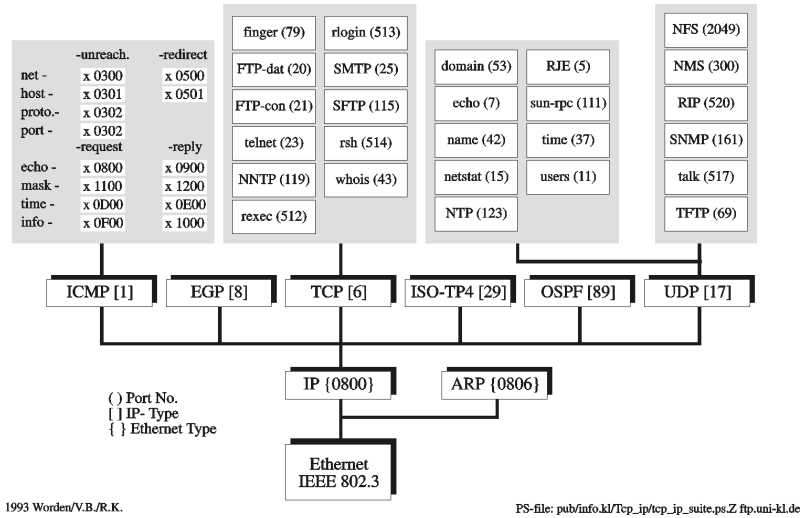
Protocols

Tasks

- End-to-end connectivity (host-to-host)
- Process-to-Process connectivity
- Reliable communication
  - Error detection
  - Error recovery, e.g. forward error correction or retransmission
- Resource management
  - avoid congestion, by flow control
    - within the network
    - within end systems
  - Priorisation
  - Resource reservation
- Support for specialized media types
  - Content description
  - Timing / Synchronization Information
- And more ...

4.8 TCP/IP Suite



4.8 IP Protocol (RFC 791)

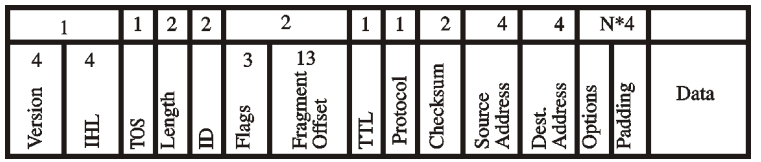
Development of IP

- DARPA: Defense Advanced Research Projects Agency
  - The research aim was to build a network that is tolerant to extensive damage, e.g. by a nuclear strike
  - 1973/1974 development of TCP/IP, a replacement of NCP (Network Control Protocol)
  - Since 1975 the ARPANET was controlled by the DoD
  - In the early 80'th the military part was extracted from the ARPANET
  - Since 1983 exclusive use of TCP/IP, defining the term Internet
- IP is specified in RFC 791
  - „This document is based on six earlier editions of the ARPA Internet Protocol Specification ..."

IP characteristics

- Provides end-to-end communication
- Connection less, i.e. state less protocol
- Provides unreliable transfer of packets
- Packets may be reordered during transmission
- Error messages are handled by the separate ICMP (Internet Control Message Protocol) protocol

IP Header



Version: version of IP header

IHL : IP header length in 32 bit words (5+ no. of options)

TOS: Type Of Service

0	3	4	5	6
precedence	D	T	R	reserved

precedence ~ priority; D,T,R if set optimize for Delay, Throughput, Reliability

Length: length in bytes including the IP header

ID: serial number

Flags, Frag. Offset: used for fragmentation

TTL: Time To Live, decremented by each machine to pass the packet

Protocol: layer 4 protocol, e.g. 1=ICMP, 6=TCP, 17=UDP

Checksum: checksum for the IP header

Options: security, record route, timestamp, source routes

IP Addresses

An IP address is a unique identifier for an end system

Each Address consists of two parts:

- a Network address
- a host address

→ a hierarchy of two levels

8 bit	8 bit	8 bit	8 bit

Example : 131 . 246 . 9 . 50

Class **A** Network

0	7-bit Network	24-bit Host

128 Networks with **17.000.000** Hosts

Address		Status
0.0.0.0		*Reserved*
1.0.0.0	- 126.0.0.0	Available
127.0.0.0		*Reserved*

Class **B** Network

1	0	14-bit Network	16-bit Host

16.000 Networks with **65.000** Hosts

Address		Status
128.0.0.0		*Reserved*
128.1.0.0	- 191.254.0.0	Available
191.255.0.0		*Reserved*

Class **C** Network

1	1	0	21-bit Network	8-bit Host

2.000.000 Networks with **255** Hosts

Address		Status
192.0.0.0		*Reserved*
192.0.1.0	- 223.255.254.0	Available
255.255.255.255		*Reserved*

1993 Worden/V.B./R.K.

PS-file: pub/info.kl/Tcp_ip/Internet_Addressing.ps ftp.uni-kl.de

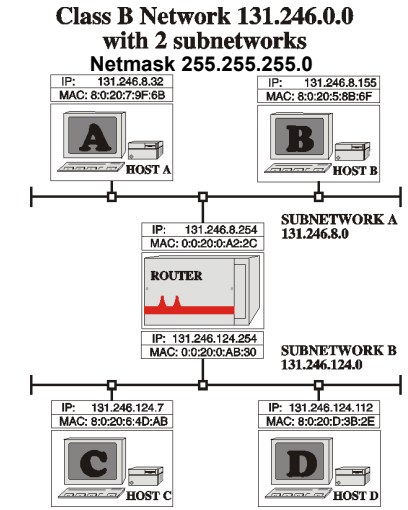
IP Addresses and Netmasking

IP-Address	131 • 246 • 9 • 5	dotted decimal format
	83 • F6 • 9 • 5	hexadecimal format
	10000011 • 11110110 • 00001001 • 00000101	binary format
Network-Mask	11111111 • 11111111 • 11111111 • 00000000	binary format
	FF • FF • FF • 0	hexadecimal format
	255 • 255 • 255 • 0	dotted decimal format

network | host

→ different network masks enable multiple level of hierarchies

Example: end-to-end communication



1994 Worden/V.B./R.K.

PS-file: pub/info.kl/route.ip.ps.Z ftp.uni-kl.de

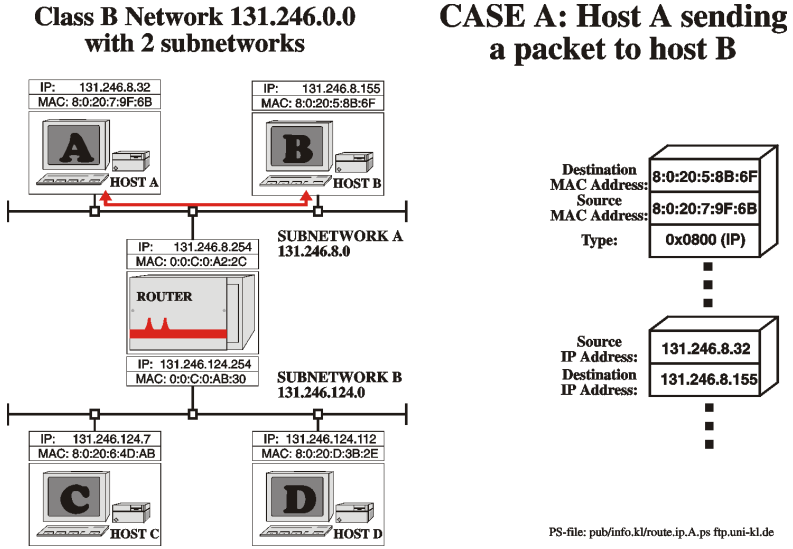
1. Communication within the same network
2. Communication to another network

Address Resolution Protocol (ARP)

ARP provides a mapping of IP addresses to MAC addresses

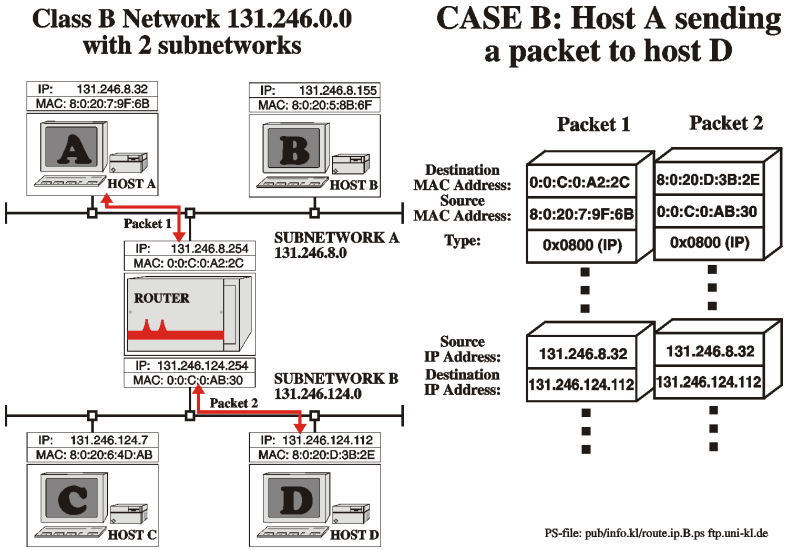
- The source station sends an ARP Request, this is a broadcast, containing:
  - The source IP address
  - The source MAC address
  - The destination IP address
- The destination sends an ARP Reply, this is an unicast message containing:
  - All addresses of the ARP Request
  - The destination MAC address
- Caching of address mapping
- ARP extensions
  - Inverse ARP, provides a mapping of MAC addresses to IP addresses
  - Proxy ARP, e.g. switches may send an ARP Reply instead of forwarding a broadcast

Example: communication within a network



PS-file: pub/info.kl/route.ip.A.ps ftp.uni-kl.de

Example: communication to another network



PS-file: pub/info.kl/route.ip.B.ps ftp.uni-kl.de

Routing

Routers are used to connect Networks

Routing: making a local forwarding decision based on global topology information

How to get topology information:

- Default routing
- Static or non adaptive routing
  - Does not take into account changes of the network
- Dynamic or adaptive routing, requires routing protocols
  - Identifies the topology of the network automatically
  - Different metrics are used to describe distances
    - Number of hobs
    - Available bandwidth
    - Error rates
    - ...

Hierarchy of routing

Interior Gateway Protocol (IGP)

- Routing within a LAN or administrative domain
- Each node may know the whole (local) topology
- Examples: Routing Information Protocol (RIP), IGRP, Hello, OSPF

Exterior Gateway Protocol (EGP)

- Between different providers
- Whole LANs are represented as a single node
- Examples: Border Gateway Protocol (BGP), EGP

Example: RIP (1)

RIP uses a very simple metric: the number of hops

- a maximum of 15 hops is allowed
- hop count 0: directly connected network
- hop count 16: unreachable

A router sends its routing table periodically to all connected networks, except

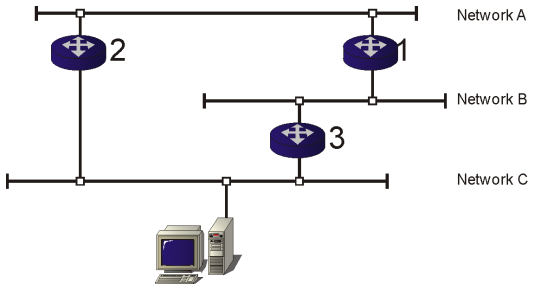
- Information about directly connected networks are not send to those networks
- Information is not send to the direction where the information came from

Update of routing information

- IP-Broadcasts to UDP port 520
- Every 30 sec. an update process sends routing information to all connected networks
- A table entry which is not updated within 90 sec is marked to be unusable
- A table entry will be deleted after 240 sec. If there is no update

Example: RIP (2)

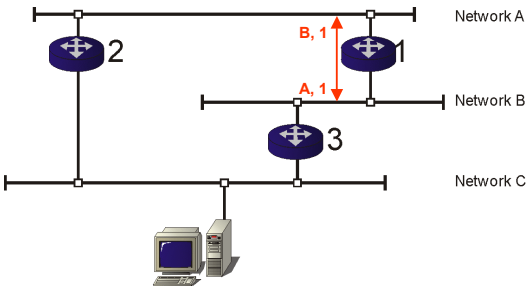
Every Router has knowledge about directly connected networks only



Routing tables network, distance (router)			
Router 1	Router 2	Router 3	Host
A, 0 B, 0	A, 0 C, 0	B, 0 C, 0	C, 0

Example: RIP (3)

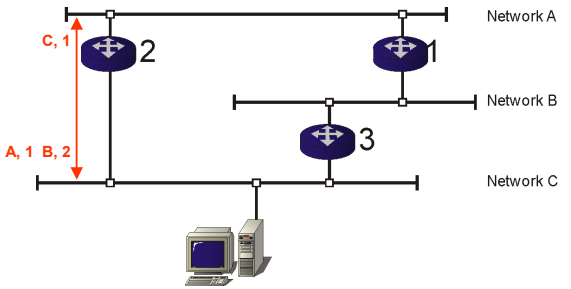
Router 1 sends an update



Routing tables network, distance (router)			
Router 1	Router 2	Router 3	Host
A, 0 B, 0	A, 0 B, 1 (R1) C, 0	A, 1 (R1) B, 0 C, 0	C, 0

Example: RIP (4)

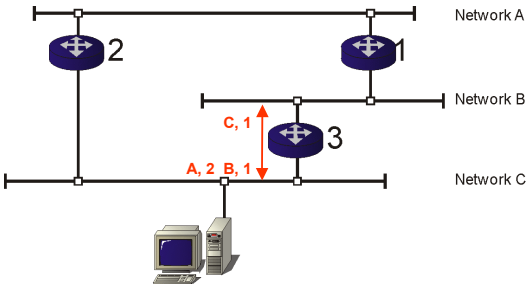
Router 2 sends an update



Routing tables network, distance (router)			
Router 1	Router 2	Router 3	Host
A, 0 B, 0 C, 1 (R2)	A, 0 B, 1 (R1) C, 0	A, 1 (R1) B, 0 C, 0	A, 1 (R2) B, 2 (R2) C, 0

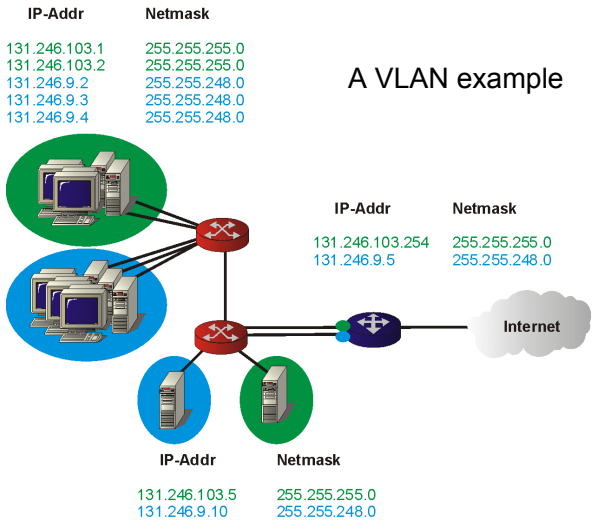
Example: RIP (5)

Router 3 sends an update, routing tables have converged



Routing tables network, distance (router)			
Router 1	Router 2	Router 3	Host
A, 0	A, 0	A, 1 (R1)	A, 1 (R2)
B, 0	B, 1 (R1)	B, 0	B, 1 (R3)
C, 1 (R2)	C, 0	C, 0	C, 0

Virtual LAN (1)



Virtual LAN (2)

VLANs are connected by routers

- Hosts of the same LAN which are configured to be in different VLANs must use a router to communicate with each other
- Routers must have a port in each VLAN

Switches should support VLANs

- Broadcast are not flooded over the whole network
- Hosts of different VLANs do not „see“ each other

VLAN advantages

- Within large networks hosts could be grouped by organizational requirements rather than physical location
- Its possible to use centralized routers
  - Fewer routers needed
  - Fewer expensive router ports needed

UDP (RFC 768) / TCP (RFC 793)

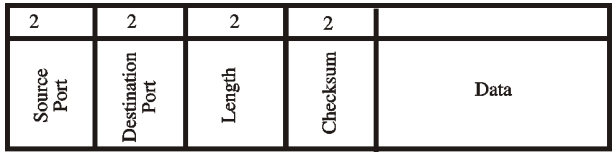
Transport protocols (Layer 4)

- Provides process to process connectivity
- Uses port number to identify processes. An IP address and a port number is an unique identifier for a service.

Characteristics

- Closely related to IP
- UDP offers a connectionless and unreliable transport service
  - Nearly the same service than IP
  - Data unit name: datagram
- TCP offers a connection oriented and reliable transport service
  - Recognition of lost data
  - Retransmission of lost data
  - Reorder data
  - Delete duplicated data
  - Flow control
    - With respect to network congestion
    - With respect to buffer overflow at the receiver side
  - User data is handled as a stream of bytes
    - User data is split into segments
  - Data unit name: segment

UDP Header



Source Port,
Destination Port: port number of sender and receiver

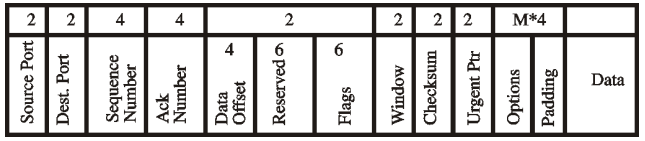
Length: length of the UDP datagram

Checksum: the checksum covers the UDP pseudo header and the UDP data

The UDP pseudo header includes the UDP Header and 12 bytes of the IP header:

- 2*4 byte IP source and destination address
- 1 byte 0
- 1 byte protocol
- 2 byte IP header length

TCP Header



Source Port,
Destination Port: port number of sender and receiver

Sequence Number: TCP counts every byte of a stream
this is the number of the first data byte

Ack Number: if the ACK control bit is set this field contains the value of the next sequence number the sender expects to receive

Data Offset: number of 32 bit words in the TCP header

Flags:

urg	ack	psh	rst	syn	fin

urg: urgent pointer is valid
psh: push data to level above
syn: synchronize seq. Number
ack: Ack Number is valid
rst: reset connection
fin: close this side of the connection

Window: amount of data the sender is willing to accept (flow control)

Urgent Pointer: points to last byte of urgent data

Options: e.g. maximum segment size

TCP Service Mechanisms

Retransmission

- GO-Back-N strategy
  - Simple and robust mechanism
  - Resource consuming
  - Causes unpredictable delay

Flow control

- Slow start and congestion avoidance realize considerate resource usage
  - This enables fair and cooperative bandwidth sharing
  - May cause high jitter

TCPs service mechanisms were designed for reliable data transfer

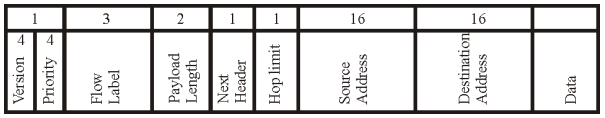
 TCP is not suitable for real-time communication

IPv6

Development of IPv6

- 1993 the IETF called for the development of an IP next generation IPng (RFC 1550)
- Improvements required
  - Larger address space
  - Reduce size of routing tables
  - Simplification of the protocol, to allow routers to process packets faster
  - Better security
  - Pay more attention to Type of Service
  - Aid multicasting
  - Support roaming
  - Easier extension of the protocol
  - Coexistance with the old IPv4
- 1995 the IETF agreed to specification named IPv6 (RFC1883)
  - Changes to other protocols of the TCP/IP suite are specified in RFC 1884-1887

Ipv6 Header



Version: version of IP header

Priority: 0-7 for non real time data, 8-15 for real-time data

Flow Label: may be used to identify a flow, RFC 1809 dicusses how the flow label could be used

Payload length: length of the datagram without the header

Next header: options are placed in separate extension header next header identifies an option or the protocol above IPv6

Hop limit: same as Time To Live of IPv4

Addresses: there are 7*10^23 IPv6 addresses per square meter of the enabling well structured addresses

- Support of provider based addresses
- And geographic based addresses

IP Header Extension

Hop-by-Hop options
- extending the payload length, enabling datagrams of more than 65536 bytes (so called jumbograms)
- Information exchange between routers

Routing
- Full or partial routing path of a datagram

Fragmentation
- Similar of IPv4 fragmentation but fragmentation is handled only by endsystems

Authentication
- Identification of the sender

Encryption security payload
- Information about encrypted payload

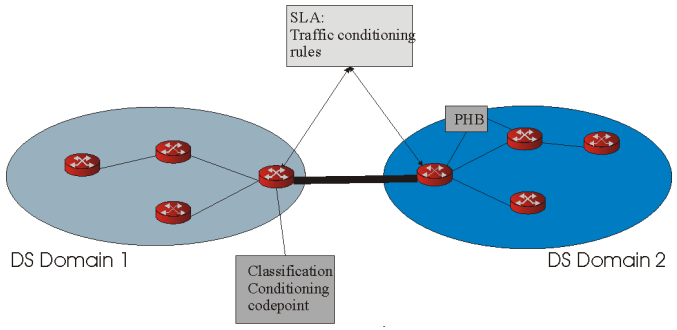
Destination options
- Information that should be interpreted by the destination only

4.9 QoS in Data Networks

	IP Service Models			*Traffic-Engeneering Concept*	*Network Technology*
	Best-Effort	DiffServ	IntServ	MPLS	ATM
QoS Guarantees	**No**	**Aggregated**	**Flow based**	**Flow based and aggregated**	**Flow based and aggregated**
QoS Parameter	**No**	**Long term Static Within a domain**	**Per flow Dynamic End-to-end**	**Support for: • DiffServ • IntServ • ATM**	**Per flow (channel) or per path Dynamic or static End-to-end or within a domain**

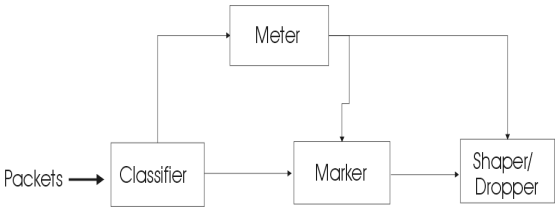
Diffserv Concepts

- The network traffic is classified at the entry into a small number of classes(behavior aggregate), each class is assigned a single DS codepoint, packets forwarding are based on codepoint
  - SLA: Service Level Agreement
  - PHB: Per Hop Behavior
  - Boundary and interior nodes



Classification and Conditioning

- Network traffic is classified based on the packet header
  - Complex Classification and traffic Conditioning functions are normally only located in boundary nodes
    - Behavioral Aggregate Classifier
    - Multi Field Classifier
- Traffic conditioning are control functions performed to enforce rules specified in a Traffic Conditioning Agreement
    - Meter
    - Marker
    - Shaper
    - Dropper



Codepoint

- Each behavior aggregate is identified by a single DS Codepoint, within the core of the Domain, packets are forwarded according to the per-hop-behavior associated with the DS Codepoint
  - DiffServ Codepoint is carried in Type of Service (ToS) field of IPV4 header and the Priority field of IPv6 header
  - No signaling protocol is required

DSCP (DiffServ Codepoint) - 6 bits
CU (Currently unused) - 2 bits
DSCP = 101110 indicates EF PHB

Per-Hop-Behavior

- PHB specifies the externally observable forwarding behavior applied at a DS node to a DS behavior aggregate
  - PHBs may be specified in terms of their resource (buffer, bandwidth) priority relative to other PHBs, or in terms of their relative observable traffic characteristics (delay, loss)
  - DiffServ codepoint of each packet is mapped into the set of PHBs
  - The mapping from DiffServ codepoint to PHB maybe 1 to 1 or N to 1
- Standard PHBs
  - Default
    - No special treatment
  - Expedited forwarding
    - With minimal delay and loss
  - Assured forwarding
    - Forwarding of IP packets in 4 independent classes
    - 3 different dropping levels within each class

Integrated Services (RFC 1633)

Integrated Services (IntServ) is a reservation based model. The intend is to guarantee different QoS profiles for each flow.

What is a flow?

- A flow is a stream of packets originated from the same application session
- A flow describes semantically coherence of data

Categories of applications

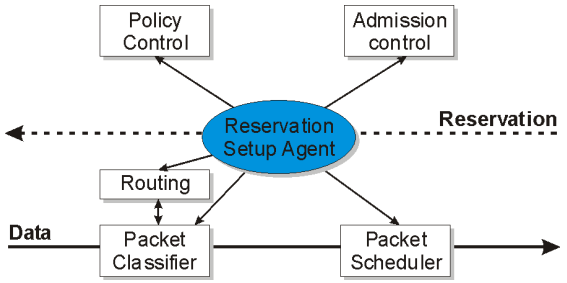
- Elastic applications, no delivery requirements as long as the packets reach the destination, e.g. TCP traffic (machine to machine)
- Real Time Tolerant (RTT) applications, demand weak bounds for the maximum transfer delay, also some packet loss is acceptable, e.g. streamed Video (machine to human)
- Real Time Intolerant (RTI) applications, demand minimal delay and jitter, e.g. interactive application or videoconferences (human to human)

IntServ: Service Classes

Service Classes

- Guaranteed Services for RTI applications
  - Guaranteed amount of bandwidth
  - Deterministic upper bound for delay
- Controlled Load Service for RTT applications
  - Provides a service equivalent to an unloaded network
  - Most packets will reach the destination
  - The average delay is guaranteed
- Best Effort Service for all other applications
  - standard use of IP

IntServ: Components



Component model for IntServ nodes (routers)

The Reservation Setup Agent is the only component that communicates with other nodes.

The Resource Reservation Protocol (RSVP) was designed to meet the signaling requirements of IntServ.

IntServ: Packet Classifier

Packet Classifier

- Determines the QoS class for each packet
- May cooperate with routing mechanisms, packets of the same flow should always use the same path
- Arbitrary parts of the packet header may be used for classification
  - IP-Addresses and port numbers of the sender and/or receiver
  - Application data like frame types of a video stream (e.g. MPEG)
- For IPv6 the flow label should be used

Potential problems:

- IP fragmentation must be avoided (RSVP compute a minimum MTU for a multicast tree)
- Variable header length may require the interpretation of protocols
- IP-level security prevents access to higher layer protocols

IntServ: Packet Scheduler

Packet Scheduler

- The QoS capabilities of a Layer-2 network could be utilized to implement the packet scheduler functionality
- Manages access to the Layer-2 network in order to guarantee the requested QoS
- Some methods for the QoS implementation
  - Priority queuing (there is one queue for each QoS class)
  - Weighted fair queuing (place a packet within a queue depending on the QoS class)
  - Packet level traffic shaping (avoid peaks of traffic)

Potential problem:

- The IntServ model can not guarantee QoS if anywhere in the data path
  - A non IntServ capable router must be passed
  - A non QoS capable Layer-2 network must be shared with a non IntServ capable node

IntServ: Admission and Policy Control

Admission Control

- resource management
- decides if the local system is able to support the requested traffic flow

Policy Control

- is optional
- determines whether the requester has administrative permissions to make the reservation

If either check fails the request is rejected PDUs of that flow may be discarded or forwarded with a lower priority

RSVP Characteristics

RSVP is a general signaling protocol for QoS control services (RFC 2210)

- The signaled objects are opaque for RSVP

A main focus of RSVP is to support multicast communication

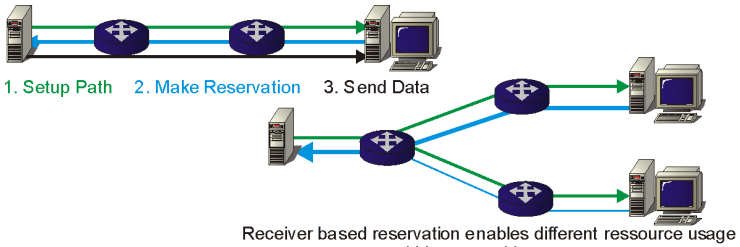
- Unicasts are treated as special cases of multicast only
- RSVP performs receiver oriented reservations
  - Support different requirements of many receivers
  - Support heterogeneous networks

RSVP uses soft-states, i.e. signaled information is valid for a fixed time interval only

- Simplifies cooperation with connection less IP
- Reservations must be refreshed periodically
  - Resource consuming
  - Prevents usage of RSVP in large networks

Merging of reservation supports multiple senders in a multicast environment

RSVP Receiver oriented Reservation



Receiver based reservation enables different ressource usage within one multicast tree

- A PATH message is send from sender to receiver
  - The sender specifies its traffic characteristic
  - Detection of path characteristics
    - Detected bandwidth limitations, minimum packet size (MTU)
    - RSVP capable nodes get to know their RSVP capable neighbors
  - RSVP does not perform routing, routing is done by standard components which do not know anything about QoS
- A RESV message is send from receiver to sender
  - The RESV message travels the path backward
  - The receiver determines the required resource reservation
    - Traffic specification (Tspec)
    - Requested Service Specification (Rspec)

RSVP Reservation Types & Merging

Distinct Reservation (or Fixed Filter Style)

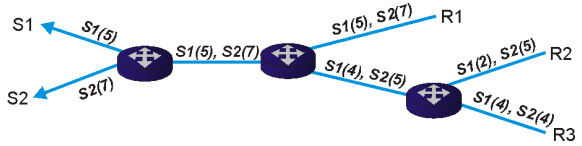
- A receiver requests reservations for each sender
- Example: all senders send different types of data in parallel

Shared Reservation

- Wild card Filter Type:
  - A receiver requests one reservation for all senders within a multicast tree
  - Senders may be added or removed without changing the reservations for the resource tree
  - Example: all senders require the same resources, but not at the same time
- Shared Explicit Reservation:
  - A receiver requests one reservation for an explicit defined set of senders
  - Shared explicit reservation may be combined with distinct reservations

Reservations are merged, dependent on their reservation type

- Example for distinct reservation:

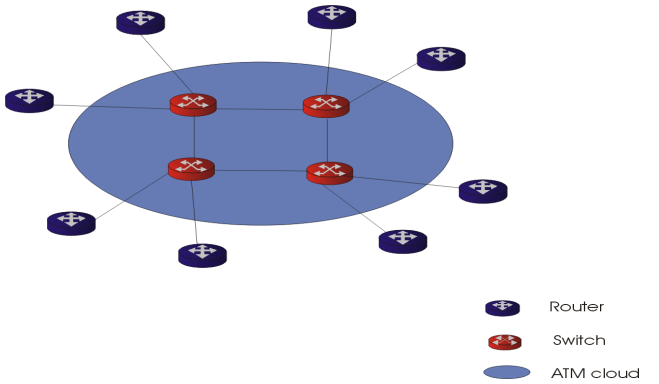


Routing Functional Components

- Forwarding
  - forwarding component using information in forwarding table and packet itself to forward packets from input to output across a switch or router
- Control
  - Control component is responsible for construction and maintenance of the forwarding table

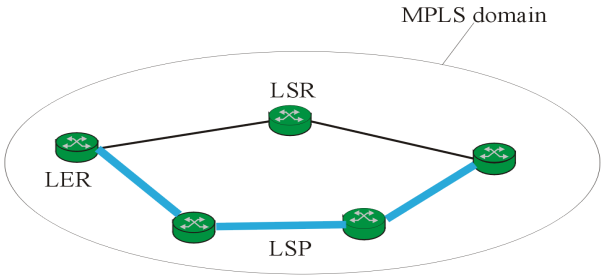
Backbone

- Forwarding IP packets is the major function of many ATM networks
- Mapping problem solutions like LAN emulation and Classical IP over ATM are complex



Multiprotocol Label Switching (MPLS)

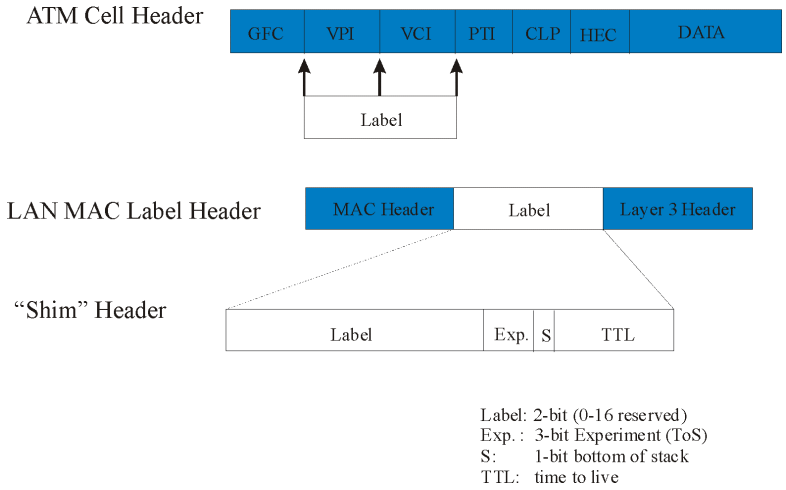
- MPLS combines the best of both worlds – ATM´s circuit switching and IP´s packet routing. It is a hybrid technology which enables very fast forwarding in the core and conventional routing at the edges. (e.g. IP routing protocol runs directly on ATM hardware)
  - Using standard IP control protocol like routing protocol and RSVP
  - Using packet switching
  - IP routing protocol runs directly on (e.g.) ATM hardware



MPLS: Concepts

- MPLS architecture is defined in draft-ietf-mpls-arch
- Packets are classified in Forwarding Equivalence Class (FEC) only once in ingress Label Switching Router (LSR)
  - All packets within a given FEC are treated in the same way
  - The same path through the network
  - The same QoS
- LSR assign a label to each FEC
  - A label is a short, fixed-length entity, with no structure
  - Labels have local significance, a label switching device usually replace the label with a new one (label swapping)
  - More than one label is allowed (label stack)
- Packet forwarding based on the label value (Label switching)
  - Forwarding decision are always made on the label at the top of the stack

MPLS: Label



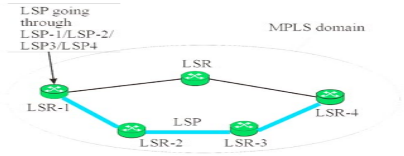
Label: 2-bit (0-16 reserved)
Exp. : 3-bit Experiment (ToS)
S: 1-bit bottom of stack
TTL: time to live

Label Distribution Protocol (LDP)

- Piggyback on top of existing routing protocols like Broder Gateway Protocol (BGP)
- LDP
  - Four classes of Message
    - *DISCOVERY* message
    - *ADJACENCY* message, deal with initialization, keepalive and shutdown between LSRs
    - *LABEL ADVERTISEMENT* message, deal with label binding advertisements, requests, withdrawal, release
      - » *LABEL MAPPING* message
      - » *LABEL WITHDRAWAL* message
    - *NOTIFICATION* message, provide advisory and error information
  - Running over TCP

MPLS: QoS

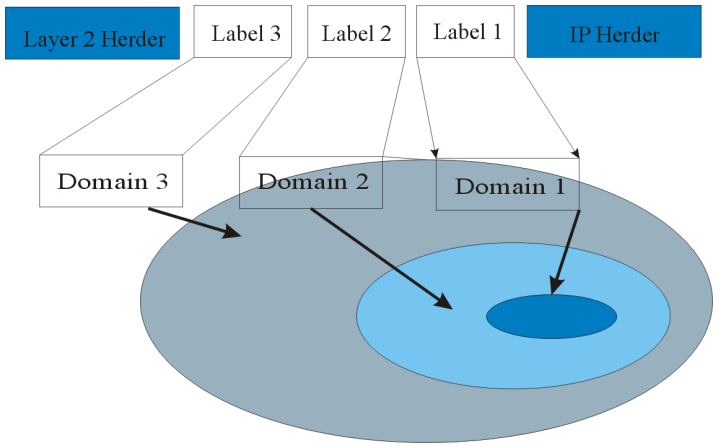
- MPLS supports IP QoS Models
  - MPLS is not an end-to-end protocol
- MPLS support of IntServ/RSVP
  - Distributing bindings between flows and labels
  - Labels are provided in the RESV messages
  - Explicit routing



- MPLS support of DiffServ
  - Using the experimental field
  - label is bound to <prefix, PHB>

MPLS: Scalability

- Network scalability
  - The label stack is useful to implement hierarchy



MPLS: Traffic Engineering

- What is constraint-based routing?
  - Compute a path from one node to another, that the path doesn´t violate the constraints and is optimal with respect to some scalar metric
- Constraint-based routing component
  - The ability to computer a path at the source
  - The ability to distribute the information about network topology and attributes associated with links throughout the network
  - Explicit routing
  - Network resources can be reserved and link attributes can be modified
- Plain IP routing is not enough for Traffic Engineering
  - IP routing takes into account only the destinations of packets
- Solving Traffic Engineering with MPLS Constraint-based routing

4.10 RTP - Real-Time Transport Protocol

Consists of two closely-linked parts:

- the real-time transport protocol (RTP), carries data with real-time properties
- the RTP control protocol (RTCP), monitors QoS and distributes this information to all participants of a session

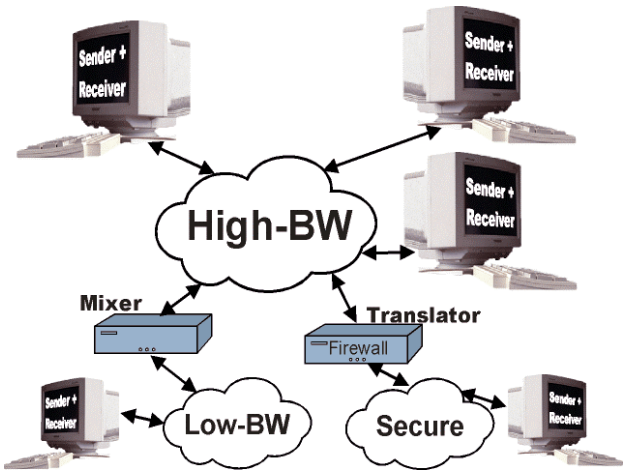
RTP makes no reservations and does not guarantee any service

RTP is a protocol framework, not a complete protocol

- a profile specification defines payload types and may extend RTP
- a payload specification defines payload formats and encoding types must be specified
- therefore RTP will typically be part of an application

Specified in RFC 1889

RTP - Scenario



many-to-many Communication (e.g. video conference)

RTP - Definitions

RTP-Session

- is a set of participants
- each participant is identified by a host and a destination port address
- each medium is carried in a separate RTP-session

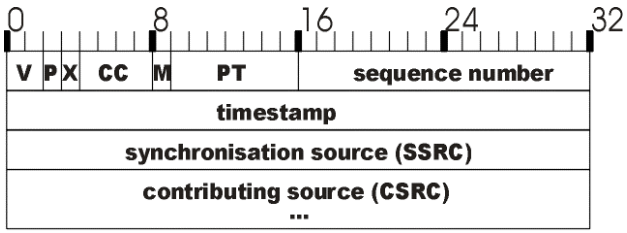
Mixer

- is an intermediate system that receives data from one or more sources, possibly changes the data format and combines packets in some manner
- a mixer will make timing adjustments and generate an own timing for combined data

Translator

- is an intermediate system that forwards data without changing media or synchronization
- encryption and addresses may be changed
- multicast may be mapped to unicast and vice versa

RTP - Header



V = Version (default = 2)
P = Padding
X = Extension
CC = CSRC count
M = Marker
PT = Payload type

- the payload type is defined by the applications profile
- the sequence number enables receivers to detect lost RTP-PDUs
- the timestamps reflects a sampling instant. i.e. the timestamp unit depends on the encoding and need not correspond to the system clock
- SSRC identifies the last sync. entity, it is unique within a session
- CCSR identifies the contributor of a source

RTCP - RTP control protocol

RTP enables receiver to monitor the QoS:

- Delay, Jitter, PDU loss rate

RTCP periodically transmits control packets between all participants of a RTP session:

- the primary function is to provide feedback about the QoS
- carries transport-level identifiers for RTP sources, the canonical name (the SSRC may change over the time the canonical name is fixes, e.g. a user name)
- the rate of sent RTCP packets depends on the number of participants, in order to make RTCP scalable
- optionally further information about the participants could be distributed, to realize a simple session control