Farhan
160304

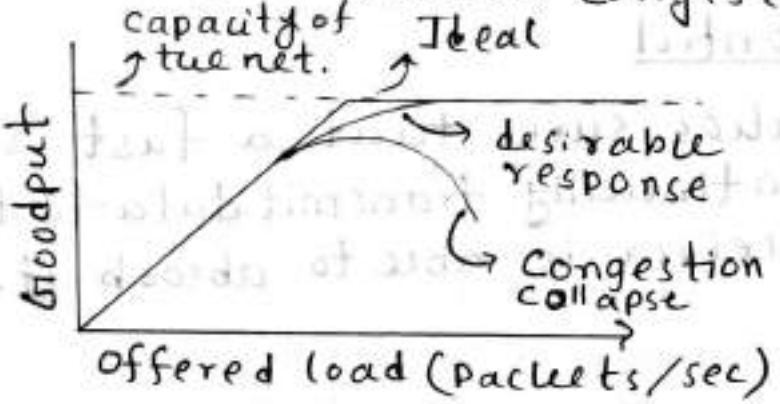
*Congestion Control Algorithms,

*Quality of Service

Congestion Control Algorithms

Congestion

↳ Too many packets present in the network causes packet delay and loss that degrade performance. This situation is called congestion.



Goodput

↳ The rate at which useful packets are delivered by the network.

Why infinite memory don't work?

- Adding more memory may help to a point. But with infinite memory congestion gets worse. Because, by the time the packets get to the front of the queue, they have already timed out repeatedly and duplicates have already been sent. It leads to congestion collapse.

Congestion Control

↳ Makes sure the network is able to carry the offered load.

Flow control

↳ Makes sure that a fast sender cannot continually transmit data faster than the receiver is able to absorb it.

Approaches to congestion control

(1) Network provisioning
(2) Traffic aware routing
(3) Admission control
(4) Traffic throttling
(5) Load Shedding

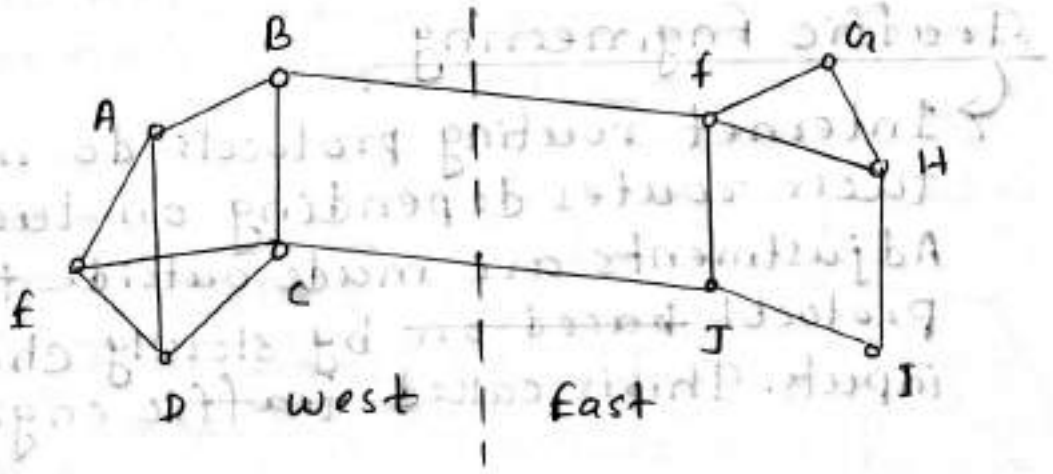
Network Provisioning

- The most basic way to avoid congestion is to build a network that is well matched to the traffic it carries.
- Sometimes resources can be added dynamically when there is a serious congestion. For example – turning on spare routers or enabling lines that are used only as backups.

- More often, links and routers that are regularly heavily utilized are upgraded at the earliest opportunity. This is called provisioning and happens on a time scale of months, ~~but~~ driven by long term traffic trends.

□ Traffic Aware Routing

- The most direct way to do this is to set the link weight to be a function of the link bandwidth and propagation delay plus the measured load or average queuing delay. Least weight paths then will favor paths that are more lightly loaded, all else being equal.



- However, there is a peril. In figure, east and west are connected by BF and CJ. Suppose most of the traffic are through BF. So, this link is heavily loaded and long delays. Including queuing delay in the weight for calcu-

lating shortest ~~delay~~ path, CJ will become more attractive. After new routing table update, most traffic will be through CJ. Consequently, in the next update, BF will be selected as new shortest path. As a result, the routing tables will oscillate widely and lead to errors.

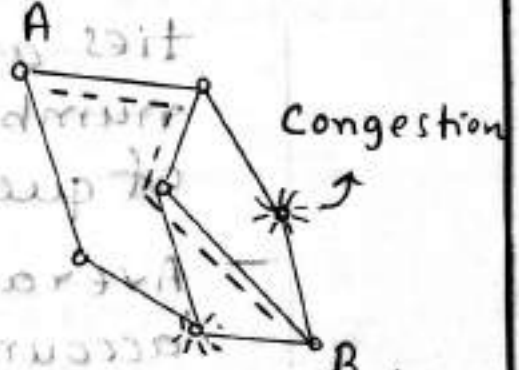
- Two techniques can be used as solution. The first is multipath routing. There can be multiple paths from source to destination. The second solution is for the routing scheme to shift traffic accross routes slowly enough that it is able to converge.

Traffic Engineering

↳ Internet routing protocols do not adjust their routes depending on the load. Adjustments are made outside the routing protocol ~~based on~~ by slowly changing its inputs. This is called traffic engineering.

Admission Control

- The idea is simple, do not set up a new virtual circuit unless the network can carry the added traffic without becoming congested. Thus, attempts to setup a virtual circuit may fail.
- However, virtual circuits in computer networks come in all shapes and sizes. Thus the circuit must come with some characterization of its traffic if we want to apply admission control. Traffic is often described in terms of its rate and shape. A commonly used description is the leaky bucket or token bucket. Armed with traffic descriptions, the network can decide whether to admit the new virtual circuit.
- Admission control can also be combined with traffic aware routing by considering routes around traffic hotspots as a part of the setup procedure.



- The example shows that there are congestion on routes that would normally used by A to send packet to be. The dashed line shows a possible route.

Traffic Throttling

- Senders adjust their transmissions to send as much traffic as the network can readily deliver. When congestion is imminent, it must tell the senders to throttle back their transmissions and slow down. This is called traffic throttling.
- Each approach must solve two problems. firstly, routers must determine when congestion is approaching, ideally before it has arrived. Secondly, routers must deliver timely feedback to the senders that are causing the congestion.

Determine Congestion

- Each router can continuously monitor the resources it is using. Three possibilities are— utilization of output links, number of lost packets, and the buffering of queued packets inside the router.
- Averages of utilization do not directly account for the burstiness of most traffic. Count of packet losses come too late.

- The queuing delay inside routers directly captures any congestion experienced by packets. It should be low most of the time. But will jump when there is a burst of traffic that generates a backlog.

$$d_{new} = \alpha d_{old} + (1-\alpha)s$$

d = queuing delay
s = queue length

α = how fast the router forgets recent history. Exponentially Weighted Moving Average (EWMA)

- whenever d moves above the threshold, the router notes the onset of congestion.

Deliver feedback

- Different schemes use different feedback mechanisms, as we will now describe.

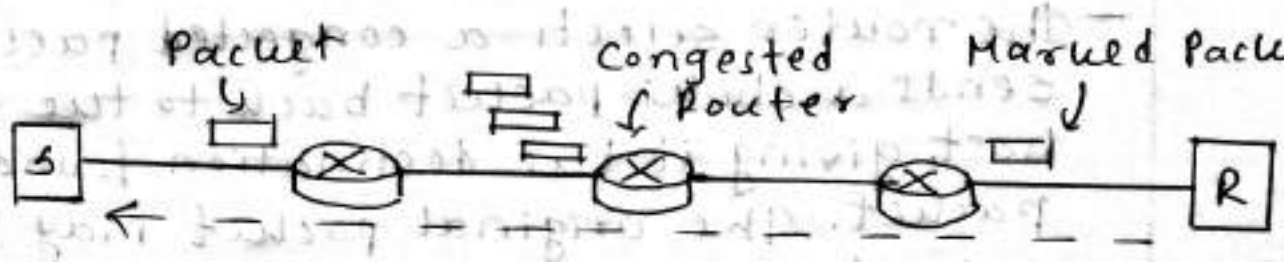
* Choke Packets

- The router selects a congested packet and sends a choke packet back to the source host, giving it the destination found in the packet. The original packet may be tagged so that it will not generate any more choke packet along the path. The router sends choke packets in low rate.

- When the source host gets the packet, it is required to reduce the traffic sent to the specific destination. The host should ingnore additional chokes for the fixed time interval. After the period, further choke packet indicate that the network is still congested.

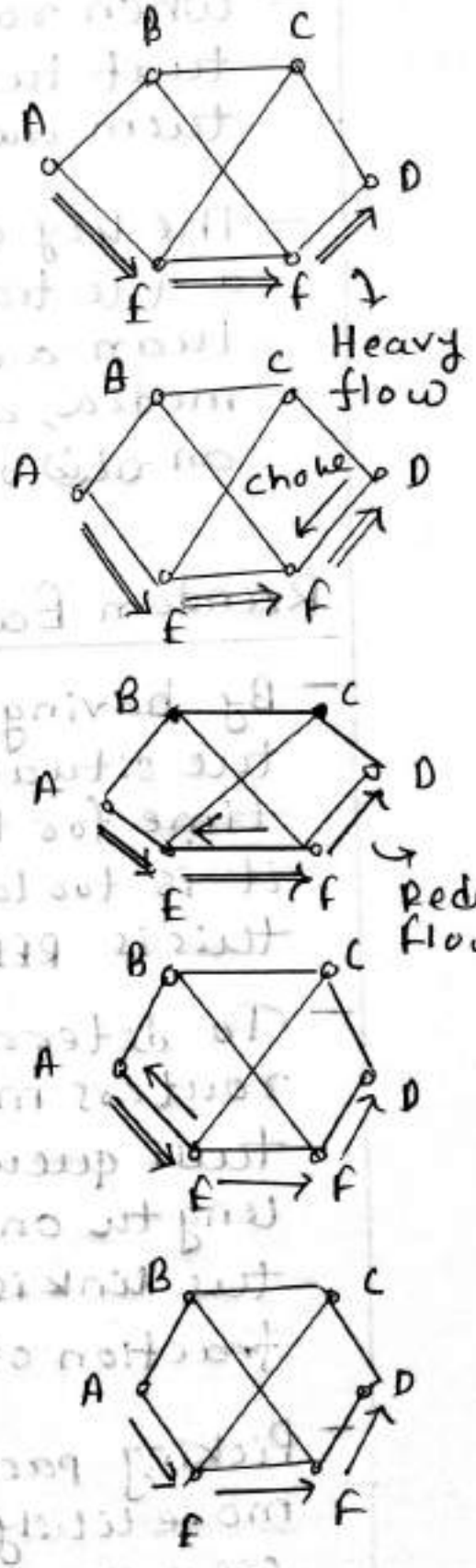
* Explicit Congestion Notification (ECN)

- Instead of generating additional packets to warn of congestion, a router can tag any packet it forwards, by setting a bit in the packet's header, to signal that it is experiencing congestion. When the network delivers this packet, the destination can note that there is congestion and inform the sender by a reply packet. The sender can then throttle its transmission as before.



* Hop-by-hop Backpressure

- An alternative approach is to have the choke packet take effect at every hop it passes through.
- Here as soon as the choke packet reaches F, F is required to reduce the flow to D.
- Next the packet reaches E and E slows down the flow.
- finally, the packet reaches the sender A, and the flow actually slows down.



Load Shedding

- When routers are being inundated by packets that they cannot handle, they just throw them away. It is called load shedding.
- The key question is which packet to drop. For a file transfer, an old packet is worth more than a new one. In contrast, for a real time media, a new packet is worth more than an old one.

Random Early Detection (RED)

- By having routers drop packets early, before the situation has become hopeless, there is time for the source to take action before it is too late. A popular algorithm for doing this is RED.
- To determine when to start discarding, routers maintain a running average of their queue lengths. When the avg. queue length on some links exceeds a threshold, the link is said to be congested and a small fraction of packets are dropped at random.
- Picking packets at random ~~with~~ makes it more likely that the fastest sender will see a packet drop. This is the best option since the router cannot tell which router is causing

the most trouble in a datagram network. The affected sender will notice the loss when there is no acknowledgement and then the transport protocol will slow down.

- The lost packet is thus delivering the same message as a choke packet, but implicitly, without the router sending any explicit signal.
- ECN is preferred option if it is available. RED is used when hosts cannot receive explicit signals.

Quality of Service

- There are applications that demand stronger performance gurantee from the network than the best that could be done under the circumstances. Quality of service mechanisms let a network with less capacity meet application requirements just as well at a lower cost.
- four issues must be addressed to ensure quality of service -
  (1) What applications need from the network
  (2) How to regulate the traffic
  (3) How to reserve resources at the routers to gurantee performance.
  (4) whether the network can seafely accept more traffic.

Application Requirements

- A stream of packets from a source to a destination is called a flow. The needs of each flow can be characterized by four primary parameters - bandwidth, delay, jitter and loss. Together these determine the QoS the flow requires.
- QoS of some applications are shown in table below.

Application	Bandwidth	Delay	Jitter	Loss
Email	Low	Low	Low	Mid.
File Sharing	High	Low	Low	Mid.
Web Access	Mid.	Mid.	Low	Mid.
Remote Login	Low	Mid.	Mid.	Mid.
Audio on demand	Low	Low	High	Low
Video on demand	High	Low	High	Low
Telephony	Low	High	High	Low
Videoconference	High	High	High	Low

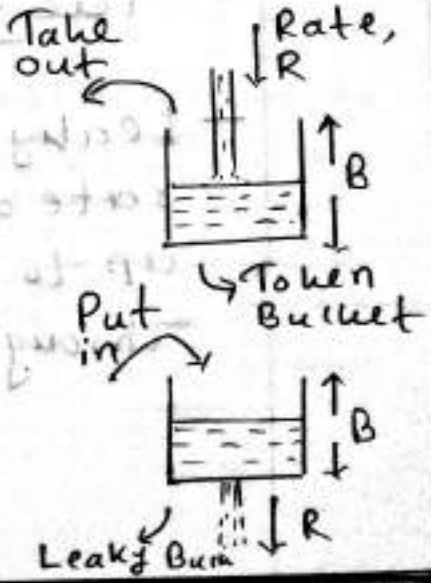
Traffic Shaping

- Bursts of traffic are more difficult to handle than constant rate traffic because they can fill buffers and cause packets to be lost.
- Traffic shaping is a technique for regulating the average rate and burstness of a flow of data that enters the network. The goal is to allow applications to transmit a wide variety of traffic that suits their needs. And also to have a simple and useful way to describe the possible traffic patterns to the network.

- When a flow is set up, the user and the network agree on a certain network pattern for that flow. This agreement is called Service Level Agreement (SLA). As long as the customer fulfills its part of the bergain and only sends packets according to the agreement, the provider promises to deliver them in a timely fasion.
- How the provider can tell if the customer is following the agreement and what to do if the customer is not? Packets in excess of the agreed pattern might be dropped by the network or they might be marked as having low priority. Monitoring a traffic flow is called traffic policing.

Leaky and Token Bucket:

- The bucket can be used to shape or police packets entering the network.
- Conceptually, each host is connected to the network by an interface containing a leaky bucket. To send a packet into the network it must be possible to put more water



into the bucket. If a packet arrives when the bucket is full the packet must either be queued until enough water leaks out to hold it or be discarded. The former might happen at a host shaping its traffic. The later might happen at a provider network interface that is policing traffic. This technique is called leaky bucket algorithm.

- A different but equivalent formulation is to imagine the network interface as a bucket that is being filled. The tap is running at rate R and the bucket has a capacity of B, as before. Now, to send a packet, we must be able to take water, or tokens out of the bucket. No more than a fixed number of tokens, B, can accumulate in the bucket. If the bucket is empty, we must wait until more tokens arrive before we can send another packet. The algorithm is called the token bucket algorithm.

- Leaky and token buckets limit the long term rate of a flow. But allow short term bursts up to a maximum regulated length to pass through unaltered and without delays.

Packet Scheduling

- To provide a performance guarantee, we must reserve sufficient resources along the route that the packet takes through the network.
- Algorithms that allocate router resources among the packets of a flow and between competing flows are called packet scheduling algorithms.
- Three different kinds of resources can be potentially reserved for different flows—
  (1) Bandwidth
  (2) Buffer space
  (3) CPU cycles
- If a flow requires 1Mbps and the outgoing line is 2Mbps, trying to direct 3 flows is not going to work. Thus, reserving bandwidth means not oversubscribing any output line.
- When a packet arrives, it is buffered inside the router until it can be transmitted on the chosen outgoing line. The purpose of the buffer is to absorb small bursts of the traffic as the flows contend with each other. If no buffer is available, the packet has to be discarded.

- For good QoS, some buffers might be reserved for a specific flow so that flow does not have to compete for buffers with other flows.
- It takes router CPU time to process a packet. So, a router can process only a certen number of packets per second. Making sure that the CPU is not overloaded is needed to ensure timely processing of these packets.

Packet scheduling Algoritums

- Packet scheduling algoritums allocate bandwidth and other router resources by determining which of the buffered packets to send on the output line next.

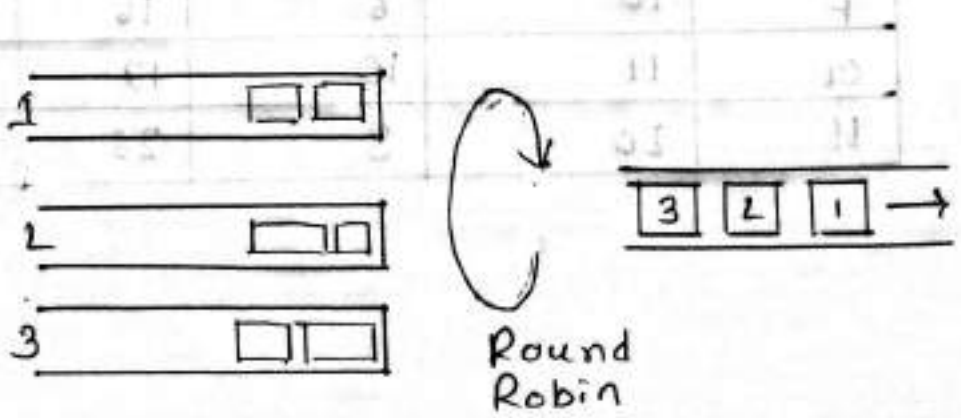
* FIFO

- Each router buffers packet in a queue for each output line until they can be sent. And they are sent in the same order that they arrived. This algoritum is known as FIFO.
- FIFO routers usually drop newly arriving packets when the queue is full. This behaviour is called Tail drop.
- RED algoritum drops newly arrived packets at random.

- FIFO scheduling is easy to implement, but is not suited to providing good QoS. Because when there are multiple flows, one can easily affect the performance of others. If the first flow sends large bursts of packets, they will lodge in the queue. Processing packets in the order of arrival means that the aggresive sender can hog most of the capacity, starving other flows. and reducing their QoS.

* Fair Queuing

- Routers have separate queues, one for each flow for a given output line. When the line becomes idle, the router scans round-robin. It then takes the first packet on the next queue. With n hosts, each host gets to send one packet of every n packets.
- It is fair in the sense that all flows get to send packets at the same rate. Sending more packets will not improve this rate.



- Problem with the algorithm is - it gives more bandwidth to hosts that use larger packets than to hosts that use small packets.
- An improvement can be done in such a way as to simulate a byte-by-byte round robin instead of packet-by-packet round robin. The trick is to compute a virtual time that is the number of the round at which each packet would finish being sent. Each round drains a byte from all of the queues that have data to send. The packets are then sorted in order of their finishing times and sent in that order.

Packet	Arrival Time	Length	Finish Time	Output Order
A	0	8	8	1
B	5	6	11	3
C	5	10	10	2
D	8	9	20	7
E	8	8	14	4
F	10	6	16	5
G	11	10	19	6
H	20	8	28	8

- One shortcoming of this algorithm is that it gives all hosts the same priority. In many situations, it is desirable to give, for example, video servers more bandwidth than, say, file servers.

* Weighted Fair Queuing (WFQ)

- As a solution to fair queues problem, for example, we can give the video server two or more bytes per round. This modification algorithm is called WFQ.

$$F_i = \max(A_i, F_{i-1}) + L_i / \omega$$

ω = weight as no of bytes per round
f_i = finish time
A_i = Arrival time
L_i = Length of packet

- With N flows, WFQ is at best $O(\log N)$ per packet, which is difficult to achieve for many flows in high speed routers.
- Deficit round robin can be implemented very efficiently with only $O(1)$ per operations per packet. WFQ is widely used given this approximation.

* Priority Scheduling

- Each packet is marked with a priority. High priority packets are always sent before any low priority packets that are buffered. With in a priority, packets are sint in FIFO order.
- However, priority scheduling has the disadvantage that a burst of high priority packets can starve low priority packets.
- A high and low priority system is essentially a two-queue WFQ system. In which, the high priority has infinite weight.

* Timestamp Scheduling

- Packets carry a timestamp. The timestamp records how far the packet is behind or ahead of the schedule as it is sent through a sequence of routers on the path.
- Packets that have been queued behind other packets at a router will tend to be behind schedule. And the packets that have been serviced first will tend to be ahead of the schedule.
- Sending packets in order of their timestamps has the benefical effect of speeding up slow packets while at the same time slowing down fast packets. So, all packets are delivered with more consistent delay.

Admission Control

- The user offers a flow with an accompanying QoS requirement to the network. The network then decided whether to accept or reject the flow based on its capacity and the commitments it has made to other flows. If it accepts, the network reserves capacity in advance at routers to gurantee QoS.
- The reservation must be made at all of the routers along the route that the packets take through the network. Many routing algoritums find the single best path between each source and each destination and send all traffic over the best path. This may cause some flows to be rejected if there is not enough capacity. QoS gurantees for new flows may still be accommodated by choosing a different route that has excess capacity. This is called QoS routing.
- It is also possible to split the traffic for each distination over multiple paths to more easily find excess capacity. A simple method is for routers to choose equal-cost ~~the~~ paths and to devide the traffic equally or in proportion to the capacity of the outgoing links.

- Because many parties may be involved in the flow negociation, flows must be described accurately in terms of specific parameters that can be negociated. A set of such parameters is called a <u>flow specification.</u>
- As the specification propagate along the route, each routers examine the specs. and modifies parameters if needed. The modifications can only reduce the flow, not increse it. When it gets to the other end, the parameters can be established.
- The first two parameters use a token bucket to give the maximum sustained rate the sender may transmit, averaged over a long time interval and the largest burst it can send over a short time interval
- The peak data rate is the max transmission rate tolerated. The sender must never exceed this rate.
- The last two parameters specify the max and min packet size, including transport and network layer headers. The min size is useful because processing each packet takes some fixed ~~site~~ time no matter how short.

Parameter	Unit
Token bucket rate	Bytes/sec.
Token bucket size	Bytes
Peak data rate	Bytes/sec
Minimum pack. size	Bytes
Maximum pack. size	Bytes

Versions of QoS for the Internet

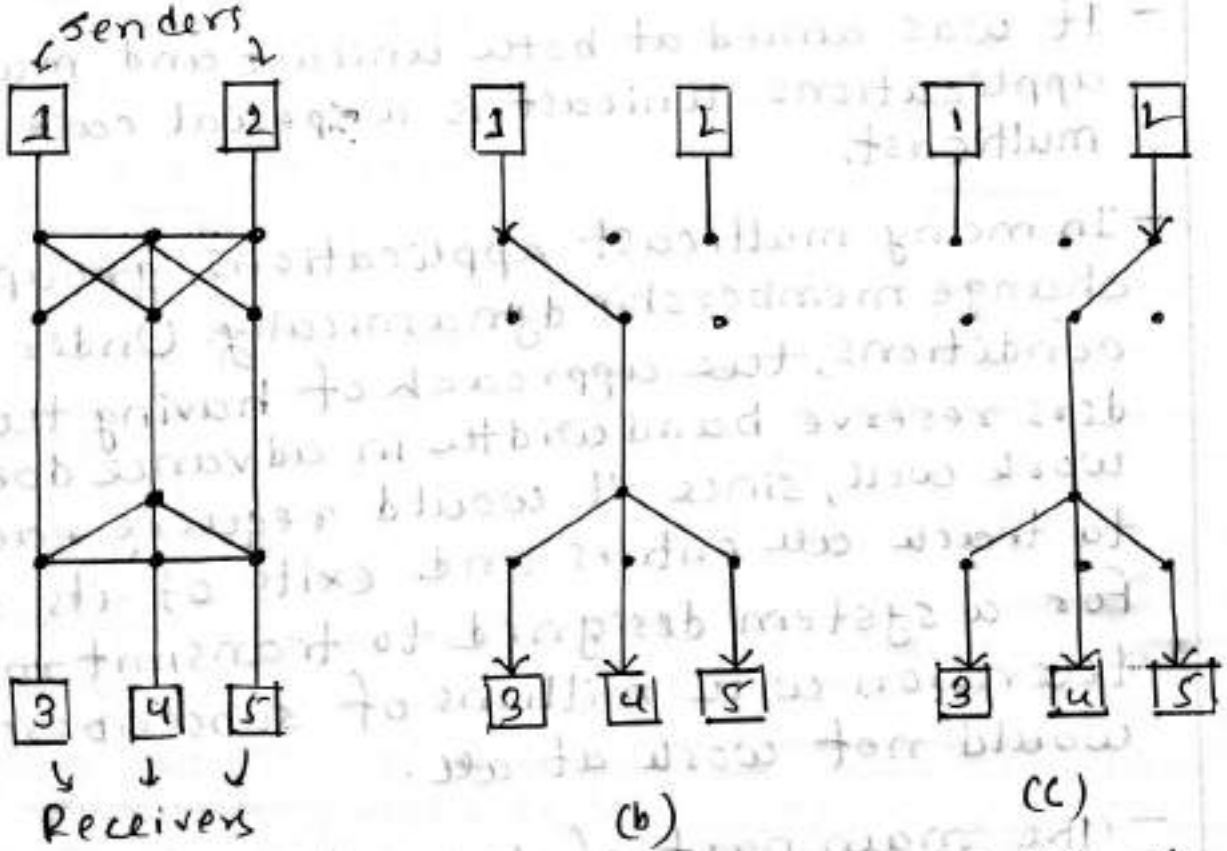
- We will describe two versions of QoS for the Internet called the Integrated Services and the Differentiated Services.

Integrated Services

- Between 1995 and 1997, IETF put a lot of effort into devising an architecture for streaming multimedia. The generic name for this work is integrated services.
- It was aimed at both unicast and multicast applications. unicast is a special case of multicast.
- In many multicast applications, groups can change membership dynamically. Under these conditions, the approach of having the senders reserve bandwidth in advance does not work well, since it would require each sender to track all entries and exits of its audience. For a system designed to transmit ~~noitions~~ television with millions of subscribers, it would not work at all.
- The main part of the integrated services architecture that is visible to the users of the network is RSVP.

The Resource reservation Protocol (RSVP)

- This protocol is used for making the reservations. Other protocols are used for sending the data.
- RSVP allows multiple senders to transmit to multiple groups or receivers, permits individual receivers to switch channels freely and optimize bandwidth use while at the same time eliminating congestion.



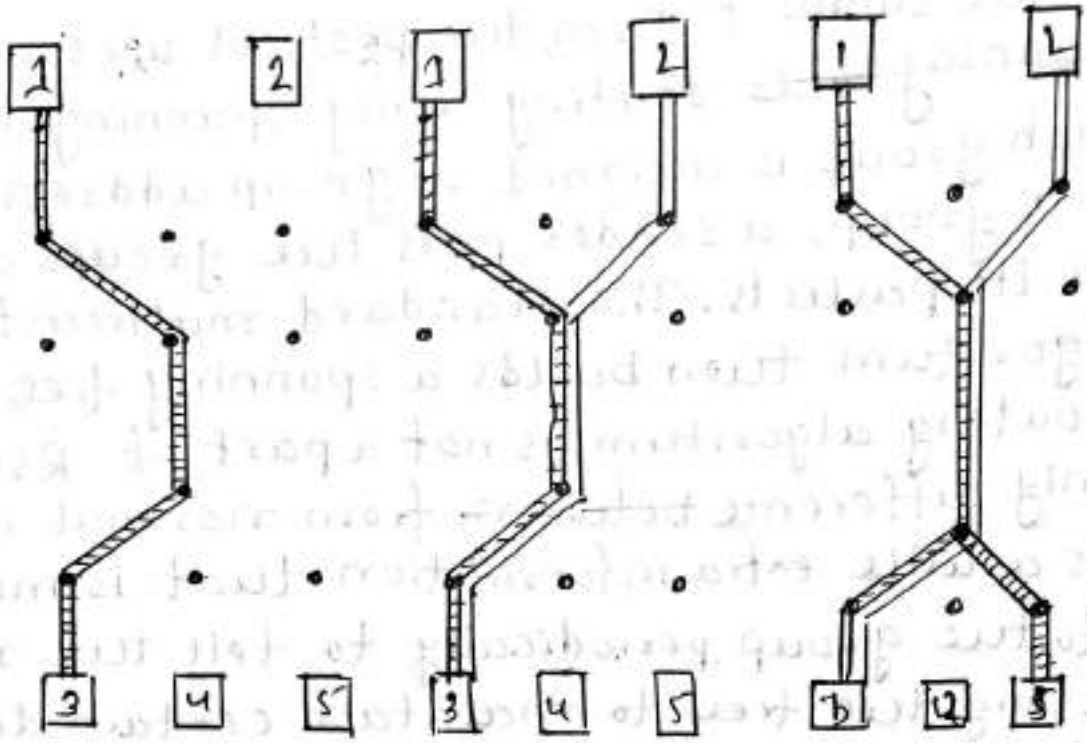
(a) A network

(b) Multicast spanning tree for host 1

(c) Multicast Spanning Tree for host 2

- In the simplest form, the protocol uses multicast ~~spanning tree~~ routing using spanning trees. Each group is assigned a group address. To send to a group, a sender puts the group's address in its packets. The standard multicast routing algorithm then builds a spanning tree. The routing algorithm is not a part of RSVP. The only difference ~~between~~ from normal multicast is a little extra information that is multicast to the group periodically to tell the routers along the tree to maintain certain data structure in their memories.

- To get better reception and eliminate congestion, any of the receivers in a group can send a reservation message to the sender. The message is propagated using the reverse path forwarding algorithm. At each hop, the router notes the reservation and reserves the necessary bandwidth. WFQ can be used to make this reservation. If insufficient bandwidth is available, it reports back failure. By the time, the message gets back to the source, bandwidth has been reserved all the way from the sender to the receiver making the reservation request along the spanning tree.



- In fig-(a), host 3 has requested a channel to host-1. Once it has been established, packets can flow without congstion.
- In fig-(b), host 3 has requested a channel to host-1. As a result, a second path is reserved. Because, two independent streams are being transmitted.
- In fig-(c), host-5 decides to watch program being transmitted by host-1. At first, dedicated bandwidth has been reserved. But after reaching the second router, it sees, it already has a feed. So, it does not have to reserve anymore.

- Note that, hosts 3 and 5 might have asked for different amounts of bandwidths. For example, host 3 is playing on small screen and wants the low resolution info. So, the capacity reserved must be large enough to satisfy the greediest receiver.
- When making a reservation, a receiver can specify one or more sources that it wants to receive from. It can also specify whether these choices are fixed for the duration of reservation or the receiver wants to keep open the option of changing sources later.
- In particular, two receivers can only set up to share a ~~link~~ path if they both agree not to change the source later on.
- Once a receiver has reserved bandwidth, it can switch to another source and keep that portion of the existing path that is valid for the new source.
- If host 2 is transmitting several video streams in real time, host 3 may switch between them without changing its reservation.
- Flow based algorithms offer good QoS to one or more flows. because they can reserve ~~however~~ whatever resources are needed.

- However, they also have downside. They require an advanced setup to establish each flow. That does not scale well when there are thousands or millions of flows. Also, they maintain interna per flow state in the routers, making them vulnerable to router crashes. Finally, the changes required to the router code are substancial and involve complex routes-to-router exchanges for setting up the flows.

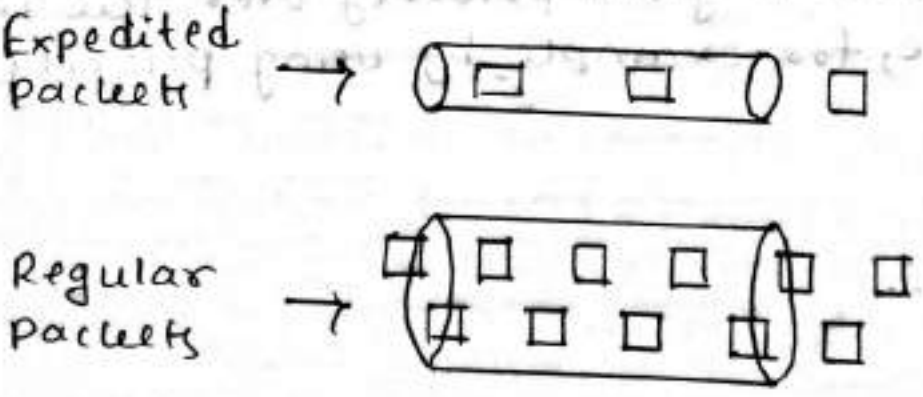
☐ Differentiated Services

- IETF has also devised a simpler approach to QoS, that can be largely implemented locally in each router without advance setup and without having whole path involved. This approach is known as class-based approach. IETF has standarited it as differentiated services.

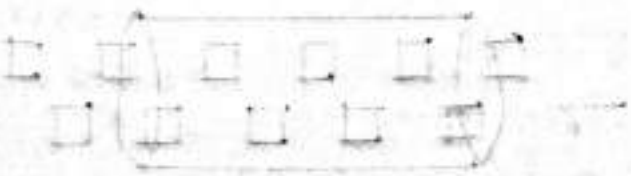
- Differentiated service can be offered by a set of routers forming an administrative domain, for example-ISP. The administration defines a set of servie classes with corres-ponding forwarding rules. If a customer subscribes to differentiated services, customer packets entering the domain are marked with the class. The information is carried in differentiated services field of IPv4 or IPv6 packets.

Expedited forwarding

- The choice of service classes is up to each operator, but since packets are often forwarded between networks run by different operators. IETF has defined some network independent service classes. The simplest class is expedited forwarding.
- Two classes of service are available - regular and expedited. Most of the traffic are regular but limited fraction are expedited. The expedited packets should be able to transit the network as though no other packets were present. So, they will get low loss, low delay and low jitter that is needed for VoIP.
- A symbolic representation of this two tube system is shown on figure. Note that, there is still just one physical line. The two logical pipes represent a way to reserve bandwidth for different classes of service not a second physical line.

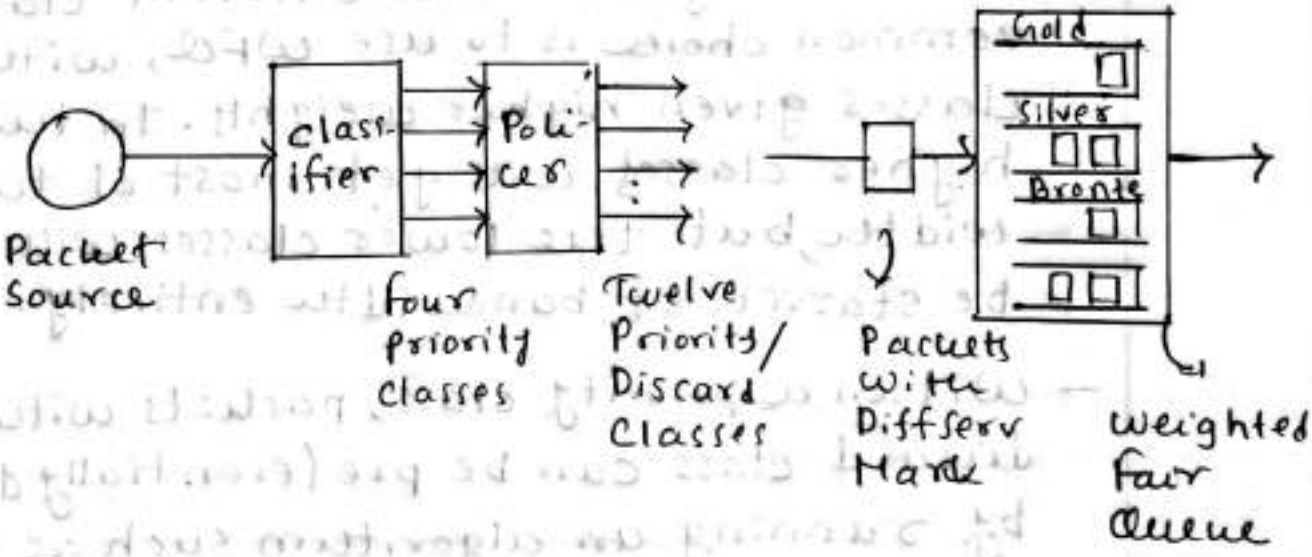


- Packets are classified as expedited or regular and marked. This might be done at host or the ingress (first) router. The advantage of doing classification on the sending host is that more information is available about which packets belong to which flows.
- If the packets pass through a network that supports expedited service, they will receive preferential treatment. If does not support, no harm is done.
- Of course, if the marking is done by the host, the ingress router will police the traffic to make sure that the customers are not sending more expedited ~~serv~~ traffic than they have paid for.
- Within the network, the routers may have two output queues for each outgoing line for two classes. Whenever a packet arrives, it is queued accordingly. Expedited queue is given priority over the regular one, for example, by using priority scheduler.



Assured forwarding

- A somewhat more elaborate scheme is called assured forwarding.
- Assured forwarding specifies that there shall be four priority ~~que~~ classes, each class having its own resources. In addition, it defines three discard classes. for packets that are experiencing congestion - low, medium and high. Taken together, these two factors define 12 service classes.



- The first step is to classify the packets into one of the four priority classes. The step might be done at the host or ingress router.
- The next step is to determine the discard class for each packet. This is done by passing the packets of each priority ~~queues~~ class through a traffic policer such as a token bucket.

- The policer lets all of the traffic through, but it identifies packets that fit within the small burst as low discard, packets that exceed small burst as medium discard and packets that exceed large bursts as high discard. The combination of priority and discard class is then encoded in each packet.
- Finally the packets are processed by routers in the network with a packet scheduler that distinguishes the different classes. A common choice is to use WFQ, with higher classes given higher weights. In this way, higher classes will get most of the bandwidth, but the lower classes will not be starved of bandwidth entirely.
- Within a priority class, packets with higher discard class can be preferentially dropped by running an algorithm such as RED. RED will start to drop packets as congestion builds but before the router has run out of buffer space. At this stage, there is still buffer space with which to accept low discard packets while dropping high discard packets.