Vorlesung Telematik (Computer Networks)

WS2004/05

QoS, Traffic Engineering and Control-Plane Signaling in the Internet

Dr. Xiaoming Fu

Telematics group
University of Göttingen, Germany

Overview

- Recent trends in network traffic and capacity
- QoS principles: "Beyond Best Effort"
- QoS architecture: IntServ, DiffServ
- MPLS and Traffic Engineering
- Control-plane Signaling: RSVP(-TE), (CR-)LDP & NSIS

Credits:
- Kurose and K. Ross, Computer Networking: A Top-Down Approach Featuring the Internet, 2nd Ed., Addison-Wesley, 2002
- Raj Jain, Ohio State University
- Chunming Qiao, State University of New York
- Olof Hagsand, Royal Institute of Technology.

Trend: Traffic > Capacity

Expensive Bandwidth	Cheap Bandwidth
❑ Sharing	❑ No sharing
❑ Multicast	❑ Unicast
❑ Virtual Private Networks	❑ Private Networks
❑ More efficient use (L3)	❑ Less efficient use
❑ Need QoS	❑ QoS less of an issue
❑ Likely in WANs	❑ Possible in LANs

Telco vs Data Networks

Telco Protocols	Data Protocols
QoS	Simplicity
Reliability	
Protection	Need QoS, ...

Traditional IP Networking

- Connectionless **best effort** service
- Each packet treated independently by routers (**stateless**)
- No bandwidth, delay and loss guarantees for packet; delay variations (jitters) introduced along the path of a packet (**network QoS issue**)
- Route lookup based on dst IP address and longest prefix match (**no distinction** of QoS requirements)
- Routing messages can be dropped without **priority**

Improving QoS in IP Networks

Thus far: "making the best of best effort"

Future: next generation Internet with QoS guarantees

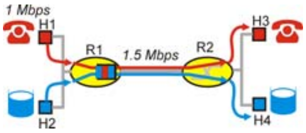
- Classification: Identifying packet flows in the network
- Signaling for resource reservations: RSVP (or NSIS)
- Better than best effort treatment for each flow:
  - Differentiated Services: differential guarantees
  - Integrated Services: firm guarantees

Functions in IP QoS

- Classification
  - Identifying the packets belonging to a certain traffic flow
- Policing and shaping
  - Policing: ensure that the flow conforms to a traffic specification
- Scheduling
  - Manage packets in queues so that they receive desired service
- Admission control and signaling
  - Signal a traffic flow's requirements and check that there are enough resources to accept this flow

Principles for QoS Guarantees

- Example: 1MbpsI P phone, FTP share 1.5 Mbps link.
  - bursts of FTP can congest router, cause audio loss
  - want to give priority to audio over FTP



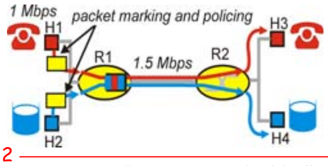
Principle 1

packet **marking and classification** needed for router to distinguish between different classes; and new router policy to treat packets accordingly

Telematics group
University of Göttingen, Germany

Principles for QoS Guarantees (more)

- what if applications misbehave (audio sends higher than declared rate)
  - **policing**: force source adherence to bandwidth allocations
- marking and policing at network edge:
  - similar to ATM UNI (User Network Interface)



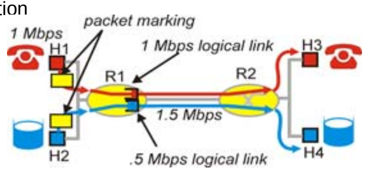
Principle 2

Policing: to force source adherence to bandwidth allocations and provide protection (*isolation*) for one class from others

Telematics group
University of Göttingen, Germany

Principles for QoS Guarantees (more)

- Allocating *fixed* (non-sharable) bandwidth to flow: *inefficient* use of bandwidth if flows doesn't use its allocation



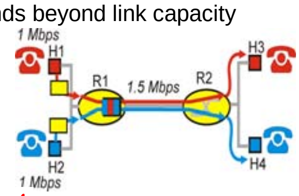
Principle 3

While providing isolation, it is desirable to use resources as efficiently as possible

Telematics group
University of Göttingen, Germany

Principles for QoS Guarantees (more)

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

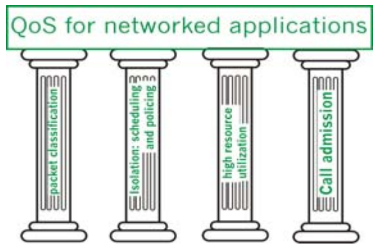


Principle 4

Call Admission: flow declares its needs (**signaling**), network may block call (e.g., busy signal) if it cannot meet needs

Telematics group
University of Göttingen, Germany

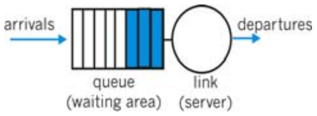
Summary of QoS Principles



Let's next look at mechanisms for achieving this ….

Scheduling And Policing Mechanisms

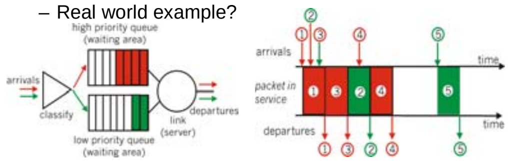
- scheduling: choose next packet to send on link
- FIFO (first in first out) scheduling: send in order of arrival to queue
  - real-world example?
  - discard policy: if packet arrives to full queue: who to discard?
    - Tail drop: drop arriving packet
    - priority: drop/remove on priority basis
    - random: drop/remove randomly



Scheduling Policies: more

Priority scheduling: transmit highest priority queued packet

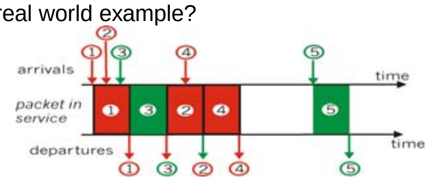
- multiple *classes*, with different priorities
  - class may depend on marking or other header info, e.g. IP source/dest, port numbers, etc..
  - Real world example?



Scheduling Policies: still more

round robin scheduling:

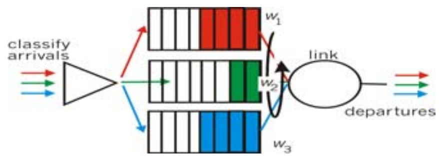
- multiple classes
- cyclically scan class queues, serving one from each class (if available)
- real world example?



Scheduling Policies: still more

Weighted Fair Queuing:

- generalized Round Robin
- each class gets weighted amount of service in each cycle
- real-world example?



Policing Mechanisms

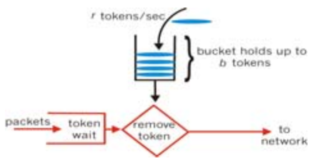
Goal: limit traffic to not exceed declared parameters

Three common-used criteria:

- *(Long term) Average Rate:* how many pkts can be sent per unit time (in the long run)
  - crucial question: what is the interval length: 100 packets per sec or 6000 packets per min have same average!
- *Peak Rate:* e.g., 6000 pkts per min. (ppm) avg.; 1500 ppm peak rate
- *(Max.) Burst Size:* max. number of pkts sent consecutively (with no intervening idle)

Policing Mechanisms

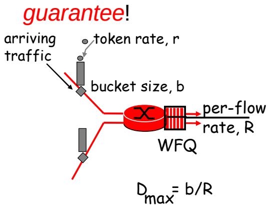
Token Bucket: limit input to specified Burst Size and Average Rate.

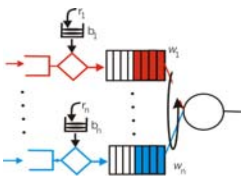


- bucket can hold b tokens
- tokens generated at rate *r token/sec* unless bucket full
- *over interval of length t: number of packets admitted less than or equal to (r t + b).*

Policing Mechanisms (more)

- token bucket, WFQ combine to provide guaranteed upper bound on delay, i.e., *QoS guarantee*!





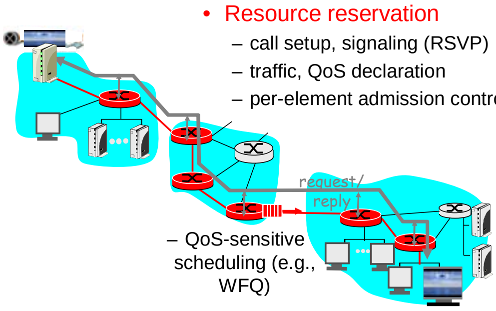
IETF Integrated Services

- architecture for providing QoS guarantees in IP networks for individual application sessions
- resource reservation: routers maintain state info (a la VC) of allocated resources, QoS req's
- admit/deny new call setup requests:

Question: can newly arriving flow be admitted with performance guarantees while not violated QoS guarantees made to already admitted flows?

Intserv: QoS guarantee scenario

- Resource reservation
  - call setup, signaling (RSVP)
  - traffic, QoS declaration
  - per-element admission control



- QoS-sensitive scheduling (e.g., WFQ)

Call Admission

Arriving session must :

- declare its QoS requirement
  - R-spec: defines the QoS being requested
- characterize traffic it will send into network
  - T-spec: defines traffic characteristics
- signaling protocol: needed to carry R-spec and T-spec to routers (where reservation is required)
  - RSVP, NSIS

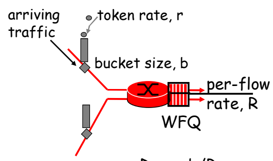
Intserv QoS: Service models [rfc2211, rfc 2212]

Guaranteed service:

- worst case traffic arrival: leaky-bucket-policed source
- simple (mathematically provable) *bound* on delay [Parekh 1992, Cruz 1988]

Controlled load service:

- "a quality of service closely approximating the QoS that same flow would receive from an unloaded network element."



$D_{max} = b/R$

IETF Differentiated Services

Concerns with Intserv:

- Scalability: signaling, maintaining per-flow router state difficult with large number of flows
- Flexible Service Models: Intserv has only two classes. Also want "qualitative" service classes
  - "behaves like a wire"
  - relative service distinction: Platinum, Gold, Silver

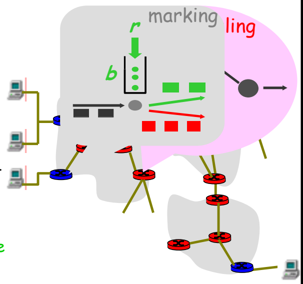
Diffserv approach:

- simple functions in network core, relatively complex functions at edge routers (or hosts)
- Don't define define service classes, provide functional components to build service classes

Diffserv Architecture

Edge router:
- per-flow traffi nagement
- marks packets as in-profile and out-profile

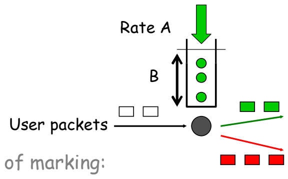
Core router:
- per class traffic management
- buffering and scheduling based on marking at edge
- preference given to in-profile packets
- Assured Forwarding



Telematics group
University of Göttingen, Germany

Edge-router Packet Marking

- profile: pre-negotiated rate A, bucket size B
- packet marking at edge based on per-flow profile



Possible usage of marking:
- class-based marking: packets of different classes marked differently
- intra-class marking: conforming portion of flow marked differently than non-conforming one

Telematics group
University of Göttingen, Germany

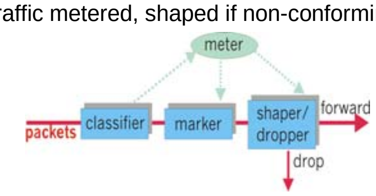
Classification and Conditioning

- Packet is marked in the Type of Service (TOS) in IPv4, and Traffic Class in IPv6
- 6 bits used for Differentiated Service Code Point (DSCP) and determine PHB that the packet will receive
- 2 bits are currently unused

Telematics group
University of Göttingen, Germany

Classification and Conditioning

may be desirable to limit traffic injection rate of some class:
- user declares traffic profile (eg, rate, burst size)
- traffic metered, shaped if non-conforming



Forwarding (PHB)

- PHB result in a different observable (measurable) forwarding performance behavior
- PHB does not specify what mechanisms to use to ensure required PHB performance behavior
- Examples:
  - Class A gets x% of outgoing link bandwidth over time intervals of a specified length
  - Class A packets leave first before packets from class B

Forwarding (PHB)

PHBs being developed:

- Expedited Forwarding: pkt departure rate of a class equals or exceeds specified rate
  - logical link with a minimum guaranteed rate
- Assured Forwarding: 4 classes of traffic
  - each guaranteed minimum amount of bandwidth
  - each with three drop preference partitions

QoS Networking: Summary

- making the best of today's best effort service
- scheduling and policing mechanisms
- next generation Internet: Intserv, Diffserv, …and
- MPLS and traffic engineering
- an important component for QoS and TE: signaling protocols:
  - RSVP, RSVP-TE
  - LDP, CR-LDP
  - NSIS

Why need MPLS?

- Limitations of existing IP Network
  - Network Scaling
  - Traffic Engineering
  - Provisioning of QoS
- **We need better control over the network.**
- MPLS stands for **MultiProtocol Label Switching**.
  - Convergence of connection oriented forwarding techniques and Internet's routing protocols

Introduction

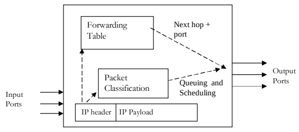
- Conventional Layer-3 (IP) forwarding

  Each router analyzes the incoming packet's header and independently chooses a next hop. Routing algorithm and adequate speed are prerequisite.

- MPLS (Layer 2.5) forwarding

  All forwarding is driven by the labels, no header analysis needed. Once a packet enters a network, it's assigned a label. Each router forwards packets according to their labels.

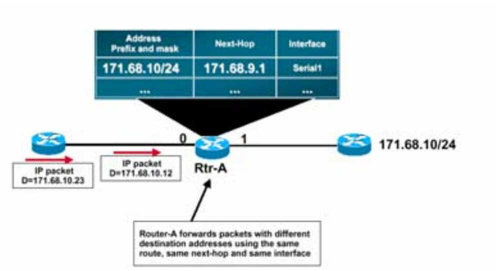
Native IP Forwarding



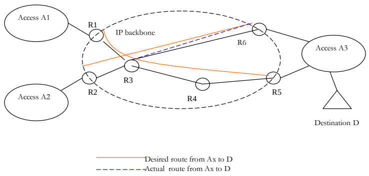
NIF node forwarding Engine

- Longest-prefix match based on packet's destination IP address

Packet IP forward

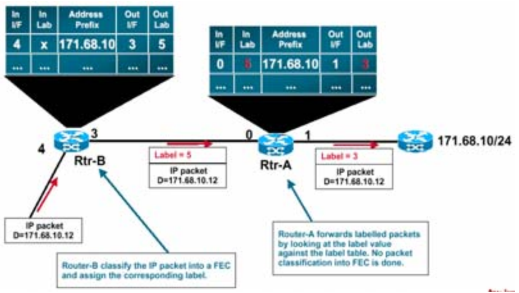


MPLS Network



Traffic Engineering to override shortest path route

MPLS Packet forward



Terminology

- **Label** - a short fixed length identifier used to identify a FEC, usually of local significance
- **FEC** - Forwarding Equivalence Class represents set of packets with common cross core forwarding requirements
- **LSR** - Label Switched Router
- **LER** - Label Edge Router
- **NHLFE** - Next Hop Label Forwarding Entry
- **ILM** - Incoming Label Map
  Maps label to a set of NHLFE entries
- **LSP** - Label Switched Path
  Path through one or more LSRs at one level of hierarchy followed by packets in a particular FEC

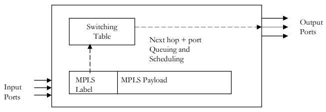
Key concept in MPLS

- Seperation of IP router’s function into Forwarding and Control
- **Forwarding** - deals with how data packets are relayed between IP routers, uses label swapping.
- **Control** - consists of network layer routing protocols to distribute routing information between LSR’s and label binding procedures for converting this routing information into forwarding tables needed for Label Switching.

MPLS Advantages

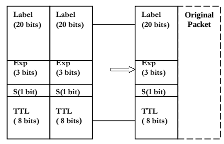
1. Router can use any information in determining label assignment, not limited to packet header;
2. How to distribute labels may become more and more complicated, without any impact on the routers that merely forward labeled packets.
3. A label can be used to represent a pre-chosen route so that the identity of explicit route need not be carried with the packet.
4. Mutiprotocol: its techniques are applicable to ANY network layer protocol.

Label Based Forwarding



- At each LSR, forwarding is done by the single index lookup into the switching table using the packet's MPLS label.
- The switching table is loaded a priori with a unique next-hop label, output port and queuing and scheduling rules.
- The establishment of mapping information is responsibility of control part - done using Label Distribution Protocols

Label Stack



MPLS stack encoding for packet-oriented transport

- MPLS allows hierarchical labels supported as LIFO stack.
- A packet is always processed based on the top label regardless of other labels that may be below it.
- Each label stack entry is 32 bits.
  - 20 bits for label
  - 3 bits for experimentation
  - 8 bits for TTL and 1 stack bit.

Control in MPLS

- Consists of
  - Network Layer routing protocols to distribute routing information between LSRs.
  - Label binding procedures to convert this routing information into the forwarding tables needed for label switching
- QoS routing requires additional information about availability of resources in the network and QoS requirements of each flow.
- A signaling protocol is needed for distributing labels (creating label/FEC mapping in LSRs)
  - and also for reserving needed resources along the a selected route
  - e.g. CR-LDP, RSVP-TE

MPLS Labels

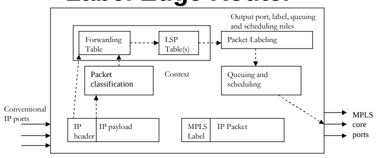
Upstream LSR *Downstream LSR*

Agreement: "binding" label L to FEC F for packets moving from Ru to Rd.

So, L becomes Ru's "outgoing label" representing FEC F, and L becomes Rd's "incoming label" representing FEC F. Note that L is an arbitrary value whose binding to F is local to Ru and Rd.

Telematics group
University of Göttingen, Germany

Label Edge Router



Ingress LER

- LER terminates and originates LSP’s and performs both label based forwarding and conventional NIF functions.
- **Ingress LER** - labels unlabelled packets and creates an initial MPLS frame by pushing one or more MPLS label entries.
- **Egress LER** - terminates LSP by popping the top MPLS stack entry

Telematics group
University of Göttingen, Germany

Recap

- Packet processing based on the top level label regardless of the label underneath
- FECs can be
  - coarse grained consisting of all the packets with same destination address
    - Allow the overall system of be scalable where it is useful to handle large bundle of flows as a single class of traffic
    - Help in rerouting in event of occurrence of a fault
  - fine grained as in packets belonging to a particular application running between two hosts.
    - Help in providing different QoS to different flows.

Telematics group
University of Göttingen, Germany

Recap Contd.

- Mapping of packets to an FEC done only once at the Ingress router upon entry into an MPLS domain
- Subsequent packets are forwarded strictly according to their labels
- label is removed by egress LSR
- Each LSR maintains label to NHLFE mapping giving a set of entries for each FEC.
- Mapping can be changed for
  - load balancing over multiple paths
  - rerouting from a failed path to an alternate path

Telematics group
University of Göttingen, Germany

Route Selection

- Method used for selecting the LSP for a particular FEC.
- **Hop by Hop** is the same as topology driven.
- **Explicit Routing**
  - Explicit route need to be specified only at the time that labels are assigned and not with each IP packet, as in case of IP routing
- **Tunneling**
  - A router Ru takes explicit action to cause a particular packet to be delivered another router Rd even though Ru and Rd are not consecutive routers on the hop-by-hop path for that packet and Rd is not the packet’s ultimate destination. This concept is called tunneling.
  - **Hop-by-Hop routed tunnel**
  - **Explicitly routed tunnel**

LSP Tunnels

- Implement a tunnel as a LSP, and use label switching rather than network layer encapsulation to cause a packet to travel through the tunnel.
- Set of packets sent through the LSP tunnel constitutes a FEC and each LSR in the tunnel must assign a label to that FEC.
- If a tunnel from Ru to Rd, then
  – Ru is transmit endpoint of the tunnel
  – Rd is receive endpoint of the tunnel

Traffic Engineering

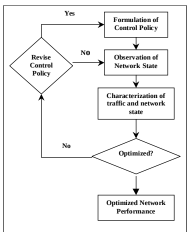
Application of technology and scientific principles to the measurement, modeling, characterization and control of internet traffic. (Fundamentally, a control problem)

e.g mapping of traffic on IP network infrastructure

A network consists of :

- Demand System (Traffic)
- Constraint System (Interconnected N/W Elements)
- Response System (N/W Protocols and Processes)

Traffic Engineering Process Model



Traffic Engineering Process Model

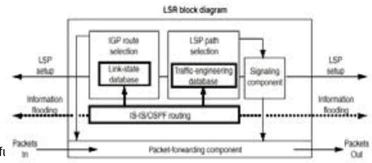
- **Performance Objective**
  – **Resource Oriented**
    - Efficient Link Utilization (Congestion Control)
  – **Traffic Oriented**
    - Packet Loss
    - Delay / Delay Variation
    - Throughput
- *Adaptive And Iterative Process*

MPLS and Traffic Engineering

- MPLS supports origination connection control through explicit LSP's
- Traffic trunk : Aggregation of traffic belonging to the same class.
- Mapping of traffic trunks on to the network topology is done by selection of routes for explicit LSP's.
- LSP tunnels provide
  - rerouting in congested conditions
  - Flexible cost effective survivability
  - Provide statistics for Traffic Matrix
  - Parameterized resource allocation

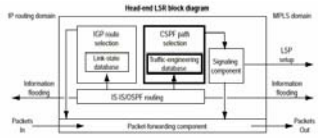
Components of MPLS TE Model

- **Network State Information Dissemination**
  - **Extending conventional IGP's link state advertisements**
    - OSPF extensions implemented with Opaque LSAs
    - IS-IS extensions implemented using Type Length Values (TLVs)
  - **Traffic Engineering Database(TED) maintained by each LSR**



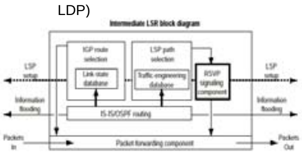
Components of MPLS TE Model

- **Path Management**
  - **Selection**
    - Explicit route for LSP tunnel generated
    - Strict or Loose path(Abstract node) is specified
    - May be defined administratively or computed automatically by a constraint-based routing entity(CSPF).



Components of MPLS TE Model

- **Path Management**
  - **Instantiation or Placement**
    - Signaling Protocol which serves as an Label Distribution Protocol
      - » Resource Reservation Protocol (RSVP) extensions
      - » Constraint Routed Label Distribution Protocol (CR-LDP)



  - **Maintenance of LSP tunnels**
    Sustain, Reroute or Terminate LSP tunnel

Components of MPLS TE Model

- **Traffic Assignment**
  - All aspects associated with allocation of traffic to established LSP's
    - Partitioning Function
- **Network Management**
  - *Online management is Non-deterministic*
  - *Offline management tools interfaced with MPLS to provide external feedback*