CSIT985

Strategic Network Design

Spring 2023

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Lecture week 8:

Network Performance

Presented by: Dr. Chau Nguyen

Lecturer, School of Computing and Information Technology, UOW

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Outline

- ❑ Background
- ❑ Definition
- ❑ Developing Goals for Performance
- ❑ Performance Mechanisms
  - ➢ Quality of Service
  - ➢ Prioritization, Traffic Management, Scheduling, and Queuing
  - ➢ Resource control
  - ➢ Service Level Agreements
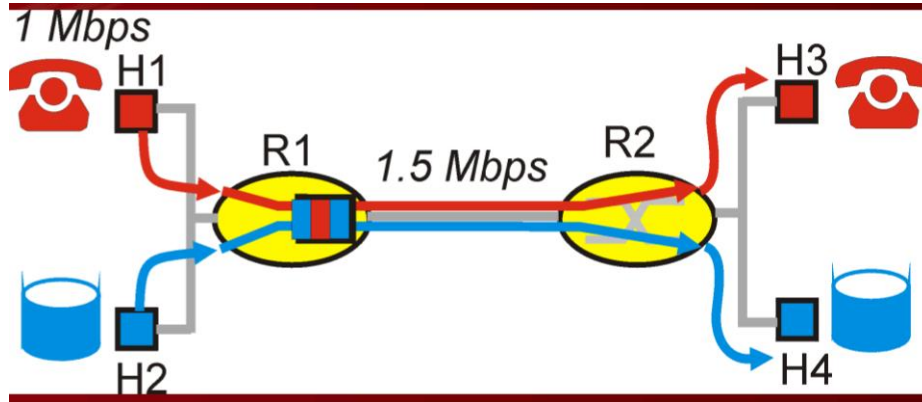 - ➢ Policies
- ❑ Architectural Considerations

Background

- To be able to say we have a network that performs well, we need to have measurable goals

  o Improve overall network performance
    - Response times, throughput improved for all users

  o Support a particular group of users or application

  o Control resource allocation for accounting billing and/or management
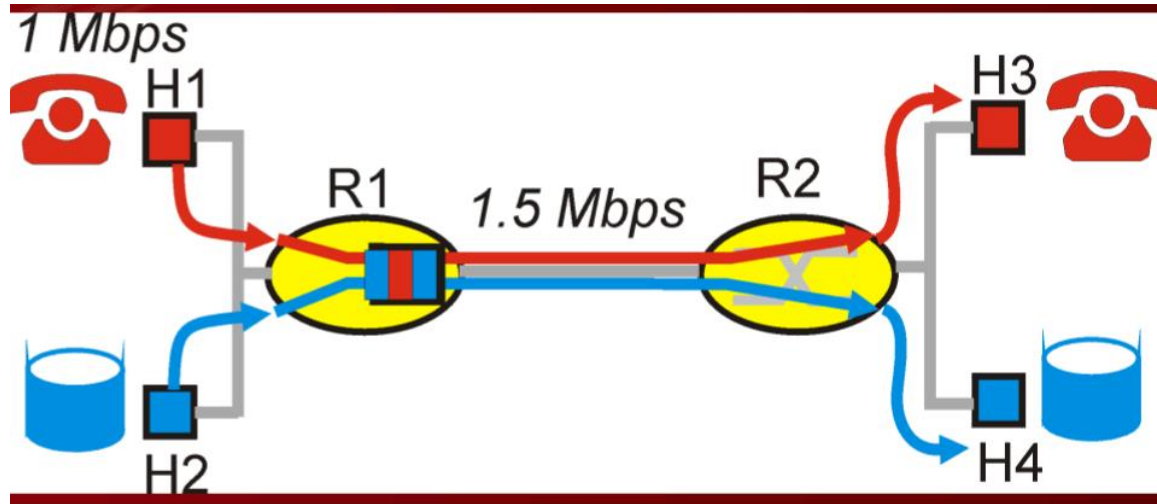
Background

- Example: 1Mbps IP phone, FTP share 1.5 Mbps link.

  o bursts of FTP can congest router, cause audio loss

  o want to give priority to audio over FTP

Background

- Basic fact of life: the network can not support traffic demands beyond link capacity

Definition

What is performance?

- *"**Performance** is the set of levels for capacity, delay, and RMA in a network. It is usually desirable to optimize these levels, either for all (user, application, and device) traffic flows in the network, or for one or more sets of traffic flows, based on groups of users, applications, and/or devices." (McCabe, 2010, p.334)*
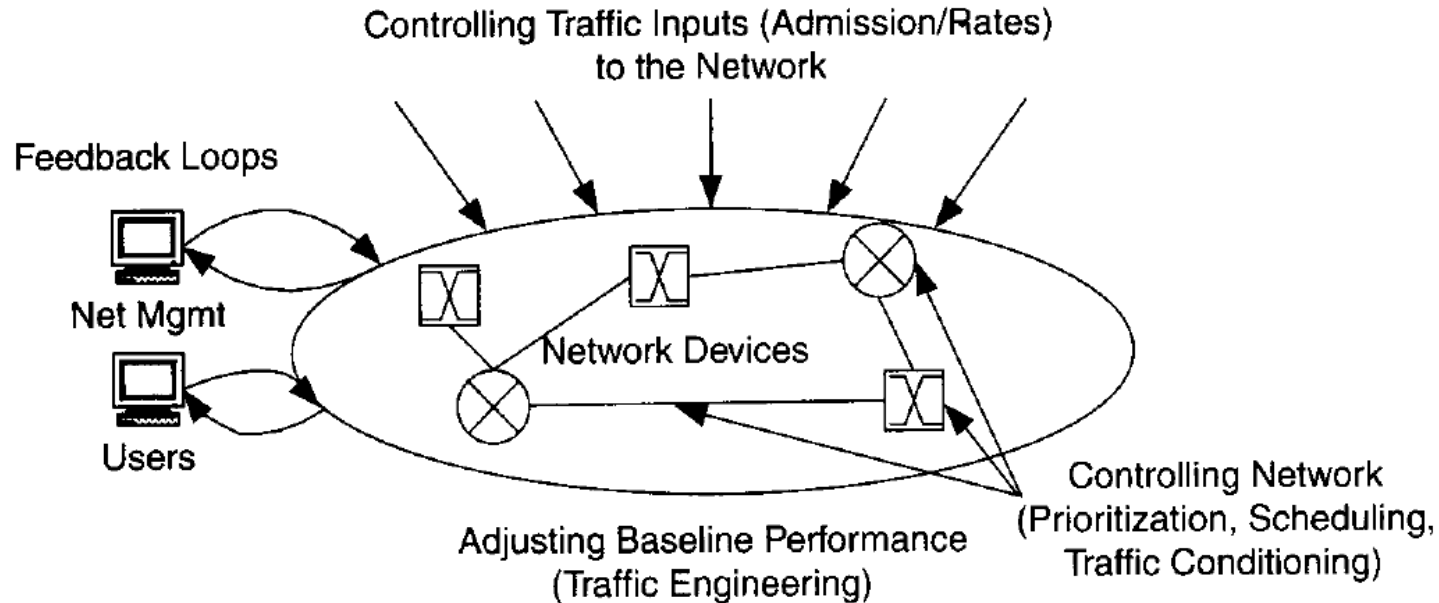
What is performance?

- Performance consists of one or more of the followings:

    o controlling traffic inputs of the network
    - admission and rate controls

    o adjusting the baseline performance of the network
    - traffic or capacity engineering

    o controlling all or part of the network for delivery of specific services
    - prioritizing, scheduling, and conditioning traffic flows

    o implementing a feedback loop to users, applications, devices, and management to modify controls as necessary

Overall Performance Mechanism

Controlling Traffic Inputs (Admission/Rates) to the Network

Feedback Loops

Net Mgmt

Users

Network Devices

Controlling Network (Prioritization, Scheduling, Traffic Conditioning)

Adjusting Baseline Performance (Traffic Engineering)

FIGURE 8.1 General Mechanisms for Performance

(McCabe, 2010, p.335)

Developing Goals for Performance

Developing Goals for Performance

- The process of developing performance goals begins during the requirements phase

  o The requirements and flow specifications and maps provide inputs

- Although performance is always desirable

  o Need to ensure that the performance mechanisms are both necessary & sufficient to meet the goals set

Developing Goals for Performance

- In developing performance goals, the following questions should be answered

  o Are performance mechanisms necessary for this network?

  o What are we trying to solve, add, or differentiate by adding performance mechanisms to this network?

  o Are performance mechanisms sufficient for this network?

Developing Goals for Performance

- Start simple and work towards a more complex architecture

  o Implement in selected areas of the network

  o Use only a few mechanisms

  o Select mechanisms that are easy to implement, operate and maintain

Developing Goals for Performance

- We should avoid implementing performance mechanisms just because they are interesting or new
  - There should be a clear need and want by customers
  - Performance mechanisms need to be constantly maintained

Developing Goals for Performance

- A well-designed performance architecture will address the following

  o Improving overall network performance

  o Improving the performance for selected users, applications and/or devices

  o Changing the network from a cost center to profitability

  o Merging multiple traffic types over a common network infrastructure

  o Differentiating customers for multiple levels of service

Developing Goals for Performance

- Single tier performance where capacity delay and RMA, are optimized for all traffic flows

- Multi-tier where capacity, delay and RMA are optimized for one or more groups of flows

Developing Goals for Performance

- A defined level of performance in a data communications system

  o E.g., to ensure that real-time voice and video are delivered without annoying blips,

  o a guarantee of bandwidth is required.

- When data is broken into packets that travel through the routers of LANs and WAN with all other data, QoS mechanisms are one way to give real-time data priority over non-real-time data.

- The only other option is to overbuild the network so there is always sufficient bandwidth.

Performance Mechanisms

Performance Mechanisms

- Quality of service (QoS)

- Resource control
  - Prioritization
  - Traffic management
  - Scheduling
  - Queuing

- Service-level agreements (SLAs)

- Policies

Performance Mechanism - Quality of Service (QoS)

Quality of Service (QoS)

- Determining, setting and acting on priority levels for traffic flows

- Includes

  o **IP QoS**

  o Type of Service (ToS)

  o ATM Class of Service (CoS)

  o Frame Relay Committed Information Rate (CIR)

Quality of Service (QoS)

- **Focus on IP QoS**
  - DiffServ
    - Aggregation of traffic flows on a per-hop basis
    - Good scalability
    - Three traffic classes: Best-effort, Assured forwarding, Expedited forwarding
  - IntServ
    - End-to-end support for individual flows

Quality of Service (QoS)

- MPLS is now the QoS mechanism of choice for telecommunication providers and is making its presence felt in corporate environments

  o MPLS – Multiprotocol label Switching

  o MPLS is an ethernet based technology

MultiProtocol Label Switching

- MPLS is used to ensure that all packets in a particular flow take the same route over a backbone.

- Deployed by many telcoms and service providers

- MPLS can deliver the quality of service (QoS) required to support real-time voice and video as well as service level agreements (SLAs) that guarantee bandwidth. (e.g., able to deliver hard service metrics like ATM)

- Large enterprises may also use MPLS in their national networks

MultiProtocol Label Switching

- A MPLS header is attached to a IP packet by a Label Edge Router (LER)

- The entry and exit points of an MPLS network are defined by LERs.

- Routers in between are called Label Switch Routers (LSR) which use the MPLS header to route packets

- VLSI technology (i.e. hardware as opposed to CPU and Router lookup tables) use the MPLS header to make routing decisions.

Quality of Service – IntServ

- Sometime traffic flows need to be treated individually

    - IntServ defines value and mechanisms for allocating resources across end-to-end paths.

    - Signaling mechanism is usually provided by RSVP – Resource ReSerVation Protocol (RFC 2205)

    - Source routers request capacity via a RESV request. Once other routers have agreed to this request, they must carry the traffic.

ReSerVation Protocol

- A communications protocol that signals a router to reserve bandwidth for real-time transmission.

- RSVP is designed to clear a path for high priority traffic

- Works at Layer Transport Layer – is similar to ICMP and IGMP
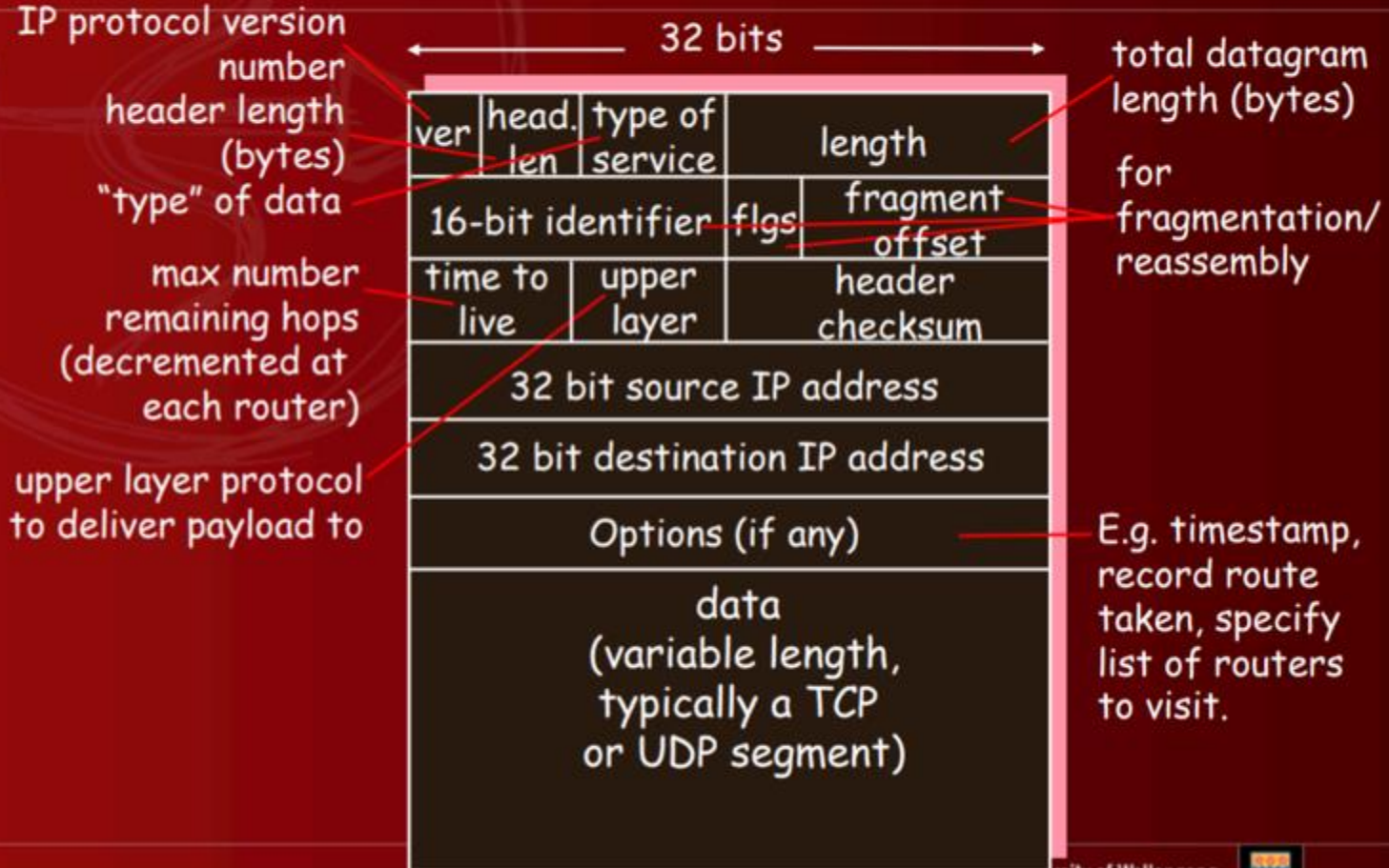
Quality of Service – DiffServ

- While IntServ is a "fine-grained" QoS system DiffServ's is a "coarse-grained" control system.

  o IntServ – individual flows are treated specially

  o DiffServ – flows are aggregated into common classes and are treated specially

DIFFerentiated SERVices (DiffServ)

- Can be invoked when IP packets are marked in the Type of Service (ToS) byte for IPv4 (or traffic class in IPv6).

- DiffServ enables three (or more) classes of flow to be handled
  - Best effort
  - Assured Forwarding (AF)
  - Expedited Forwarding (EF)

IPv4 datagram format

IP protocol version number

header length (bytes)

"type" of data

max number remaining hops (decremented at each router)

upper layer protocol to deliver payload to

total datagram length (bytes)

for fragmentation/ reassembly

E.g. timestamp, record route taken, specify list of routers to visit.

←————— 32 bits —————→

ver	head. len	type of service	length
16-bit identifier		flgs	fragment offset
time to live	upper layer		header checksum
32 bit source IP address			
32 bit destination IP address			
Options (if any)			
data (variable length, typically a TCP or UDP segment)			

DIFFerentiated SERVices (DiffServ)

- Second generation DiffServ uses the Differentiated Services Code Point (DSCP) in the ToS byte to specify precedence "class".

- DiffServ

  - Aggregation of traffic flows on a per-hop basis

  - Good scalability

  - Best-effort, Assured forwarding, Expedited forwarding

Quality of Service – DiffServ – RFC 2745

- DiffServ operates over a DiffServ domain that is defined by routers designated as "boundary nodes"
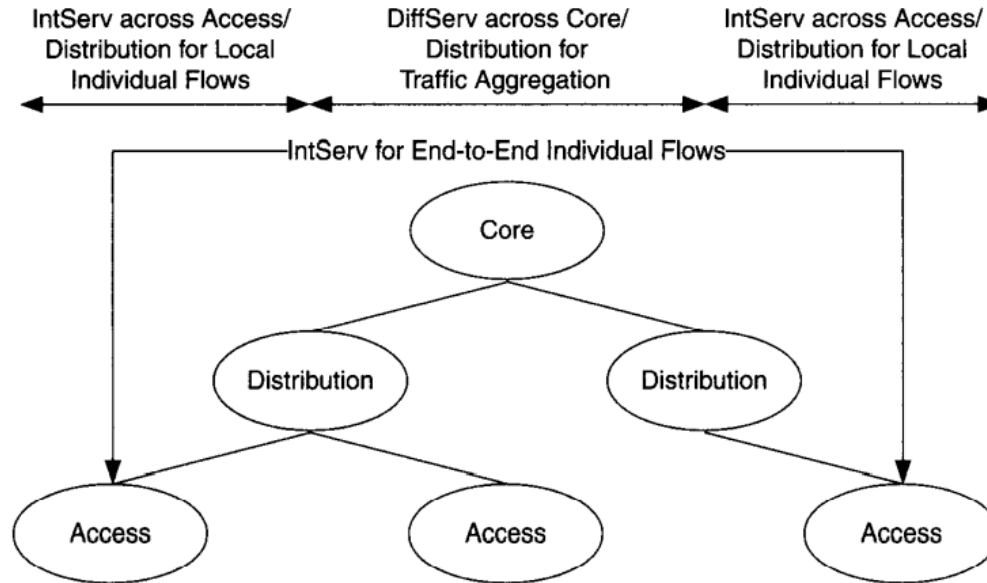 - RFC 2745 2.1

Quality of Service – DiffServ

- It is expected that service providers will use MPLS within their networks and use Diffserv at the edges of the network for classification and assignment to the right connection

QoS – Before MPLS

- IntServ is best implemented in resources the network administrator has direct control over.

35

Quality of Service

- IntServ (RSVP) does not scale well and is not commonly found in corporate networks

- DiffServ works on all routers

- RSVP has been re-purposed by the IETF to work on MPLS (see RSVP-TE)

Quality of Service

- A Comparison of DiffServ and IntServ

Function/Feature	DiffServ	IntServ
Scalability	Scalable to Large Enterprise of Service-Provider Networks	Limited to Small or Medium-Size Enterprise Networks
Granularity of Control	Traffic Aggregated into Classes	Per-flow or Groups of Flows
Scope of Control	Per Network Device (Per-Hop)	All network Devices in End-to-End Path of Flow

(McCabe, 2010, p.340)

Performance Mechanisms - Resource Control

Resource Control

- Prioritization

- Traffic management

- Scheduling

- Queuing

Resource Control - Prioritization

- Prioritization

  o A ranking based upon importance and urgency relative to other network entities

  o Users, applications, devices, flows and connections

  o Priorities should be determined during the requirements and flow analysis phases

  o May be based on
    - Type of protocol (TCP vs UDP)
    - Port number

Resource Control – Traffic Management

- **Traffic management**

  o **Admission control**
  - Ability to refuse access
  - Uses priority level to change access behaviour

  o **Traffic conditioning**
  - Modify performance to traffic flows
  - Classification
  - Metering
  - Temporal characteristics of traffic flow
  - Shaping
  - Delaying non-conforming traffic to match performance characteristics
  - Dropping

Resource Control – Traffic Management

- Traffic conditioning

- Set of mechanisms that increase or decrease performance to traffic flows

- Taken in order of nominal traffic flow across the network, traffic conditioning consists of:

  1. Classification

  2. Metering

  3. Shaping

  4. Dropping

Resource Control – Traffic Management

- Classification

  o Packets can be marked with Differentiated Services Code Points (DSCPs) for Best Effort, Assured Forwarding or Expedited Forwarding

- Metering

  o Temporal characteristics of traffic flow (traffic rates or burst sizes) are measured

- Metering is performed in routers.

- This information is compared with the traffic profile detailed in the SLA

Resource Control – Traffic Management

- Shaping (example)

  o Consider an SLA that specifies a peak rate of 1.5Mb/s

  o A meter is measuring that traffic and calculates a rate of

    → (200 packets/sec) * (1500 byte/packets) * (8bits/byte) = 2.4 Mbits/sec

  o This non-conforming flow is subsequently forwarded to a shaper queue where packets are delayed by an average of 10ms → 100 packets/sec

    → (100 packets/sec) * (1500 byte/packets) * (8bits/byte) = 1.2 Mbits/sec

  o Shaping continues for either a set period or until flow conforms

- Dropping – Packets are discarded

Resource Control - Scheduling

- Scheduling
  - Determines the order in which traffic is processed for transmission
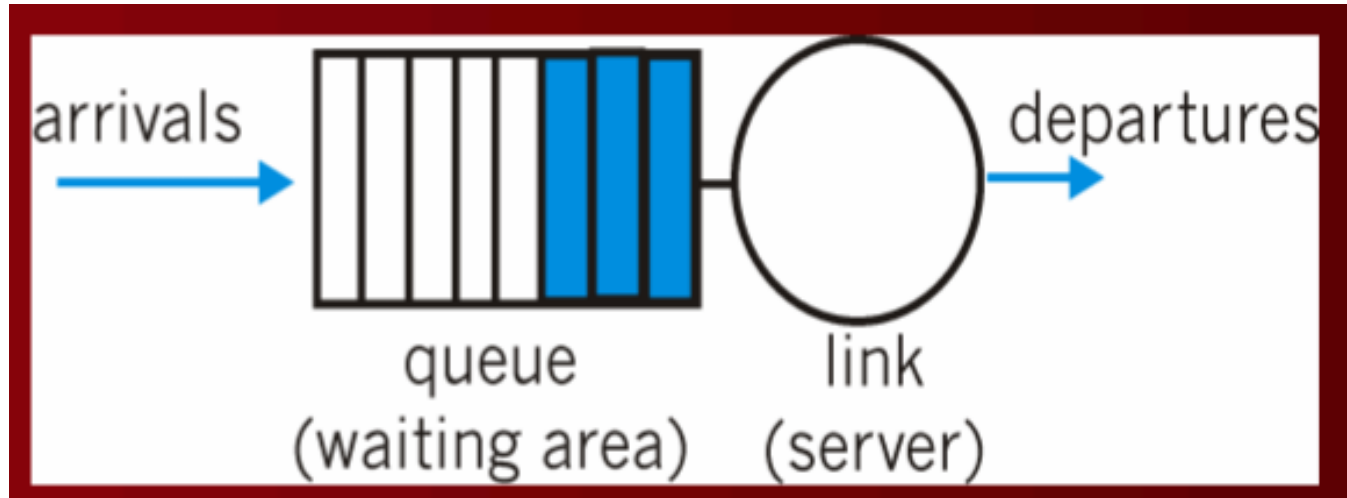 - Applied throughout a network

Resource Control – Queuing

- Queuing

  o Storing packets (cell/frames etc.) while they wait for processing

  o May use one of the following queue mechanisms

    • First In First Out (FIFO)

    • Class Based Queuing (CBQ)

    • Weighted Fair Queuing (WFQ)

    • Random Early Detect (RED)

    • Weighted RED (WRED)

Resource Control – Queuing
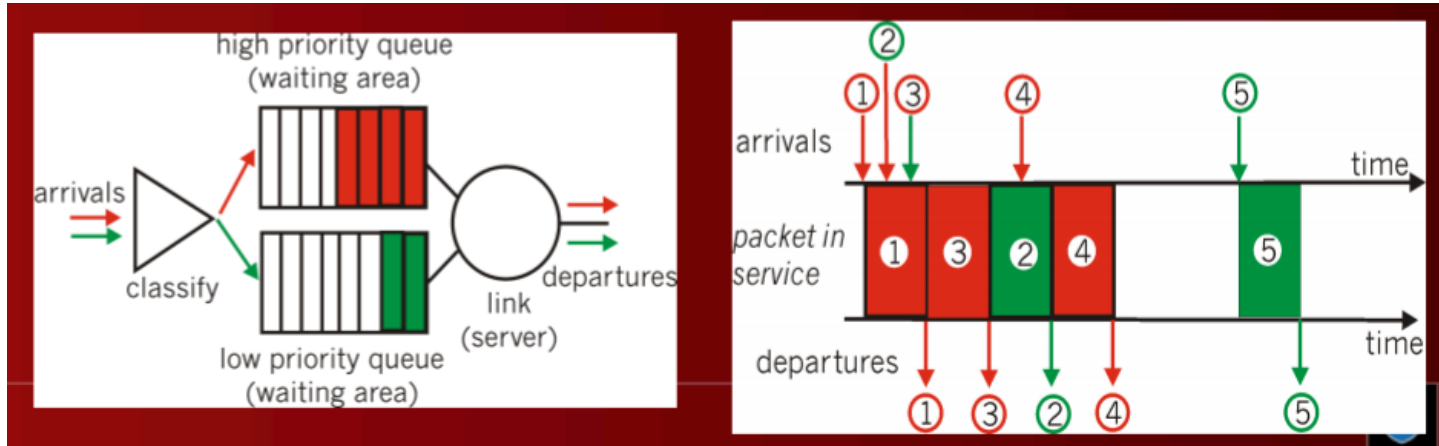
- FIFO (first in first out) scheduling: send in order of arrival to queue

Resource Control – Queuing

- Priority scheduling: transmit highest priority queued packet

- multiple classes, with different priorities

  o class may depend on marking or other header info, e.g., IP source/dest, port numbers, etc.

Scheduling Policies

- round robin scheduling:
  - multiple classes
  - cyclically scan class queues, serving one from each class (if available)

- Weighted Fair Queuing:
  - generalized Round Robin
  - each class gets weighted amount of service in each cycle

Performance Mechanisms - Service Level Agreements (SLAs)

Service Level Agreements (SLAs)

- Typically formal contracts between a provider and a user

- Define the providers responsibility to user and extent of accountability

- SLA performance elements may include

  o Data rate

  o Burst tolerance

  o Upstream/downstream

  o Delay and RMA metrics

- Can be further defined as upstream and down stream

Service Level Agreements

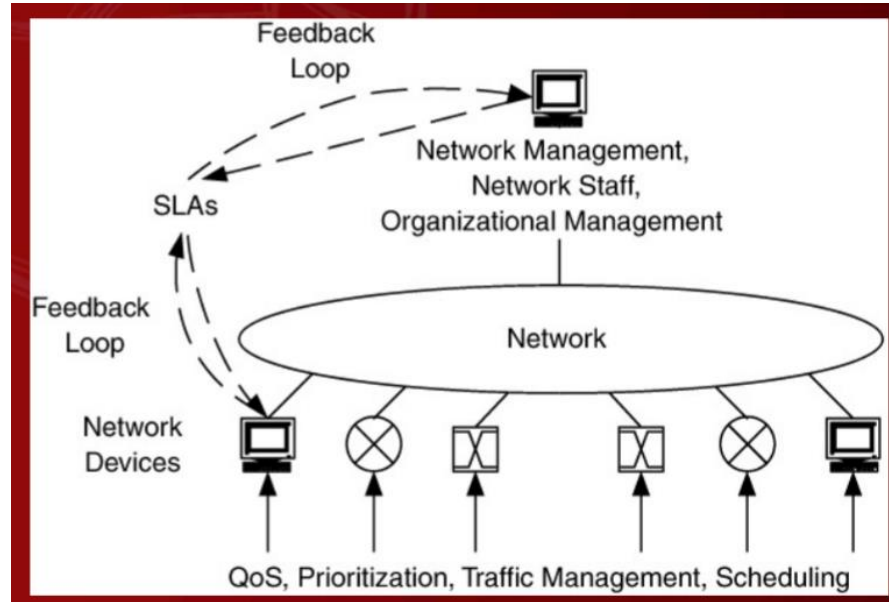| Network Service Description for My Enterprise			
Service Levels:	**Capacity Performance**	**Delay Performance**	**Reliability Performance**
Basic Service	As Available (Best Effort)	As Available (Best Effort)	As Available (Best Effort)
Silver Service	1.5 Mb/s (Bidirectional)	As Available (Best Effort)	As Available (Best Effort)
Gold Service	10 Mb/s (Bidirectional) (Burst to 100 Mb/s)	Max 100-ms Round-Trip (Between Specified Points)	As Available (Best Effort)
Platinum Service	100/10 Mb/s Up/Down (Burst to 1 Gb/s)	Max 40-ms Round-Trip (Between Specified Points)	99.999% Uptime (User-Server)

Figure 8.9: Example of enterprise SLA.   University of Wollongong
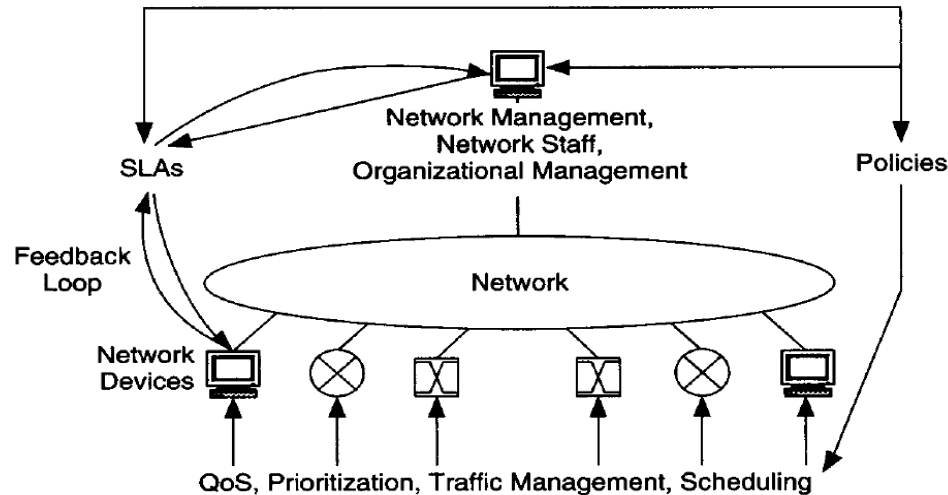
Service Level Agreements

- If used correctly, SLAs can form an important link in communicating network performance between users, staff and management

Figure. Performance mechanisms with SLAs Added (McCabe, 2010, p.351)

Performance Mechanisms - Policies

Policies

- Policies complete the framework for performance for a network by coupling the high-level management view of how the network should perform with mechanisms to implement at the network devices (QoS) and feedback loops using SLAs.

FIGURE 8.11   Performance Mechanisms with Policies Added

55

(McCabe, 2010, p.352)

Policies

- Policies describe
  - what network computing storage or other resources are available to users,
  - when these resources are available or
  - which users are permitted access to certain resources.

- Share some similar attributes to routing and security policies

Architectural Considerations

Architectural Considerations

- **Performance mechanisms should be evaluated against**

  - Requirements of the network

  - Goals

  - Type of environment

Architectural Considerations

- Performance mechanisms that operate on

  o Individual flows should be considered more when traffic flows are more likely to be individual (access networks)

  o Aggregate flows should be considered where traffic flows are more likely to aggregate (interfaces to external networks)

Architectural Considerations

- Internal Relationships

  o Trade-offs between end-to-end and per-hop scheduling and conditioning of traffic flows, admission control, and

  o are flows treated individually or aggregated into groups?

- These factors will determine the kinds of performance mechanisms you impose

Architectural Considerations

- **External Relationships**

  o **Performance and Addressing/Routing**
    - Closely coupled when performance relies on protocols such as DiffServ or MPLS
    - Access networks and switches aim to provide maximum bandwidth and minimal delay before they need to be routed elsewhere

  o **Performance and Network Management (NM)**
    - Depends on Network Management to configure, monitor, manage, verify and bill for performance
    - NM links QoS with SLAs and Policies

Architectural Considerations

- **External Relationships**
  - Performance and Security
    - Security can be the enemy of performance because it
      - ✓ Intercepts, inspects and controls network access
      - ✓ Results in reduced capacity, delays.
      - ✓ Capacity can be compensated for but delay is more difficult

References and Reading

❖ **Chapter 8** - McCabe, J. D. (2010). *Network Analysis, Architecture, and Design*. San Diego, CA, USA: Elsevier Science.

❖ *RFC's 2474, 2475* www.ietf.org

Thank you
Q&A ?