down during its journey, but one of the greatest is latency. Latency exists in primarily two forms: real latency and induced latency.

- Real latency is caused by limitations by the physical devices and speed of the media. Data traveling on fiber is “limited” to the speed of light. If the network media is copper then the network's real latency is increased. Copper is simply a slower transport medium than fiber.
- Induced latency is introduced into the network by such factors as queuing delays, processing delays and congestion at different points in the packet’s transmission path. Larger packets take longer to travel across a network than small packets, thus increasing latency. Delays are created each time a switch or router must look up addresses or change a header.

Minimizing latency across a network is an important factor in increasing the quality of service.

Jitter

Jitter is the difference in end-to-end delay caused by lack of synchronization of the data packets, which creates distorted data, especially when it occurs with real-time media, such as audio or video. From the time several packets depart from one end and arrive at the other end, there are many factors that will affect when the packets will actually arrive. As a packet travels across a network, it traverses routers and switches, entering and leaving queues at different rates. If the packets are out of order when they arrive, due to significant delays, the data can be heavily distorted.

Queue management and the layout of the network play an important role in reducing jitter.

Bandwidth

Bandwidth is the maximum data transfer rate that can be sustained between two end points. The word "sustained" is important. The total bandwidth of a connection is not the issue. The issue is how much bandwidth can a connection maintain over time for a specific data flow. The limitations to bandwidth relate to the type of media and the fact that there may be other transmissions competing for that same highway. A key process in QoS is allocating different bandwidths for different data flows.

Reliability

Reliability is concerned with the average error rate that occurs during transmission. If the error rate is at an acceptable level for the class of service, then the network and transmission path can be considered reliable. If there are more errors than an acceptable level, then this is considered a less reliable connection than is desired.

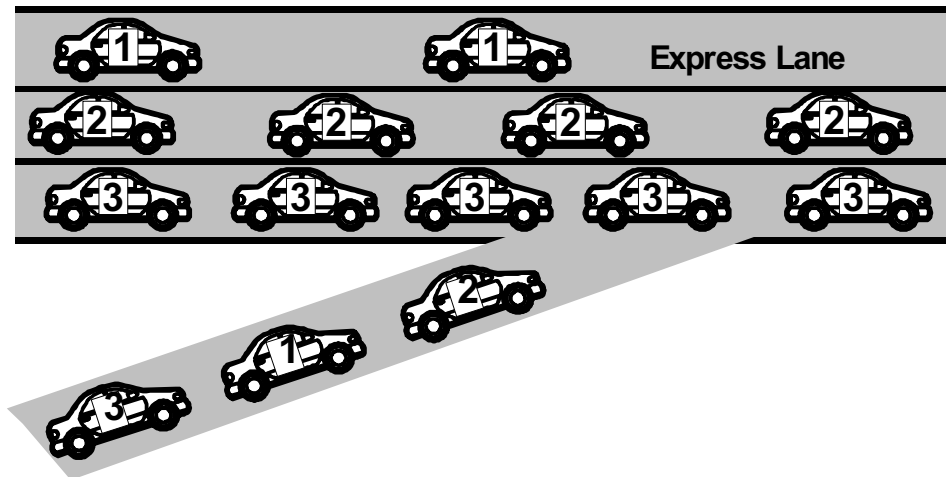
Classes of Service (CoS)

A concept included in Quality of Service is the idea that different types of transmissions belong to different "classes." The concept or CoS is similar to traveling on a plane. People boarding a plane are seated by class, and then by row. Usually those passengers who have paid more for a larger, more comfortable First Class seat are allowed to enter the plane before those who paid less for the smaller Coach or Second Class seats. First Class passengers are given "priority" over Coach class passengers. There may be further differentiation as well, when people with small children are given the opportunity to enter earlier to give them time to settle their children.

Thus, classes allow the airline to differentiate between passengers, or more specifically, "types" of passengers. They do not want to know what the passenger's eye color is, or where they were born. What is important to the airlines is what level of service is expected. This is the concept behind Classes of Service; it provides the network manager who is implementing Quality of Service processes to differentiate between expected levels of support, speed, consistency and accuracy.

If an expected level of service can categorize all network traffic, then the CoS can establish rules and algorithms to handle data transmissions. A packet that has a high priority because it is considered critical to the mission of the organization will be given as much bandwidth as is needed and travel the most dependable, collision-free paths across the network to its destination. Essentially, it is given the right to use the "express" lane of the highway. There is no need for the software processes or the network manager to know what exactly is in these packets while they are processed, only which class of service is expected.

Class of Service



Class of Service is often used synonymously with Quality of Service, but in fact CoS is part of the definition of QoS. Quality of Service measures how well a network behaves and it includes the categorizing of network traffic. Class of Service is the actual process of service categorization of distinct classes that can be treated individually or in relation to each other.

Congestion Management Service Levels

An integral part of QoS is congestion management across a network. QoS has three basic levels of service for decreasing congestion across a network.

- **Best-Effort Service**—This service is also known as lack of QoS. There are no guarantees made about what will happen to packets transmitted. No priorities are set and there is no guarantee of available bandwidth. As network congestion increases the performance decreases. Service can only be improved by adding more bandwidth and connections. Think of it as adding 4 more lanes to a 2-lane highway, which provides more space in which to travel for each car. Dense Wavelength Division Multiplexing (DWDM) and Gigabit Ethernet are being developed for this simplest form of congestion management.
- **Tiered Services**—Tiered services treats some traffic better than the rest. The service provided is based on statistical probability, not hard-and-fast rules. In other words, some traffic will be given preference if it occurs in a predictable frequency, but there are still no guarantees of service quality. It allows higher priority traffic to be moved ahead of lower priority traffic. One tiered service, Assured Forwarding (AF), does include four priorities of traffic, each receiving different allocations of bandwidth. However, during times of high utilization, some packet delay or loss may still occur. This service is useful for packetized voice and video or mission-critical applications.
- **Guaranteed Service**—With guaranteed service, resources are reserved for specific traffic, no matter what may happen to compete with it. This service guarantees that the data flow should experience no problems with performance. ATM's Constant Bit Rate traffic classification guarantees service. At times bandwidth can go unused in network configurations using guaranteed services. Guaranteed services are excellent for applications such as voice, which require no packet loss, minimum latency and jitter, or both.

Deciding which level of service is needed should be based on the application or problem presented to the network, whether the network infrastructure can realistically be upgraded, and, of course, the cost of upgrading service to higher levels. The higher the level of guaranteed service, the higher the implementation cost.

Check Your Understanding

- ◆ Create a diagram that demonstrates the concepts behind "best effort" service.
- ◆ Create a diagram that demonstrates the concepts behind "guaranteed" service.

Congestion Management Tools

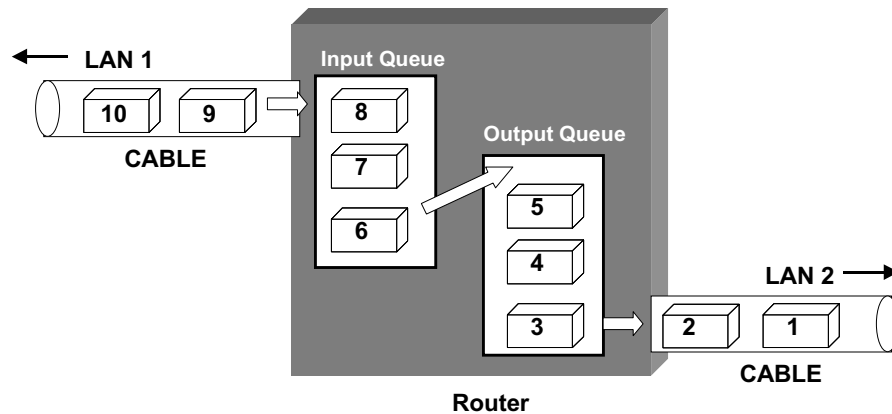
In order to manage traffic throughout the network, tools have evolved in the form of queuing algorithms. Queuing algorithms are congestion management tools that sort network traffic and then determine some way of prioritizing them for travel across a network.

"Queuing" means to store packets for later transmission. Queues exist for both input and output of packets. Network managers are typically more concerned with the removal of packets from output queues and the effect queue depth has on the entire process. Queue depth is the maximum number of packets that can reside in a queue. When a queue is too deep (a lot of packets are stored in the queue), excess delays may occur causing transmission sessions to timeout. If a queue is not deep enough, the network device may not have the ability to process packets fast enough, and packets may be discarded.

First-In, First-Out Queuing (FIFO)

First-In, First-Out queuing is the simplest and standard method of queuing. All packets leave the queue in the same order they arrive into the queue. There are no decisions made about packet priority, and processing solely depends on the order of arrival. It is much like shopping at the supermarket. Customers line up to pay for their groceries in a FIFO line. FIFO is a very simple means of sequencing packets when bandwidth is plentiful, but as traffic increases and bandwidth resources shrink, the queue becomes full and packets are dropped as they arrive.

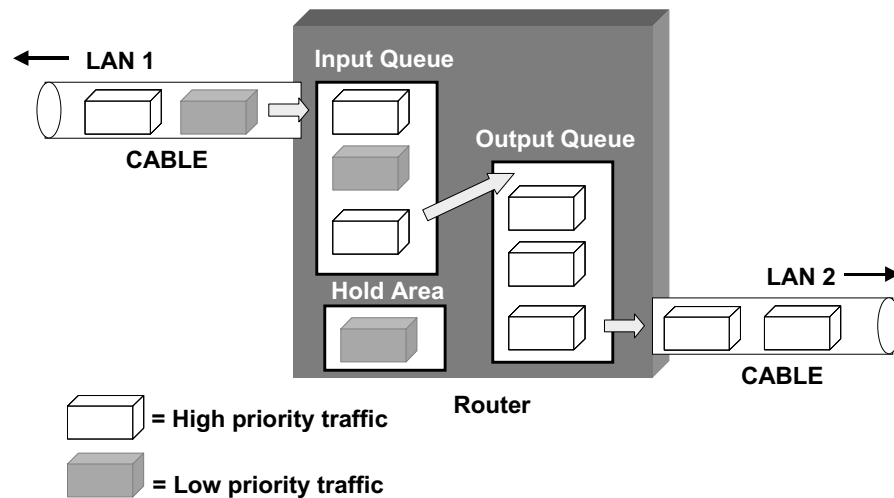
FIFO



Priority Queuing (PQ)

Priority queuing gives important traffic the fastest path through the network, regardless of what other traffic may be competing for the same resources. This hierarchy can be based on several characteristics of the packets, such as packet size (e.g., smallest packets travel first), source/destination address, or tolerance for delay (e.g., less tolerant travel first). When no more high priority packets arrive, and there is room in the output queue, the other packets are sent. In PQ, each packet is placed in one of four queues: high, medium, normal or low. The default for any packet not tagged by PQ is the normal queue. The biggest problem with PQ is that high priority traffic could be transmitted at the expense of lower priority traffic. Over time, lower priority traffic that has not left the queue could timeout and be discarded. PQ also requires more processing time, which may increase latency.

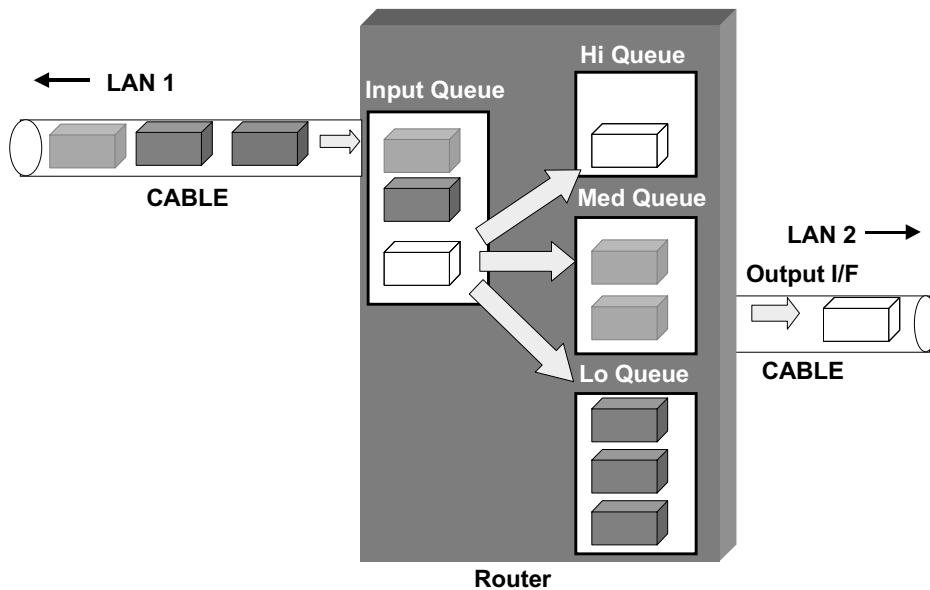
PQ



Class-based Queuing (CBQ)

Class-based queuing uses multiple output queues and packet classification. Packets are assigned a queue based on their class of service. Queues are typically serviced by removing a set number of bytes from a queue, moving from queue to queue until all queues are serviced. The network manager can set what number of bytes is removed from each queue. For example, the network manager configure the queues so that 500 bytes is removed from the high queue, 300 bytes from the medium queue, and 100 bytes from the low queue. All queues are serviced, but the high queue has 5 times as many bytes removed as the low queue. Even though all the queues are serviced, during periods of high traffic, it is possible none of the queues will be serviced fast enough and packet may be dropped.

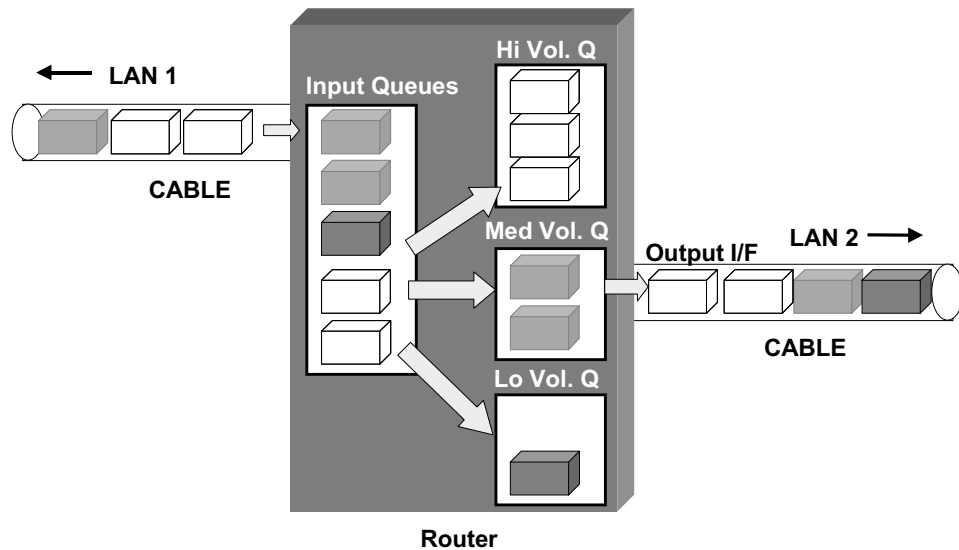
CBQ



Weighted Fair Queuing (WFQ)

Weighted Fair queuing takes the unique approach of favoring low-volume traffic over medium-volume traffic, in order to reduce the interference of low volume transmissions. High volume traffic is still allowed to transmit in a reasonable period of time. WFQ sorts and re-distributes packets based on the volume of particular types of traffic. This method of controlling traffic varies from manufacturer to manufacturer. Just as with CBQ and PQ, the time required measuring and maintaining flow information and then altering the placement of packets in the queues could create significant delays.

WFQ

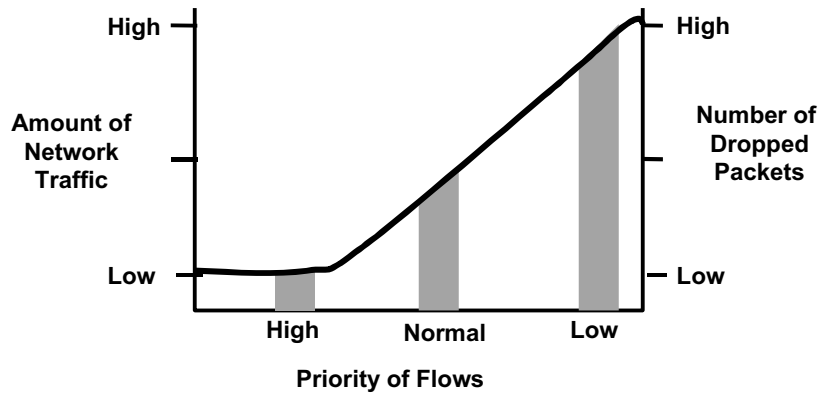


Weighted Random Early Detection

Finally, the alternative to congestion management is congestion avoidance, and a popular method of avoiding congestion is Weighted Random Early Detection (WRED). With the earlier Random Early Detection (RED) algorithm, network managers set queue depths for each flow. As congestion increased and the queues exceeded their thresholds, packets were randomly discarded. No special consideration was given to any traffic.

WRED combines this random process with some prioritization. Certain flows are assigned priority and become less likely to be discarded, based on an algorithm devised by the network manager. For example, when congestion occurs, packets with no priority begin to be discarded. As congestion increases, packets with no or low priority may be discarded, and so on.

WRED



Check Your Understanding

- ◆ What is the difference between Priority Queuing and Class-based Queuing?
- ◆ In some books, the concept of queuing is discussed in terms of "fairness" and "unfairness." Speculate how these words might be used when discussing queuing.

Emerging QoS Protocols

Along with traffic prioritization and congestion management techniques, there are several emerging protocols that are expected to provide tools for network managers to improve the quality of service across their networks.

RSVP

The Resource Reservation Protocol provides a way for an application to communicate its QoS requirements to the network. As a network control protocol it allows network applications to request priority service for transmitting its data. This requires the reservation of bandwidth across the path. Every router along the path must be RSVP enabled in order to support the QoS requests.

The receiver initiates all reservations. When a router receives a reservation request, it checks to see if the outgoing line has enough bandwidth to accommodate the request. If there is enough bandwidth, then the router reserves the bandwidth specifically to pass on the data flow to the next router along the path.

Reservation requests are allocated on a first-come, first-served basis. Once all the routers between the sender and the receiver have allocated the requested bandwidth, that amount of bandwidth is used for that application flow until the application finishes, or there is a topology change (e.g., a router is removed).

Quality of Service Routing (QoSR)

RSVP is not a routing protocol. It depends on the routing protocols, such as the Routing Information Protocol, to determine the topology of the network. RSVP can not initiate changes in the routing tables used by the routers to determine the best path for delivery of packets. Each time there is a change in the topology, the reservation for a specific application must be refreshed. RSVP must start over. This is considered a drawback to RSVP.

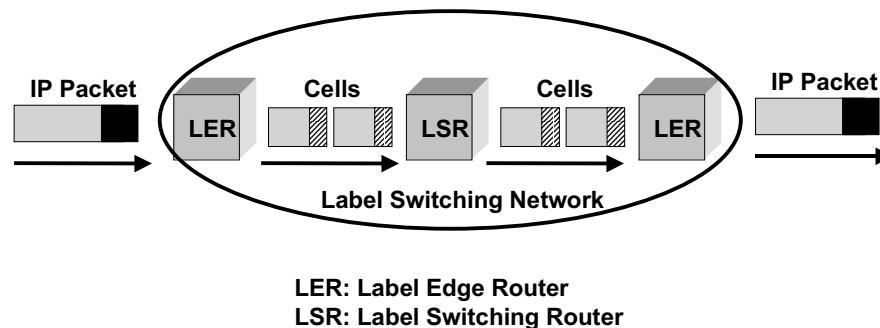
The Internet Engineering Task Force (IETF) is working on a solution that combines some of the benefits of RSVP with the abilities of existing routing protocols. Existing routing protocols forward packets based on the shortest path from the source to the destination. The new routing protocols, Quality of Service Routing, will allow the forwarding of packets based on available resources (such as bandwidth). QoSR will forward a packet along a path that meets the class and quality of service requested by the user.

Multi-Protocol Label Switching (MPLS)

Routers maintain information about the network topology in routing tables. When an IP packet arrives, the router uses the Internet Protocol to get the destination address from the IP packet header. The router looks up that address in its routing table and forwards the packet to its destination along the designated shortest path. The routing algorithms used by routing protocols tend to converge traffic onto the same network links or interfaces, which contributes significantly to congestion. This results in poor quality of service.

IETF is developing a new protocol, Multi-Protocol Label Switching (MPLS), that will significantly improve traffic engineering and improve quality of service. Traffic engineering can be defined as the ability to efficiently direct traffic onto an existing network topology in such a way as to optimize the use of network resources.

MPLS



There are several steps to MPLS.

1. Information about the network topology is maintained by a standard routing protocol. MPLS is implemented within an interior domain of the entire network. The domain is called a Label Switching network. Non-MPLS routers send IP packets to MPLS routers located at the edge of the domain. The edge routers are called Label Edge Routers (LER). Routers lying along the route between edge routers are called Label Switching Routers (LSR).
2. For each route and links in the network, a label is assigned.
3. Each router maintains a record of the incoming label and its associated interface (the port on the router), and the corresponding outgoing label and interface. The interface associations are based on the information in the routing table.

4. When an IP packet arrives from a non-MPLS router, the LER
 - receives the packet and examines its routing table to find the associated outgoing interface and label,
 - segments the IP packet into cells,
 - inserts the correct outgoing label in front of each cell in the header,
 - and forwards the packet onto the Label Switching network..
5. The cells pass through LSRs along the route, each LSR
 - receives the cells,
 - examines the routing table to find the associated outgoing interface and label
 - inserts the correct outgoing label for the incoming label,
 - and then forwards the cells over the outgoing interface.

The switching of labels in the packet header is called Label Switching. The path through the network created in this manner is called a Label Switched Path.

6. When the cells arrive to another LER, the LER
 - receives the cells,
 - extracts the labels,
 - reassembles the IP packet,
 - and then forwards the packet on to a non-MPLS router. From there the packet is forwarded on to its final destination.

MPLS also provides QoS using packet-forwarding classes, called Forward Equivalence Class. Members of a forward equivalence class may have the same destination, or specified service requirements. Each router builds and maintains a forwarding information base, a table with one row per forwarding equivalence class. Using class information in the packet header, the router refers to the forwarding information base to determine the path on which to forward the packet based on the packets QoS requirements.

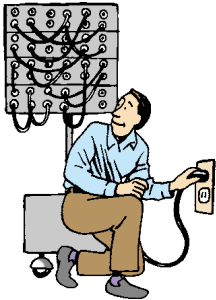
Check Your Understanding

- ◆ Based on your knowledge of RSVP from Unit 5, Lesson 2, and the information presented in this lesson about QoS, diagram how a packet will cross a network using QoS.
- ◆ What aspects of MPLS are similar to ATM?

Try It Out: Network Reliability and Availability

Materials Needed:

- Windows 95 PC
- Internet Connection (optional)
- Any Word Processor (e.g., MS Word)
- Pen/Pencil and Paper
- Poster Board



In this activity, you will learn about three important measures of performance in data communication networks: reliability, availability, and serviceability. You will manipulate some of the numbers that are used to measure reliability and availability, and will apply your investigations to discussions about improvements to your school network.

Part 1: Measures of Communication System Performance.

There are three different ways to measure the performance of any communications system:

- **Reliability**—The probability that a system, or its components, will not fail. Reliability is measured in MTBF (mean time between failures). An example of reliability is that your telephone system rarely goes down while you are making a call.
- **Availability**—The probability that the system's services will be available at any given time. This is often expressed as the probability that an individual user requesting service will receive it. An example of availability is that you almost always can get a dial tone when you pick up the telephone.
- **Serviceability**—The ease with which a system can be repaired in the event of failure. This is often expressed as MTTR, (the mean time to repair). An example of serviceability is the 1 hour turnaround that phone companies give to problems on T1 data lines.

Make a list of factors you believe affect Reliability, Availability, and Serviceability of networking systems.

Part 2: Reliability of a Network System.

Each device in a communications system has a mean time between failures. These MTBFs are usually in the order of 50,000 hours. If we plug 10 of them in today and log on, they will run for around 50,000 hours before they will fail.

Why do communications devices fail at all? They fail because their components fail. Heat and other environmental factors usually cause failures. Also, if we choose higher quality components they will last longer.

1. If a communications device has a MTBF of 50,000 hours, and we plug a new one in today, when can we expect it to fail (on average)? HINT: How many hours are there in a year? Show your work.

The probability that a device will fail today is usually stated as a number like .0001. That means that there is one chance in 10,000 that it will fail. Another way to say that is that it has .9999 reliability.

Because communications networks depend on many devices working together, and the path between two users on a network depends on NICs, hubs, switches, routers, and other devices the calculation of reliability is complex.

2. What do you think happens when we use four devices with .99 reliability in a simple network system? How reliable will the whole system be? Show your work.

To calculate the total reliability of a system with four serial components, all needed for it to operate successfully, we multiply their individual reliabilities together. So, $.91 \times .91 \times .91 \times .91 = .69$. It doesn't take long for complex systems to become unreliable.

3. Calculate the total serial reliability of six devices with the following individual reliabilities: .997, .998, .92, .9999, .995, .9999999. Show your work.
4. Research reliability on the Internet, specifically looking for strategies for improving reliability in networks.
5. Working in a group of five, discuss strategies to improve reliability in networks.
6. Of the strategies discussed, choose one and create a poster for the classroom that outlines the strategy

Part 3: Availability of a Network System.

Availability is a complex concept. You can have a system that is reliable (it seldom fails while you are using it), or that is not very available (it is hard to get a connection, but when you do, it is a good one).

You can have a system that is available (it is easy to get a connection), or that is not very reliable (it almost always fails while you are using it).

1. Research the issue of availability in networks. Use the Internet or any other resource you find helpful. Answer the questions below.
2. What are some of the things we could do to improve the availability of a network?
3. When would we want a network to be unavailable?
4. Which is worse, a network that is unreliable, or one that is unavailable?
5. What do you think are some of the tradeoffs that affect network performance?

Part 4: Serviceability of a Network System.

Serviceability is measured in mean time to repair, MTTR. Because we know that all components fail with some frequency, we have to plan service strategies to repair them efficiently.

One strategy to reduce MTTR is to stock spare parts at various places throughout a network. Another strategy is to place repair people at various places throughout the network. Another strategy is to maintain a hot standby unit (one already plugged in and ready to go).

Before devices that fail can be repaired, the failures need to be diagnosed. In large complex networks this is not an easy task. Diagnosing failures (MTTD) is also an important factor in network design. Automatic diagnostics, loopback test scenarios, and automated test systems are used to reduce MTTR statistics.

The service standard for T1 lines is that they be down no more than an hour a year. That represents about 99.995% availability. If a T1 is down for longer than an hour Bell Atlantic will refund the customer a full month's charges. This may be \$ 300-900.

1. Research Internet for strategies for improving MTTR.
2. Using the information from your research outline strategies to improve MTTR in your school network.
3. Share your strategies with the class during class discussion.

Rubric: Suggested evaluation criteria and weightings:

Criteria	%	Your Score
Thoughtful answers to activity questions	20	
Quality poster covering reliability strategies	50	
Meaningful participation in class discussion	30	
TOTAL	100	

Stretch Yourself: Models for Implementing QoS

Materials Needed:

- Windows 95 PC
- Internet Connection
- Any Word Processor (e.g., MS Word) or Spreadsheet (e.g., MS Excel)
- Pen/Pencil and Paper
- Student Portfolio



On the Internet and in other networks, QoS is the idea that transmission rates, error rates, and other characteristics can be measured, improved, and, to some extent, guaranteed in advance. QoS is of particular concern for the continuous transmission of high-bandwidth video and multimedia information. Transmitting this kind of content dependably is difficult in public networks using ordinary "best effort" protocols.

Using the Internet's Resource Reservation Protocol (RSVP), packets passing through a gateway host can be advanced based on policy and reservation criteria arranged in advance. Using ATM, allows a company or user to pre-select a level of quality in terms of service. QoS can be measured and guaranteed in terms of the average delay at a gateway, the variation in delay in a group of cells (cells are 53-byte transmission units), cell losses, and the transmission error rate.

1. Identify some of the network applications that require a guaranteed level of service quality. Think about applications that might be possible if QoS was available on public networks.
2. What do we mean when we talk about "best effort" protocols?
3. Research on the Internet the subject of QoS and identify companies who have announced, or who provide, products that implement Quality of Service models.

4. For each of the companies, identify specifically what they call their service model, what its features are, and how it works. Create a chart that summarizes each model.
5. Bring to class discussion two QoS implementations that you found particularly interesting. Describe the strengths and weaknesses of each model.
6. Document your resources and include your research in your portfolio.

Rubric: Suggested evaluation criteria and weightings:

Criteria	%	Your Score
Thoughtful answers to questions	10	
Documented resources and inclusion in portfolio	20	
Thorough chart of QoS models	50	
Meaningful participation in class discussion demonstrating preparation of two QoS implementations	20	
TOTAL	100	

Network Wizards: Network Design Portfolio Case Study

Materials Needed:

- Windows 95 PC
- Any Word Processor (e.g., MS Word) or Spreadsheet (e.g., MS Excel)
- Pen/Pencil and Paper
- Student Network Design Proposal Working Draft



Part One: Quality of Service Issues

Your case study organization must certainly have quality of service issues that need resolving. Every network does.

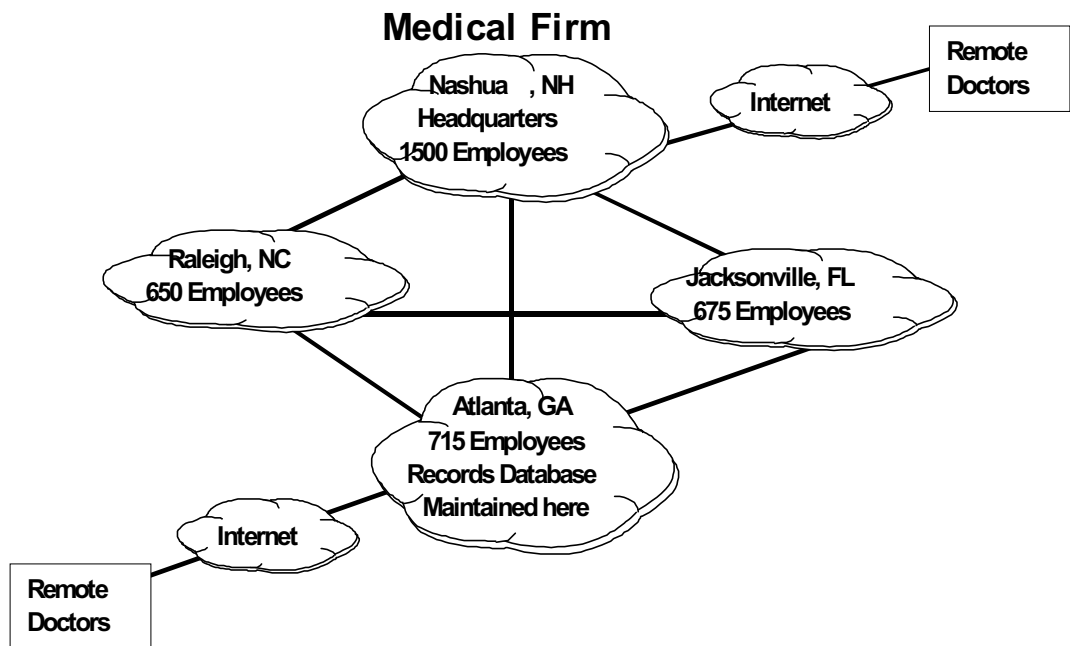
1. Review the data you gathered about the organization's network and future requirements.
2. From the data, evaluate the QoS issues that are applicable to your case study. Do they use special applications, such as VoIP or Video, which are critical to the mission of the organization? Are they planning to implement new applications that have special QoS requirements?
3. Create a chart of the QoS issues pertinent to your case study organization. Include in your chart solutions that could be implemented to address the QoS issues you listed. You may have to do some additional research. Be as thorough as possible, given your understanding of quality of service.
4. Share your chart with your teacher and your case study contact for review. Make any changes they suggest.
5. Include this chart with your network design proposal. Remember your chart should reflect careful and neat preparation.

Rubric: Suggested evaluation criteria and weightings:

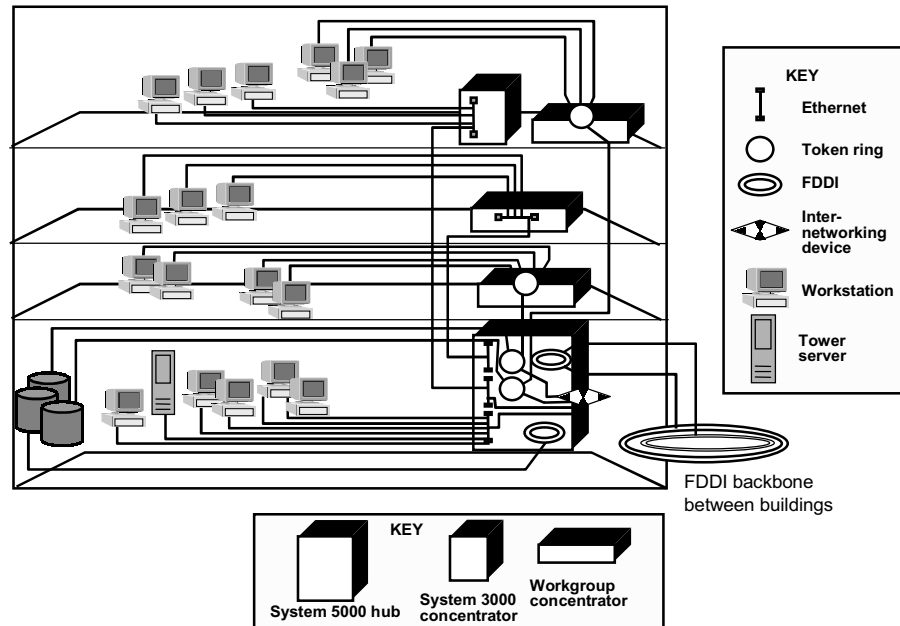
Criteria	%	Your Score
Thorough list of QoS issues based on gathered data	30	
Viable solutions to resolve QoS issues listed	30	
Carefully prepared chart suitable for presentation to client	40	
TOTAL	100	

Part Two: Network Architecture and Network Diagram

There are essentially two graphical representations of the network a designer creates; the network architecture and the network diagram. The network architecture represents the overall high-level, long-term view of the network. It is the basis for creating the network design. Think of it as the "grand plan," showing the overall direction of the network infrastructure. It typically does not include placement of routers, switches, or other network components. The network architecture diagram below represents a medical firm with four locations and remote doctors who need to connect to the firm's network. Notice that does not include information about line speeds or the placements of network components. It is just a high-level view of the expected network.



The network diagram, on the other hand, is very detailed. It gives the exact details of the network with specifics about products, technologies, installation, and configuration requirements. The future network is implemented based on the network diagram. It contains the exact topologies of the network. The network design is derived from the network diagram. The diagram below represents only one of the network design diagrams that might be prepared by the designer.



1. You are not expected to have the expertise to create a complete network diagram. To do so requires years of experience and training. However, you will create a network architecture diagram. This is a much simpler task and one that you have enough knowledge to produce.
2. With the business requirements and data you have gathered, create a high-level architecture diagram. It should be very neat and clear.
3. Write a short explanation of the diagram, including any justifications for how you laid out the network.
4. Present your diagram to your teacher for review and comment. Incorporate any changes and then present the diagram to your contact for review and comment. Incorporate any changes.

5. Request to see the network diagram created for the existing network at your case study organization. Have your contact go over the fine points of the diagram, explaining how to read the diagram and why components are placed where they are in the diagram.
6. Take notes about this diagram. Try making a network diagram for your network design proposal using the information you gained from viewing the network diagram from the existing network. Do not worry about getting every component in place. This is just an exercise, it is not expected that you will create a masterpiece.
7. Present your network diagram to both your teacher and your contact for review and comment. Incorporate their comments and include the diagram in your network design proposal.

Summary

Quality of Service

In this lesson, you learned the following:

- The definition of QoS and why it is important
- The key concepts in QoS
- The congestion services and management tools
- The overview of RSVP, QoS, and MPLS
- How to create a network architecture and a network diagram

Review Questions

Quality of Service

Part A:

1. What does Quality of Service mean?

2. Why is QoS so important for the networks of today and in the future?

Part B:

1. A stream or set of packets that have similar characteristics are called
 - a. Jitter
 - b. Cells
 - c. Flow
 - d. Packets
 - e. None of the above
2. Real latency is caused by
 - a. the end user.
 - b. queuing delays.
 - c. processing delays.
 - d. the limitations of the physical devices and media.
 - e. None of the above
3. The difference in end-to-end delay caused by lack of packet synchronization is called
 - a. Jitter
 - b. Flow
 - c. Deterioration
 - d. Latency
 - e. None of the above

4. The elapsed time for a packet to travel from the sender to the receiver is called
 - a. Jitter
 - b. Flow
 - c. Deterioration
 - d. Latency
 - e. None of the above
5. Induced latency is caused by
 - a. the physical limitations of the devices and media.
 - b. queuing delays.
 - c. congestion along the packet's transmission path.
 - d. A, B, and C
 - e. B and C only
6. What is meant by reliability?
 - a. The performance of a network based on the measurement of the average error rate during transmission.
 - b. The performance of a network based on the maximum data transfer rate that can be sustained between two end points.
 - c. The performance of a network based on the applications used by the end user.
 - d. All of the above
 - e. None of the above

7. When a router keeps no record of the history of the data flow it is operating in what mode?
 - a. Stateful
 - b. Flow controlled
 - c. Stateless
 - d. Best Effort
 - e. Guaranteed
8. What factors does QoS monitor?
 - a. Reliability and Bandwidth
 - b. Flows
 - c. Jitter
 - d. Latency
 - e. All of the above
9. Class of Service is another name for Quality of Service.
 - a. True
 - b. False
10. The actual process of service categorization for individual packets is called
 - a. Flow Management.
 - b. Bandwidth Management.
 - c. Class of Service.
 - d. Traffic Management.
 - e. None of the above.

Part C:

Identify each statement about congestion management tools and service levels as either True or False.

1. _____ WFQ favors medium-volume traffic over all other types of traffic in the queue.
2. _____ Best-Effort service is known as lack of QoS.
3. _____ FIFO processes packets solely on the order of arrival in the queue.
4. _____ CBQ uses a single queue and processes the highest priority packets over other packets.
5. _____ Guaranteed service reserves resources for specific traffic, no matter what may compete for the resources.
6. _____ PQ forwards packets out of the queue on a first come, first serve basis.
7. _____ Queuing means to store packets for later transmission.
8. _____ Tiered service is based on hard fast rules established by the network manager.
9. _____ ATM is an example of Best-Effort service.
10. _____ WRED places fairness back into Random Early Detection.

Part D:

1. How will QoS improve on current routing protocols?

2. Briefly describe how RSVP works.

3. Diagram MPLS and explain how this protocol works.

Scoring

Criteria	%	Your Score
Part A: Define QoS and describe why it is important.	30	
Part B: Identify key concepts in QoS.	20	
Part C: Identify congestion services and management tools.	20	
Part D: Compare RSVP, QoS, and MPLS.	30	
TOTAL	100	
Try It Out:	100	
Stretch Yourself:	100	
Network Wizards: Create a network architecture and a network diagram.	100	
FINAL TOTAL	400	

Resources:

Ferguson, P. & Huston, G. (1998). Quality of Service: Delivering QoS on the Internet and in Corporate Networks. New York: John Wiley & Sons, Inc.

Ferguson, P. & Huston, G. (1998). Quality of Service in the Internet: Fact, Fiction, or Compromise?. Available Online:
http://www.employees.org:80/~ferguson/inet_qos.htm.

Nortel Networks. (1999). IP Traffic Engineering Using MPLS Explicit Routing in Carrier Networks: Between the Signaling Approaches- CR-LDP and RSVP. Available Online:
http://www.nortelnetworks.com/products/announcements/mpls/collateral/ip_traffic_eng.pdf.

Nortel Networks. (1999). Quality of Service Technical Tutorial: Student Guide. Billerica, Massachusetts: Nortel Networks.

