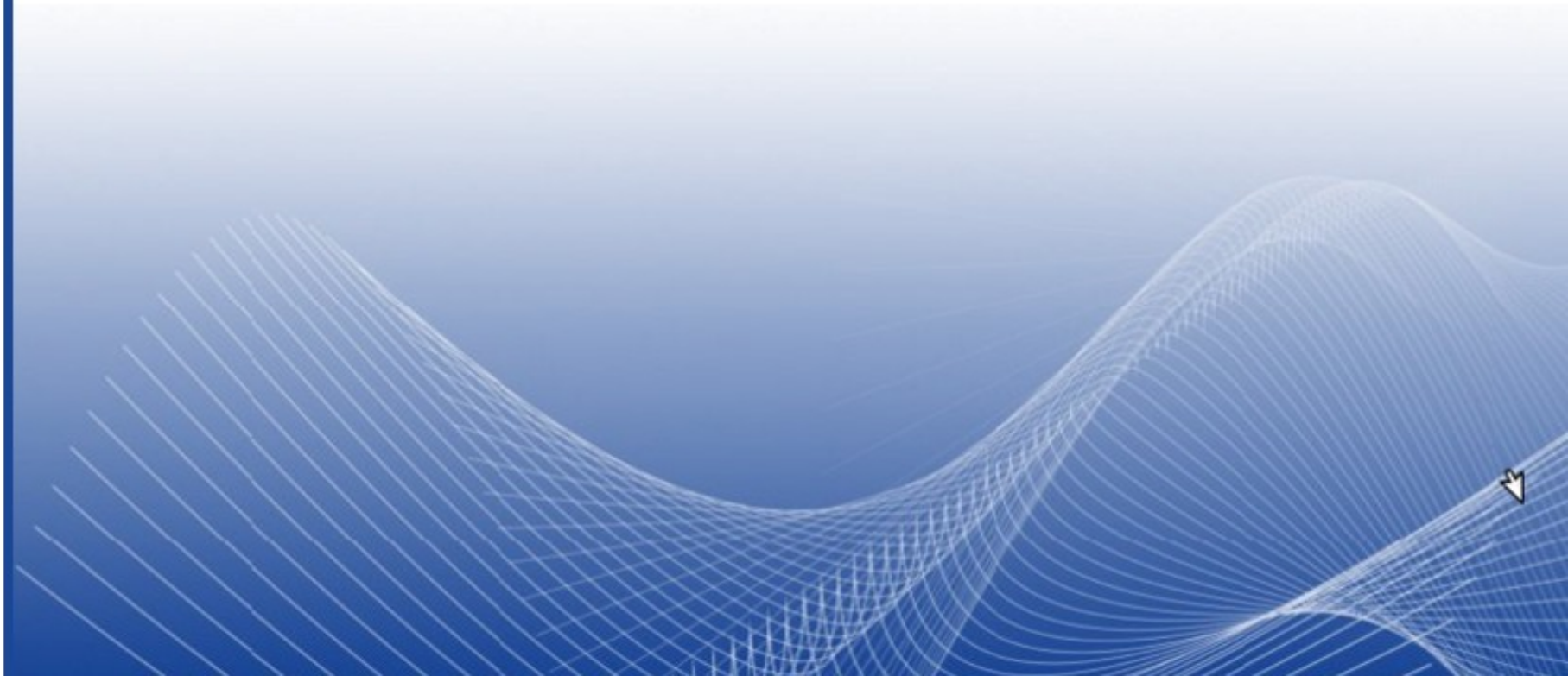
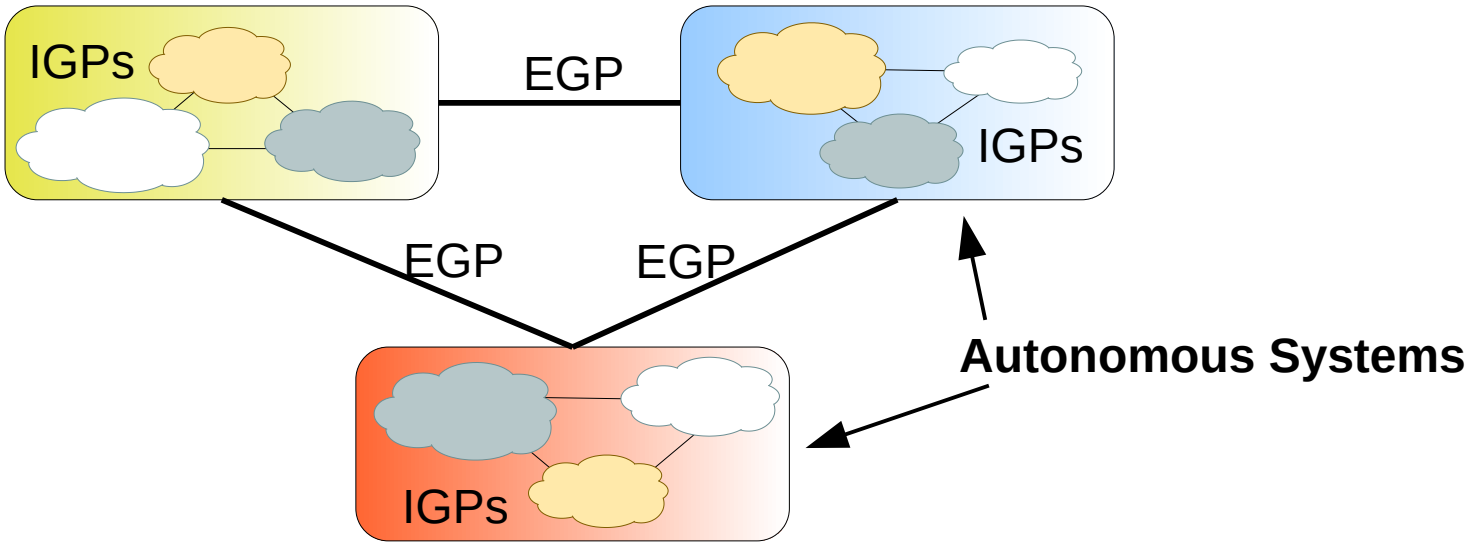
External Routing (BGP and MP-BGP)

Arquitectura de Comunicações



universidade de aveiro
deti.ua.pt

Border Gateway Protocol (BGP)



- Border Gateway Protocol Version 4 of the protocol (BGP4) was deployed in 1993 and currently is the protocol that assures Internet connectivity
- BGP is mainly used for routing between Autonomous Systems
- Autonomous System (AS) is a network under a single administration
  - One or more network operators with a common well defined global routing policy

AS Numbers

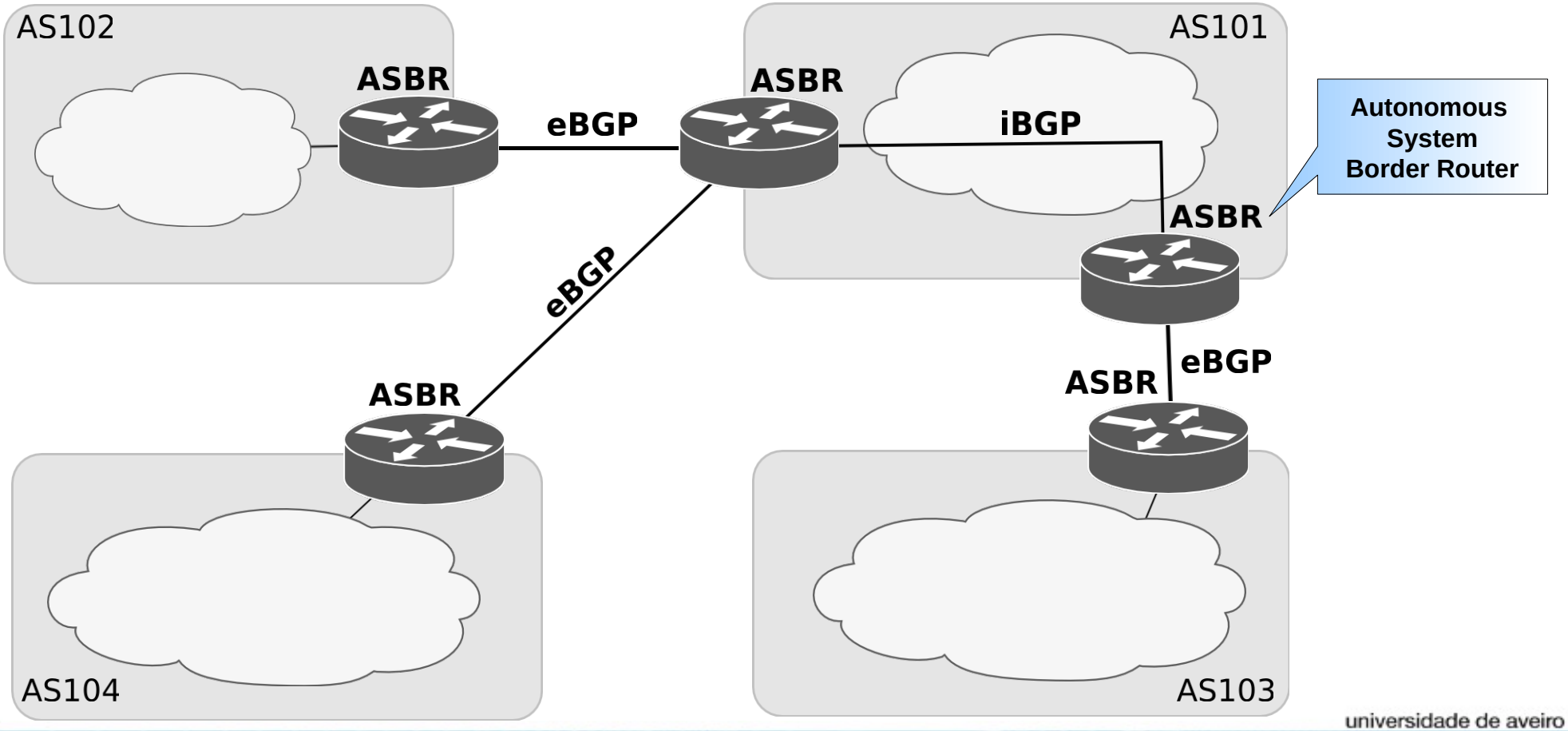
- Allocated ID by InterNIC and is globally unique
- RFC 4271 defines an AS number as 2-bytes
  - Private AS Numbers = 64512 through 65535
  - Public AS Numbers = 1 through 64511
    - 39000+ have already been allocated
    - We will eventually run out of AS numbers
- Need to expand AS size from 2-bytes to 4-bytes
- RFC4893 defines BGP support for 4-bytes AS numbers
  - 4,294,967,295 AS numbers
  - As of January 1, 2009, all new Autonomous System numbers issued will be 4-byte by default, unless otherwise requested.
  - The full binary 4-byte AS number is split two words of 16 bits each
    - Notation:
    - <higher2bytes in decimal>.<lower2bytes in decimal>
    - Example1: AS 65546 is represented as “1.10”
    - Example2: AS 50000 is represented as “0.50000”
  - Cannot have a “flag day” solution

BGP Neighbor Relationships

- Often called peering
  - Usually manually configured into routers by the administrator
- Each neighbor session runs over TCP (port 179)
  - Ensures reliable data delivery
- Peers exchange all their routes when the session is first established
- Updates are also sent when there is a topology change in the network or a change in routing policy
- BGP peers exchange session KEEPALIVE messages
  - To avoid extended periods of inactivity.
  - Low keepalive intervals can be set if a fast fail-over is required

Internal BGP (iBGP) & External BGP (eBGP)

- Neighbor relations can be established between
  - Same AS routers (Internal BGP – iBGP).
  - Different AS routers (External BGP – eBGP).
- Routers that implement neighbor relations are called an Autonomous System Border Router (ASBR).

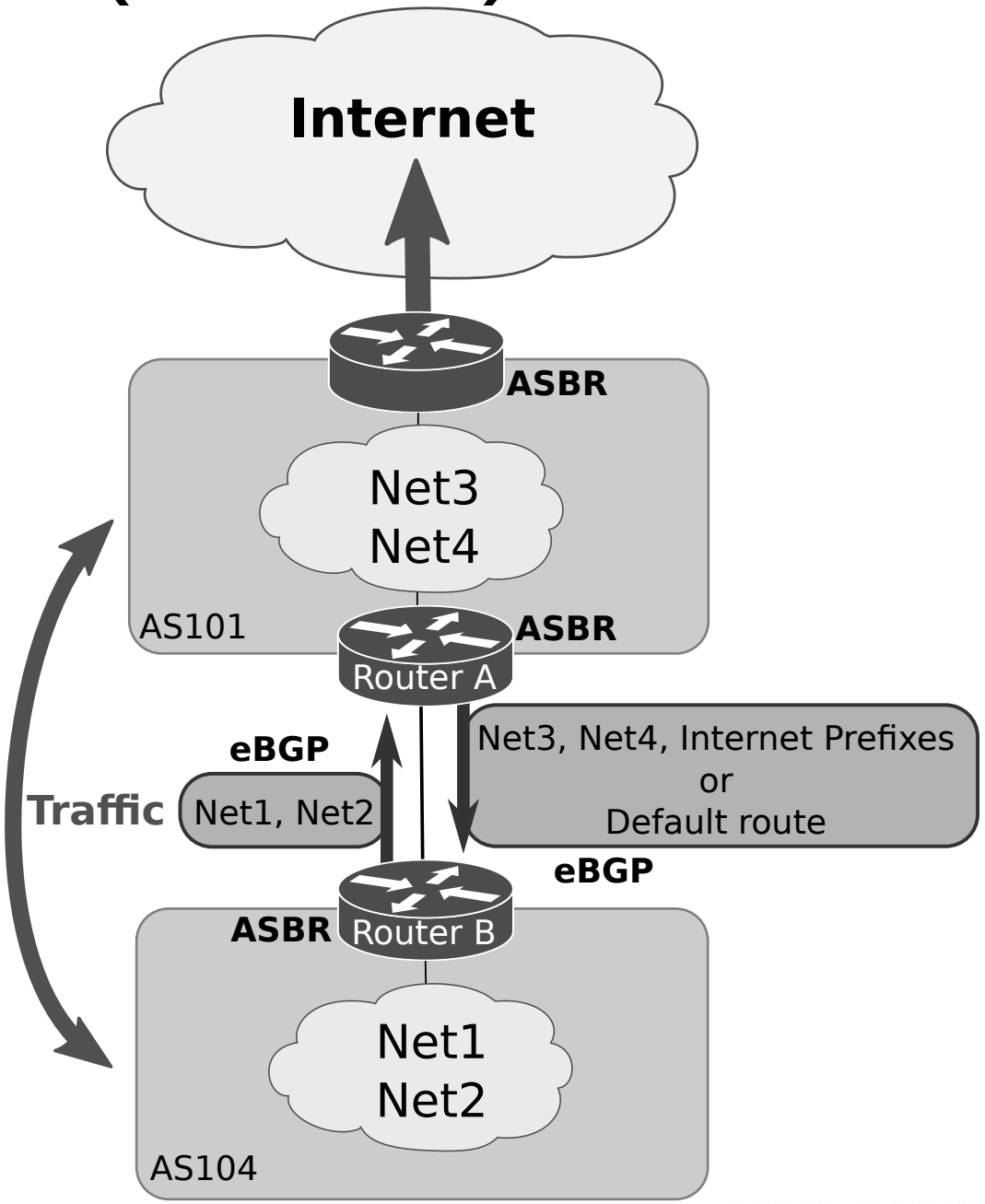


External and Internal BGP

- External BGP (eBGP) is used between AS.
- Internal BGP (iBGP) is used within AS.
- A BGP router never forwards a path learned from one iBGP peer to another iBGP peer even if that path is the best path.
  - An exception is when a router is configured as route-reflector.
- A BGP forward the routes learned from one eBGP peer to both eBGP and iBGP peers.
  - Filters can be used to modify this behavior.
- iBGP routers in an AS **must maintain an iBGP session with all other iBGP routers** in the AS (iBGP Mesh).
  - To obtain complete routing information about external networks.
  - Most networks also use an IGP, such as OSPF.
  - Additional methods can be used to reduce iBGP Mesh complexity.
    - Route reflectors, private AS, ...

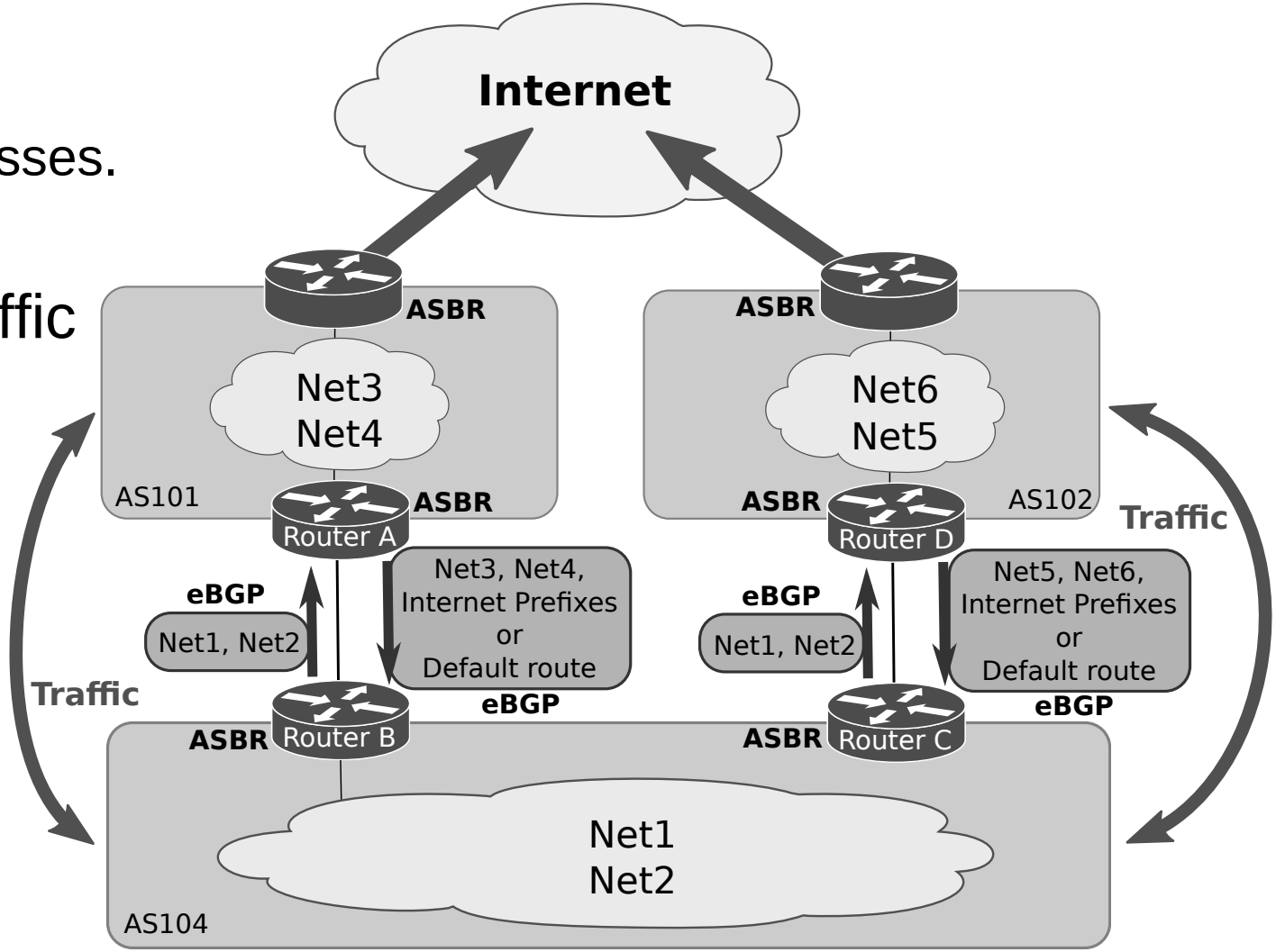
Single-homed (or Stub) AS

- AS has only one border router (ASBR)
  - Single Internet access.
  - Single ISP.



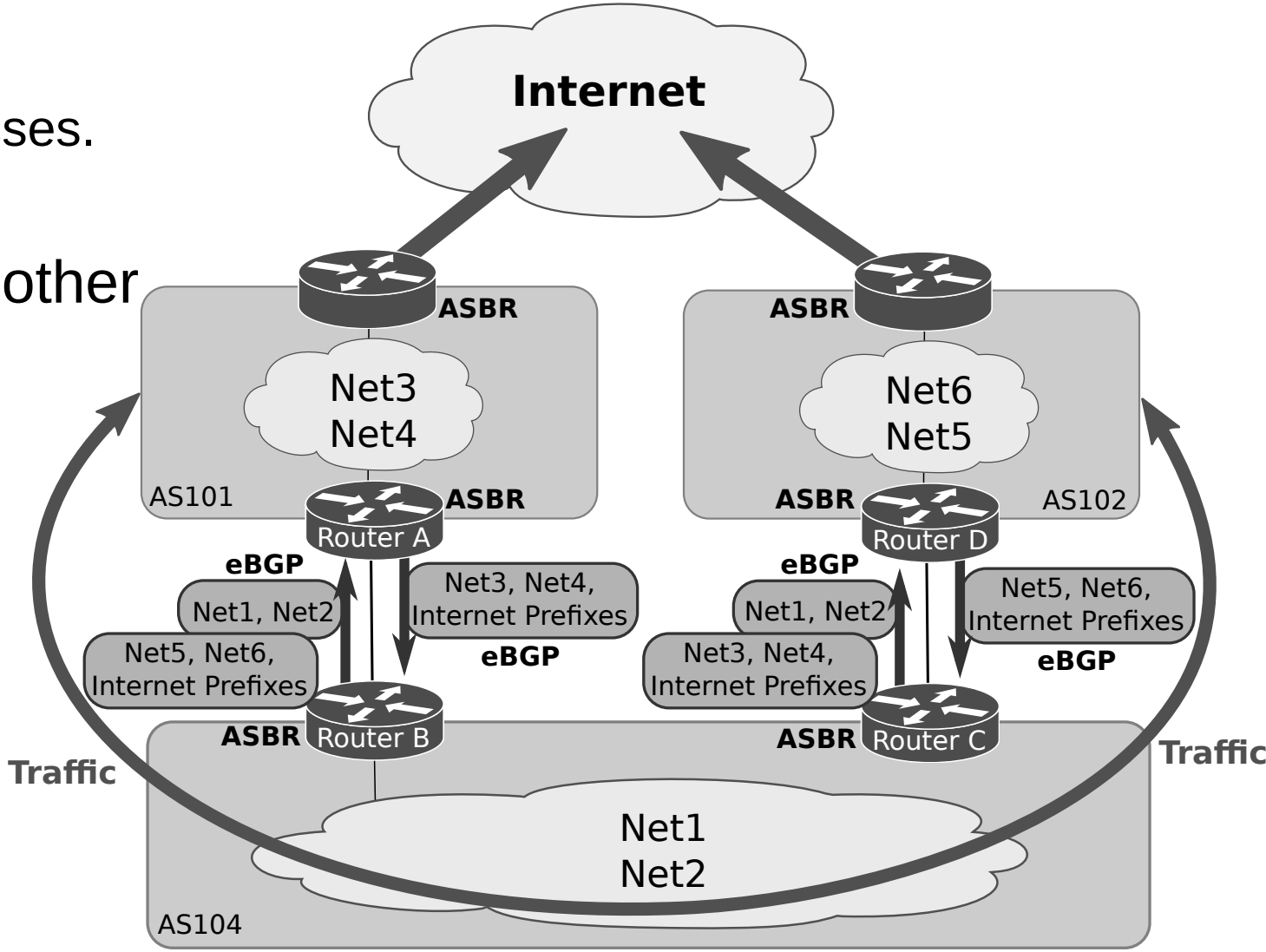
Multi-homed Non-transit AS

- AS has more than one border router (ASBR)
  - Multiple Internet accesses.
  - Multiple ISP.
- Does not transport traffic from other AS.



Multi-homed Transit AS

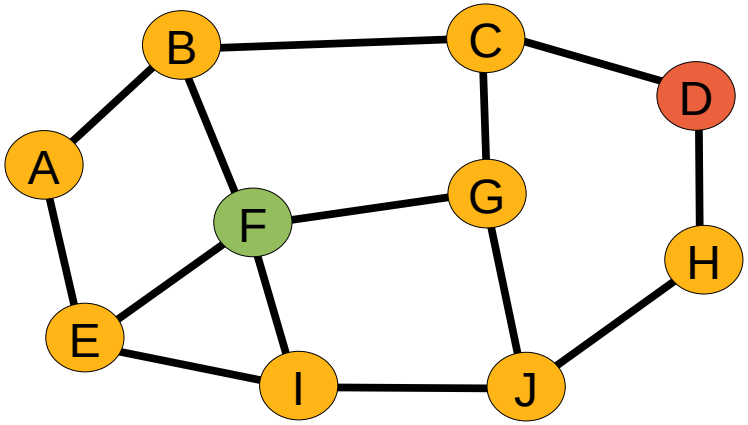
- AS has more than one border router (ASBR).
  - Multiple Internet accesses.
  - Multiple ISP.
- Transports traffic from other AS.



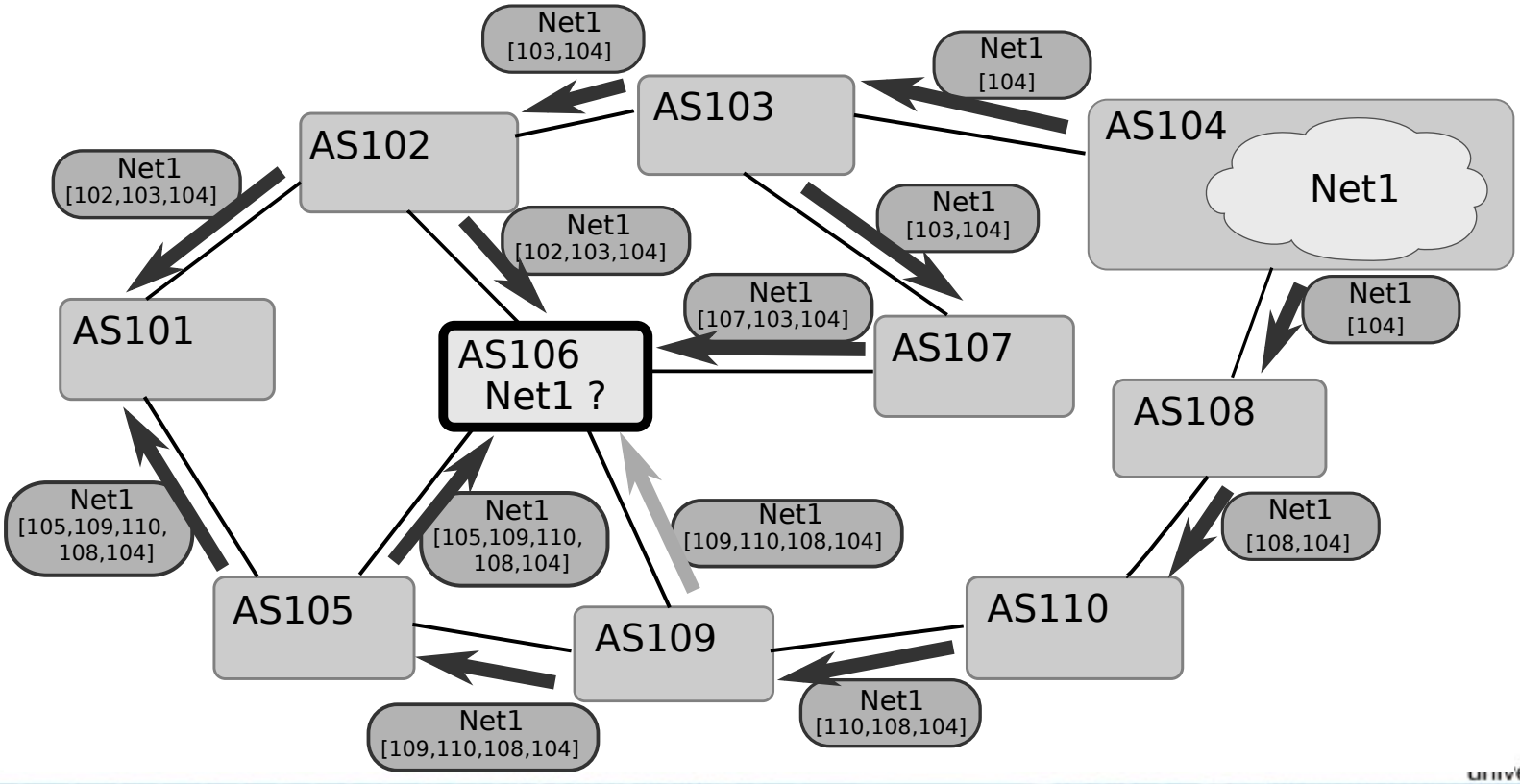
Path-vector

- BGP is a path-vector protocol
- Although it is essentially a distance-vector protocol that carries a list of the AS traversed by the route
  - Provides loop detection
- An EBGP speaker adds its own AS to this list before forwarding a route to another EBGP peer
- An IBGP speaker does not modify the list because it is sending the route to a peer within the same AS
  - AS list cannot be used to detect the IBGP routing loops

Path vector

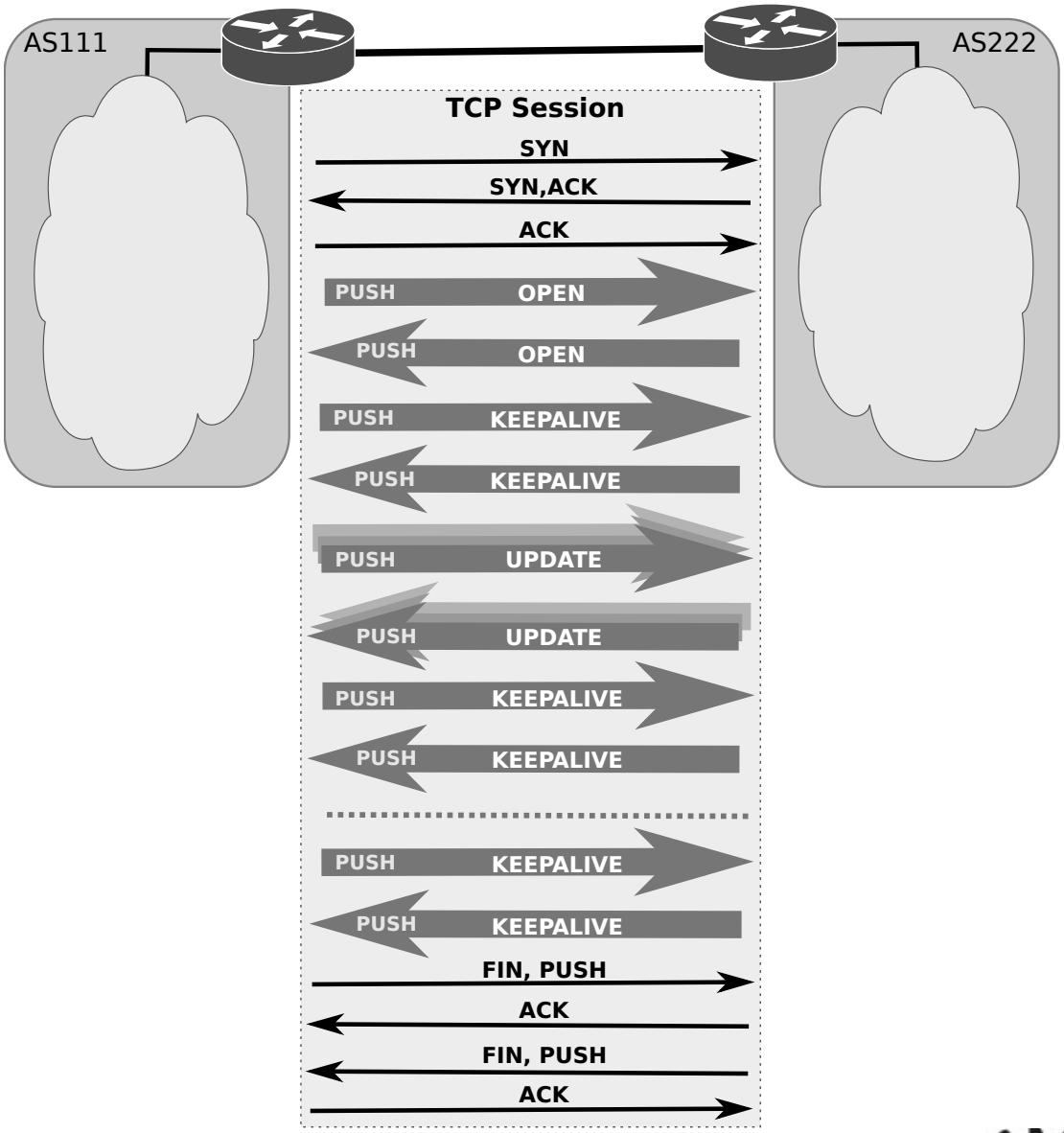


- F receives from its neighbors different paths to D:
  - De B: “I use BCD”
  - De G: “I use GCD”
  - De I: “I use IFGCD”
  - De E: “I use EFGCD”



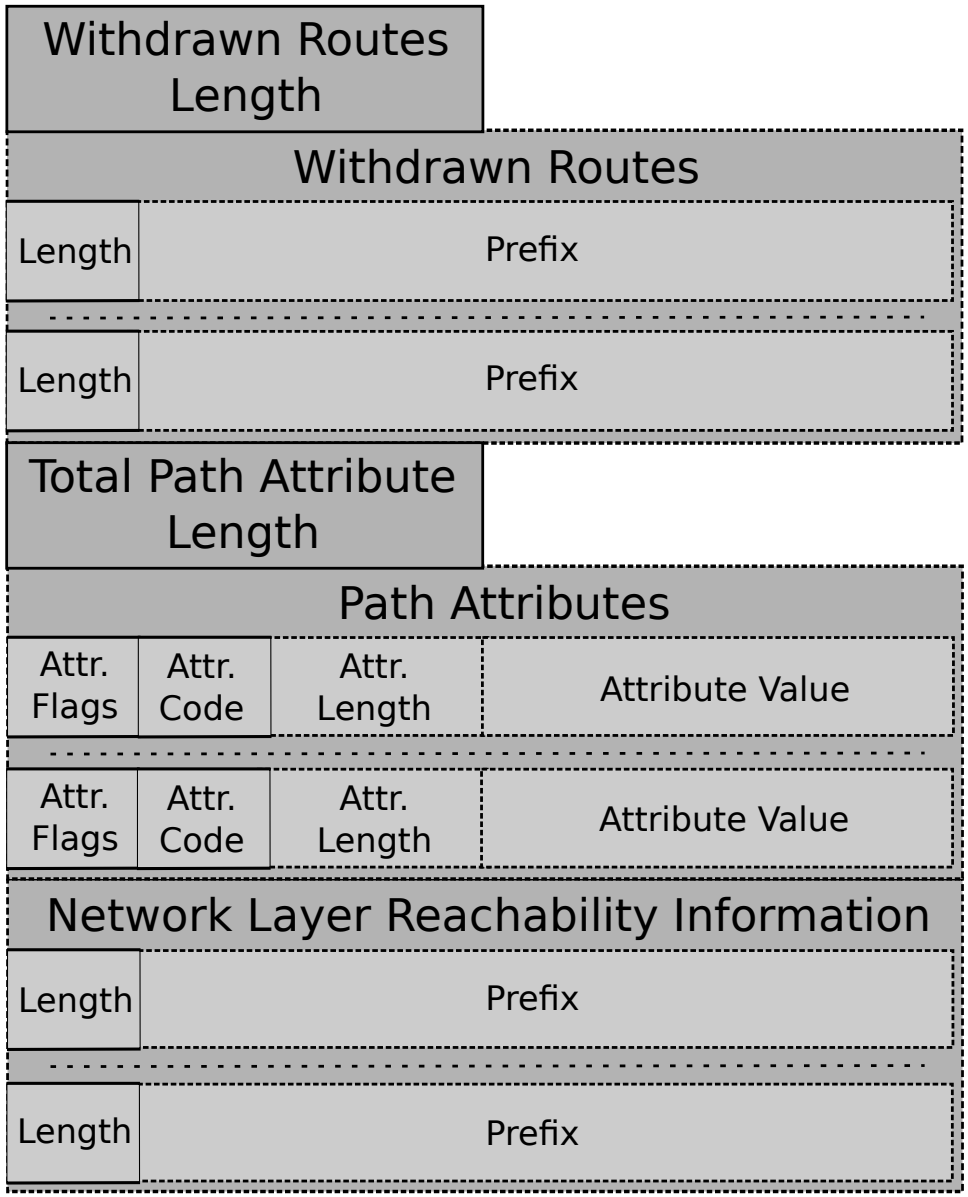
BGP Messages

- OPEN messages are used to establish the BGP session.
- UPDATE messages are used to send routing prefixes, along with their associated BGP attributes (such as the AS-PATH).
- KEEPALIVE messages are exchanged whenever the keepalive period is exceeded, without an update being exchanged.
- NOTIFICATION messages are sent whenever a protocol error is detected, after which the BGP session is closed.

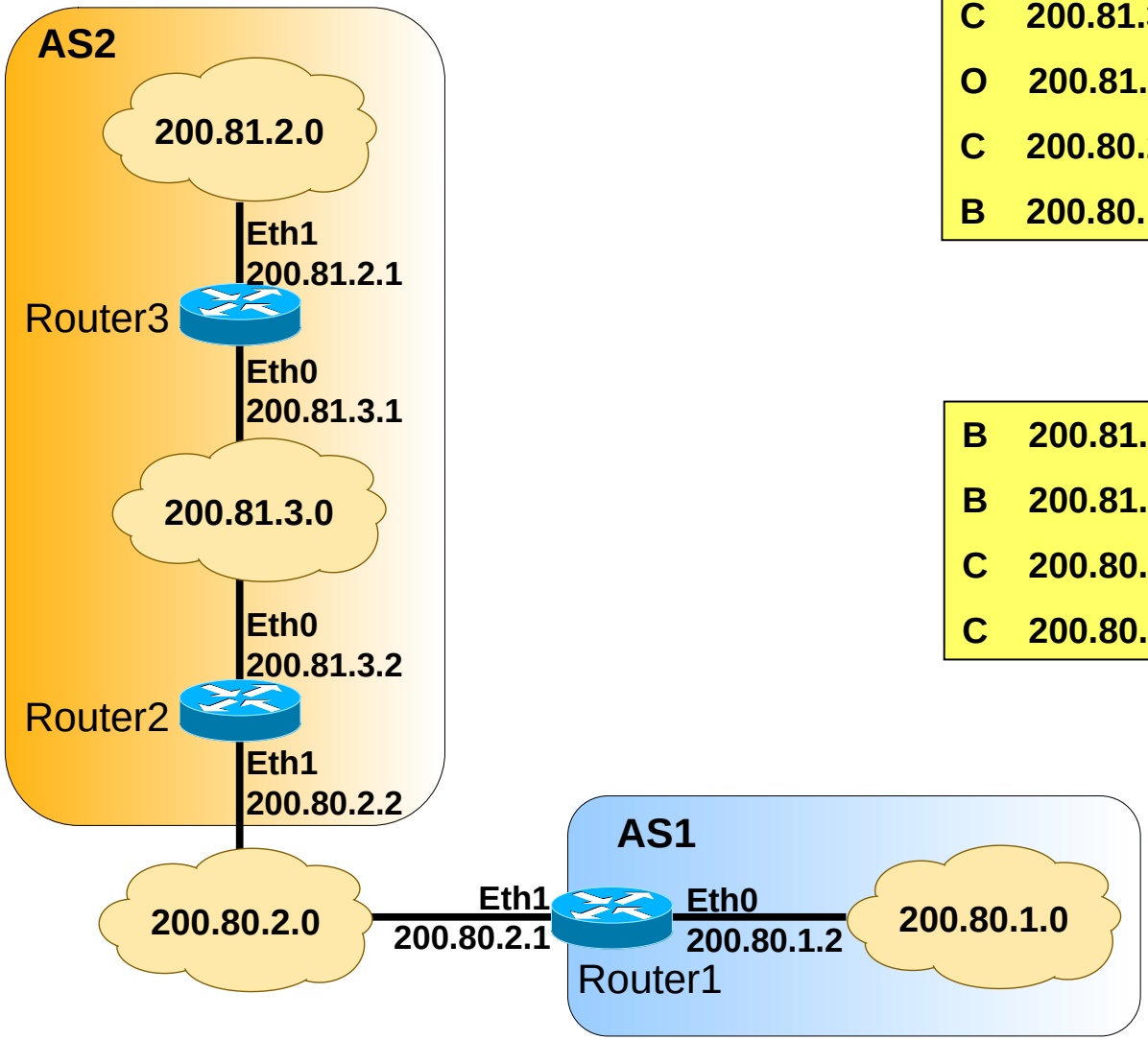


Update Message

- Withdrawn routes – List of IP networks no longer accessible.
- Path attributes – parameters used to define routing and routing policies.
- Network layer reachability information – List of IP networks with connectivity.



Example



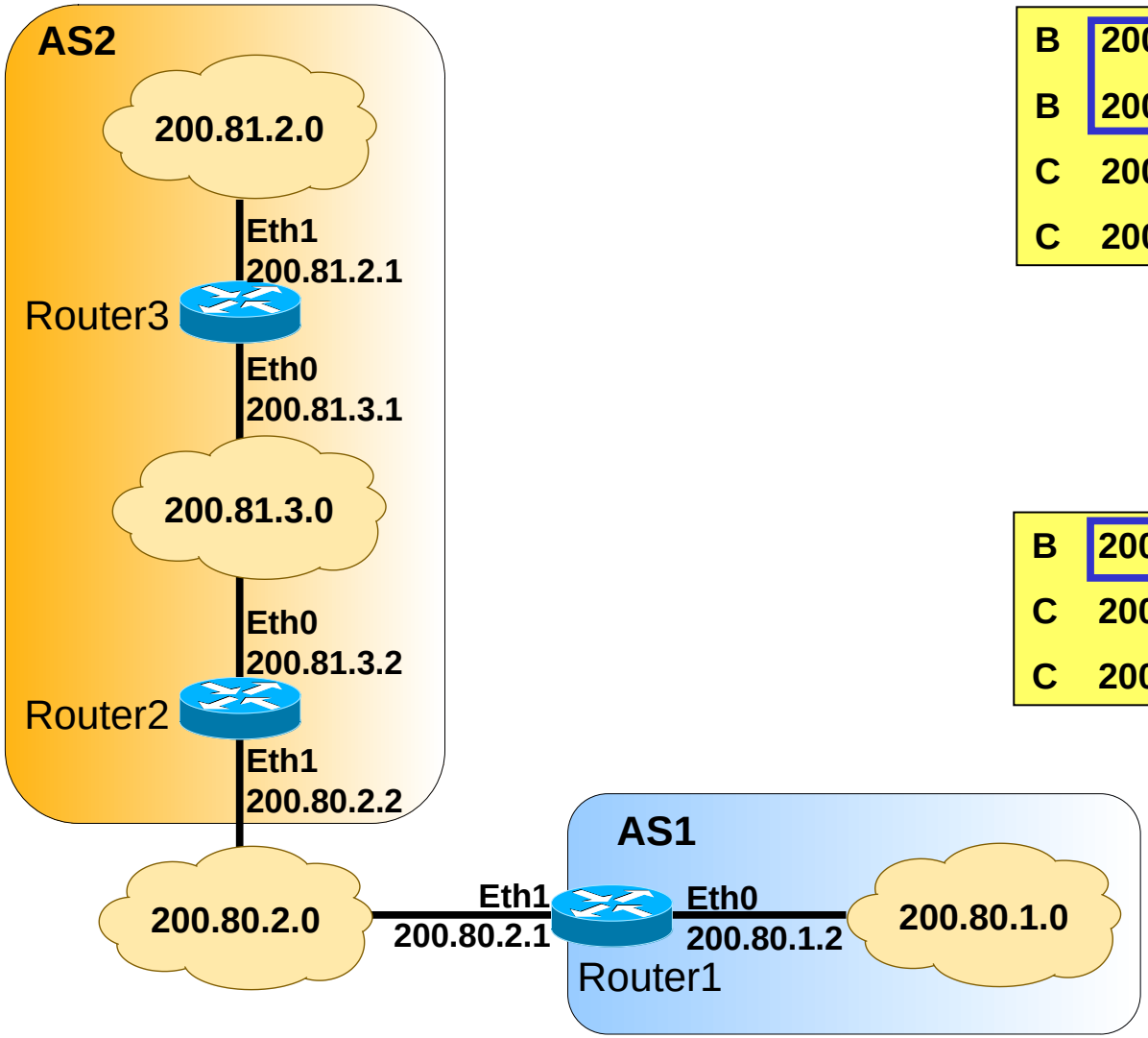
C	200.81.3.0/24 is directly connected, Ethernet0
O	200.81.2.0/24 [110/20] via 200.81.3.1, 00:01:12
C	200.80.2.0/24 is directly connected, Ethernet1
B	200.80.1.0/24 [20/0] via 200.80.2.1, 00:00:29

Router 2's routing table

B	200.81.3.0/24 [20/0] via 200.80.2.2, 00:01:58
B	200.81.2.0/24 [20/0] via 200.80.2.2, 00:01:57
C	200.80.2.0/24 is directly connected, Ethernet1
C	200.80.1.0/24 is directly connected, Ethernet0

Router 1's routing table

Example – BGP networks aggregation



Before aggregation

```
B   200.81.3.0/24 [20/0] via 200.80.2.2, 00:01:58
B   200.81.2.0/24 [20/0] via 200.80.2.2, 00:01:57
C   200.80.2.0/24 is directly connected, Ethernet1
C   200.80.1.0/24 is directly connected, Ethernet0
```

Router 1

After aggregation

```
B   200.81.2.0/23 [20/0] via 200.80.2.2, 00:01:06
C   200.80.2.0/24 is directly connected, Ethernet1
C   200.80.1.0/24 is directly connected, Ethernet0
```

Router 1

BGP Attributes

- A BGP attribute, or path attribute, is a metric used to describe the characteristics of a BGP path.
- Attributes are contained in update messages passed between BGP peers to advertise routes. There are 4+1 categories of BGP attributes.
  - Well-known Mandatory (included in BGP updates)
    - AS-path, Next-hop, Origin.
  - Well-known Discretionary (may or may not be included in BGP updates)
    - Local Preference, Atomic Aggregate.
  - Optional Transitive (may not be supported by all BGP implementations)
    - Aggregator, Community, AS4_Aggregator, AS4_path.
  - Optional Non-transitive (may not be supported by all BGP implementations)
    - If the neighbor doesn't support that attribute it is deleted
    - Multi-exit-discriminator (MED).
  - Cisco-defined (local to router, not advertised)
    - Weight

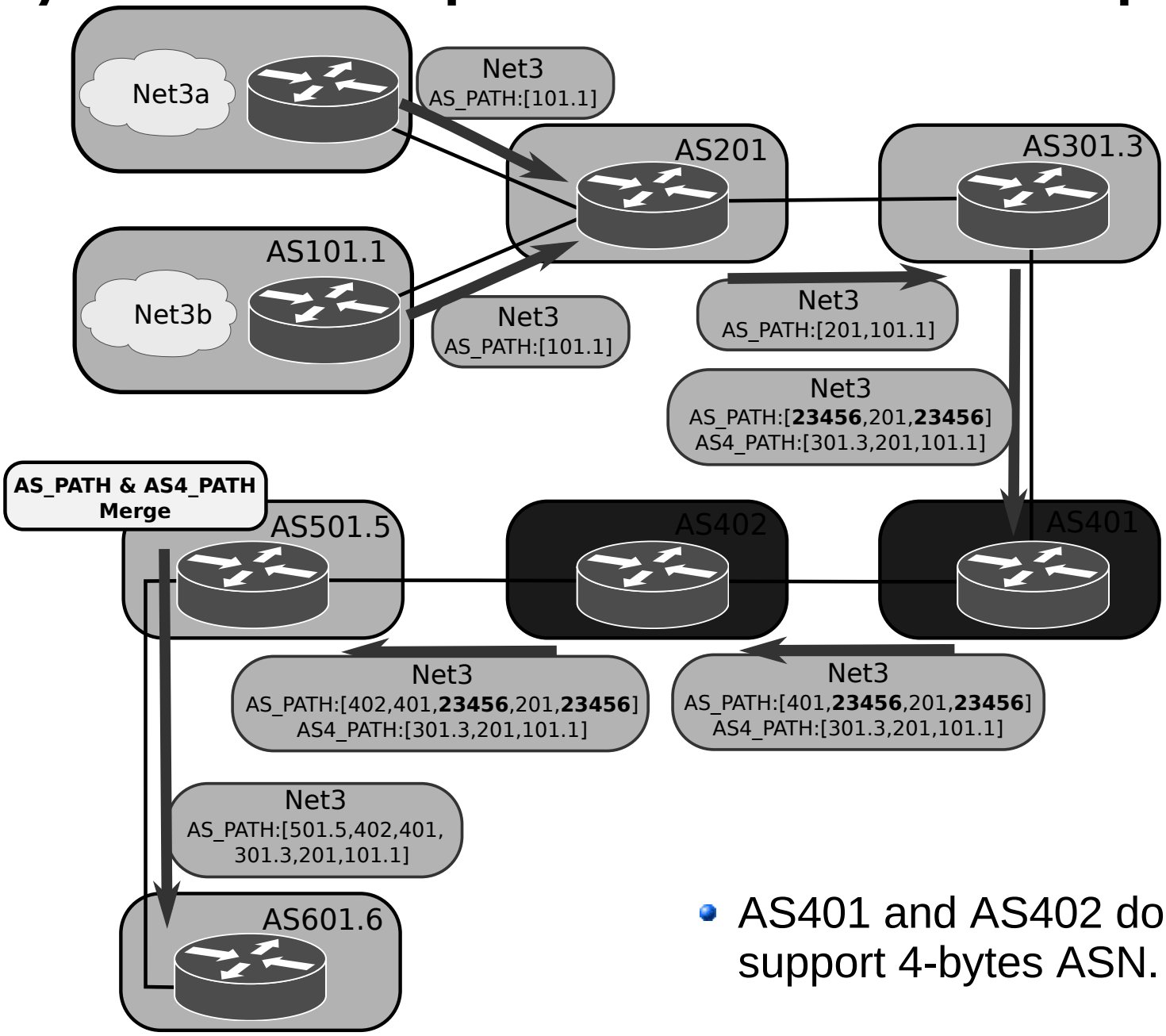
AS-PATH and ORIGIN Attributes

- AS-PATH
  - When a route advertisement passes through an autonomous system, the AS number is added to an ordered list of AS numbers that the route advertisement has traversed.
- ORIGIN
  - Indicates how BGP learned about a particular route. Can take three possible values:
    - IGP (0) value is set if the route is interior to the originating AS, resulting from an explicit inclusion of a network within the BGP routing process by means of manual configuration.
    - INCOMPLETE (2) value is set if the route is learned by other means, namely, route redistribution from other routing processes into the BGP routing process.
    - EGP (1) is no longer used in modern networks.

AS4_PATH & AS4_AGGREGATOR

- AS4_PATH attribute has the same semantics as the AS_PATH attribute, except that it is optional transitive, and it carries 4-bytes AS numbers.
- AS4_AGGREGATOR attribute has the same semantics as the AGGREGATOR attribute, except that it carries a 4-bytes AS number.
- 4-byte AS support is advertised via BGP capability negotiation
  - Speakers who support 4-byte AS are known as NEW BGP speakers
  - Those who do not are known as OLD BGP speakers
- New Reserved AS number
  - AS_TRANS = AS 23456
    - 2-byte placeholder for a 4-byte AS number
    - Used for backward compatibility between OLD and NEW BGP speakers
- Receiving UPDATEs from a NEW speaker
  - Decode each AS number as 4-bytes
  - AS_PATH and AGGREGATOR are effected
- Receiving UPDATEs from an OLD speaker
  - AS4_AGGREGATOR will override AGGREGATOR
  - AS4_PATH and AS_PATH must be merged to form the correct as-path
- Merging AS4_PATH and AS_PATH
  - AS_PATH → [275 250 225 23456 23456 200 23456 175]
  - AS4_PATH → [100.1 100.2 200 100.3 175]
  - Merged AS-PATH → [275 250 225 100.1 100.2 200 100.3 175]

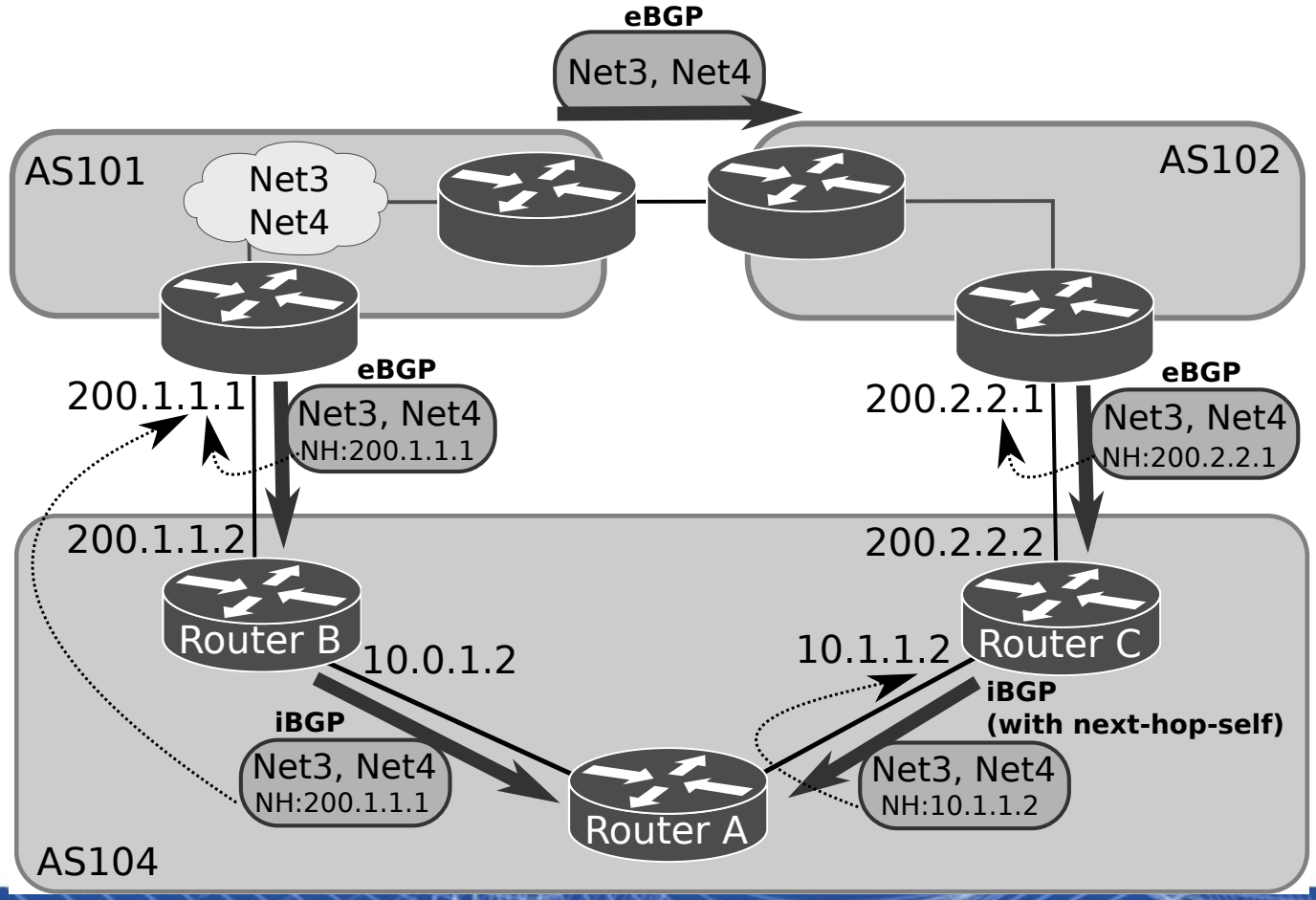
4-bytes AS Operational Example



- AS401 and AS402 do not support 4-bytes ASN.

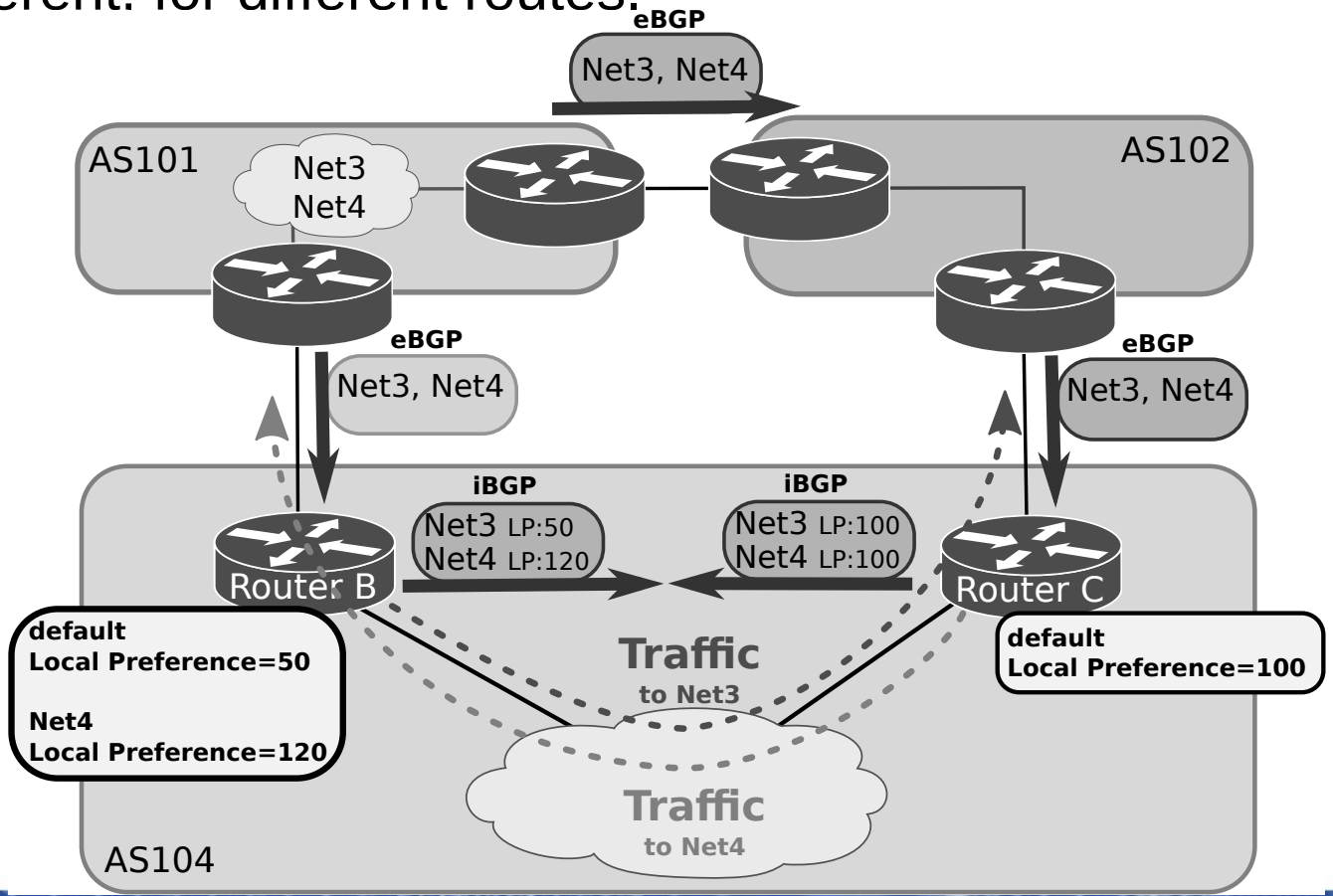
Next-Hop Attribute

- The eBGP next-hop attribute is the IP address that is used to reach the advertising router
- For eBGP, the next-hop address is the IP address of the connection between the peers
- For iBGP, the eBGP next-hop address is carried into the local AS
  - By configuration the AS border router can be the next-hop to iBGP neighbors



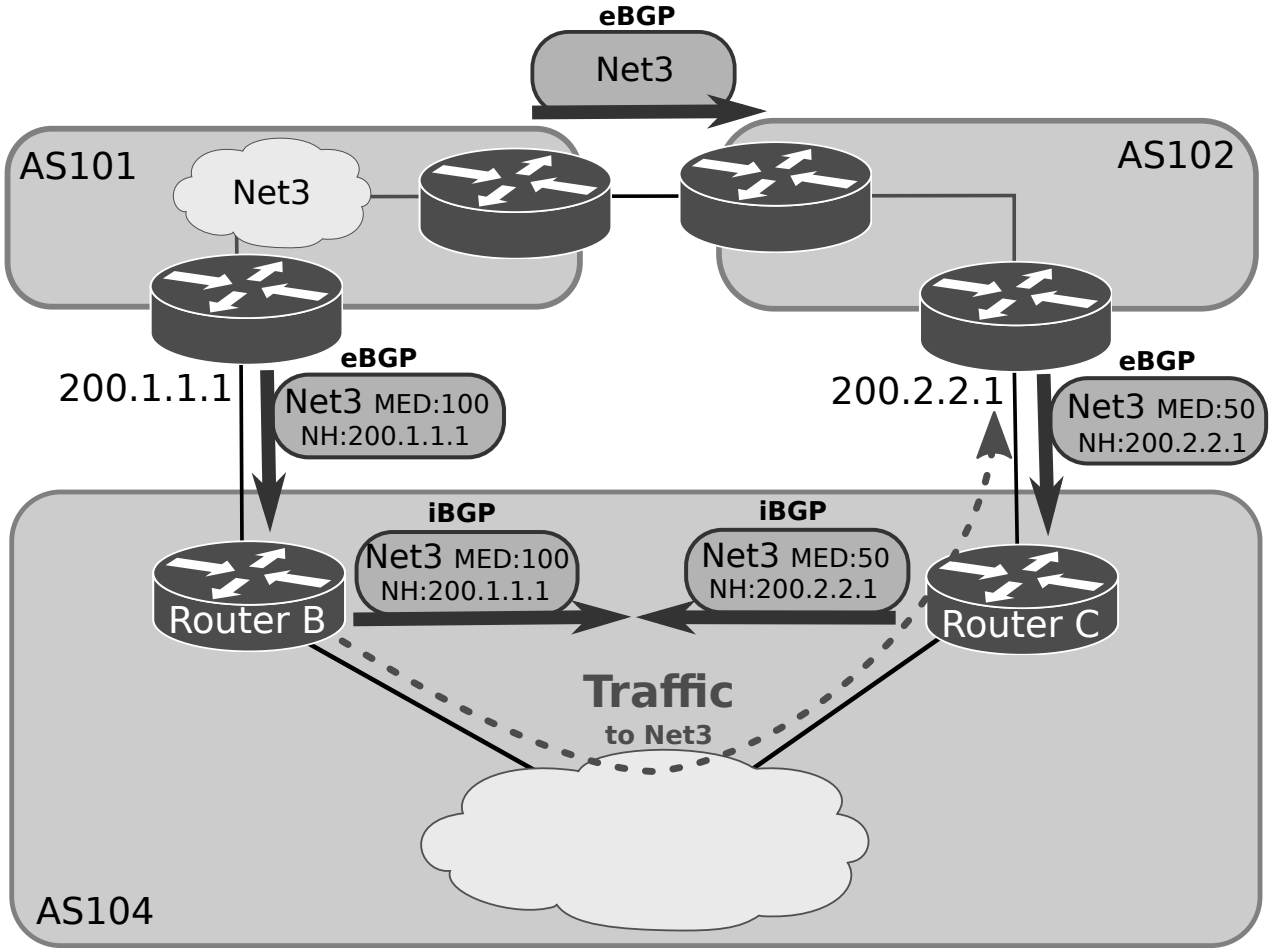
Local Preference Attribute

- The local preference attribute is used to choose an exit point from the local autonomous system (AS).
  - **Higher value** is preferred.
- The local preference attribute is propagated throughout the local AS.
- Can be different. for different routes.



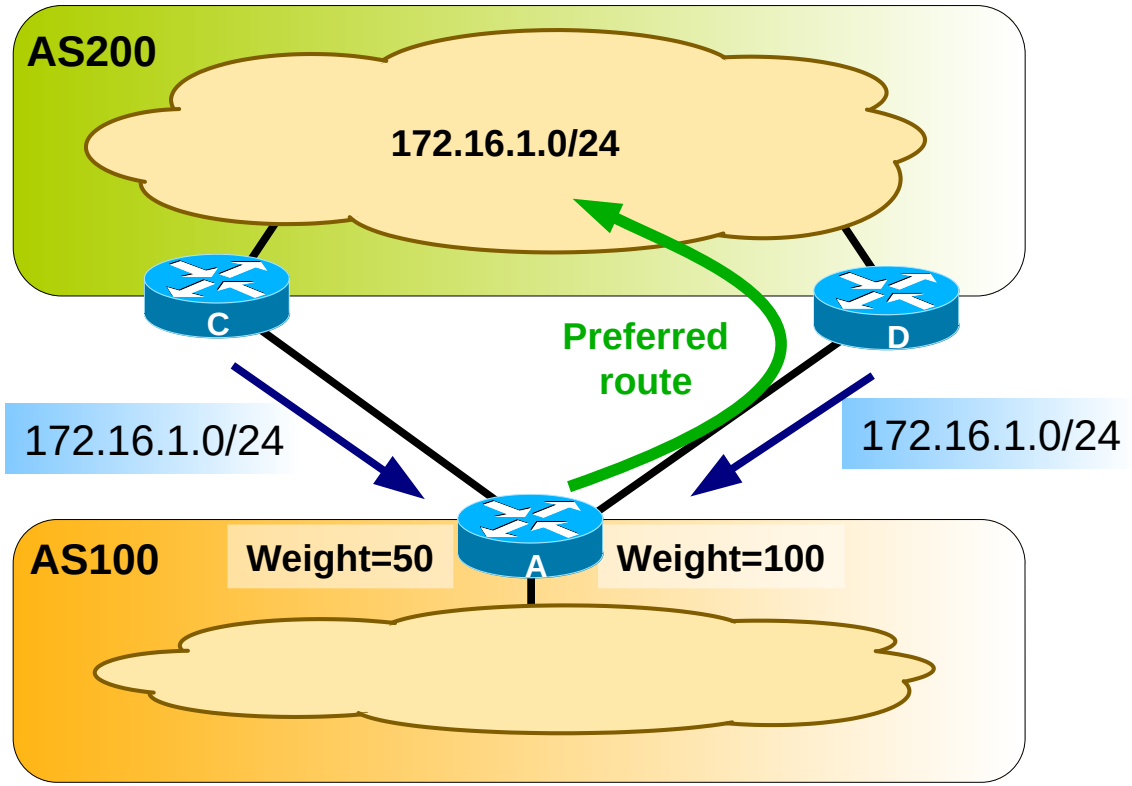
Multi-Exit Discriminator Attribute (MED)

- The multi-exit discriminator (MED) or metric attribute is used as a suggestion to an external AS.
- The external AS that is receiving the MEDs may be using other BGP attributes for route selection.
- The **lower value** of the metric is preferred.
- MED is designed to influence incoming traffic.

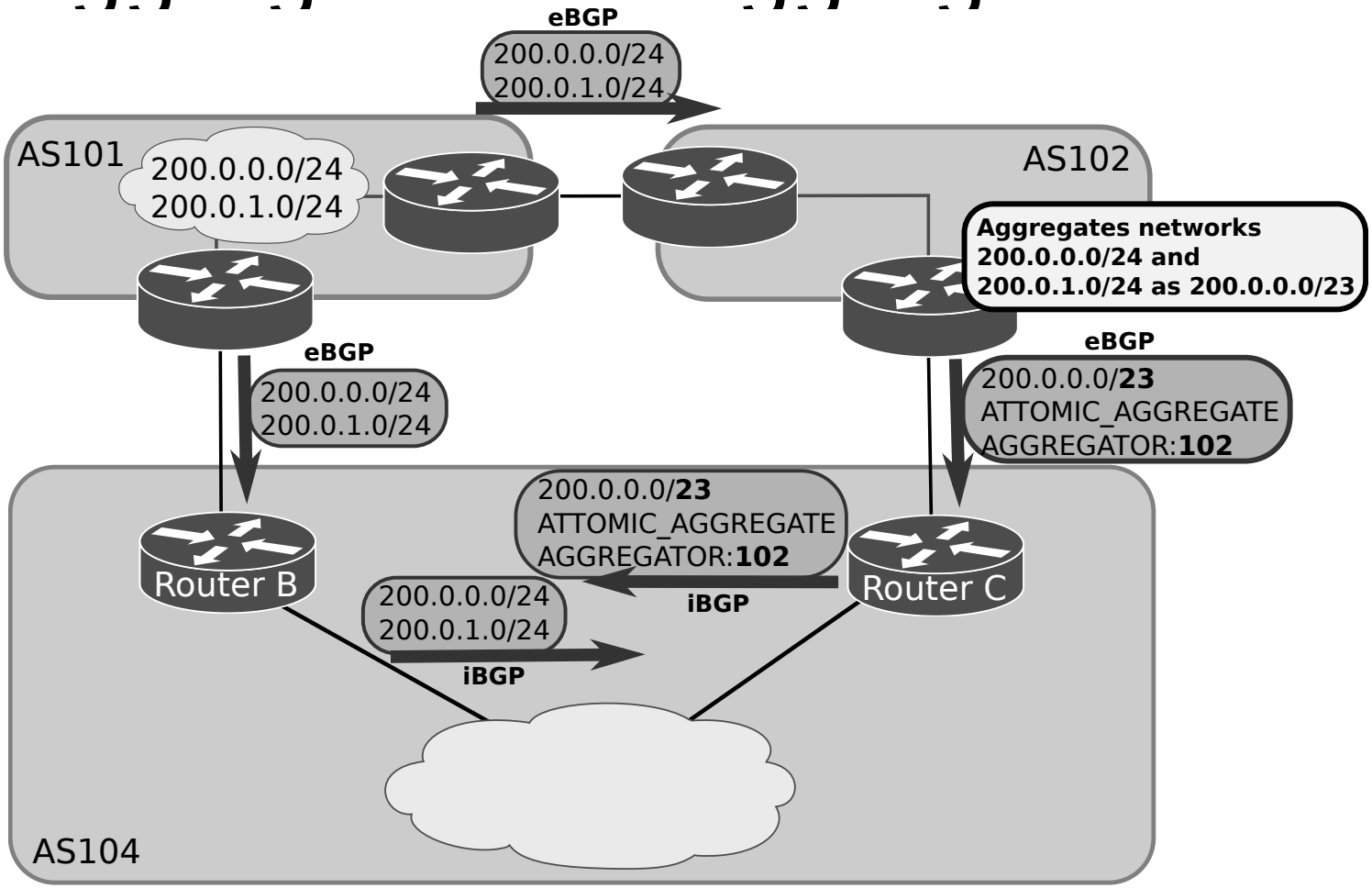


Weight Attribute

- Weight is a Cisco-defined attribute that is local to a router.
- The weight attribute is not advertised to neighboring routers.
- If the router learns about more than one route to the same destination, the route with the **highest weight** will be preferred.

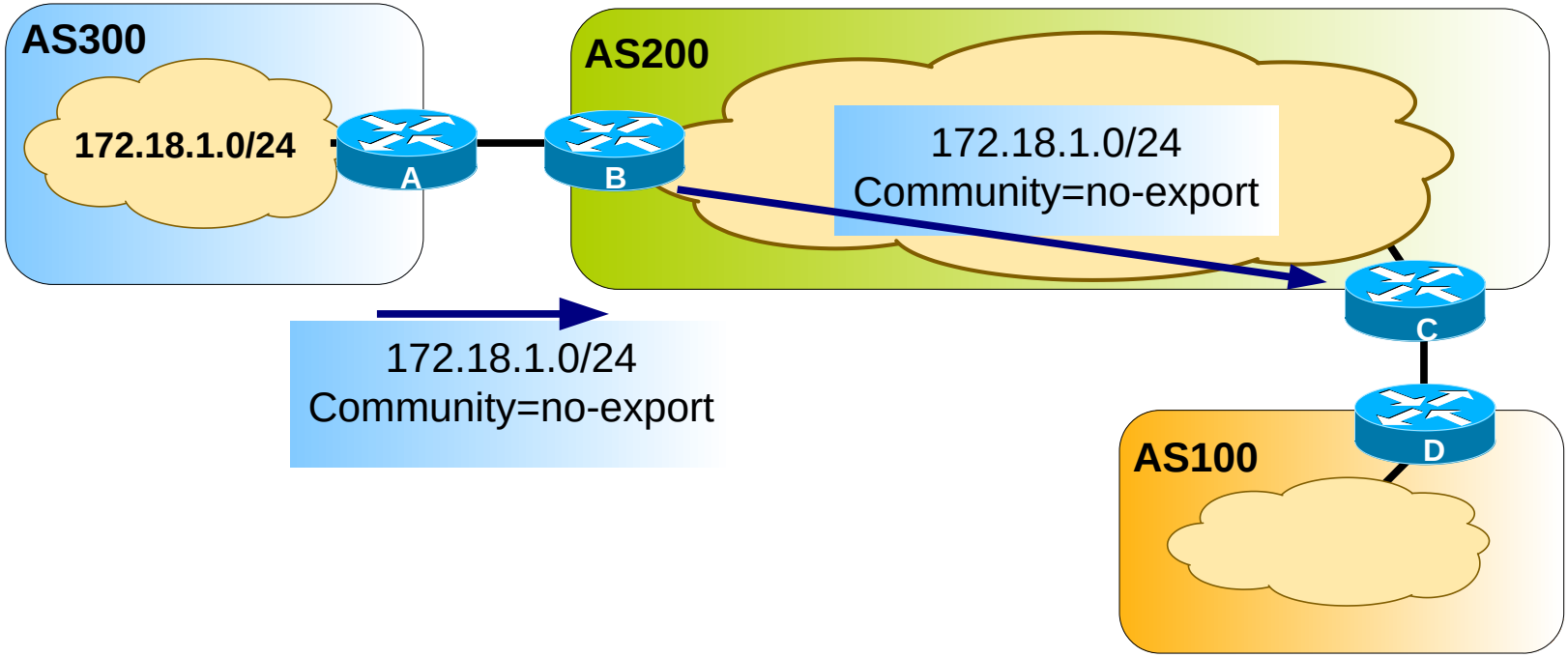


Atomic Aggregate and Aggregator Attributes



- Atomic Aggregate
  - Is used to alert routers that specific routes have been aggregated into a less specific route.
  - When aggregation like this occurs, more specific routes are lost.
- Aggregator
  - Provides information about which AS performed the aggregation.
  - And the IP address of the router that originated the aggregate.

Community Attribute



- Used to group routes that share common properties so that policies can be applied at the group level
- Predefined community attributes are:
  - no-export - Do not advertise this route to EBGP peers
  - no-advertise - Do not advertise this route to any peer
  - internet - Advertise this route to the Internet community; all routers in the network belong to it
- General communities format is ASnumber:Cnumber
  - e.g. 300:1, 200:38, etc...

BGP Path Selection

- BGP may receive multiple advertisements for the same route from multiple sources.
- BGP selects only one path as the best path.
- BGP puts the selected path in the IP routing table and propagates the path to its neighbors. BGP uses the following criteria, in the order:
  - Largest weight (Cisco only)
  - Largest local preference
  - Path that was originated locally
  - Shortest path
  - Lowest origin type (IGP lower than EGP, EGP lower than incomplete)
  - Lowest MED attribute
  - Prefer the external path over the internal path
  - Closest IGP neighbor

Multi-Protocol Border Gateway Protocol (MP-BGP)

MP-BGP Description

- Extension to the BGP protocol
- Carries routing information about other protocols/families:
  - IPv6 Unicast
  - Multicast (IPv4 and IPv6)
  - 6PE - IPv6 over IPv4 MPLS backbone
  - Multi-Protocol Label Switching (MPLS) VPN (IPv4 and IPv6)
- Exchange of Multi-Protocol Reachability Information (NLRI)

MP-BGP Attributes

- New non-transitive and optional attributes
  - MP_REACH_NLRI
    - Carry the set of reachable destinations together with the next- hop information to be used for forwarding to these destinations
  - MP_UNREACH_NLRI
    - Carry the set of unreachable destinations
- Attribute contains one or more triples
  - Address Family Information (AFI) with Sub-AFI
    - Identifies protocol information carried in the Network Layer Reachability Information
  - Next-hop information
    - Next-hop address must be of the same family
- Reachability information

MP-BGP Negotiation Capabilities

- MP-BGP routers establish BGP sessions through the OPEN message
  - OPEN message contains optional parameters
  - If OPEN parameters are not recognized, BGP session is terminated
  - A new optional parameter: CAPABILITIES
- OPEN message with CAPABILITIES containing:
  - Multi-Protocol extensions (AFI/SAFI)
  - Route Refresh
  - Outbound Route Filtering

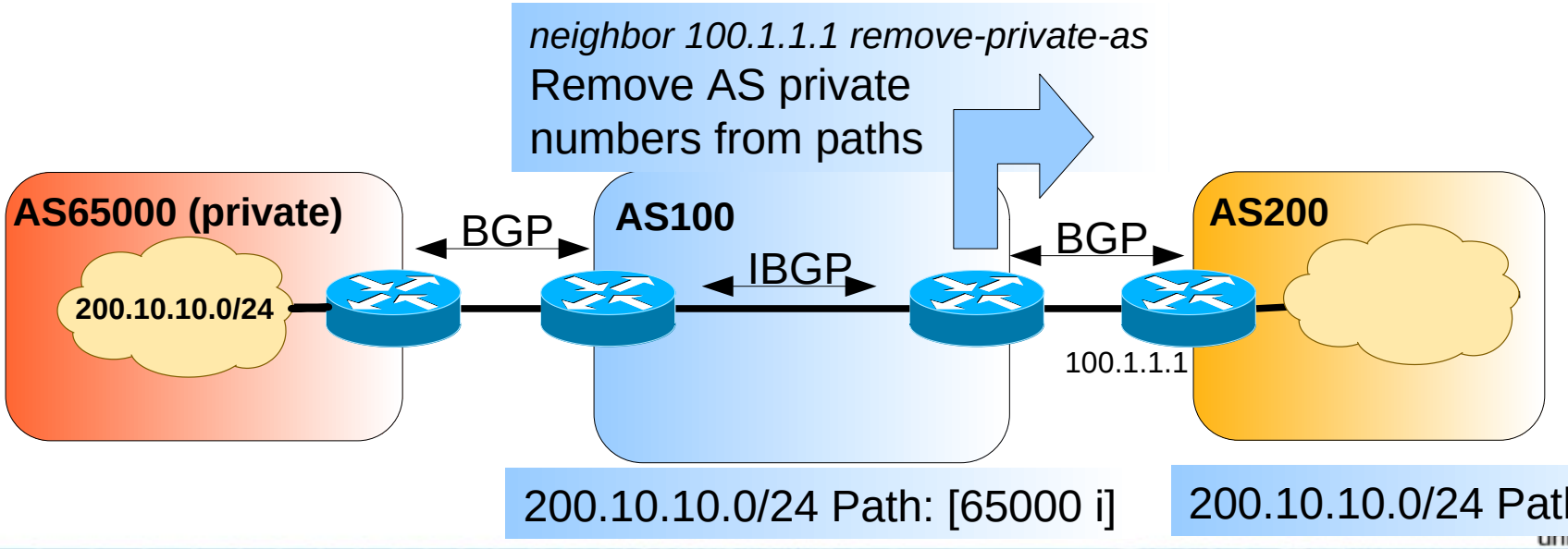
MP-BGP New Features for IPv6

- IPv6 Unicast
  - MP-BGP enables the creation of IPv6 Inter-AS relations
- IPv6 Multicast
  - Unicast prefixes for Reverse Path Forwarding (RPF) checking
  - RPF information is disseminated between autonomous systems
  - Compatible with single domain Rendezvous Points or Protocol Independent Multicast-Source Specific Multicast (PIM-SSM)
  - Topology can be congruent or non-congruent with the unicast one
- IPv6 and label (6PE)
  - IPv6 packet is transported over an IPv4 MPLS backbone
- IPv6 VPN (6VPE)
  - Multiple IPv6 VPNs are created over an IPv4 MPLS backbone
- Layer 2 VPN

Advanced BGP

Private BGP AS

- Private autonomous system (AS) numbers range from 64512 to 65535
- When a customer network is large, the ISP may assign an AS number:
  - Permanently assigning a **Public** AS number in the range of 1 to 64511
    - Should have a unique AS number to propagate its BGP routes to Internet
    - Done when a customer network connects to two different ISPs, such as multihoming
  - Assigning a **Private** AS number in the range of 64512 to 65535.
    - It is not recommended that you use a private AS number when planning to connect to multiple ISPs in the future



BGP AS Routing Policies

```
aut-num:        AS15525
as-name:        PTPRIMENET
descr:          PT Prime Autonomous System
descr:          Corporate Data Communications Services
descr:          Portugal
import:         from AS1930 action pref=100;
                accept AS-RCCN # RCCN
import:         from AS3243 action pref=200;
                accept AS-TELEPAC # Telepac
import:         from AS5516 action pref=100;
                accept AS5516 # INESC
import:         from AS5533 action pref=100;
                accept AS-VIAPT # Via NetWorks Portugal
import:         from AS8657 action pref=300;
                accept ANY # CPRM
import:         from AS12305 action pref=100;
                accept AS12305 # Nortenet
import:         from AS1897 action pref=100;
                accept AS1897 AS9190 AS13134 AS15931 # KPN Qwest
import:         from AS13156 action pref=100;
                accept AS13156 # Cabovisao
import:         from AS8824 action pref=100;
                accept AS8824 AS15919 # Eastecnica
.................
```

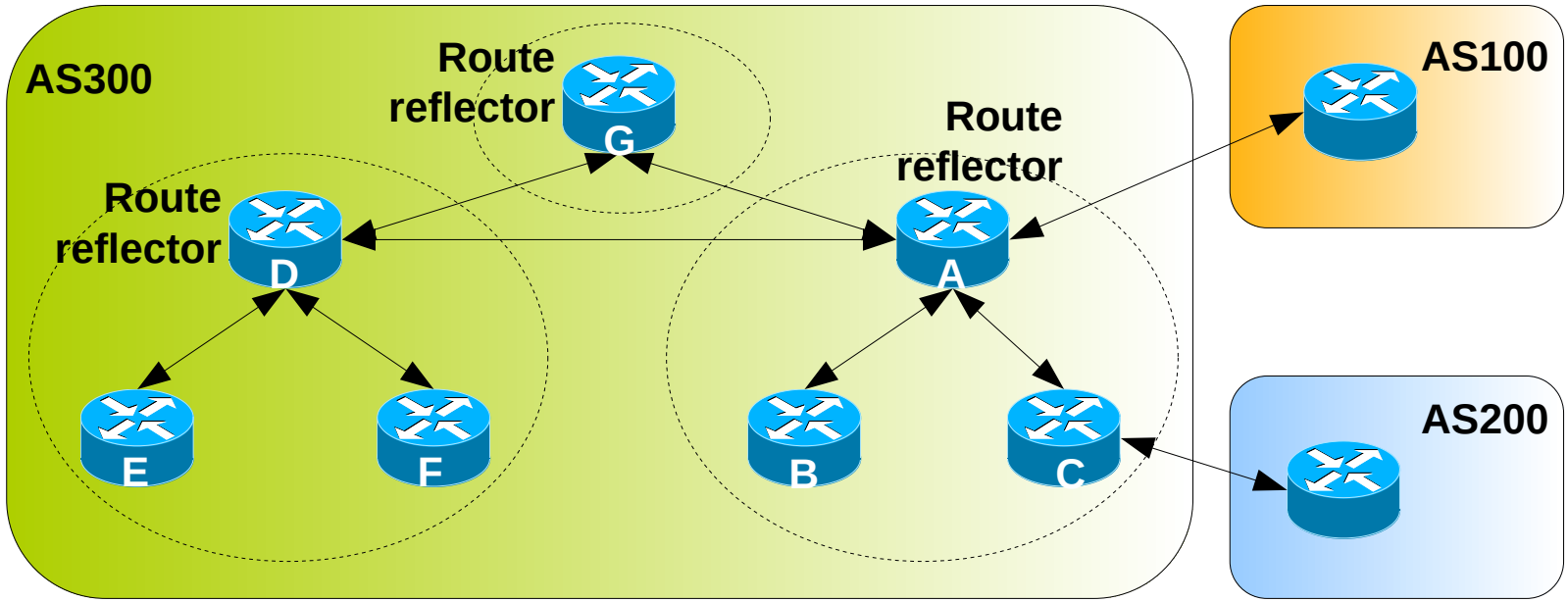
```
export:         to AS1897 announce RS-PTPRIME # KPNQwest
export:         to AS1930 announce RS-PTPRIME # RCCN
export:         to AS3243 announce RS-PTPRIME # Telepac
export:         to AS5516 announce {0.0.0.0/0} # INESC
export:         to AS5533 announce RS-PTPRIME # Via NetWorks Portugal
export:         to AS8657 announce RS-PTPRIME # CPRM
export:         to AS8824 announce RS-PTPRIME  # Eastecnica
export:         to AS8826 announce {0.0.0.0/0} # Siemens
export:         to AS9186 announce RS-PTPRIME # ONI
export:         to AS12305 announce RS-PTPRIME # Nortenet
export:         to AS12353 announce RS-PTPRIME # Vodafone Portugal
export:         to AS13156 announce RS-PTPRIME # Cabovisao
export:         to AS13910 announce ANY # register.com
export:         to AS15931 announce ANY # YASP Hiperbit
export:         to AS24698 announce RS-PTPRIME # Optimus
export:         to AS25005 announce ANY # Finibanco
export:         to AS25253 announce {0.0.0.0/0} # CGDNet
export:         to AS28672 announce ANY # BPN
export:         to AS31401 announce {0.0.0.0/0} # SICAMSERV
export:         to AS39088 announce {0.0.0.0/0} # Santander-Totta
export:         to AS41345 announce RS-PTPRIME  # Visabeira
export:         to AS43064 announce  RS-PTPRIME # Teixeira Duarte
export:         to AS43643 announce  ANY # TAP
.....................
```

From RIPE database
http://www.db.ripe.net

BGP Synchronization

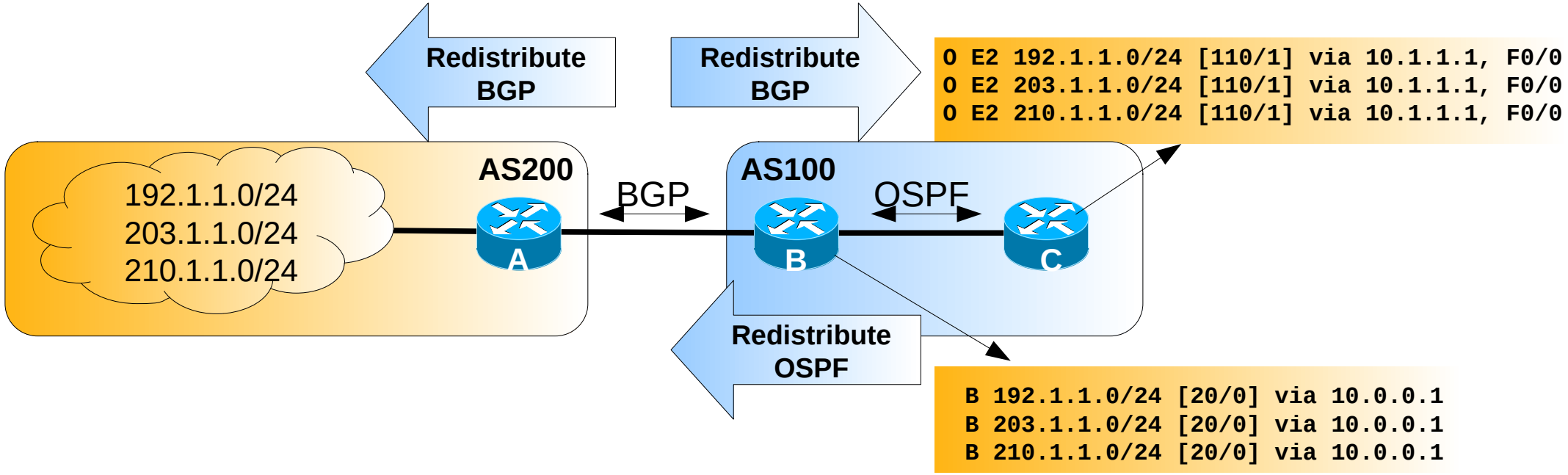
- Synchronization states that, if your AS passes traffic from another AS to a third AS, BGP should not advertise a route before all the routers in your AS have learned about the route via IGP.
- BGP waits until IGP has propagated the route within the AS. Then, BGP advertises the route to external peers.

BGP Route Reflectors



- Without a route reflector, the network requires a full iBGP mesh within AS300.
- The route reflector and its clients are called a cluster.
  - Router A is configured as a route reflector, iBGP peering between Routers B and C (and others) is not required.
  - Router D is configured as a route reflector, iBGP peering between Routers E and F (and others) is not required.
- Full IBGP mesh between route reflector Routers.

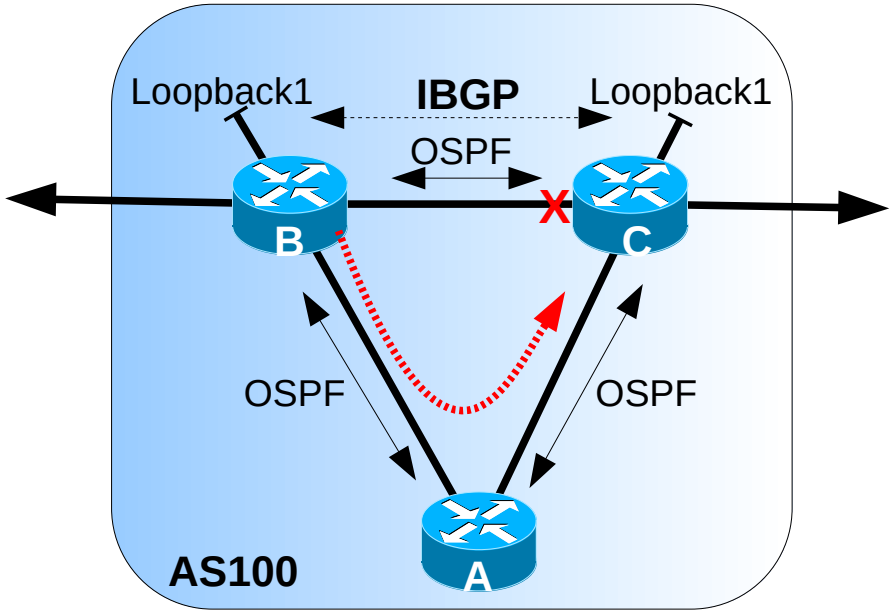
Routes Redistribution



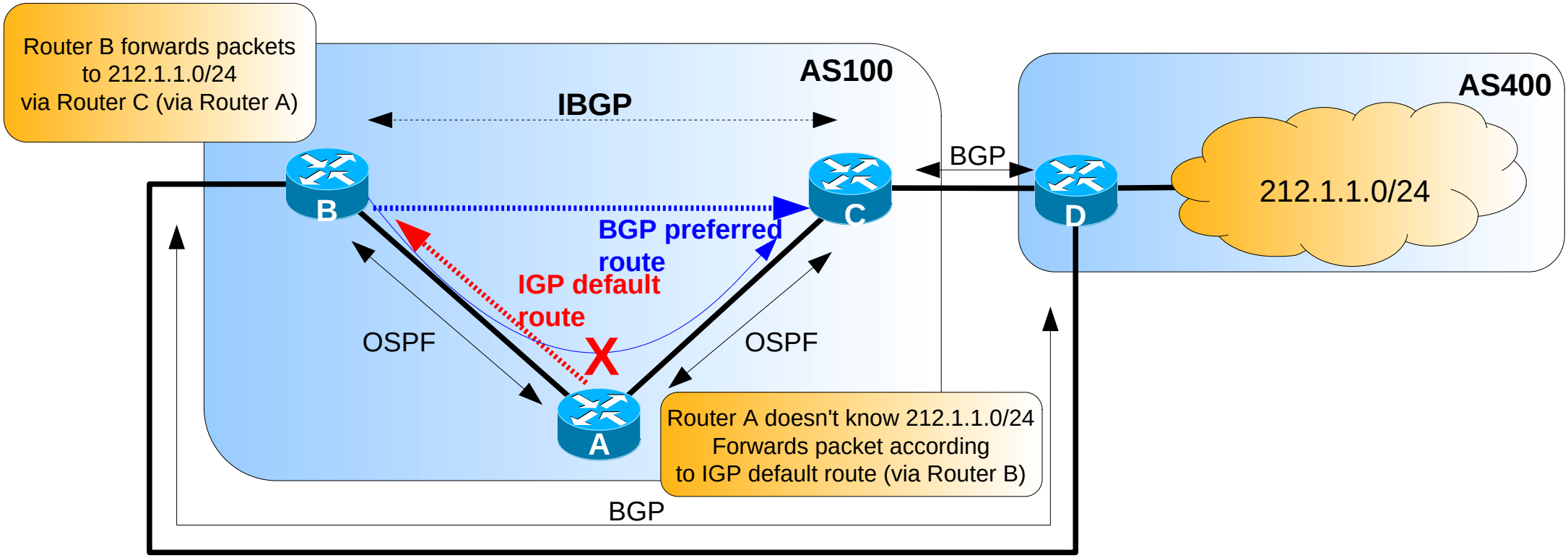
- Redistributing IGP routes by BGP will:
  - Simplify BGP configuration (advantage)
  - And BGP will announce only internal networks with connectivity (advantage)
- Redistributing BGP routes by IGP protocols will:
  - Make internal routes know all external routes (disadvantage/advantage?)
  - Increase routing tables size in internal routers (disadvantage)
    - Decrease routing time, imposes memory requirements, ...
  - Avoid the usage of internal default routes (disadvantage/advantage?)

BGP Neighborhood Resilience

- BGP neighbor relations between physical interfaces are dependent on interface stability/status
- (Virtual) neighbor relations using Loopback interfaces/addresses
  - Loopback interfaces are virtual and software based
    - If the router is active Loopback interfaces are always active
  - Neighbor relation is active while a path exists between the virtual networks
    - (Alternative) Routing provided by IGPs

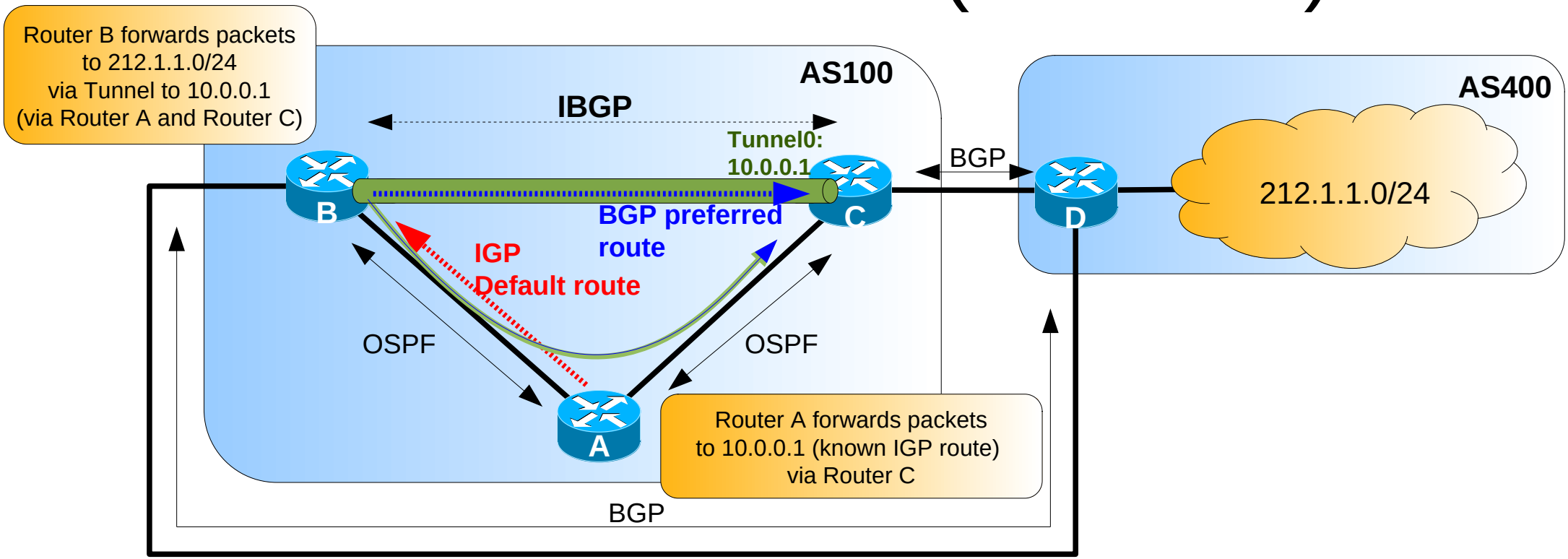


BGP and IGP conflicts



- Routing conflicts may arise with
  - Internal routers without BGP
  - No redistribution of BGP routes by IGP
  - IGP default routes
  - BGP preferred routes (with no agreement with IGP default routes)
- Solutions
  - Adjust IGP default routes
  - Adjust BGP preferred routes (e.g. with local preference)
  - BGP neighborhood and Internal routing via IP-IP tunnels

BGP over Tunnels (over IGP)

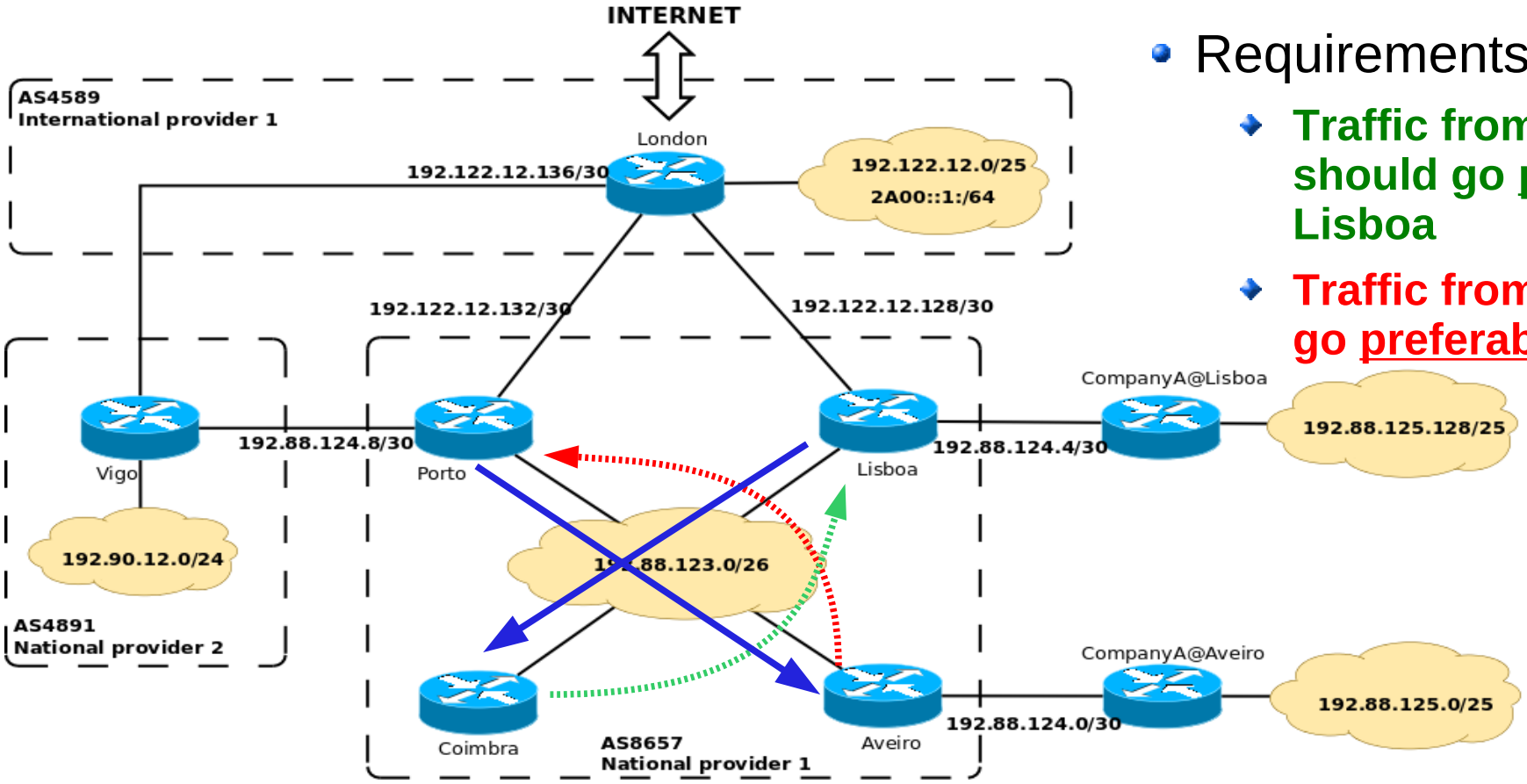


- IP-IP tunnels to solve BGP/IGP routing conflicts
  - Tunnels manually configured
    - Between physical or Loopback interfaces
  - BGP neighborhood via Tunnel
  - BGP routes learned via Tunnel (next hop is remote Tunnel end-point)
  - Tunnel “network” distributed internally via IGP
- In Router A, to any packet destined to an outside network it's forwarded via Tunnel
  - A new IP header is added, new IP destination address is the remote Tunnel end-point
  - Internally, packet is routed according to the new IP header (Tunnel end-points IP addresses)

BGP Filtering and Route Maps

- Sending and receiving BGP updates can be controlled by using a number of different filtering methods.
- BGP updates can be filtered based on:
  - Route information,
  - Path information,
  - Communities.
- Route maps are used with BGP to
  - Control and modify routing information.
  - Define the conditions by which routes are redistributed between routing domains.

BGP Case Studies

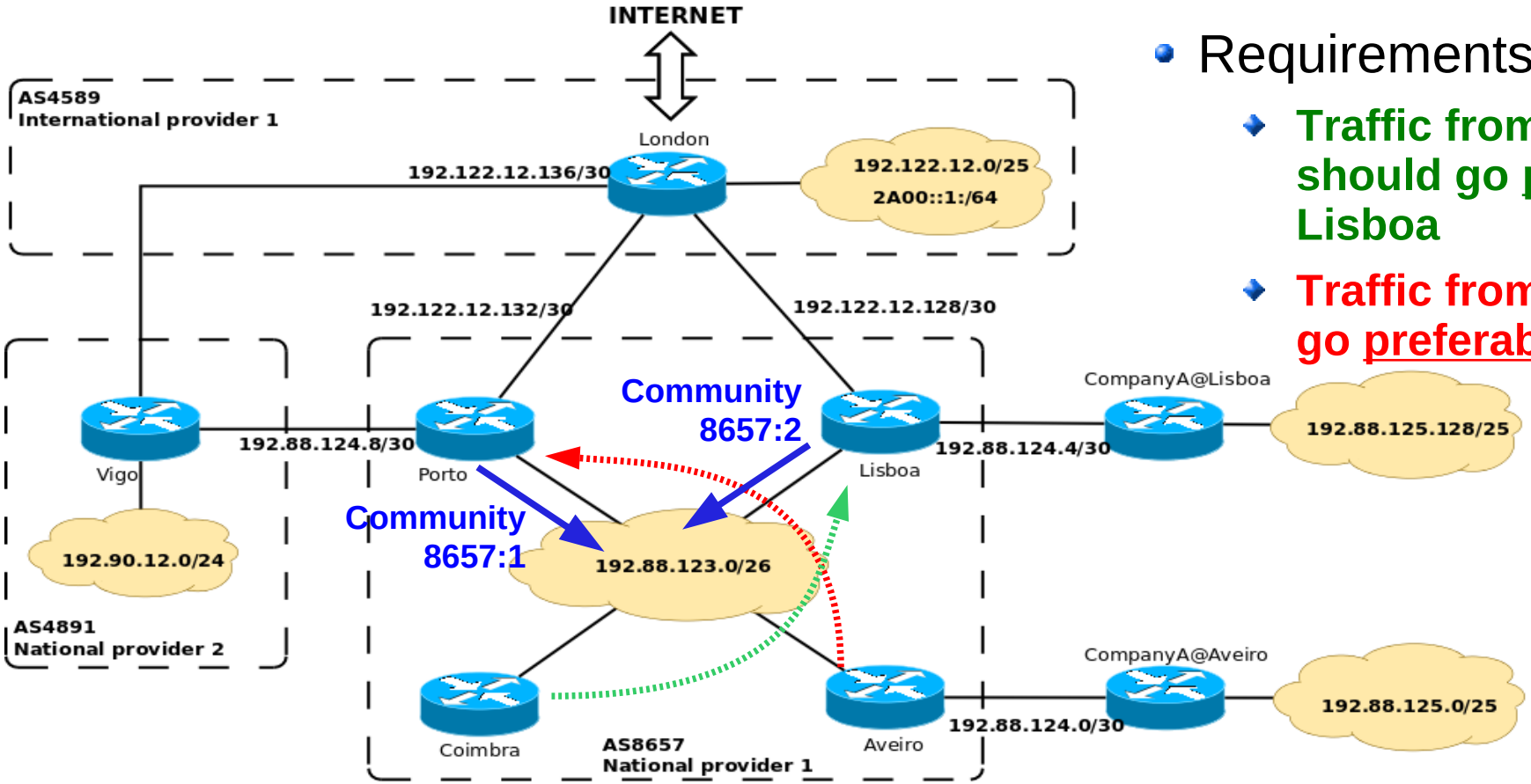


- Requirements
  - **Traffic from Coimbra should go preferably by Lisboa**
  - **Traffic from Aveiro should go preferably by Porto**

- @Porto
  - Nothing to do besides peering!
- @Lisboa
  - Nothingto do besides peering!

- @Aveiro
  - If UPDATE is from Porto → **Local-preference 200**
  - If UPDATE is not from Porto → **Local-preference 100**
- @Coimbra
  - If UPDATE is from Lisboa → **Local-preference 200**
  - If UPDATE is not from Lisboa → **Local-preference 100**

BGP Case Studies



- Requirements
  - **Traffic from Coimbra should go preferably by Lisboa**
  - **Traffic from Aveiro should go preferably by Porto**

- @Porto
  - Route-map applied to all BGP announced external routes/nets
  - Adds BGP attribute: **Community 8657:1**
- @Lisboa
  - Route-map applied to all BGP announced external routes/nets
  - Adds BGP attribute: **Community 8657:2**

- @Aveiro
  - Route-map applied to all BGP received routes/nets
  - If **Community 8657:1** → **Local-preference 200**
  - If **Community 8657:2** → **Local-preference 100**
- @Coimbra
  - Route-map applied to all BGP received routes/nets
  - If **Community 8657:1** → **Local-preference 100**
  - If **Community 8657:2** → **Local-preference 200**

BGP Community Attribute (real data)

- TeliaNet Global Network

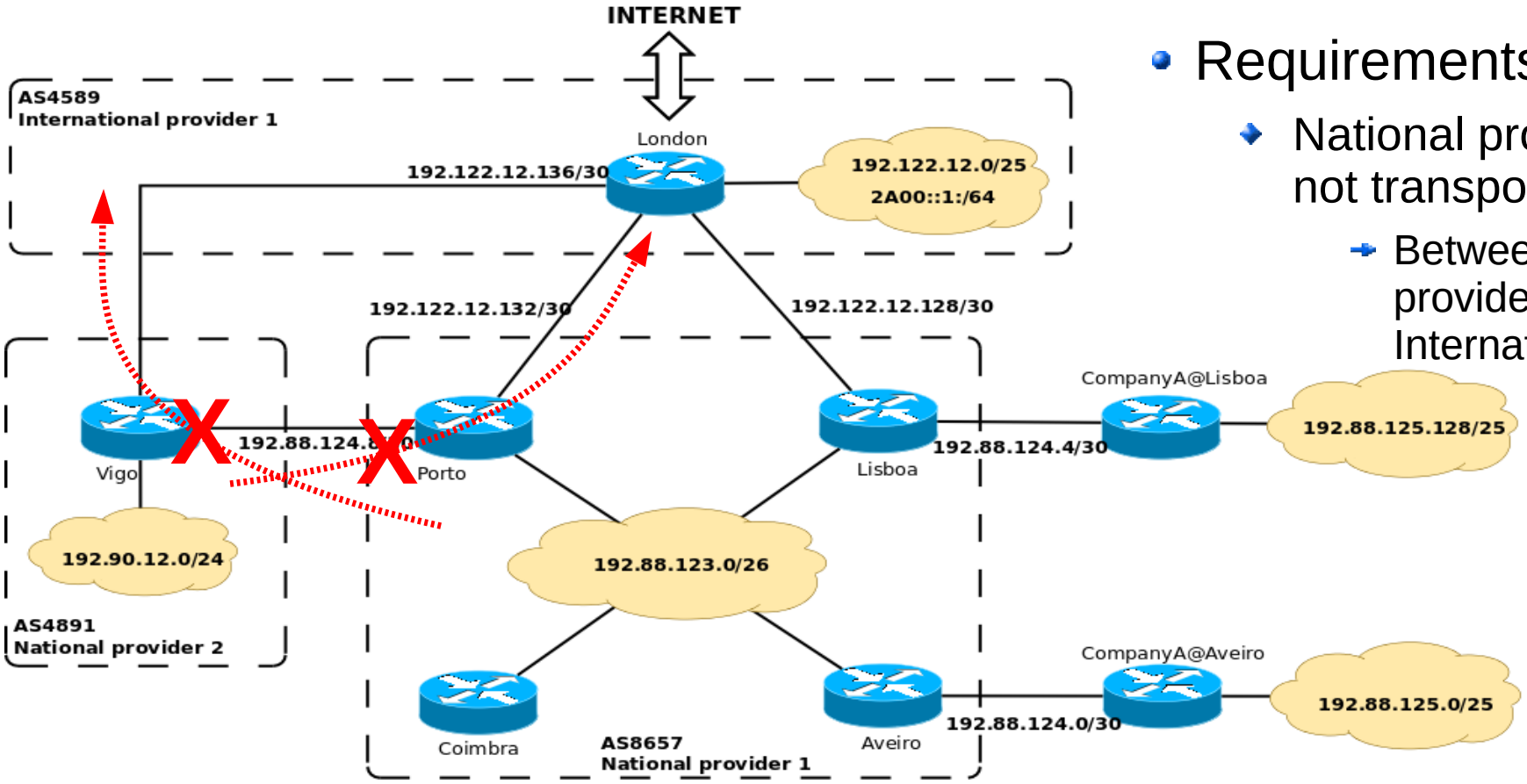
```
remarks:      BGP COMMUNITY SUPPORT FOR AS1299 TRANSIT CUSTOMERS:
remarks:
remarks:      Community Action
remarks:      --------- ------
remarks:      1299:50   Set local pref 50 within AS1299 (lowest possible)
remarks:      1299:150  Set local pref 150 within AS1299 (equal to peer, backup)
remarks:
remarks:      European peers/ix-points          US peers/ix-points              Asia peers/ix-points
remarks:      Community  Action                 Community  Action               Community  Action
remarks:      ---------  ----------------------  ---------  ----------------------  ---------  ----------------------
remarks:      1299:200x  All peers Europe incl:  1299:500x  All peers US incl:      1299:700x  All peers Asia incl:
...                  -
remarks:      1299:250x  Sprint/1239             1299:550x  Sprint/1239             -
remarks:      1299:251x  Savvis/3561             1299:551x  Savvis/3561             -
remarks:      1299:252x  Verio/2914              1299:552x  Verio/2914              -
remarks:      1299:253x  Abovenet/6461           1299:553x  Abovenet/6461           -
remarks:      1299:254x  FT/5511                 1299:554x  FT/5511                 1299:754x  FT/5511
remarks:      1299:255x  GBLX/3549               1299:555x  GBLX/3549               1299:755x  GBLX/3549
remarks:      1299:256x  Level3/3356             1299:556x  Level3/3356             -
remarks:      1299:257x  UUnet/702               1299:557x  UUnet/701               -
remarks:      1299:558x  AT&T/7018               1299:758x  AT&T/2687
remarks:      1299:259x  Telefonica/12956        1299:559x  Telefonica/12956        -
remarks:      1299:260x  BT/Concert/5400         -                                  -
remarks:      1299:261x  Qwest/209               1299:561x  Qwest/209               -
remarks:      1299:263x  Teleglobe/6453          1299:563x  Teleglobe/6453          -
remarks:      1299:264x  DTAG/3320               1299:564x  DTAG/3320               -
remarks:      1299:268x  AOL/1668                1299:568x  AOL/1668                -
remarks:      1299:269x  Tiscali/3257            1299:569x  Tiscali/3257            1299:769x  Tiscali/3257
remarks:      1299:270x  UPC/6830                -                                  -
remarks:      1299:273x  Cogent/174              1299:573x  Cogent/174              -
remarks:      1299:274x  Telecom Italia/6762     1299:574x  Telecom Italia/6762     1299:774x  Telecom Italia/6762
remarks:      1299:275x  Tele2/1257              -                                  -
...
remarks:      1299:284x  Cable & Wireless DE/1273 1299:584x  Cable & Wireless DE/1273 -
remarks:      1299:286x  KPN/286                 -                                  -
remarks:      1299:287x  China Netcom/4837       1299:587x  China Netcom/4837       1299:787x  China Netcom/4837
remarks:      1299:288x  China Telecom/4134      1299:588x  China Telecom/4134      1299:788x  China Telecom/4134
```

From RIPE database

https://apps.db.ripe.net/

e.g., https://apps.db.ripe.net/db-web-ui/#/query?bflag=false&dflag=false&rflag=true&searchtext=as1299

BGP Case Studies

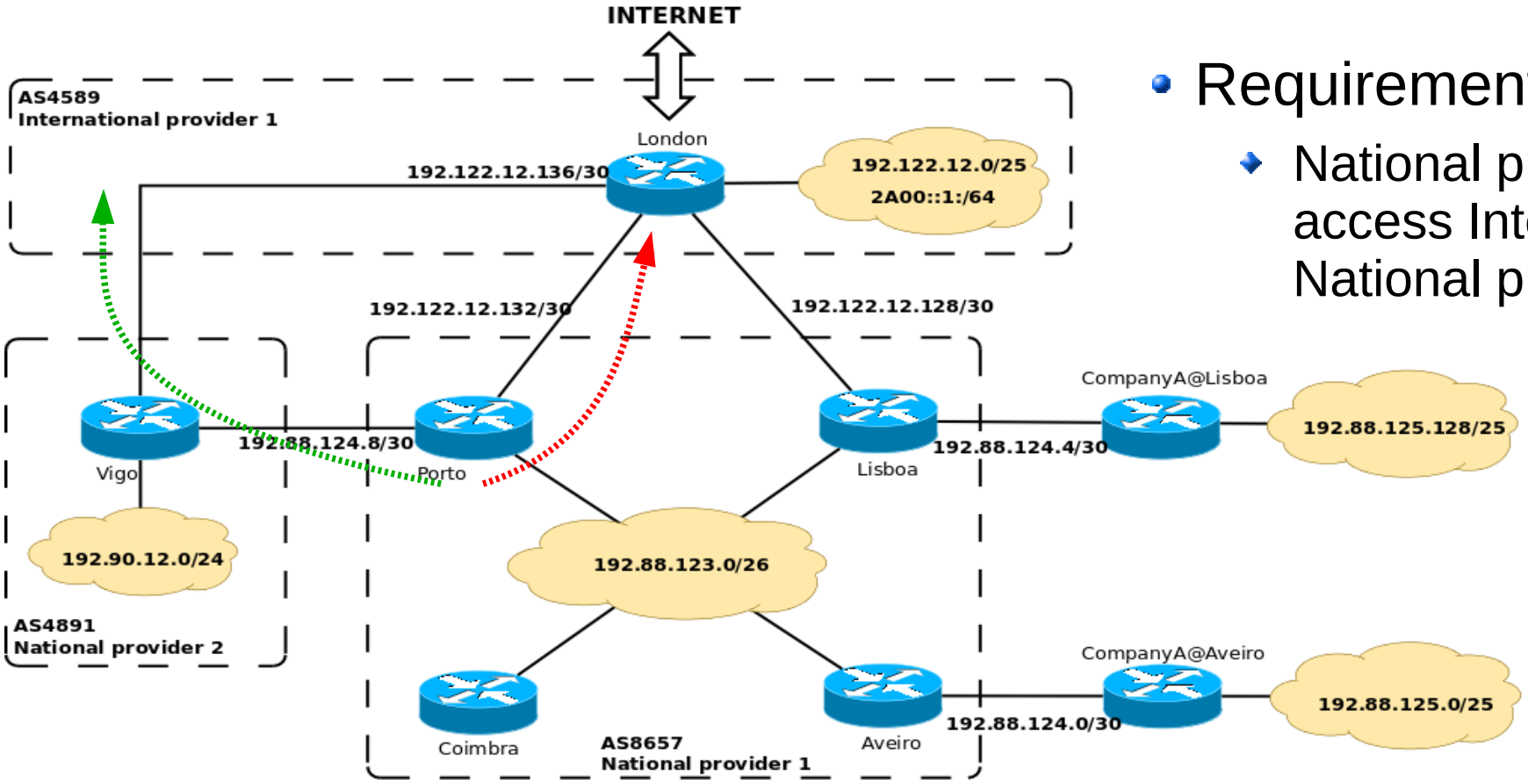


- Requirements
  - National providers should not transport traffic
    - Between other national providers and the International provider

- @Porto, @Lisboa
  - Route-map applied to all external BGP announcements
  - Announce only internal routes/nets
    - Empty path “^$”

- @Vigo
  - Route-map applied to all external BGP announcements
  - Announce only internal routes/nets
    - Empty path “^$”

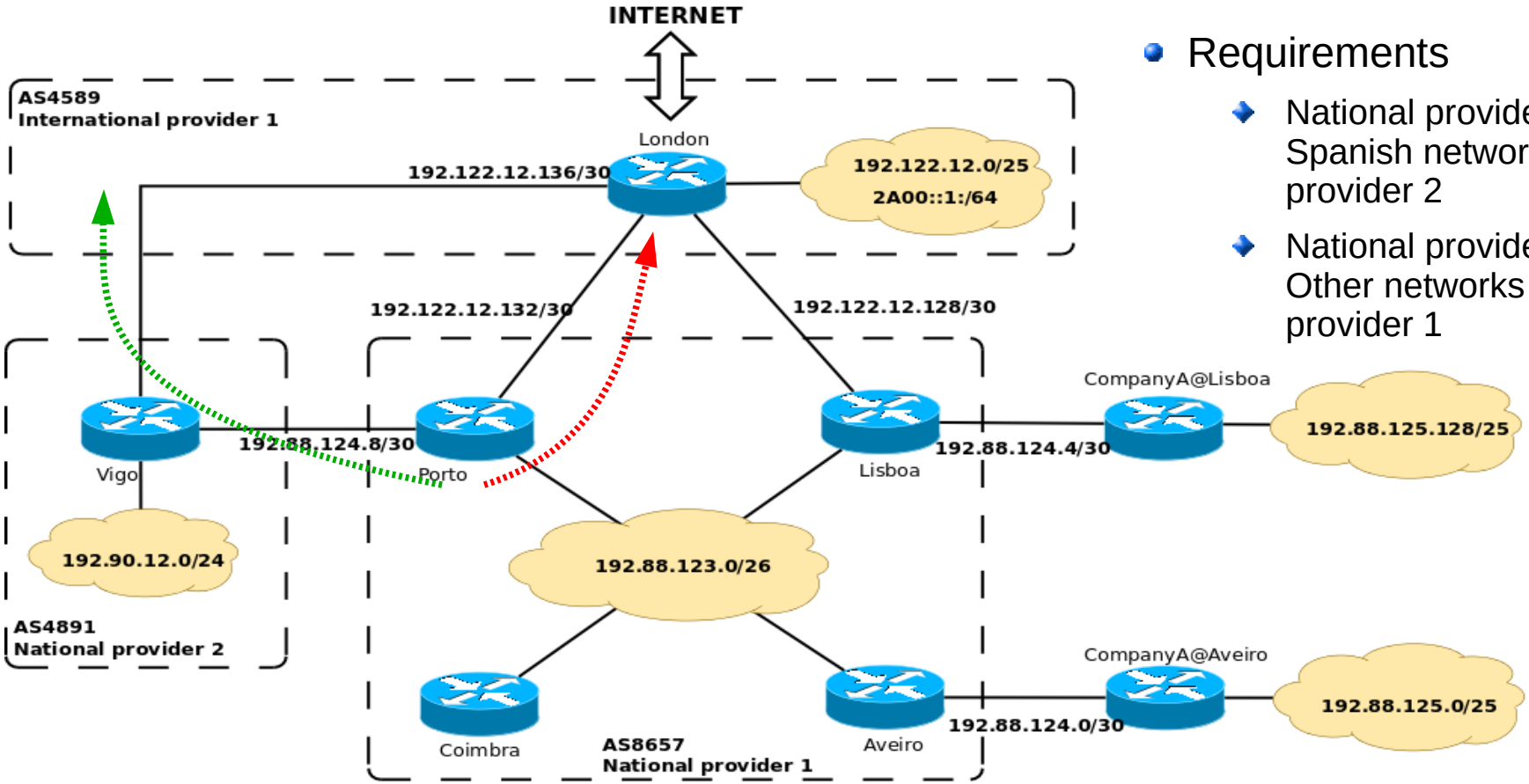
BGP Case Studies



- Requirements
  - National provider 1 should access Internet using National provider 2

- @Porto, @Lisboa
  - Route-map applied to all BGP announcements received
  - If Path contains “4891” → **Local-preference 200**
  - If Path does not contain “4891” → **Local-preference 100**

BGP Case Studies

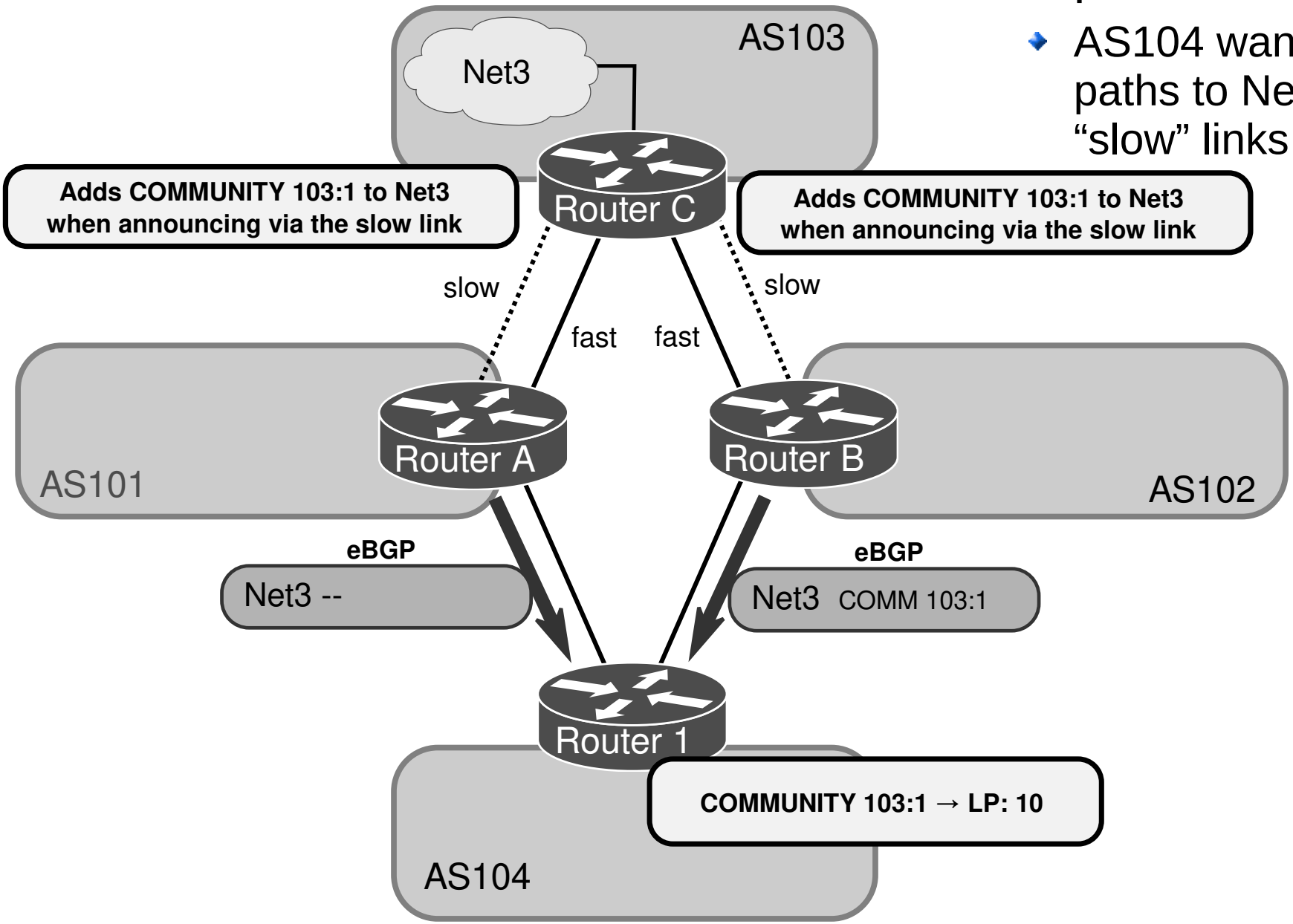


- Requirements
  - National provider 1 should access Spanish networks using National provider 2
  - National provider 1 should access Other networks using International provider 1

- @Porto, @Lisboa
  - Route-map applied to all BGP announcements received
    - E.g. known Spanish operators AS: 4891, 7654, 9876 and 3352
  - If Path starts (from right to left) with “4891$ or 7654$ or 9876$ or 3352$” and ends in “^4891” → **Local-preference 200**
  - If Path does not start with “4891$ or 7654$ or 9876$ or 3352$” and ends in “^4891” → **Local-preference 50**
  - Assuming default Local-preference 100.

BGP Case Studies

- Requirements
  - AS104 wants to avoid paths to Net3 that use “slow” links.



universidade de aveiro
theoria poiesis praxis

TCP/IP
Quality of Service

Qualidade de Serviço
2023/2024

Mestrado Engenharia Computadores e Telemática
DETI-UA

1

TODAY

- We will see mechanisms to add quality of service inside the network, providing “other” guarantees than the basic “TCP connection pipe” assurances

2

2

Multiservice Networks

Emerging services – heterogeneous requirements

QoS over IP networks

3

3

Current services

- Internet has many services beyond the basic network services
- Services
  - Interactive games
  - Audio/video
  - High definition moving image
  - Data base
  - Information storage
  - Communication networks

That require large transport systems

- Data networks can be very complex!

4

4

Services requirements

- Packet loss
  - Some applications (e.g., real-time audio/video) support losses
    - Voice supports more losses than video
    - TCP and its retransmissions
  - Other applications (e.g., file transfer, telnet) require 100% of success in transmission
    - However, they use TCP
- Bandwidth
  - Some applications (e.g., multimedia) require a minimum bandwidth
    - Buffer gets full
    - Large delays and some losses
  - Other applications ("elastic applications", e.g., email, file transfer) use the bandwidth they can get
- Timing: delay and jitter
  - Some applications (e.g., Internet telephony, multiplayer games) require low delays
  - Other applications (non-real-time) do not present strict limits on end-to-end delay
  - Some applications do not react well to delay variations (jitter)

5

5

Multimedia services

- Transmission of different types of information in the same service
  - Voice
  - Video
  - Data
- It is required synchronization between these types of information
  - Sending and reception in the terminal equipments
- It is required to handle different requirements of services in the same network
  - Interactivity vs non-interactivity
  - Bandwidth
  - Delay
  - Losses
- It is required to support interactivity in environments with variable delay

7

7

Multimedia services

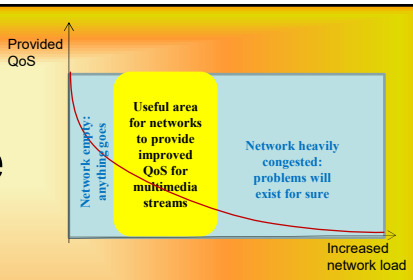
- Transmission of different types of information in the same service
  - Voice
  - Video
  - Data
- It is required synchronization between these types of information
  - Sending and reception in the terminal equipments
- It is required to handle different requirements of services in the network
  - Interactivity vs non-interactivity
  - Bandwidth
  - Delay
  - Losses
- It is required to support interactivity in environments with variable delay

Networks do not handle this usually: they are best-effort

8

8

Multimedia: quality assurance



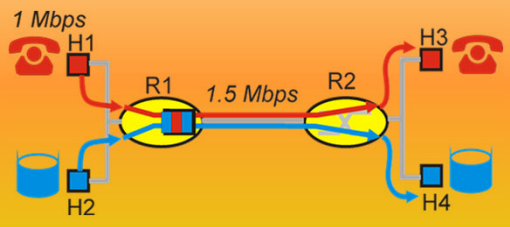
- **Multimedia applications assume best-effort networks**. As such there are interactions between the codecs and the network behaviour.
- Internet multimedia applications have several usual strategies:
  - Dynamically changing codecs, trading quality by bandwidth/resilience
  - Buffering, dynamically changing the size of the buffer
  - Progressive downloads, prefetching some seconds/minutes beforehand
- However, in networks medium loaded, some resource management is required, favouring multimedia flows over (best-effort) data flows, effectively creating an approach of weighted multiplexing gains.

9

9

What is needed to guarantee QoS?

- Example: 1Mbps IP audio/video stream and FTP transfer share a 1.5 Mbps connection.
  - FTP bursts can congest the router, causing loss in audio/video
  - It is intended to give priority to audio/video



Principle 1

Packet marking is required so that the router can distinguish between the different traffic types; new policies are needed on the router for handling packets

10

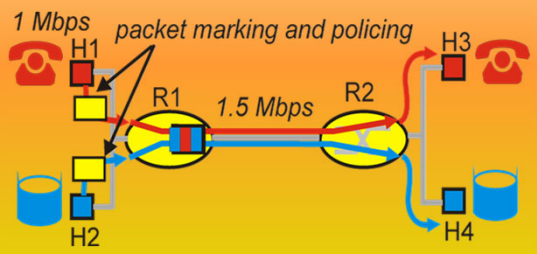
10

What is needed to guarantee QoS?

What if the applications "misbehave" (e.g. audio/video sends more than the declared bitrate)?

- policing: forces the compliance of the sources to the agreed bandwidth

- marking and policing at the network entry



Principle 2

Provide **protection** (isolation) of one traffic class in relation to the other

11

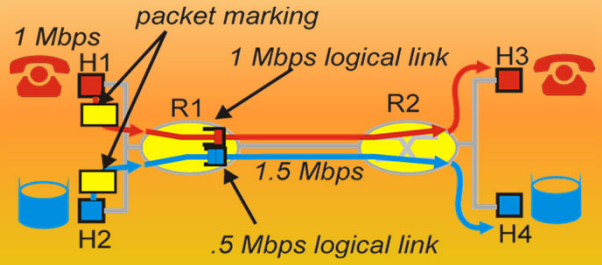
11

Para garantir a Qualidade de Serviço (QoS), é necessário ter em consideração alguns princípios. Um deles é a atribuição de uma largura de banda fixa ao stream de áudio/vídeo. No entanto, esta abordagem pode ser ineficiente se o stream não utilizar toda a largura de banda atribuída.

O princípio 3 é que, ao fornecer isolamento (reservar recursos dedicados a determinados fluxos), é desejável utilizar os recursos tão eficientemente

What is needed to guarantee QoS?

- Can we assign a Fixed BW to the audio/video stream?
  - Inefficient use of the bandwidth if the stream does not use the bandwidth that was assigned



Principle 3

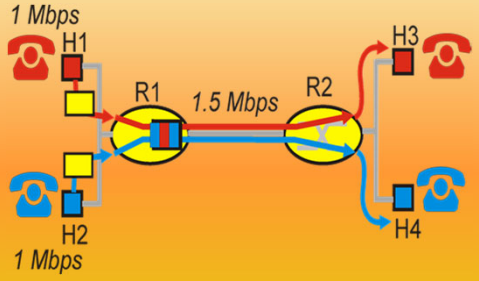
When providing isolation, it is desirable to use the resources **as efficiently as possible**

12

12

What is needed to guarantee QoS?

- It is not possible to support requests that exceed the connection capacity

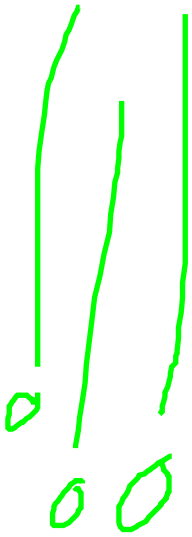


Principle 4

Call admission: the stream declares its requirements, and the network can block the call (busy signal) if it can't support them

13

13



What is needed for QoS support

1. Some form of signalling between applications and network (and internally between parts of applications)
2. Signaling for resource reservation/management (typically RSVP)
3. Ability to differentiate traffic treatment inside network equipment (typically queueing strategies in routers, see last slides)
4. Control and policing of the network usage (see last slides).

Previous slides we saw the basic concept that allow 3) and 4) inside the network. This section discusses 1) and (mostly) 2, and how the concepts work together.

14

14

Main approaches in IP networks

Basic IP service:

- Packets suffer delays, losses, jitter and reordering.

Differentiated Services

- Classes of services

Integrated Services

- Defined service levels

15

15

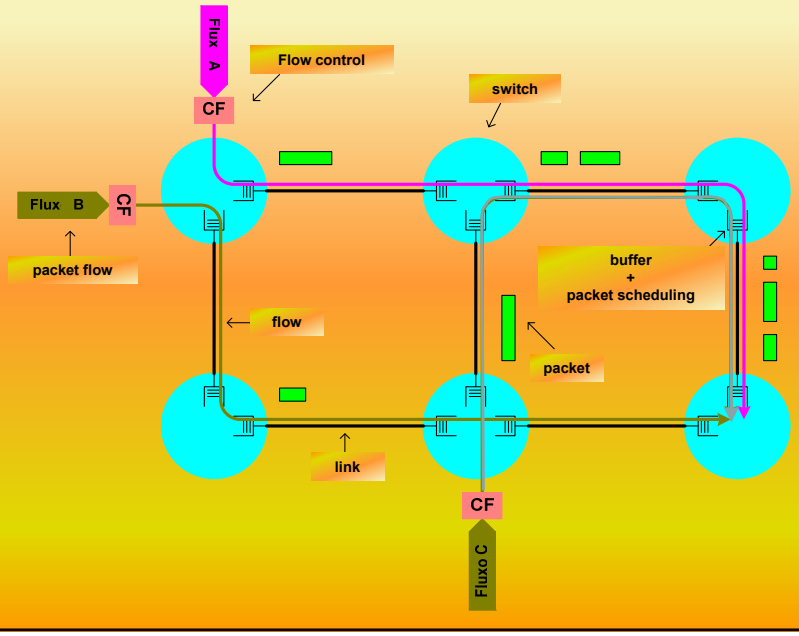
Main differences

- IntServ
  - ✓ Rely in flows, implementing two types of E2E services: GS (guaranteed service) and CL (Controlled Load).
  - ✗ Does not scale in the core!!!
- DiffServ
  - ✓ Works on simpler aggregates, with an approach effective for QoS on the core.
  - ✗ Does not provide E2E guarantees.

16

16

IP network with QoS support: summary



17

17

How to condition traffic?

Basic concepts

18

18

Policing/Shaping

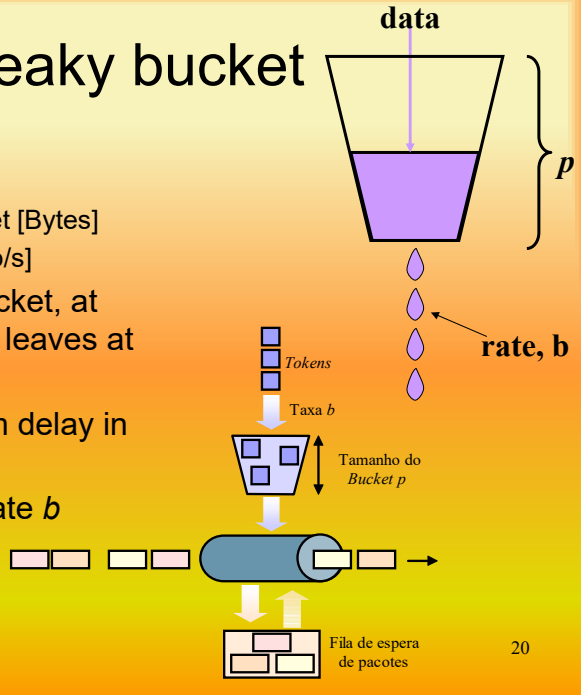
- Policing reduces the impact of excess traffic
  - Loss of excess packets
  - Tagging with lower qos
- Shaping stores traffic, smoothing bursts
  - Only allows traffic to be sent at A certain rate

19

19

Leaky bucket

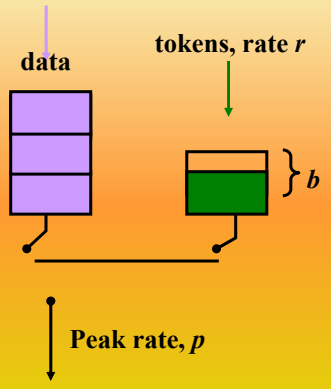
- Two parameters:
  - *p*: size of the bucket [Bytes]
  - *b*: exit rate [B/s or b/s]
- Data enters the bucket, at whatever rate, and leaves at constant rate.
- *p*/*b* is the maximum delay in transmission
- Bit rate limited at rate *b*



20

Token bucket

- Three parameters:
  - *b*: token bucket size [B]
  - *r*: token rate [B/s or b/s]
  - *p*: peak rate [B/s or b/s]
- Bucket fills with tokens at rate *r*
  - bucket starts full (supporting burst)
- If there are tokens, transmission is possible.
  - Bursts are possible, with rate *p*
- Total data sent in time t < *r*t + *b*
- Traffic limited at rate *p* when there are tokens in the bucket (tokens are generated at a smaller rate *r*)

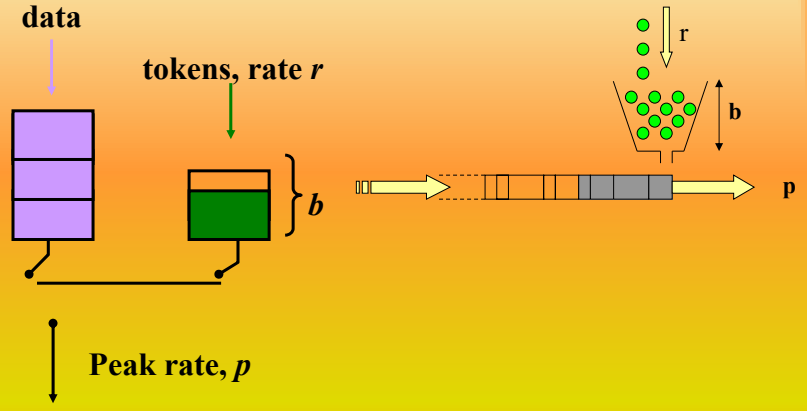


- **Average rate** never exceeds ***r***

21

21

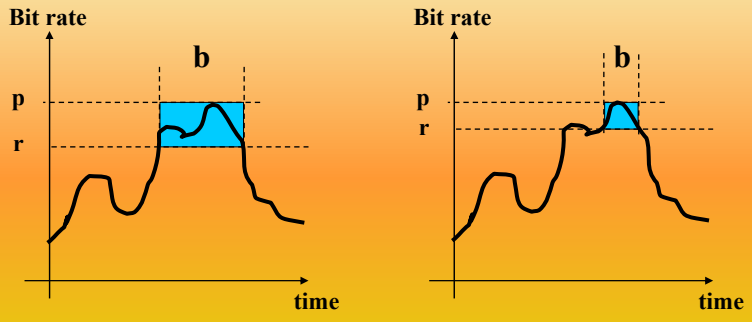
Token Bucket



22

22

Describing sender traffic



- In the time interval Δt the number of packets (bytes) accepted must be equal or less than (rΔt + b).
  - Actions to take may be policing or shaping

23

23

Token(s)

- Basic tool for "measurement"
- Can be used to:
  - Describe traffic
  - Validate traffic compliance
  - Clarify terms we use regularly ("bandwidth", "burst",...)
  - Assist in mathematical analysis of networks

24

24

Example

- **2 tokens, size 100 bytes, added every second to the token bucket, with total capacity 500 bytes. Operations are handled in bytes**
  - **Average rate = 200 bytes/sec,**
  - **burst size = 500 bytes**
  - **Packets larger than 500 bytes are never sent**
- **Is it possible to get a peak rate above 200 bytes/sec ?**
  - **Yes, any rate is possible as long as you send in each second only 500 bytes**
  - Example: we can transmit 100 bytes in 1ms, meaning 100Kbps of peak transmission…

25

25

Admission control

26

Policing (drops)

- Dropping packets is one of the possible actions to ensure that the expected network performance is not exceeded
- For a some traffic types (e.g. voice) it makes no sense to miss only "a few" packets, minimum guarantees are required
- Admission control:
  - Reject/accept flows, with a well-defined traffic, making sure that a certain QoS will be guaranteed
  - admission control may however have actions such as shaping

It is an action inherent in the operation of circuit switching networks (telephone networks) – all calls are previously "admitted" before being processed.

27

Quando ocorre uma sobrecarga nos buffers do router (filas de espera), é comum descartar pacotes para evitar a sobrecarga não só nos buffers, mas também na rede como um todo. No entanto, a questão de como descartar esses pacotes é importante.

Existem várias abordagens para descartar pacotes:

1. Último pacote a chegar: Nesta abordagem, o último pacote que chega é descartado. Isso pode ser útil quando se deseja dar prioridade aos pacotes mais

Packet drop techniques

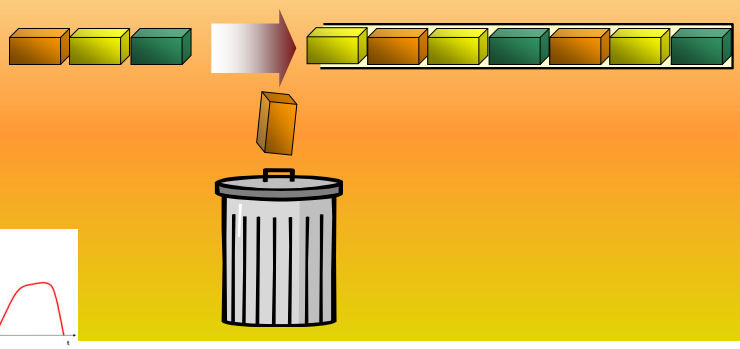
- Dropping packets to stop overload of the router buffers (waiting queues)
  - And of the network.
- How to drop?
  - Last packet to arrive?
  - First packet to arrive?
  - Any random packet?
  - Differentiated packets per class?

28

28

Congestion control (I) – *Tail Drop*

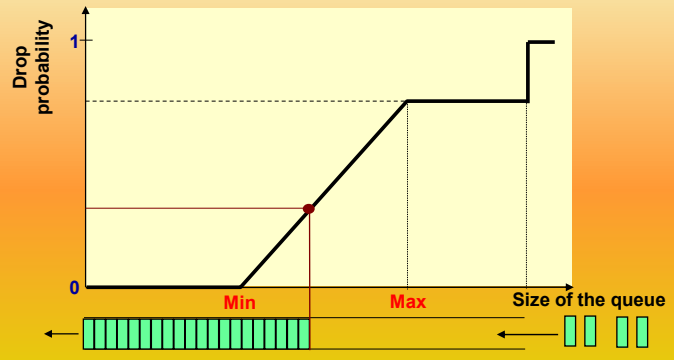
Tail drop: in each buffer, the packets received are dropped when the buffer is full



Problem: Interaction with TCP flow control mechanisms in the core routers ➔ **Global synchronization of the TCP traffic sources.**

29

Congestion control (II) – RED (Random Early Detection)



30

30

RED: Random Early Detection

- Random Early Detection:
  - Handles congestion before it appears
  - packet loss → Congestion signal
    - Source slows down
    - Prevents actual congestion
- What packets to lose?
  - Probability of packet loss ∝ queue length
  - Monitoring of flows
    - Cost in processing vs overall network performance
  - Queue length – exponential average:
    - "smooths" reaction to traffic bursts
    - Limits constant heavy traffic, being good for Intserv (Controlled Load)
  - Packets may be lost or marked as "offending"
    - RED-aware routers will lose packets "offending", when needed
  - Source should adapt:
    - TCP: OK!
    - real-time traffic - UDP ?

31

31

RED

- **RED Parameters**
  - Minimum queue threshold (minQ)
  - Maximum queue threshold (maxQ)
  - Average Queue length (AvgQ).
    - Dynamic calculated
  - Maximum drop probability (maxP)
  - Drop probability (P)
    - Dynamic calculated
- **Algorithm**
- For each incoming packet
  - If AvgQ <= minQ
    - queue packet
  - If minQ <= AvgQ < maxQ
    - Mark packet with probability P
  - If maxQ <= AvgQ
    - Mark the packet

32

32

RED

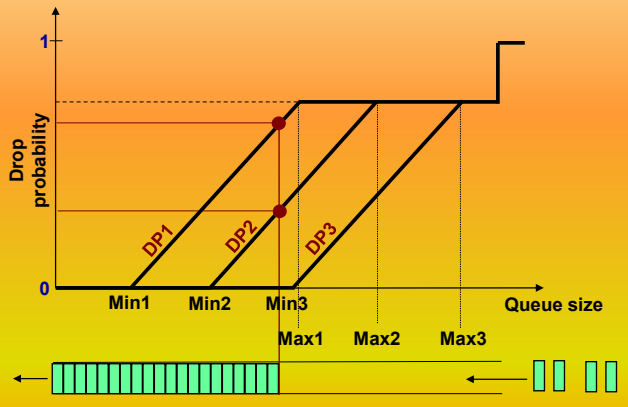
- **Size of the waiting queing**
  - **AvgQ = (1 – weight) x AvgQ + weight x currQ**
    - 0 < Weight < 1
    - currQ é o tamanho actual da fila de espera
- **Drop probability**
  - **TempP = MaxP x (AvgQ – minQ)/(maxQ – minQ)**
  - **P = TempP/ (1 – count.TempP)**
    - Count is the number of new packets that reached the queue

33

33

WRED (*Weighted Random Early Detection*)

- **3 drop levels – 3 drop levels for different sizes of que queue**
  - **The last to be discarded are those with more priority**
  - **The first to be discarded are those with lower priority**

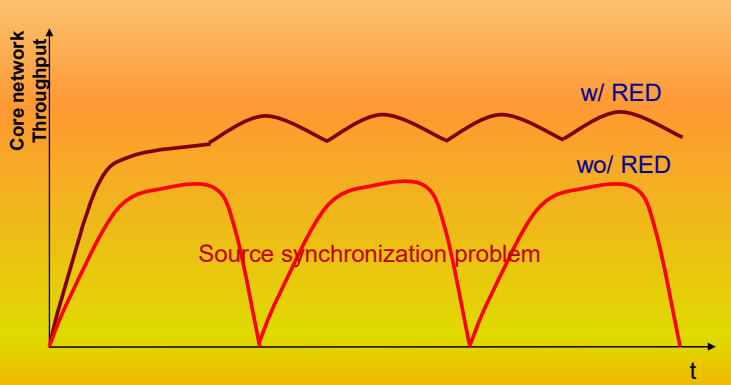


34

34

Congestion control (II) – RED impact in global synchronization

- **With RED we are able to decrease the TCP global synchronization in core routers,** created by all input flows reacting at the same time to the queue congestion in core routers



35

35

Scheduling algorithms

Scheduling algorithms: Decide the order in which packets belonging to different streams are served in a queue.

Work conserving scheduling algorithms ensure that the server is always busy if there is a packet waiting to be transmitted.

Examples of work conserving scheduling algorithms:

(1) FIFO,

(2) Strict priority,

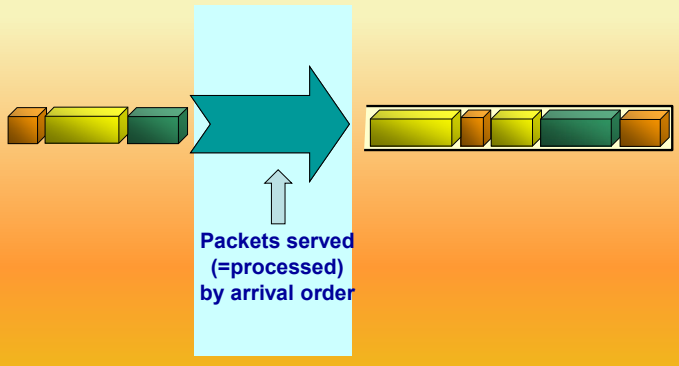
(3) Fair Queuing,

(4) Weighted Fair Queuing.

36

36

First In First Out (FIFO)

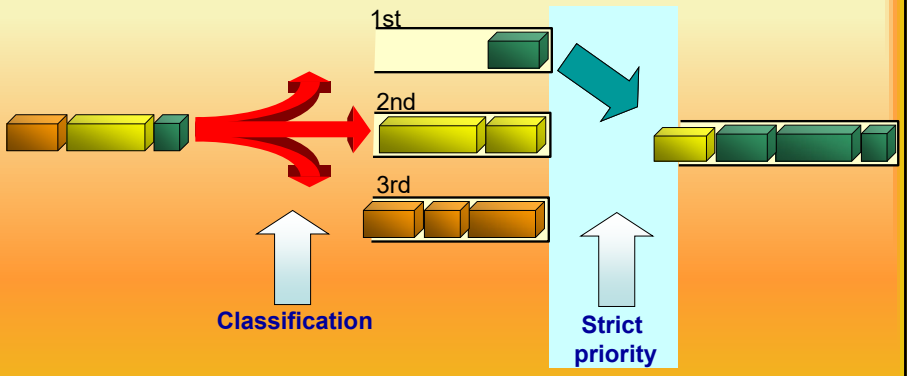


- It does not perform sort processing.
- It does not allow differentiation of quality of service.
- Flows with *n* times more traffic receive *n* times more service.
- In finite queues, streams with smaller packets, receive more service

37

37

Priority Queueing

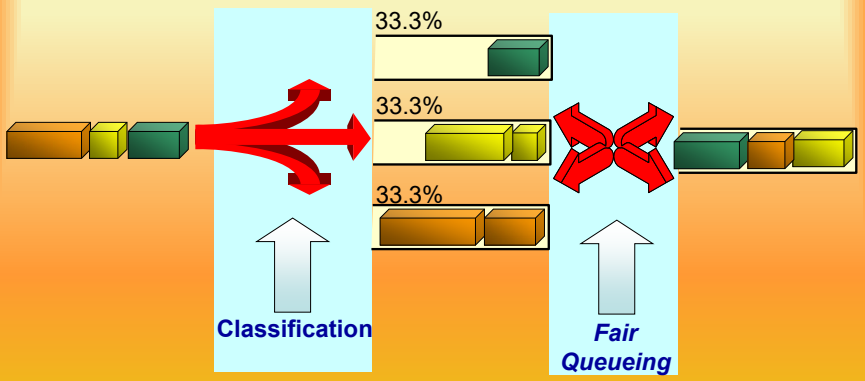


- It involves classification of traffic according to priority.
- The highest priority traffic is always served before the traffic of smaller priorities.
- Allows differentiation of the quality of service.
- Higher priority flows can prevent lower flows from receiving any service

38

38

Fair Queueing (FQ)



- Involves classification of traffic in different queues.
- Transmission bandwidth is equally distributed across non-empty queues.
- Allows quality of service allocation and reservation.

39

39

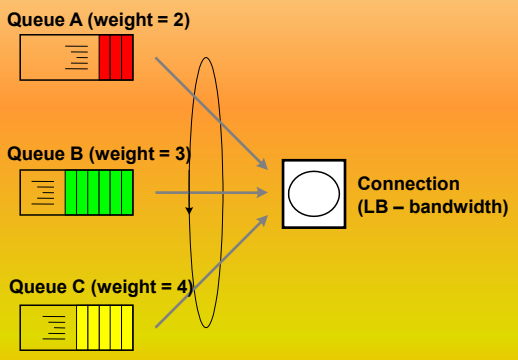
Weighted Fair Queuing (WFQ)

This algorithm ensures that each queue achieves a percentage of the connection bandwidth at least equal to its weight divided by the sum of the weights of all queues

$$R_A = \frac{2}{2+3+4} LB$$

$$R_B = \frac{3}{2+3+4} LB$$

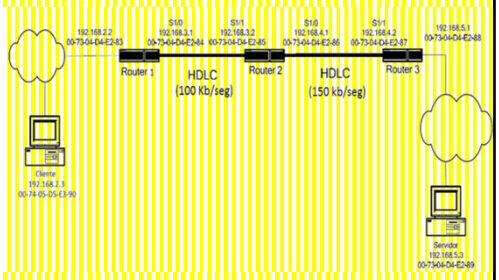
$$R_C = \frac{4}{2+3+4} LB$$



40

40

Exercise



- Consider that in this network, the serial interfaces of the connections between routers 1, 2 and 3 only have FIFO mechanisms active. However, you can configure also Random Early Detection (RED) with 2 drop probabilities: high discard probability and low probability.
  - How does this mechanism work?
  - If you had to choose priorities for a video and file transfer service, which one would you choose? Justify.
  - If routers had Weighted Fair Queueing active with 3 different queues and with weights of 2 for the voice queue, 5 for video, and 3 for data, what bandwidth is available for each service?
  - In the same case, if there are no video packets, what bandwidth is available for voice and data?
- Considering the network with an active 50 Kb/sec video service and FIFO queues with unlimited capacity, determine the delay of the 1000-bit video service packets, considering that only the serial connections have a meaningful delay.

41

41

Quality of Service

Supporting network services

42

NOW

- We will see how we can signal and provide QoS in the network, using the previous mechanism.

43

43

Integrated Services (IntServ)

44

Integrated Services:

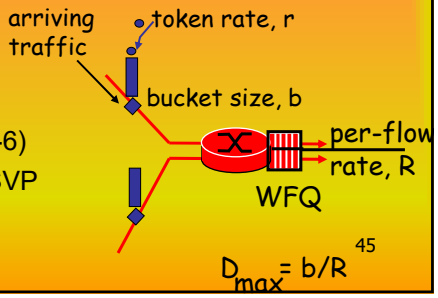
- Controlled Load CL (RFC2211)
  - Assures service end-to-end (E2E) that is as load independent as possible (*good best-effort always)*
  - End nodes will receive most packets with minimal delays in routers
- Guaranteed Service GS (RFC2212)
  - Assures E2E service in terms of delay, for a given bandwidth.
- Best Effort BE

Does not guarantee any quality of service,

only the existence of connection.

Other services are possible (RFC2215-6)

Defined signaling (RFC2205,2210): RSVP

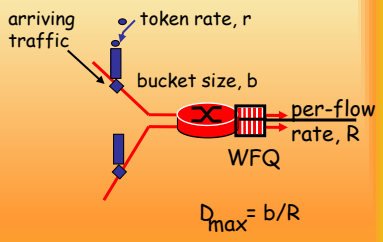


45

45

Control model for IntServ

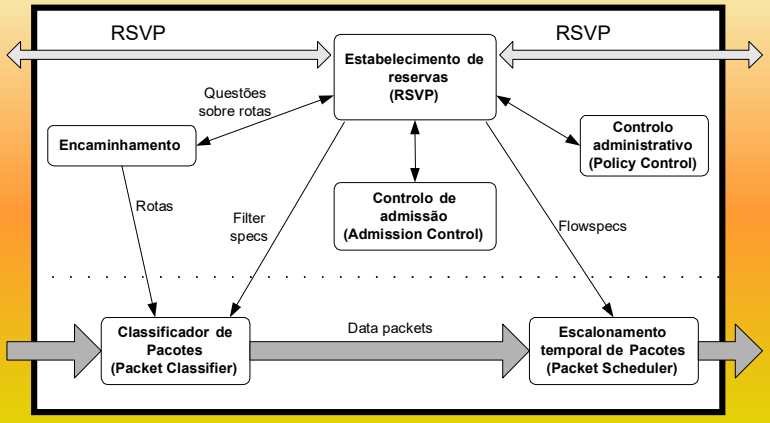
- Flow specification
  - Intserv is used for flows!
- Routing
- Admission control
  - Sender should control the sending of packets using a token bucket model
- Policing
- Resource reservation
- Packet Scheduling



46

46

Router architecture



47

47

Call admission process

The session that starts should:

- declare its QoS requirements
  - R-spec: defines the QoS that is being required
- characterize the traffic that will send to the network
  - T-spec: defines the characteristics of the traffic

A signaling protocol is required to carry the R-spec and T-spec towards routers (where the reservations are needed):

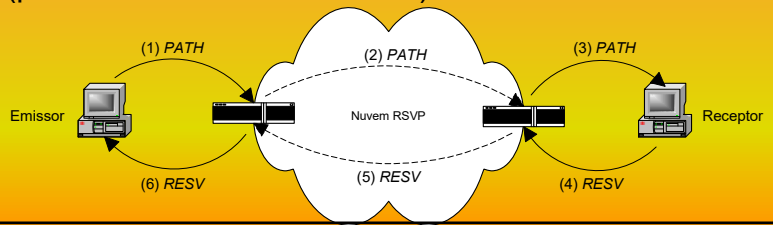
- RSVP [RFC 2205]

48

48

RSVP (*Resource Reservation Protocol*)

- RFC 2205
- Encapsulated in IP; protocol type = 46 (0x2E)
- Signalling is based on the PATH and RESV message exchange
  - PATH announces the traffic characteristics of the sender
  - RESV confirms the reservations, initiated by the receivers
  - If the reservation is not possible, the message RESV ERR is sent
- The states of routers must be refreshed periodically (process know as: soft states)



49

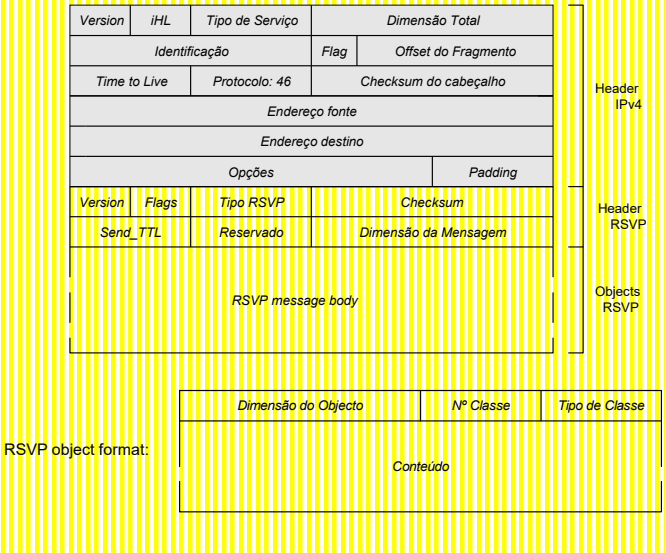
Service template: generic parameter format

- There are generic message parameters defined for IntServ operation
  - NON_IS_HOP (flag): node does not support Intserv
  - NUMBER_OF_IS_HOPS: counter of QoS-aware nodes
  - AVAILABLE_PATH_BANDWIDTH: available bandwidth (Adspec)
  - MINIMUM_PATH_LATENCY: path delay (Adspec)
  - PATH_MTU: MTU maximum transfer unit size possible to use.
  - TOKEN_BUCKET_TSPEC: traffic specifications as token bucket parameters
    - r (rate), b (bucket size), p (peak rate)
      m (minimum policed unit), M (maximum packet size)

50

50

RSVP Messages



RSVP object format:

51

51

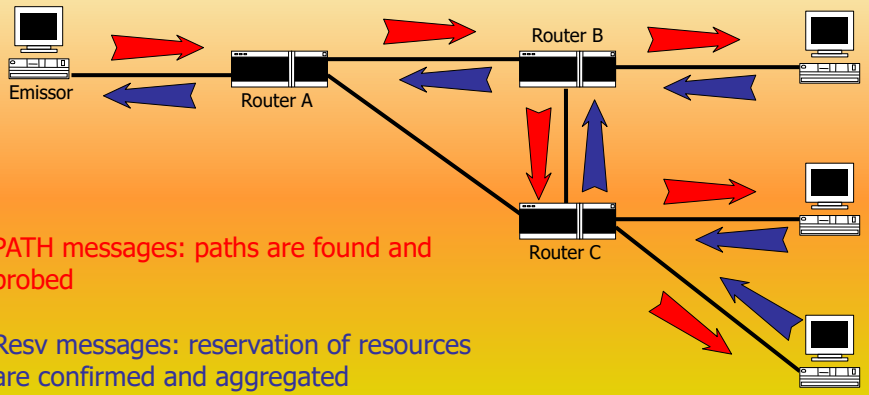
RSVP operation

- Receiver joins a (multicast) group
  - Operation outside RSVP
  - Senders do not need to join the group
- Signalling sender-network
  - *Message path:* makes the sender known to routers
  - Path erasure: removes from routers the path corresponding to a sender.
- Signalling receiver-network
  - *Message reservation:* reserves resources from the sender(s) to the receiver
  - Reservation cancelling: deletes the reservations made by receiver
- Signalling network- end-system
  - *path error*
  - *reservation error*

52

52

RSVP – Basic operation



PATH messages: paths are found and probed

Resv messages: reservation of resources are confirmed and aggregated

53

53

Reservation styles supported in RSVP (STYLE)

- "Fixed Filter" (Style Option Vector = 0x00000A)
  - Receiver specifies a reservation value per each sender
- "Wildcard Filter" (Style Option Vector = 0x000011)
  - Receiver sets a single reservation value to receive info from any sender
- "Explicit Filter" (Style Option Vector = 0x000012)
  - Receiver specifies a list of senders from which to receive information, and a single reservation number to receive traffic from those identified senders.

In RSVP RESV:

- The reservation style is set by the object STYLE
- Senders are identified by the object FILTER_SPEC

54

54

Flows

- A flow is a set of packets related by some reasons
  - In RSVP a flow is a set of packets that cross a Network Element (NE), and that are covered by the same QoS request.
- A Packet Classifier sets which packets belong to each flow
  - IPv6: Flow label helps this classification
- In ISPs…
  - Microflow: TCP or similar connection…
  - Macroflow: Large set of packets between two NEs

Flowspec define the traffic parameters

- LB, buffering needs, using token bucket specs

Filterspec identify the packets in the flow

- Basic Filter: Source, Dest address/port pair
- Advanced data filter: depends on packet content

55

55

Service models for Intserv

- There are service models that describe the semantics of service for the flow.
- Specify how the packets belonging to a flow are to be treated by the network elements.
- Parameters: general format:
  - <service_name>.<parameter_name>
    - Can have values between [1, 254]
- Services:
  - TSpec: specify the traffic pattern (CL+GS)
  - Rspec: specify the service request (GS)

56

56

IntServ General definitions

- Token bucket (rate, bucket-size):
  - Used to define data rate
- Admission control:
  - Verification before accepting a reservation
- Policing:
  - Verify if TSpec is fulfilled
  - The packet treatment may be changed if Tspec is violated (e.g. service degradation, packet discard, etc…)
- Parameters: locals and composed
  - Total path value is the combination of the local values with the path

57

57

Controlled Load

- Service that provides QoS similar to what exists in unloaded network "Best-effort in unloaded environments":
  - Statistical guarantees
  - No delay limitation
- Motivation
  - Support of applications sensitive to large delays
  - Keeping minimal functionalities
- Operation
  - Average delay in queues small or null
  - Small or null losses.
  - Analysis period: much larger than a burst period

58

58

Guaranteed Service

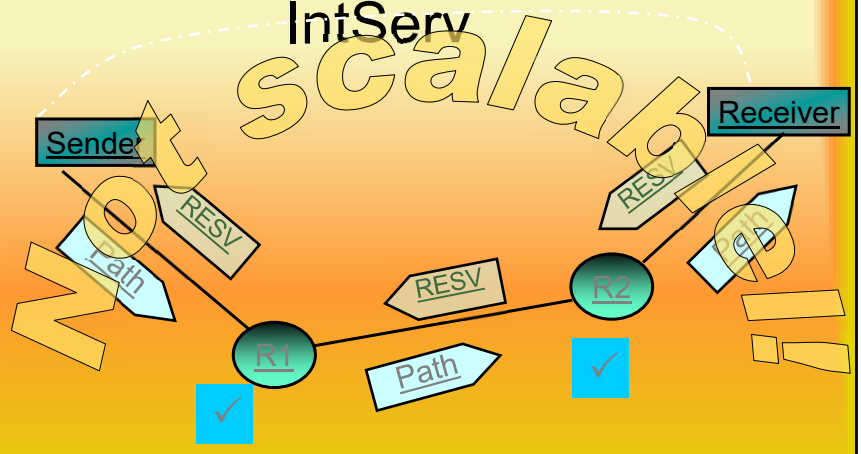
Service providing na assured bitrate, with a limit on total delay
- Deterministic guarantee
- No jitter guarantee

- Parameters used:
  - TSpec: TOKEN_BUCKET_TSPEC
  - RSpec: R (rate), S (delay slack term, μs)
    - Larger R: smaller E2E
    - Larger S: larger delays, but better reservation possibilities
- Admission control:
  - Weighted Fair Queuing (WFQ)
- Policing:
  - drop, move to best-effort; reshape (delay)

59

59

Resource Reservation in IntServ



RSVP!
Works with individual flows!

60

60

Example – Resource Reserve in IP Networks

- RSVP reservations for voice and video services over IP
  - Low quality: Reservation of 64 kb/seg for voice and 1 Mb/seg for video
  - High quality:
    - IntServ Guaranteed Service for voice with 1Mb/seg bandwidth
    - IntServ Controlled Load for video with 5 Mb/seg bandwidth

    Or
    - IntServ Guranteed Service for voice and video with 6 Mb/seg

61

61

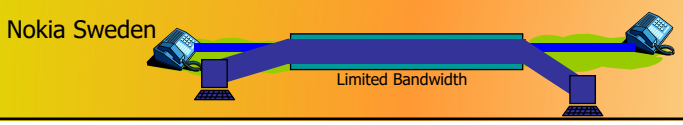
Differentiated Services (DiffServ)

62

Basic idea

The real question is to choose which packets shall be dropped. The first definition of differential service is something like "not mine."
-- Christian Huitema

- Some packets are more important than others
  - whoever "pays" for better service, should get it...
- Differentiated services should provide a mechanism to specify relative priority of packets
  - Implement simple routing operations on the network's core routers and leave complex operations to the network edge routers.



63

63

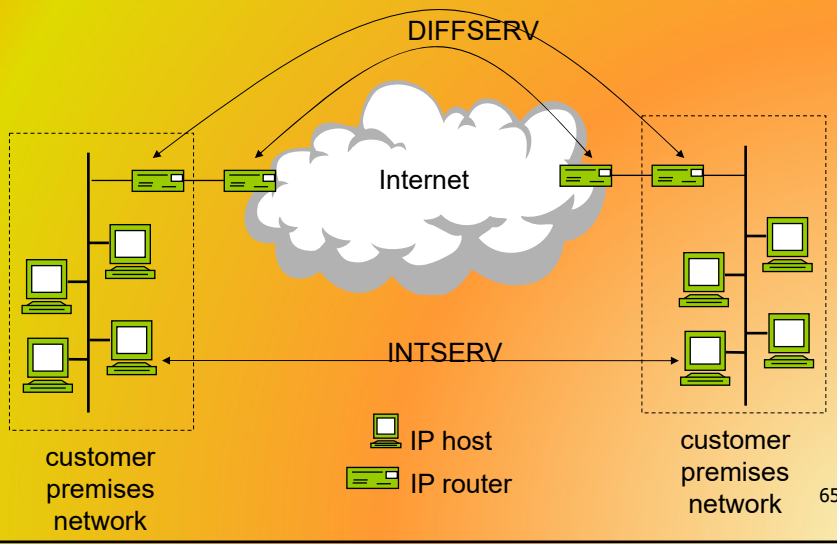
Objectives

- Ability to charge differently for different services
- Ability to discriminate services in a scalable way for the core, with low complexity
  - No per-flow state
  - No per-flow signalling
- Easy to evolve, with simple start-up implementation
  - Define only elements that may implement any class of service
- Simpler and more efficient than IntServ
  - With signalling separated by services
  - With “more-or-less” static user-services
  - With traffic aggregated in classes
- Without individual reservation per link

64

64

DIFFSERV Scope



65

65

DiffServ

- Oriented towards the core network (*core*)
- No direct E2E guarantees: service assurances are structured by layers
- No per-flow control
  - Packets marked by the network (not the app)
  - All existing applications can be supported.
- Simple control and marking tools (RFC2474)

⇒ no E2E user service model (RFC2475)

⇒ More proper to speak of CoS than of QoS

⇒Simpler to implement in core than IntServ

⇒Can be deployed in current networks!

66

66

DiffServ approach

Based in three assumptions (implicitly assumed):

- The network is overprovisioned regarding the needs of QoS traffic (⟹ small number of customers with DiffServ)
- Non-real-time traffic will be the largest load of all network traffic
  - No explicit reservation per link, and as such links in “popular” areas could have problems with congestion (when more people would like to use prioritize traffic)
- E2E services can be implemented over networks with different QoS features

67

67

Services and SLA

- Two types of services:
  - quantitative; require numerical metrics and information about entry and exit points
  - qualitative; require only entry information
- Not (necessarily) per-flux:
  - Packet treatment is performed over aggregates of a "source"
  - A service level specification (SLS) is defined
  - The service is provided by the network (and the host may not even perceive this)
- All services are unidirectional!

68

68

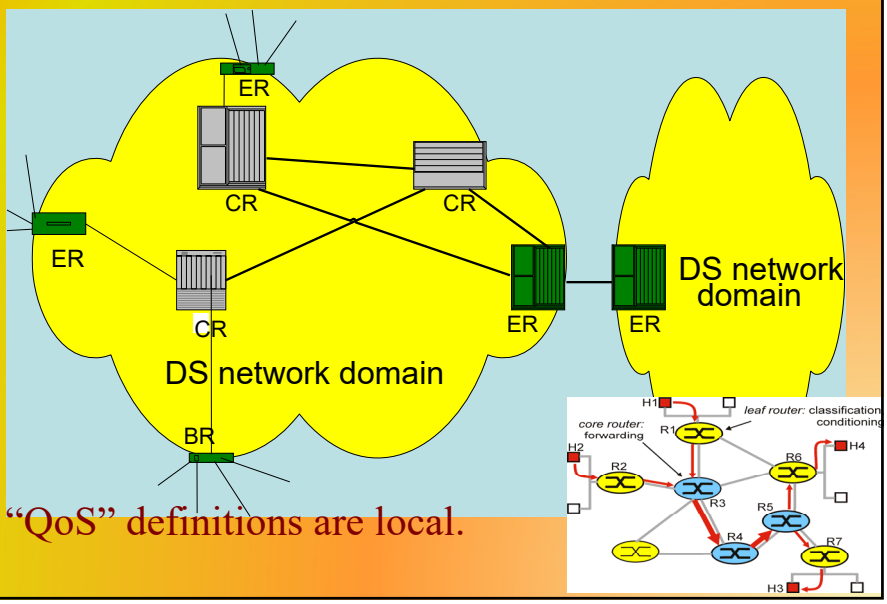
Architecture components

1. Define PHB (per-hop-behaviour) for the routers. These will operate over traffic aggregates. The PHB priotize traffic, in terms of delay and loss probability.
2. Traffic control and management is fundamentally performed at the borders, and then is aggregated.
3. Services are clearly separated from the network technical constrains.

69

69

Overview of a DiffServ network



70

Functional elements

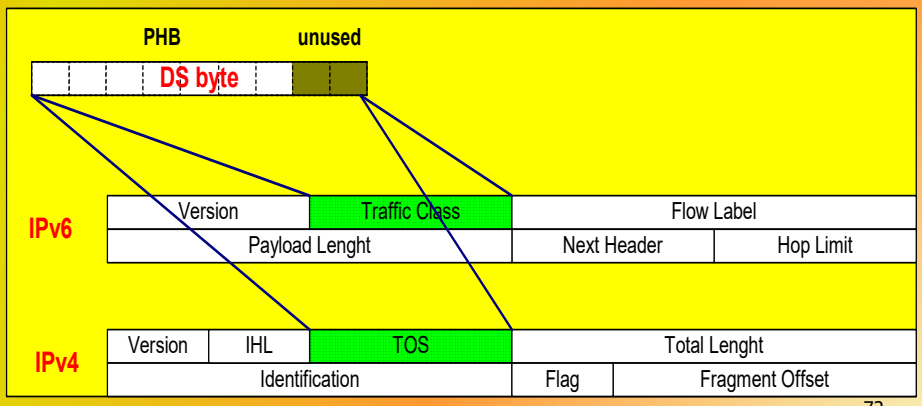
- *Edge (border) Routers:*
  - *Classify packets: Mark each packet in the Type of Service field of the IP header*
  - *Condition traffic: for example, they use a "Token Bucket" to verify that incoming traffic is contracted and*
    - *delay excess traffic or*
    - *drop non-conformant traffic*
- *Core (internal) Routers:*
  - *Identify the treatment to give to packets based on marking and according to a defined Per-Hop-Behavior (PHB)*

71

71

Field DS

Packets are marked in the Type of Service (TOS) field of the IPv4 header or Traffic Class of the IPv6:
- DSCP header - Differentiated Service Code Point, 6 bits
- rest - Currently Unused



72

72

PHB

- Per-Hop-Behaviour is the forwarding behaviour that a DS node applies to a traffic aggregate. This is perceived in terms of:

Delay and

Packet loss probability

- The specification of transmission rate, losses and delay should be done in the SLS
- Any PHB is only defined inside an administrative domain

73

73

QoS Processing in *Core Routers*

- Different Per-Hop-Behaviors (PHBs) result in different network performances that can be measurable
- PHBs do not specify which queuing mechanisms should be used
- Examples of PHBs:
  - Class A packets are assigned x% of the physical connection bandwidth during any time interval for a specified duration
  - Class A packets are always served first than Class B packets
  - Class A packets are served with twice the service bandwidth of Class B packets

74

74

DIFFSERV PHBs

- Two types of PHB already developed:
  - AF (Assured Forwarding) (RFC2597):
  - EF (Expedited Forwarding) (RFC2598):
    - virtual leased line (VLL) service

+BE (best effort)

- Coupled to different types of services:
  - Premium (low delay) - EF
  - Assured (high transmission rates, low losses) - AF

75

75

DiffServ service classes

- *Default* (DE) → DSCP = 000000
  - *best-effort* service with a single queue, FIFO managed
- *Expedited Forwarding* (EF) → DSCP = 101110
  - "Virtual leased line" service
  - provides loss, delay, and delay variance control within a given maximum bandwidth
- *Assured Forwarding* (AF) → DSCP = aaadd0
  - provides a relative Quality of Service (AFi is served with more bandwidth than AFj for i<j)
  - in each class there are 3 precedence levels for packet deletion in case of congestion

AF Codepoints	AF1	AF2	AF3	AF4
Low drop precedence	001010	010010	011010	100010
Medium drop precedence	001100	010100	011100	100100
High drop precedence	001110	010110	011110	100110

76

76

Expedited Forwarding (EF)

- For critical traffic
  - Low delay, small jitter, no losses
- Nodes must forward these packets ASAP (PQ).
- Packets cannot be lost, or reorder
  - Resources must be reserved in a conservative way, by the maximum value.
  - Agreed bandwidth is assured
  - Packets out of profile are lost: stringent policing in the border
- EF can block all other network traffic.
- Defined for quantitative services, as it require entry and exit nodes well defined.
  - VLL: maximum LB defined, available when required

77

77

Assured Forwarding (AF)

- Defines four classes, and three levels of packet loss for each class.
  - AF11 - “best”, AF13 - “worst”
- Relations between classes are not defined
  - Provisioning according with the expected usage
- Performance in each class should be degraded gradually (3 levels) in terms of packet loss, as traffic increases.
  - Packets inside profile will not be “usually” lost
  - Packets out of profile may be treated (almost) as BE, so higher bandwidths may be used if available
- Allows qualitative services, and only requires the knowledge of entry-node
  - Bandwidths are roughly respected

78

78

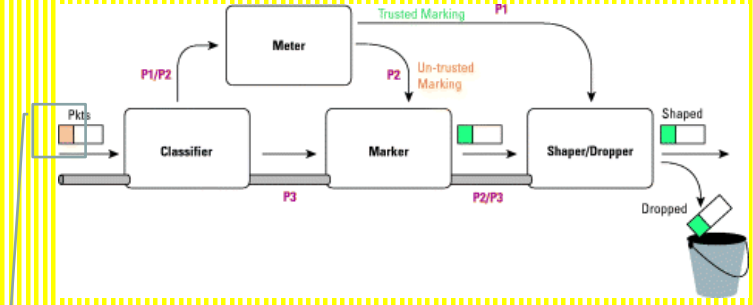
Borders: *Edge Routers*

- Control network access, policing and classifying entry traffic.
  - These is also required between DS networks, as diffserv definitions are local...
- Traffic may be in or out of profile.
- The network requires traffic conditioners at the borders (optional at *core*), that classify and act over traffic:
  - Meters – check the timing features of the flow, confronting with the SLA associated
  - Classifier – identifies the traffic class of the (for the) packet
  - Markers – set a DS codepoint to each packet (in/out profile)
    - Packets may be remarked
  - *Droppers* – remove packets out-of-profile
  - *Shapers* – delay packets out-of-profile, using buffers and smoothing methods

79

79

Border control model

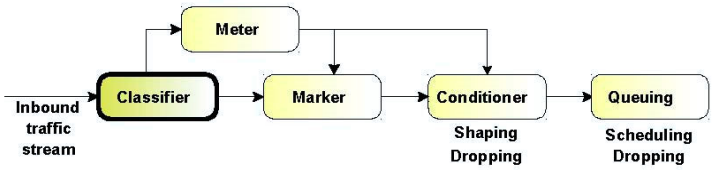


Packet may be marked already

80

80

Classification

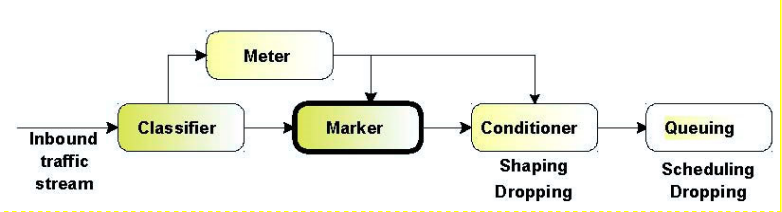


- Many traditional QoS mechanisms already include intrinsically classifiers
  - Committed Access Rate (CAR)
  - QoS policy propagations via BGP (QPPB)
  - Queuing Mechanisms
  - …

81

81

Marking

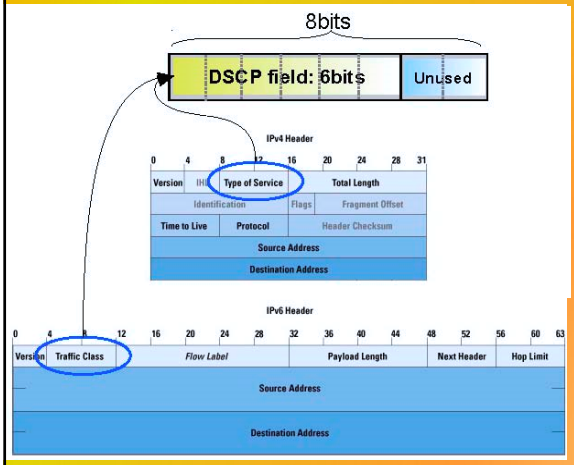


- Marking is used to set:
  - IP precedence
  - DSCP
  - QoS group
  - MPLS experimental bits
  - IEEE 802.1Q or ISL CoS
- Marking mechanisms:
  - Comitted Access Rate (CAR)
  - QoS Policy Propagation
  - through BGP (QPPB)
  - Policy-based Routing (PBR)
  - Class-based Marking

82

82

Edge Routers: packet marking

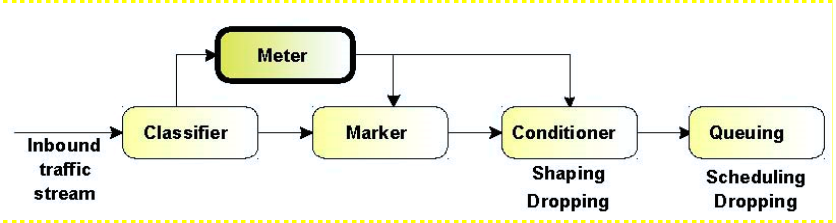


- Packets are marked in the field *Type of Service* (TOS) of IPv4 or *Traffic Class* in IPv6 DSCP – *Differentiated Service Code Point* CU – *Currently Unused*

83

83

Metering



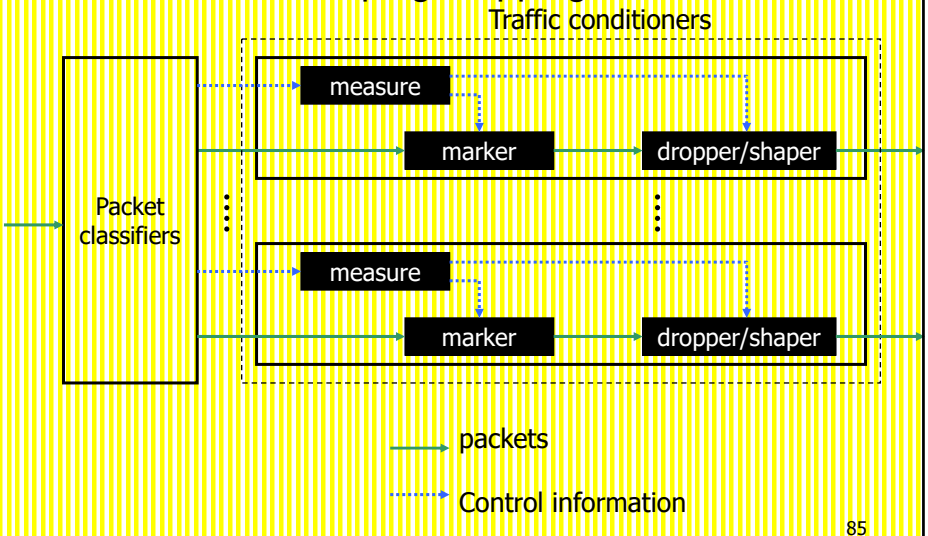
- Token Bucket models are used for metering
  - Committed Access Rate (CAR)
  - Generic Traffic Shaping (GTS)
  - Class-based Weighted Fair Queuing (CB-WFQ)
  - Class-based Policing
  - Class-based Shaping

84

84

Traffic conditioning

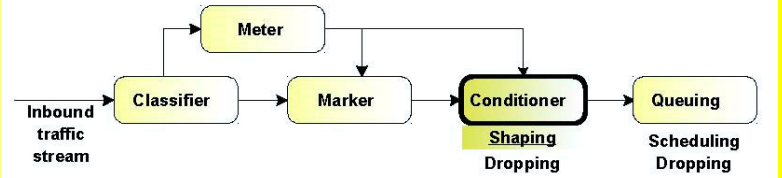
shaping/dropping



85

85

Conditioning traffic: shaping

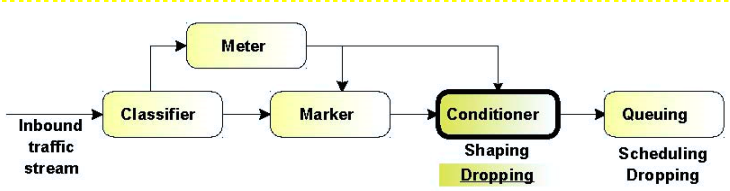


- *Traffic Shaping mechanisms*:
  - Generic Traffic Shaping (GTS)
  - Class-based Shaping

86

86

Conditioning traffic: dropping

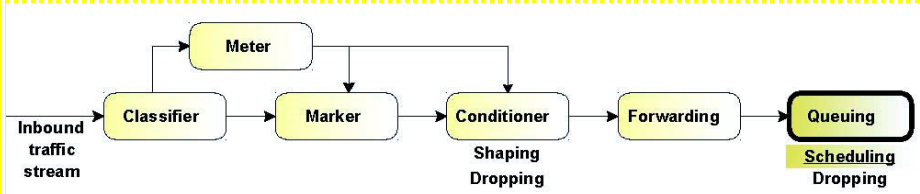


- *Dropping* traffic:
  - Committed Access Rate (CAR) and Class-based Policing may drop packets that exceed agreed rate
  - Weighted Random Early Detection (WRED) may drop packets randomly when an interface starts to reach congestion

87

87

Final step: Queuing

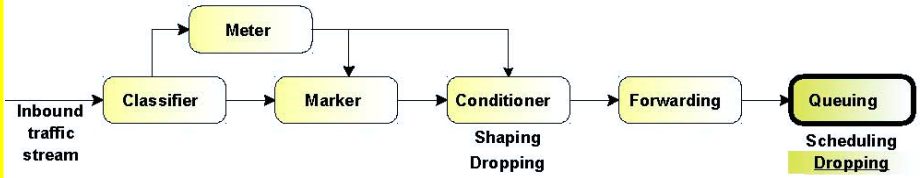


- Traditional queuing mechanisms
  - FIFO, Priority Queuing (PQ), Custom Queuing (CQ)
- Weighted Fair Queuing (WFQ) family
  - WFQ, dWFQ, CoS-based dWFQ, QoS-group dWFQ
- Advanced queueing mechanisms
  - Class-based WFQ, Class-based LLQ

88

88

Queuing

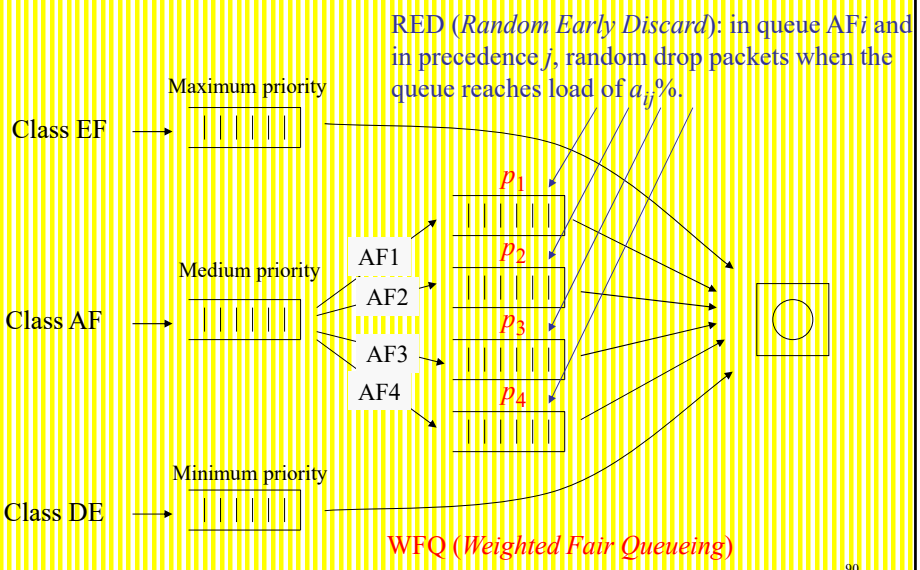


- *Dropping* mechanisms
  - *Tail drop* when there is congestion in the waiting queue
  - WFQ has an improved approach to *tail-drop*
  - Weighted Random Early Detection (WRED) may drop packets randomly when an interface starts to reach congestion

89

89

Implementation example



90

DiffServ problems

- No standards for SLAs:
  - The same DS codepoint can be used by different services, between different ISPs
  - Different networks, using the same PHB, may provide different behaviors
  - No generalized edge-to-edge semantics (PDBs: per-domain-behavior)
- Lack of symmetry:
  - Protocols like TCP work best in nearly symmetric environments
- Multicast:
  - No support for multiparty, symmetric, communications
- Network configuration, for each PHB

91

91

Multiplexing effects

- "aggregates" means that we do not see simple flows
  - Thousands of different flows reach the same core router with the same PHB
  - Different delays ⇒ variable traffic conditions

⇒QoS traffic percentage vs. Best Effort varies at each instant!!

- Total QoS bandwidth allocation must be carefully considered.
  - BW reservation for each QoS class must be assessed.
  - The buffer parameters for each queue has to be done separately.

  (RIO - RED with In and Out – is frequent in DS systems)

92

92

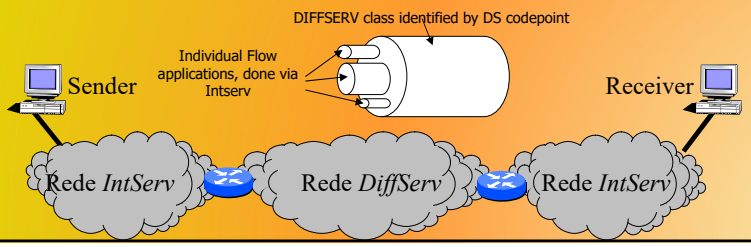
93

INTSERV vs DIFFSERV

- Complementary:
  - DIFFSERV: aggregate, per user/customer/groups/applications – oriented to the service provider
  - INTSERV: per flow – oriented to application

One can integrate:

- INTSERV reservations inside DIFFSERV "flows"
- The border routers of the two types of network:
  - classify RSVP requests in the adequate DiffServ service classes.
  - If there are insufficient resources, refuse RSVP reservation requests



93

INTSERV and DIFFSERV

	INTSERV	DIFFSERV
signaling	By the application	Network management, application
granularity	Flow	flow, source, site (aggregation)
mechanism	Endereço de destino, protocolo e número de porto	Classe de pacotes (outros mecanismos possíveis)
scope	End-to-end	Between networks, E2E

94

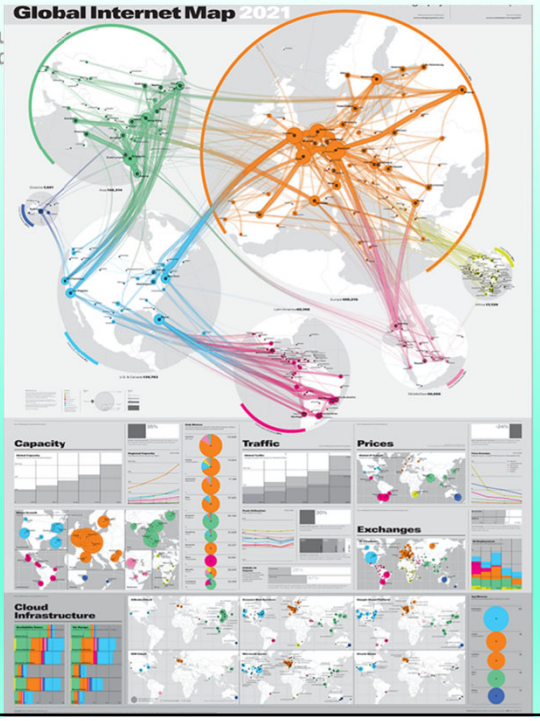
universidade de aveiro

Gestão/Management

Management of Local and Global Networks

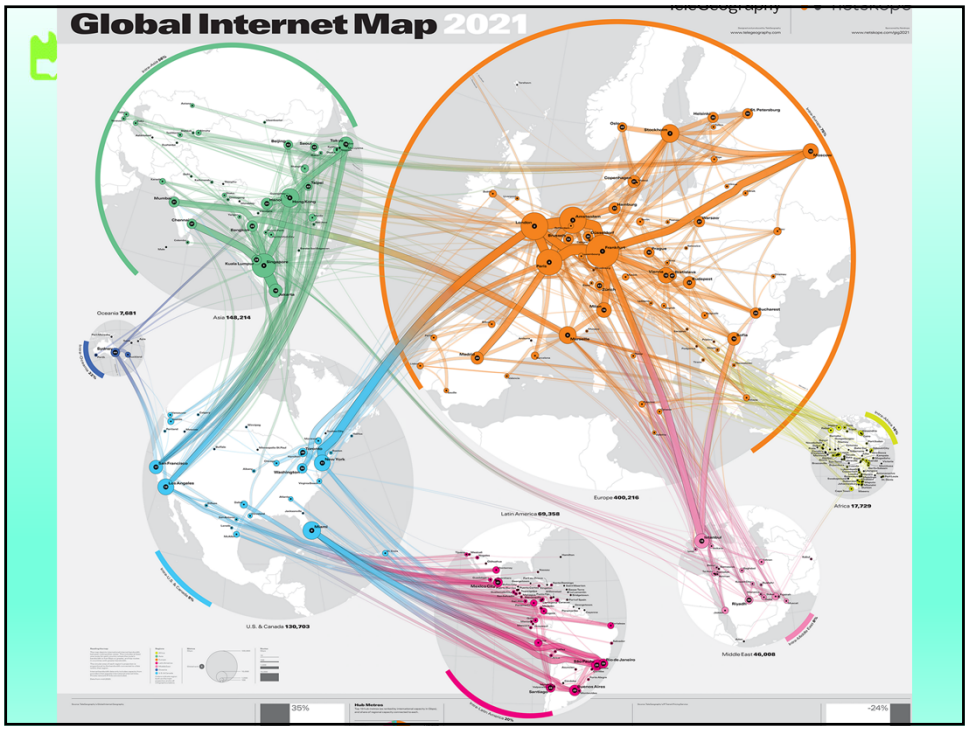
Concepts and Protocols

1

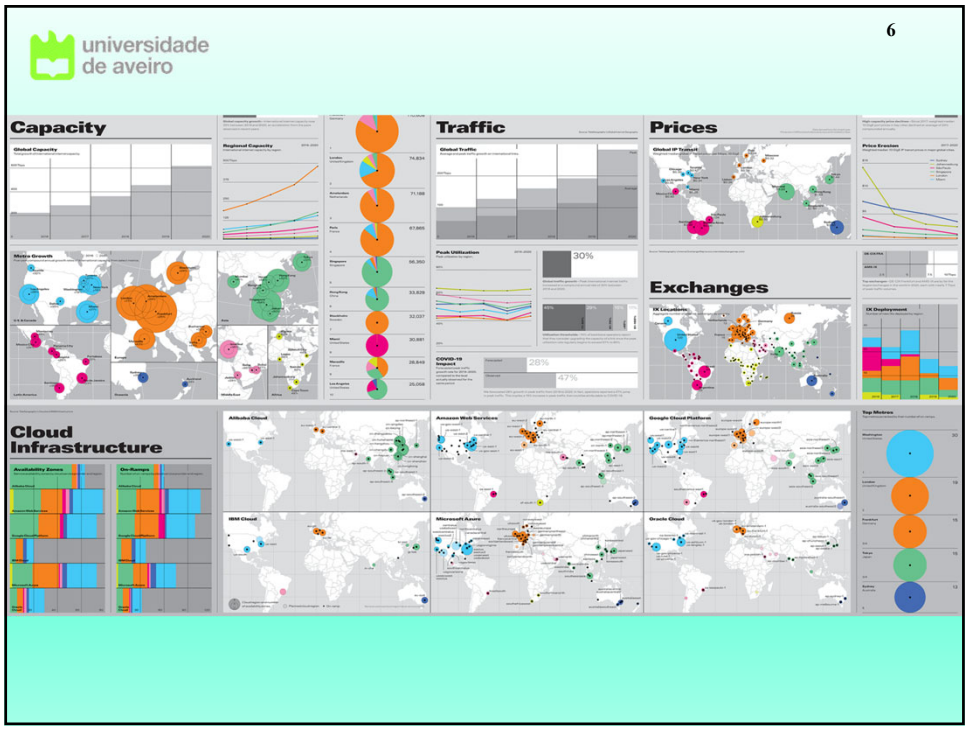


4

4


Global Internet Map 2021
Asia 140,214
Europe 400,216
Latin America 69,358
U.S. & Canada 130,703
Middle East 46,008
Africa 17,729
Oceania 7,681

5


universidade de aveiro
6
Capacity
Traffic
Prices
Exchanges
Cloud Infrastructure

6

universidade de aveiro

Why Networks and Systems Management?

7

- Lower Cost – Manual management is costly
- More efficient – Automatic systems allow an efficient planning, and mechanisms to predict the utilization trends: lower errors and faster actuation
- Better service – The manager is informed at the same time the (client) is, and can make an automatic check of the situation
- Greater knowledge – more information exists about the network, allowing better decisions and planning
- Why not human intervention?
  - **Difficult to describe responsabilities**
  - **Technology rapidly evolves**
  - **Mangement systems rapidly evolve**
  - **Lack of technical resources**

7

universidade de aveiro

Commercial perspective

8

- **Problems need to be quickly solved**
- **Management systems simplify the work of multi-functional networks (e.g. VoIP in multiple networks)**
- **Persons better used – they do not need to perform repetitive tasks**
- **Companies need to optimze their structures, and network management allow resources optimization**

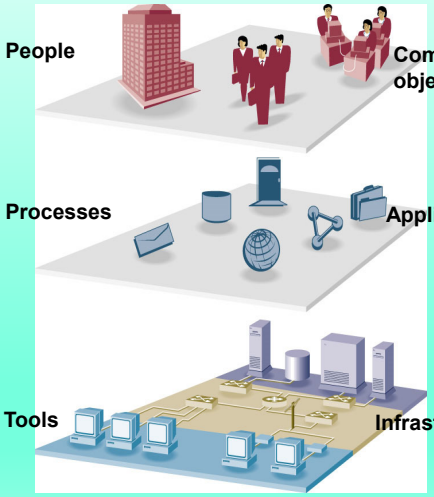
8

universidade de aveiro

9

Network management is:

Implement, integrate and coordinate resources (HW, SW and people) to plan, operate, manage, analize, test, evaluate, design and expand the system to guarantee the service objectives (temporal, performance), with a reasonable cost and capacity.

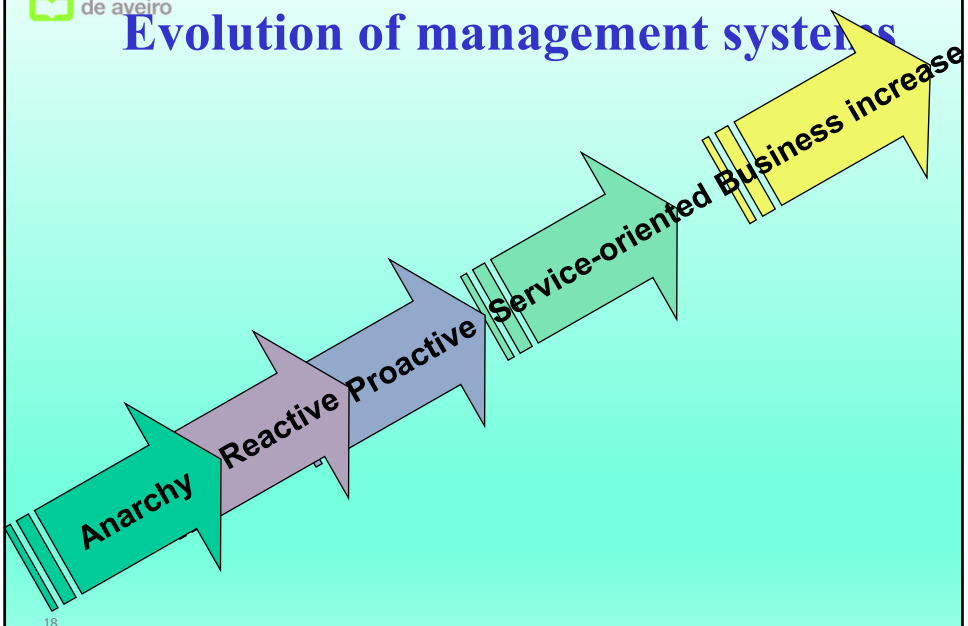


9

universidade de aveiro

10

Evolution of management systems



18

10

universidade de aveiro

Management alternatives

scope

- **Systems management** – Covers all company aspects
- **Networks management** – Covers mainly network aspects and communications systems and equipment

comunication protocol

- **Dedicated protocols** – dedicated for networks
- **Web based systems** – resort to HTTP models, recently common

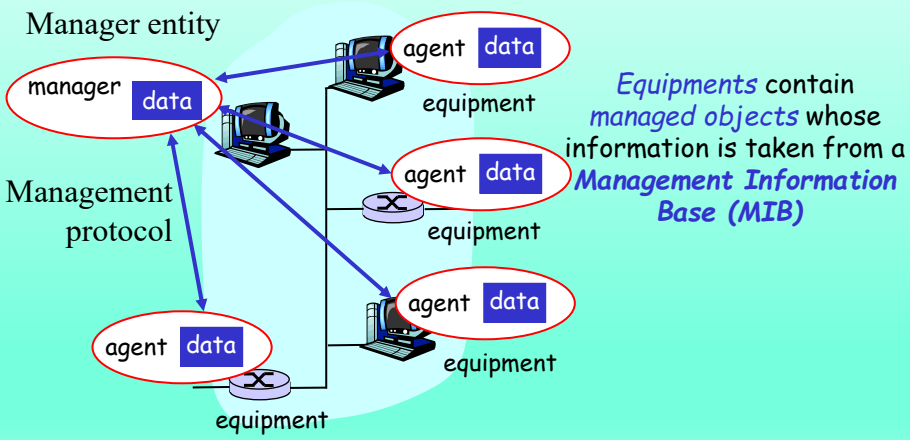
Decision model

- **Centralized models** – Agent-manager model
- **Distributed models** – Share of the management responsibilities
- **Hierarchical models** – Hierarchic structure with centralized information in the root

Current real management structures very complex, with several operational models simultaneously

universidade de aveiro

Basic Model for Network Management

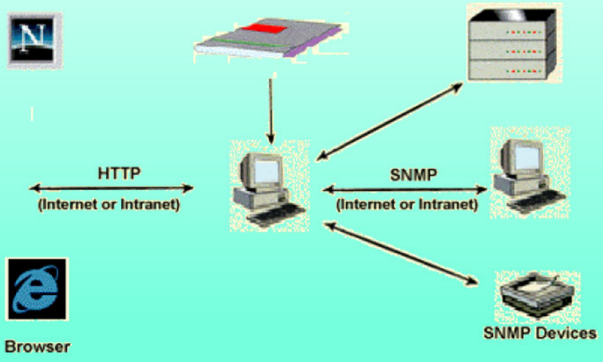


Ver SNMP

13

universidade de aveiro

WEB-based management



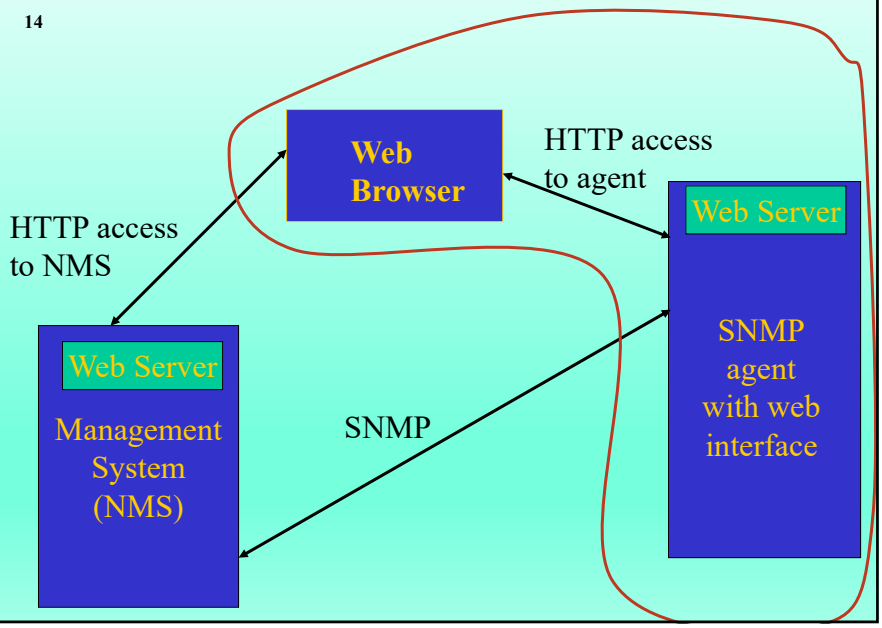
Very common:
Network and device management via web interfaces

13

universidade de aveiro

Web-based management concept

14



14

universidade de aveiro

15

Network management

- ISO defined five areas for network management
  - Fault management – detection, isolation, and correction of anomaly behaviors

FAULT

  - Configuration management – control data for the network elements / colect data from network elements

CONFIGURATION

  - Accounting management – measure network utilization and determine network costs and user accountings

ACCOUNTING

  - Performance management – evaluate/report network equipment behavior/efficiency

PERFORMANCE

  - Security management– support communications network secure management

SECURITY

15

universidade de aveiro

16

Network management

- ISO defined five areas for network management
  - Fault management – detection, isolation, and correction of anomaly behaviors

FAULT

  - Configuration management – control data for the network elements / colect data from network elements

CONFIGURATION

  - Accounting management – measure network utilization and determine network costs and user accountings

ACCOUNTING

  - Performance management – evaluate/report network equipment behavior/efficiency

PERFORMANCE

  - Security management– support communications network secure management

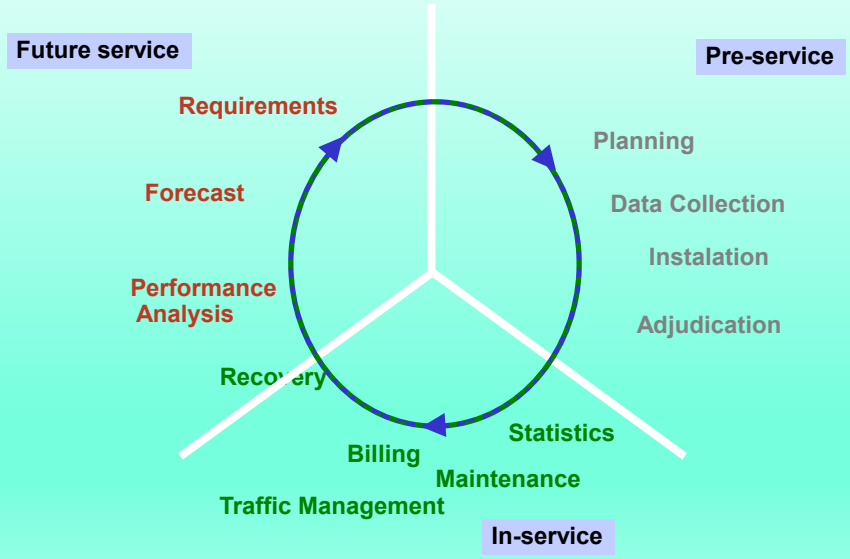
SECURITY

Isolating the management problems in distinct areas, ISO model allows conceptual solutions optimized to specific problems in each functional area

16

universidade de aveiro

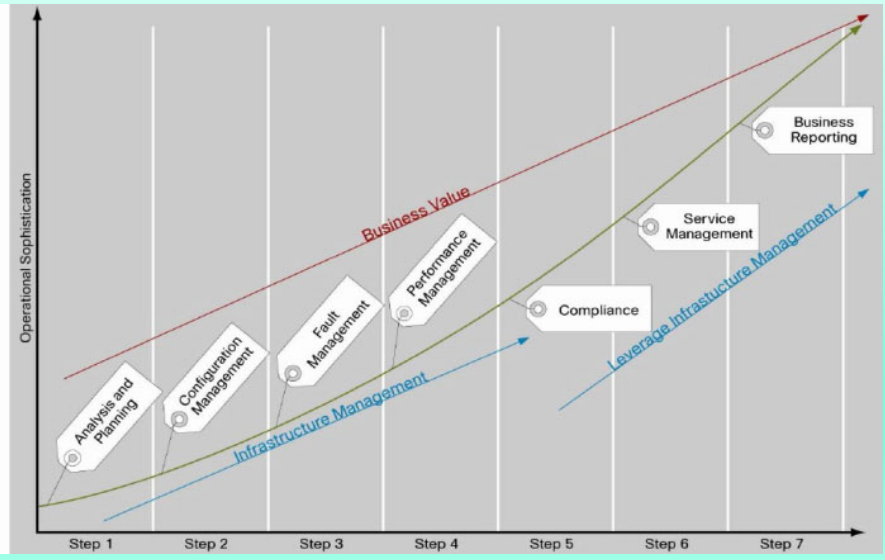
Management Life Cycle



17

universidade de aveiro

Implementation plans



18

universidade de aveiro

19

Aspects of Network Management

- What to manage?
  - Network, equipment, systems, users, services, applications
- How to manage?
  - Interfaces, actions, abstractions
- What protocol(s) format(s)?
  - Protocol abstraction, formats, messages
- What information format(s)?
  - Information type

Standards for all this – including global frameworks

19

universidade de aveiro

20

Management protocols

- Methods to monitor and configure network equipments
- Do not describe how to achieve management objectives

Simple protocols ⇒ common data and parameters formats allowing easy information transfer

Complex protocols ⇒ add flexibility and security capacity

Advanced protocols ⇒ remotely execute network management tasks, without depending on specific protocol layers

20

universidade de aveiro

21

Tools for network management

- WAN/LAN monitoring and analyzers
- Software monitors
- Security managers
- Documents, presentations and administrative instruments
- Tools for cross-analysis
- Databases, tools for information management
- Console emulator
- Tools for systems modelling
- Toolkits for development

21

universidade de aveiro

22

Network management standardization global models

- Internet Engineering Task Force (IETF)
  - Simple Network Management Protocol
    - SNMP, disman
    - *Operations and Management Area*
- International Telecommunications Union (ITU-T)
  - Telecommunications Management Network
    - *SG IV*
- International Standard Organization (ISO)
  - OSI, CMIP-CSIS
    - *ISO-IEC/JTC 1/WG 4*
- Others
  - DMTF, TM FORUM, OMG, IEEE, …

Early discussions across bodies. Now cooperation is the normal across bodies.

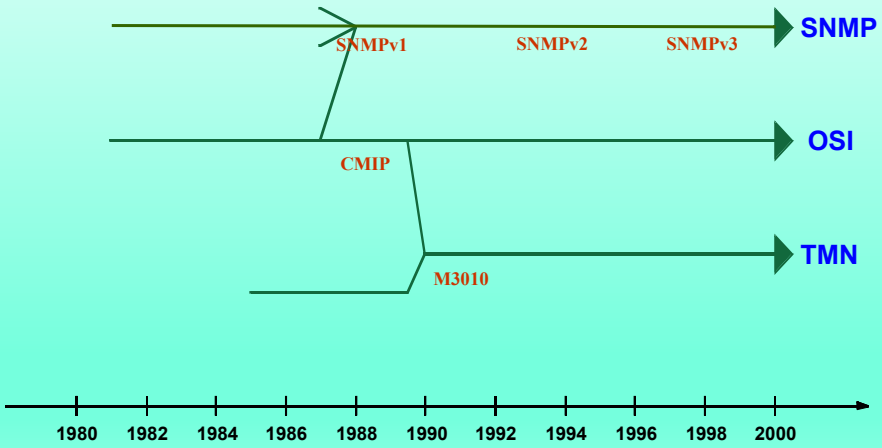
22

universidade de aveiro

23

Early Cronology

(why we reached here)



Note that concepts are different: SNMP copes with part of the OSI management framework, and the TMN expands the OSI framework into a focused environment (telecomunication operators)

23

universidade de aveiro

SNMP

(short review)

Simple Network Management Protocol

24

universidade de aveiro

25

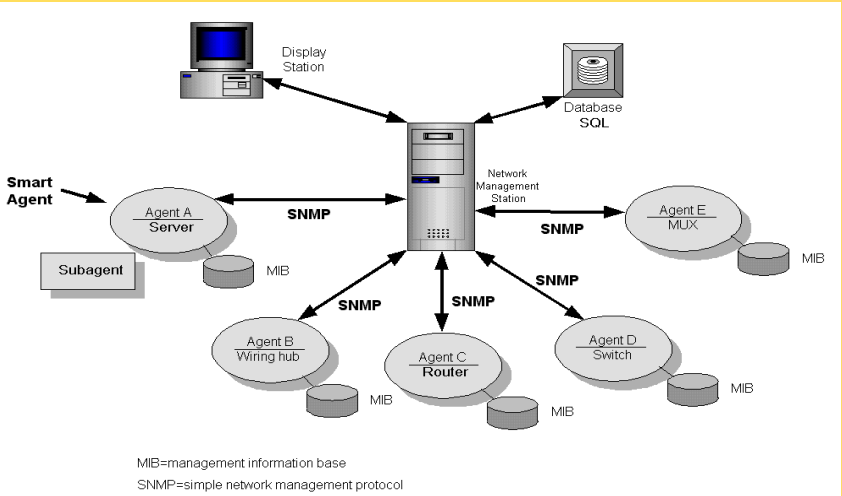
Manager/Agent Paradigm

- Manager/agent: common in all NMS (especially in SNMP/CMIP)
- Idea of a client/server, but many clients and only some servers
  - (manager ↔client; agent ↔server)
- The agent operates with the equipment
  - Reports problems to the manager, to control all the equipment information
- The manager contains the intelligence to decide what the agents should do, and gives instructions to them
  - It controls the agents and manages their interworking

universidade de aveiro

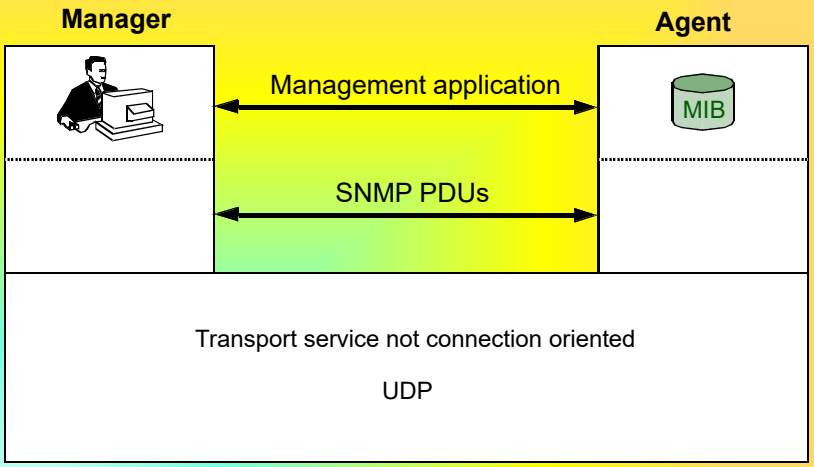
26

Structure of SNMP management



universidade de aveiro

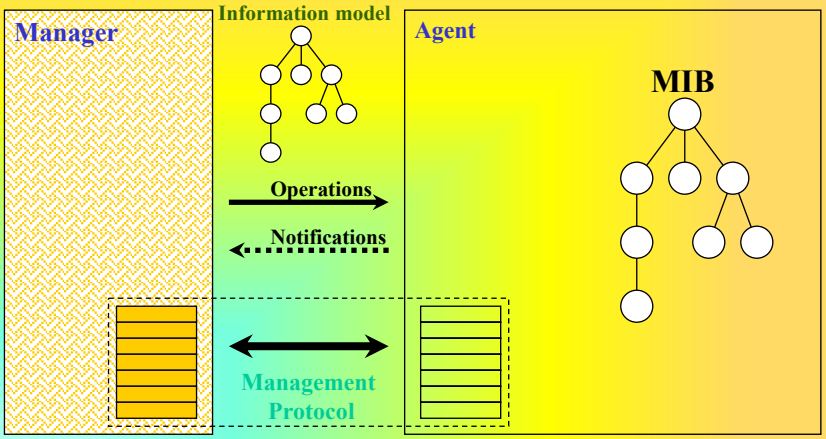
SNMP Structure



27

universidade de aveiro

Manager/Agent model



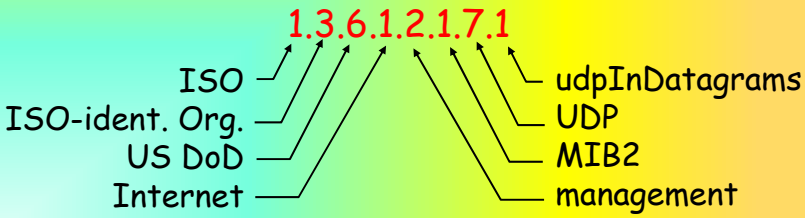
28

Names (numbers) SNMP

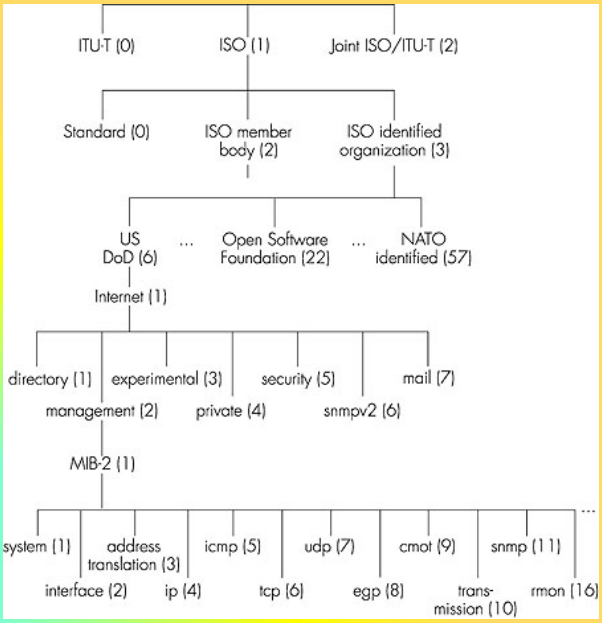
Problem: **How to name all possible objects (protocols, data, etc..) in all possible protocols??**

Answer: ***ISO Object Identifier tree:***

- **Hierarchical naming for objects**
- **Each node has a name and a number**



OSI Object Identifier Tree



www.alvestrand.no/harald/objectid/top.html

universidade de aveiro

SNMP: Polling

- Manager periodically asks the agent for new information
- ☺ Advantage: Manager completely controls the equipment, and knows all network details
- ☹ Disadvantage: delay between event and its entry in the system, and unnecessary communication overhead:
  - Slow polling, slow answer to the events
  - Quick polling, quick reaction, but large bandwidth wastage

31

31

universidade de aveiro

SNMP: Traps

- There is an event ⇒ trap is sent
- Trap contains appropriate information
  equipment name, time instant of event, type of event

☺ Advantage: information only generated when required

☹ Disadvantage:
- ☹ More resources required in the managed equipment
- ☹ Traps can be useless
  - If many events occur, bandwitdh can be wasted with all traps (thresholds can solve)
  - Since the agent has only a limited scope of the network, NMS may already know about the events.

- Traps&Polling
  - Event occurs ⇒ trap is sent
  - Manager performs polling to obtain the rest of information
  - Manager also performs periodic polling, as backup

32

32

SNMP Protocol: types of messages

Types of messages	Function
GetRequest GetNextRequest GetBulkRequest	Mgr → agent: "get me data" (instantiates, next on the list, block of information)
InformRequest	Agent → Mgr: informs the Manager of exception in a reliable way
SetRequest	Mgr → agent: defines MIB value
Response	Agent → mgr: answer value to Request
Trap	Agent → mgr: informs the manager of an exception event

33

33

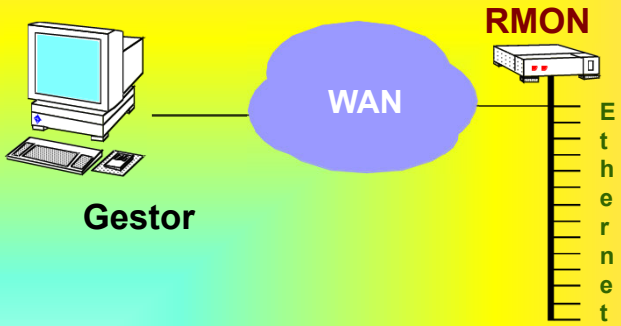
SNMP: security and authentication

- In its initial version, the authorization and authentication were based in the notion of "SNMP community string"
- The "community words" identifying the permissions of the machine that access the agente: read-only ou read-write
- By default, all systems come configured with the strings:
  - public (read-only)
  - private (read-write)
- These strings are case sensitive.

34

34

REMOTE MONITORING



- RMON1 (RFC 1757)
- Token Ring extensions to RMON (RFC 1513)
- RMON2 (RFC 2021)
- SMON (RFC 2613)

35

RMON

- **Remote monitoring MIB – measure network traffic**
  - **Agents – management interface**
  - **Probes – equipment for network analysis (promiscuous); usually configured to specific data types.**
- **Off-line operation (separated from the network)**
- **Preemptive monitoring, providing multiple information in the network.**
- **Support multiple managers and probes**
- **Detection and report of problems**
- **RMON has 9 groups:**

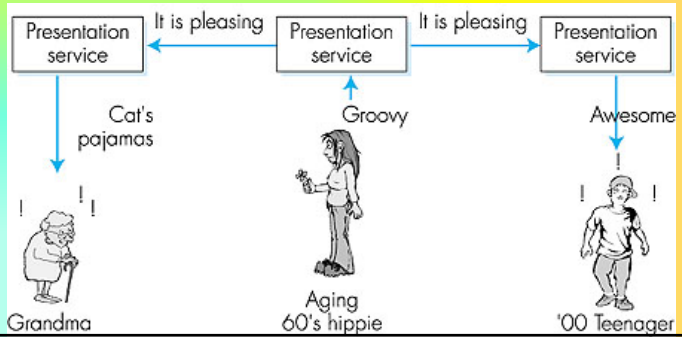
  Statistics, History, Alarm, Host, HostTopN, Matrix, Filter, Packet Capture, and Event

36

universidade de aveiro

The presentation problem?

1. Translate the local format to a host-independente format.
2. Transmit the data in a host independente format
3. Translate the host-independente format in a format adequate to the new machine adequado à nova máquina.



37

universidade de aveiro

ASN.1

- ISO X.680 standard
  - Formal language to describe SMI
  - Frequent in Internet
  - "Heavy", but essential for heterogenous environments.
- Data types, object constructors
  - As in SMI
- BER: Basic Encoding Rules
  - Specified the format as ASN.1 data should be transmitted.
  - Each transmitted object has a coding Type, Length, Value (TLV) encoding

38

TLV Coding

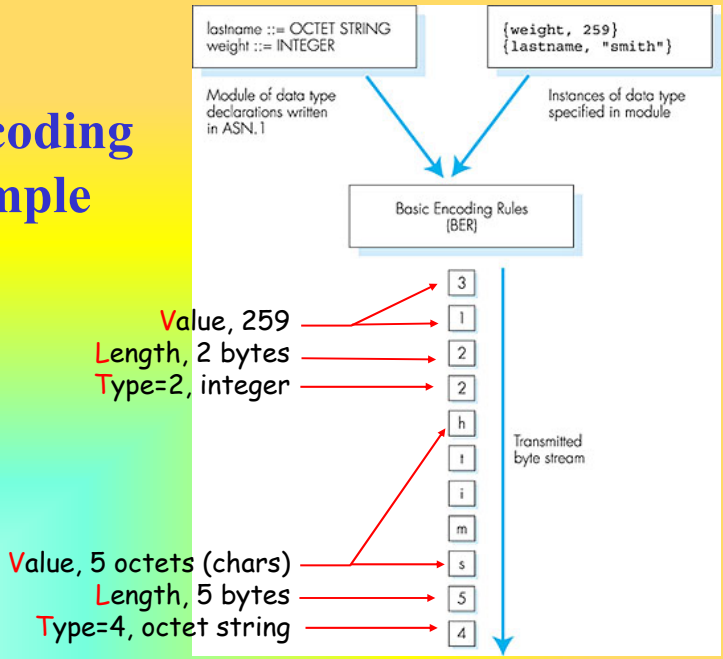
Idea: **Data must be auto-identified**

- **T: data type, (ASN.1-defined)**
- **L: data lenght in bytes**
- **V: data, coded according with ASN.1 syntax.**

Valor Tag Tipo

1	Boolean
2	Integer
3	Bitstring
4	Octet string
5	Null
6	Object Identifier
9	Real

TLV coding example



universidade de aveiro

41

SNMP: Pros and Cons

- Agents widely used/known
- Simple to implement
- Robust e extensible
- Polling approach adequate to LAN objects

Critical requirement satisfied: available to be developed in the right time

- Very simple: does not scale
- Specifc semantics make its integration with other approaches difficult
- Large communication overhead due to polling
- Many specific implementations (private MIBs)
- In several management systems, small agents may be inadequate

Note that SNMP became a misnomer, referring both to the management protocol and the management framework. These are different things.

41

universidade de aveiro

PBM and COPS

Concept: Policy Based Management

Protocol: Common Open Policy Service

42

universidade de aveiro

43

Policies - Example

- **Network with multiple services support**
  - **Differentiated QoS**
  - **Additional requirements in AAA functions**
    - Different levels
      - User
      - Service
      - QoS
- **Service authorized**
  - **only to some users**
  - **between authorized network points**
  - **with specific QoS requirements**
  - **between specific time intervals**
- **User also needs to be charged according to the service characteristics being received**

43

universidade de aveiro

44

Management based on Policies

- **Objective: globally manage the network and not its elements.**
- **Mechanism:**
  - **Define policies (rules) to inform the network of what to do – e.g:**
    - Operation center should have access to all routers
    - Charging department has priority in the last 3 months of each year
    - In the maximum, only 10% of each link can transport video.
  - **The policy rules are translated in equipment configuration changes**

44

45

Elements of systems based on policies

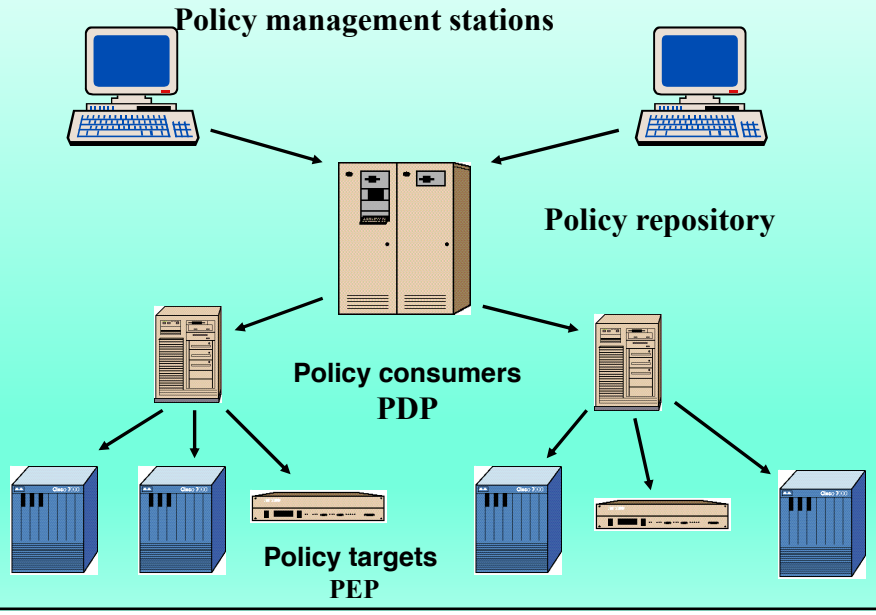
Conceptual parts:

- **Management policy tools:**
  - Used to create the policy rules
- **Policies repository**
  - Store the policy rules
- **Policy consumers – *policy decision points*, PDP**
  - Make decisions and transfer the policy rules (eventually translated) to the policy targets.
- **Policy targets, *policy enforcement points*, PEP**
  - Functional elements affected by the policy rules.

45

46

Policy management system - example



46

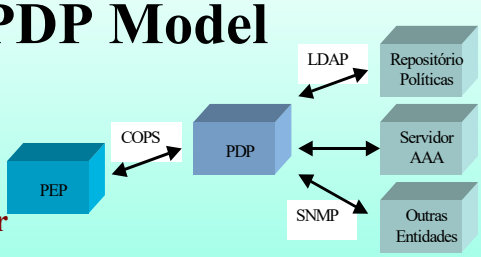
universidade de aveiro

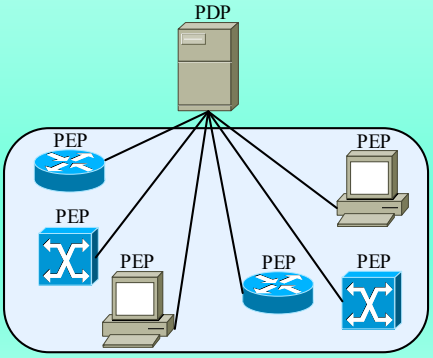
47

PEP-PDP Model

- Model uses (may use) different protocols
  - Abstraction levels differ
- Increasing trend
  - Software defined netwoking (SDN) can be seen as a variation of this concept

  (OpenFlow is a protocol for SDN)





47

universidade de aveiro

48

Processing rules - sequence

- Rules definition
  - Verify internal conflicts
  - Include in a repository (e.g. with LDAP)
- Get policies from policy consumers
  - Take decisions based on policies
  - Processed to create configurations in policy targets
  - May use temporal restrictions
- Send policies to policy targets
  - Can be “pushed” or “pulled” (e.g. by COPS or SNMP)
- Policy targets
  - Instal configurations

48

COPS – Common Open Policy Service

49

- Question/answer protocol to PDP-PEP interaction
- Based on TCP
- Maintains state synchronization
  - Recovers from fault
  - State maintenance with keep-alive
- PDP can send notifications to PEP
  - Default concept was for QoS support/control
- PDP can receive policies through LDAP and SNMP
- Supports two types of clients
  - RSVP, outsourcing model
  - Diff-serv, configuration model

49

PDP-PEP Interactions

50

- Outsourcing (RSVP)
  - PEP contacts PDP when a decision is needed
  - Request contains relevant elements for the policy, and admission control information (e.g. flowspec)
  - Best match for RSVP-based QoS systems
- Configuration requests (Diffserv)
  - PDP configures PEP with specific equipment information
  - Considers a PIB (policy information base) that maintains provisioning information
  - Best match for DiffServ-based QoS systems

50

universidade de aveiro

52

COPS Session

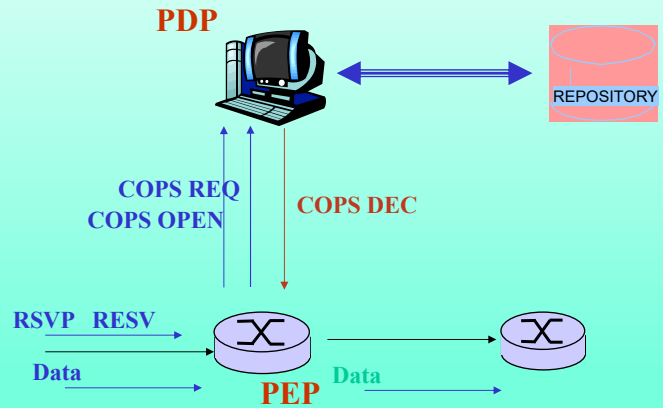
- PEP opens a COPS session
  (specifying a client type: RSVP, DiffServ)
- PEP sends requests and receives answers
- PEP can also send non-solicited commands
- PDP can change commands previously sent
- PEP sends messages related to resources utilization (charging)
- *KeepAlives* are sent if there is no activity

52

universidade de aveiro

53

COPS with RSVP

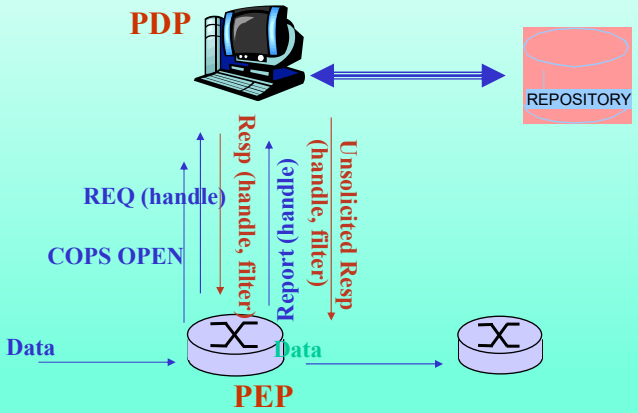


53

universidade de aveiro

54

COPS with DiffServ



It is not required end-to-end signalling
It configures routers with packets lists and actions

54

universidade de aveiro

CMIS/CMIP

Common Management Information Services/Protocol

55

universidade de aveiro

56

Management protocolos (LAN-oriented)

OSI CMIP

- Common Management Information Protocol
- Designed in 1980's: *the* unifying protocol ("advanced") to network management
- Implemented very slowly

SNMP: Simple Network Management Protocol

- Internet based (SGMP)
- Very simple in the beginning
- Rapidly spreaded
- It grew in largeness and complexity
- actual: SNMPv3
- Management protocol *de facto*

56

universidade de aveiro

57

OSI Management architecture

ITU-T	Acronym	Title
X.701		*System Management Overview*
X.710	CMIS	*Common Management Information Service*
X.711	CMIP	*Common Management Information Protocol*
X.712	CMIP-PICS	*CMIP Protocol Implementation Conformance State Proforma*
X.720	MIM	*Management Information Model (defines fundamental concepts of the objects)*
X.721	DMI	*Definition of Management Information*
X.722	GDMO	*Guideline for Definition of Management Objects (techniques for specification of objects)*

57

universidade de aveiro

58

CMIS/CMIP

- Approach object-oriented - objects
  - Have attributes
  - Generate events/notifications (reliably)
  - Execute operations
- Objects with same attributes, notifications and operations belong to the same class
- Objects inserted in multiples hierarchies, with different inherits and containers
- Intelligent agents
  - Can use rules or policies defined by the manager
  - Can be changed on-line
- Actions (verbs)

  GET, SET, CREATE, DELETE, ACTION, NOTIFICATION, CANCEL_GET
- Capacity of CMIP actions is related to scoping and filtering capacities - through GDMOs

58

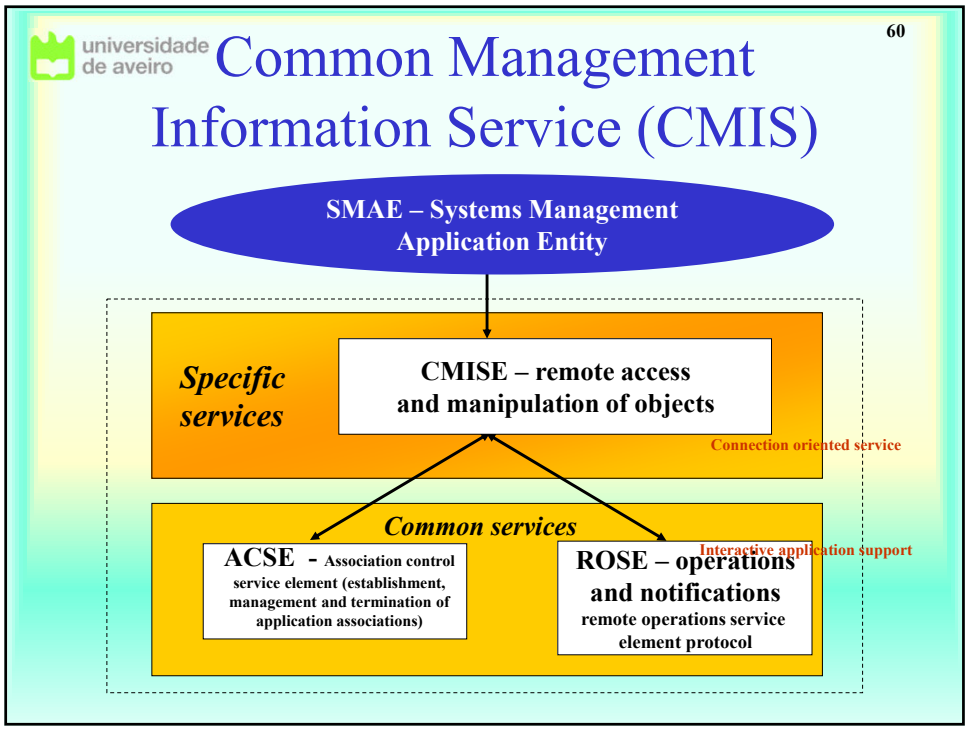
universidade de aveiro

59

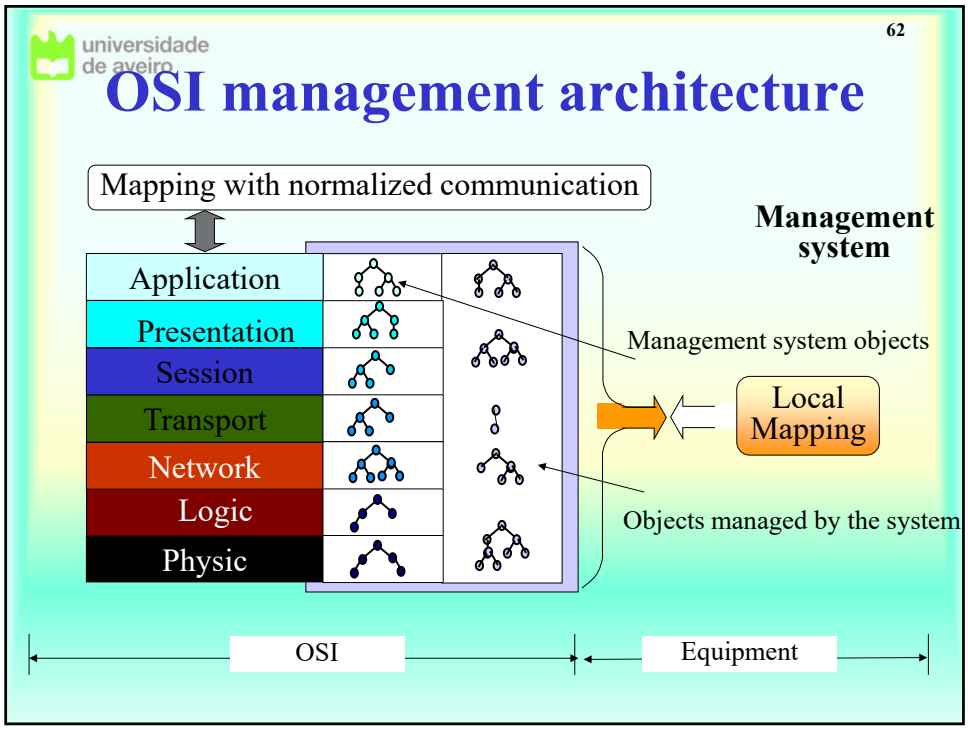
CMIP - GDMOs

- Guideline for the Definition of Managed Objects
  - The equipment through which the agent operates
- Model objects inside the equipment
  - Instantiation of GDMOs is called MIB
- Do not have well-defined behaviors, with large implementation freedom
  - Flexibility
  - Problem (complexity)
- CMIP is not polling oriented
  - Better scalability is achieved
- There are not so many defined GDMOs as MIBs

59


60
universidade de aveiro
Common Management Information Service (CMIS)
SMAE – Systems Management Application Entity
Specific services
CMISE – remote access and manipulation of objects
Connection oriented service
Common services
ACSE - Association control service element (establishment, management and termination of application associations)
ROSE – operations and notifications remote operations service element protocol
Interactive application support

60


62
universidade de aveiro
OSI management architecture
Mapping with normalized communication
Management system
Application
Presentation
Session
Transport
Network
Logic
Physic
Management system objects
Local Mapping
Objects managed by the system
OSI
Equipment

62

universidade de aveiro

CMIP: pros and cons

- CMIP advantages
  - Object-oriented approach is flexible and extensible
  - Support from telecommunications industry and international vendors
  - Support of manager-manager interaction
  - Support of automation environments
  - Imposed in some industrial areas
- CMIP disadvantages
  - Complex and multi-layer
  - Large management overhead
  - Few management systems based on CMIP
  - Few CMIP agents in use
  - Generally rejected in the Internet.

63

universidade de aveiro

Frameworks: SNMP and CMIS

SNMP	CMIS
• Static MIBs • Concepts of limited models • Non-connection oriented protocol • Polling model • Implementation-oriented • Ligth • Limited functionalities • Bulk capcacity only in new versions • Completely dominating the market • Many SNMP-based products	• Dynamic MIBs • Object-oriented models • Connection-oriented protocol • Event-oriented model • Specification-oriented • Heavy • Functionalities until the system management level • Bulk capacity with scope and filtering • Some relevance in the telecommunications market • Some CMIP-based products in the market

64

universidade de aveiro

TMN

Telecomunications Management Network

65

universidade de aveiro

66

What is TMN ?

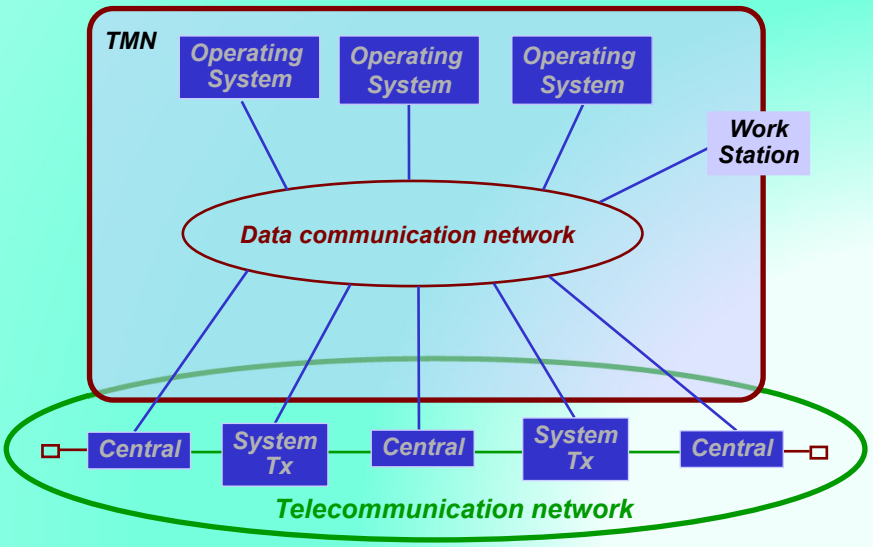
- ***Objective***
  - **Support the management of the telecommunication networks and services**
- ***Concept***
  - **Create an organized structure to allow the interconnection of several operating systems and telecommunications equipments, using a well-defined architecture, with normalized protocols and interfaces**

66

universidade de aveiro

67

TMN and the communications network



67

universidade de aveiro

68

TMN

- **TMN is the telecommunications management network.**
  - **Relies on other management protocols and concepts.**
  - **Operations systems are where the main management functionality resides**
    - Now also known as OSS operational support systems
  - **The data communications network is where the management information flows**
  - **The TMN boundary intersects NEs (network elements) as they include some CM functionality.**
  - **Workstations provides user access to management functionality.**
    - The workstation glass interface is outside the bounds of standardisation.

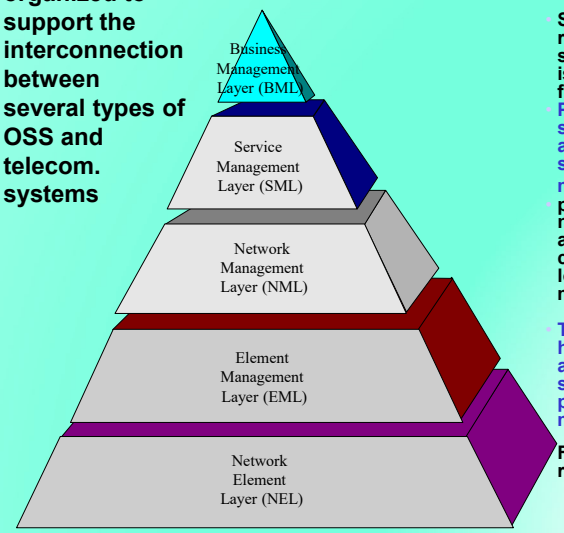
68

universidade de aveiro

69

Multi-layer model

Architecture organized to support the interconnection between several types of OSS and telecom. systems



- **Service levels, contact point for client reports, definition of policies for fault support, analyzes trends and quality issues, for example, or to provide a basis for billing and other financial reports**
- **Performs functions for the handling of services in the network: definition, administration and charging of services, scalability policies, service restore, monitoring and report of SLAs availability.**
- **performs functions for distribution of network resources: configuration, control and supervision of the network, correlation of events and filtering, fault location, problem isolation, syslog management and monitor.**
- **This includes alarm management, handling of information, backup, logging, and maintenance of hardware and software, automatic correction of faults, polling of network elements, event management.**

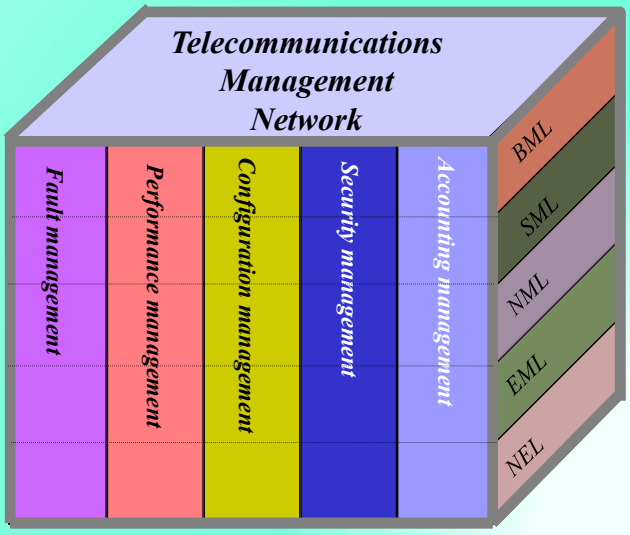
Fault detection, trap generation, restoration/substitution procedures.

69

universidade de aveiro

70

TMN Matrix



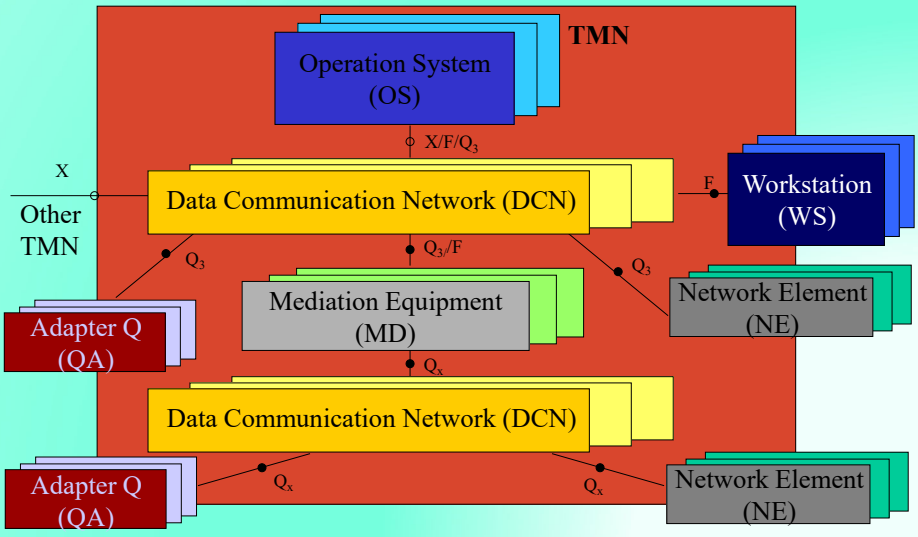
Depicts the FCAPS functionality along the TMN layers

70

universidade de aveiro

77

TMN Physical architecture



77

universidade de aveiro

78

Physical Architecture

Standard interfaces

- Interface Q_3
  - Operation System ↔ TMN elements connected to it
- Interface Q_x
  - Mediation Element ↔ Network element/adapter
- Interface F
  - Workstation ↔ Operation system / mediator (through a DCN)
- Interface X
  - TMN network ↔ TMN network / network with *interface* to TMN.

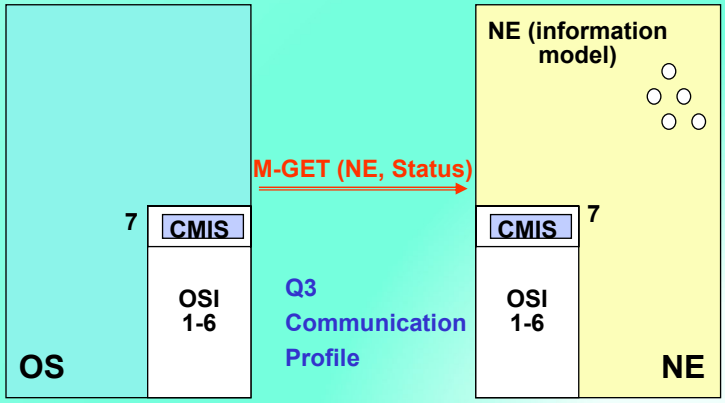
- **Communication protocols**
  - For each interface, the choice depends on the physical implementation.
  - It is normal to use CMIP protocol in the Q_3 interface.

78

universidade de aveiro

79

TMN: Interface Q3



Q3 Interface is supported by CMIS/CMIP!!

The scope of the available access is defined by the NE information model (its MIB)

79

universidade de aveiro

81

TMN and OSI

- TMN adds-on to OSI management
  - **Information model – new network**
  - **Organization model – extension through the concept of functional block**
  - **Communication model – Correspondence between interface - protocol**
  - **Functional model – new management functions (network)**

- Interworking TMN with other OSI systems
  - **Atenuate differences between protocols**
  - **Services functionalities and complementaring functions**
  - **Increase the OSI management potentialities or restrict the TMN management potentialities**

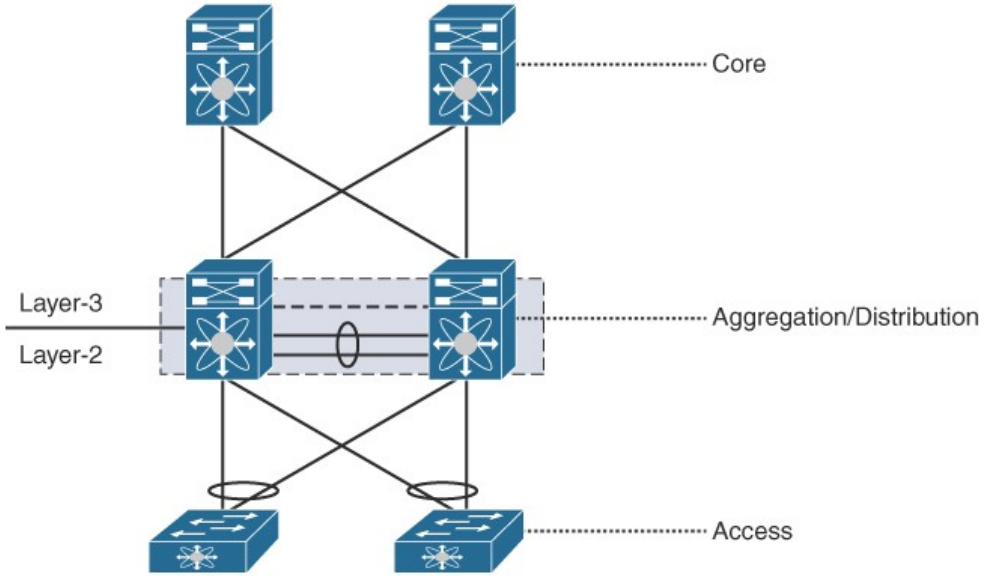
81

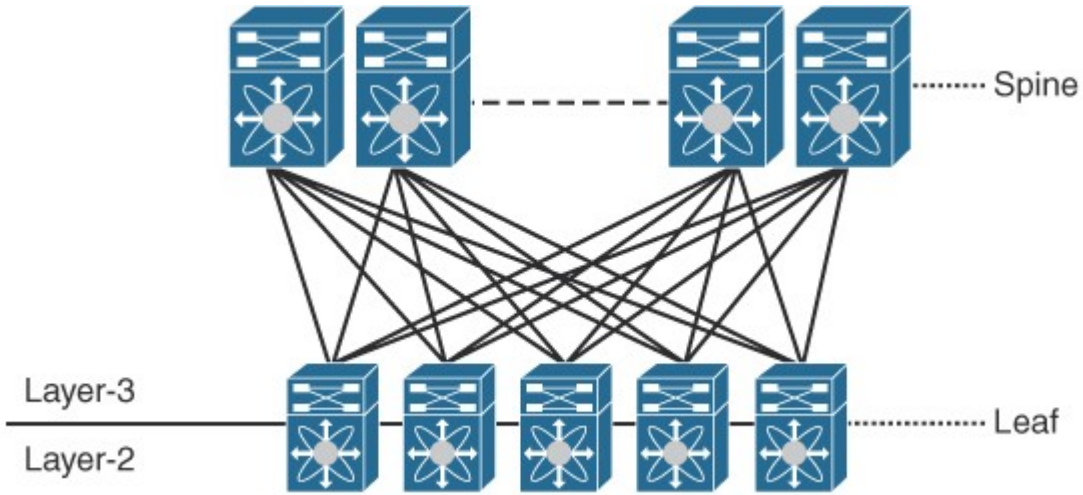
Layer 2 VPN
VXLAN and BGP EVPN

universidade de aveiro
deti.ua.pt

Datacenter CLOS Topology

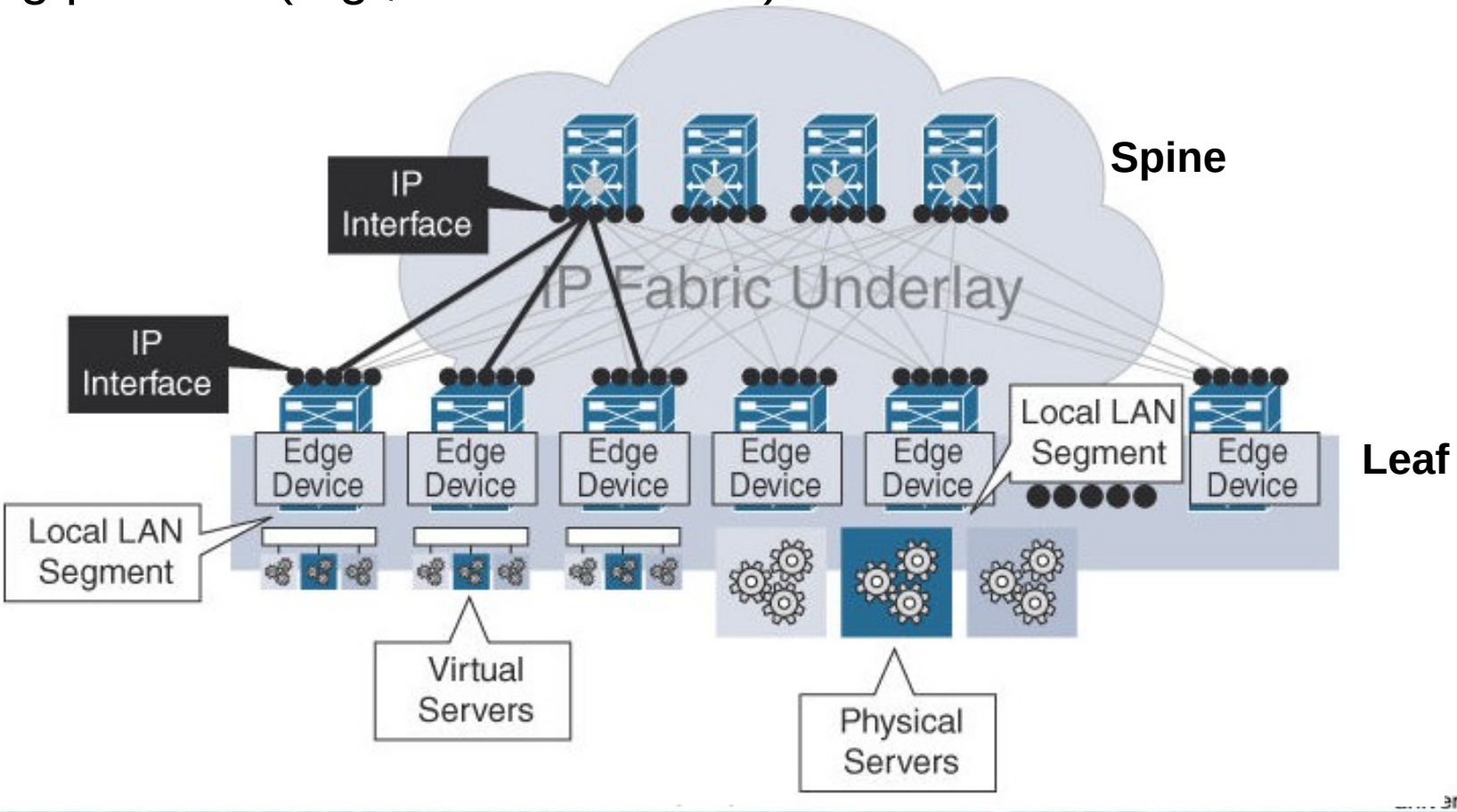
- With large-scale data center deployments, three-tier topologies have become scale bottlenecks.
- The classic three-tier topology evolved to a CLOS topology.
  - Original designed by Charles Clos in 1950 to find a more efficient way to handle telephonic call transfers.
- Eliminating the need for STP the network evolved to greater stability and scalability.
- Layer 3 moves to the Access Layer.
- Usually called Spine-and-Leaf Architecture.





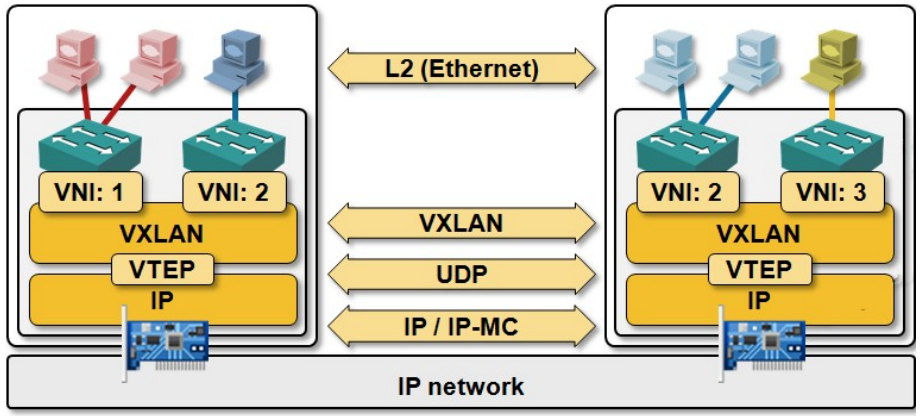
Spine-and-Leaf Architecture

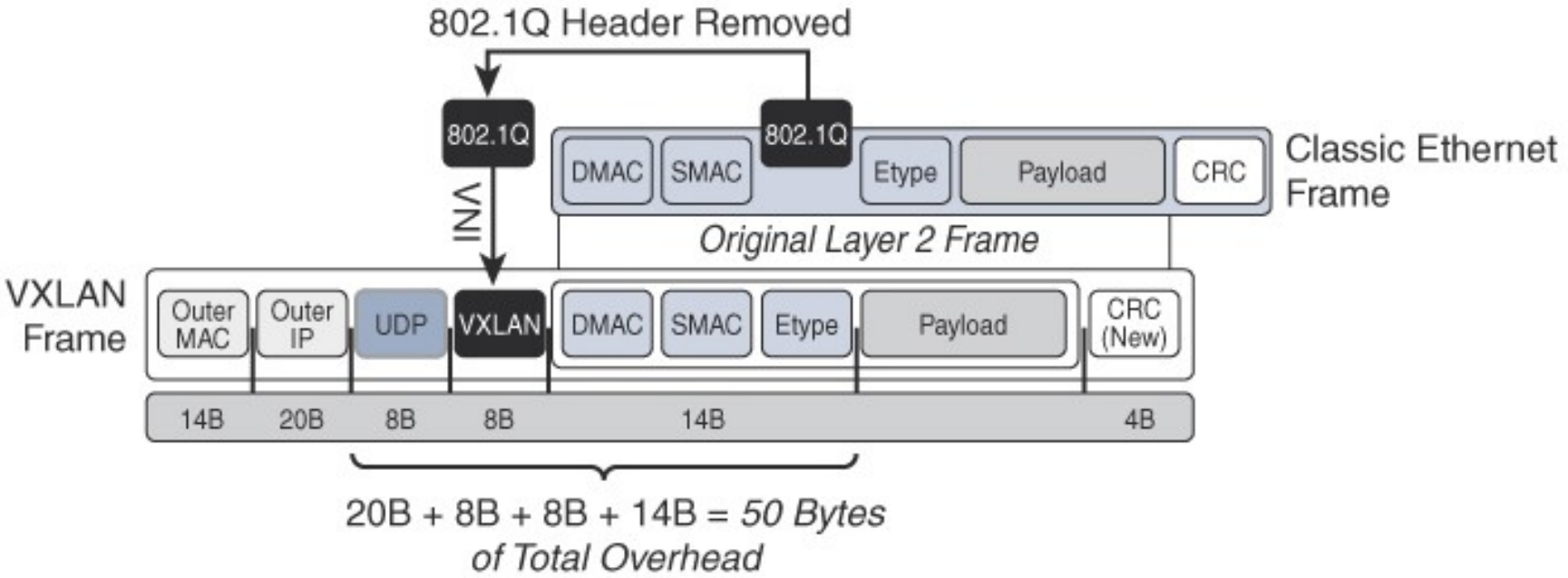
- The access layer with Layer 3 support is typically called the Leaf layer.
- The aggregation layer that provides the interconnection between the various leafs is called the Spine layer.
- The IP underlay transport between Spines and Leaves requires an IGP routing protocol (e.g., OSPF or IS-IS).



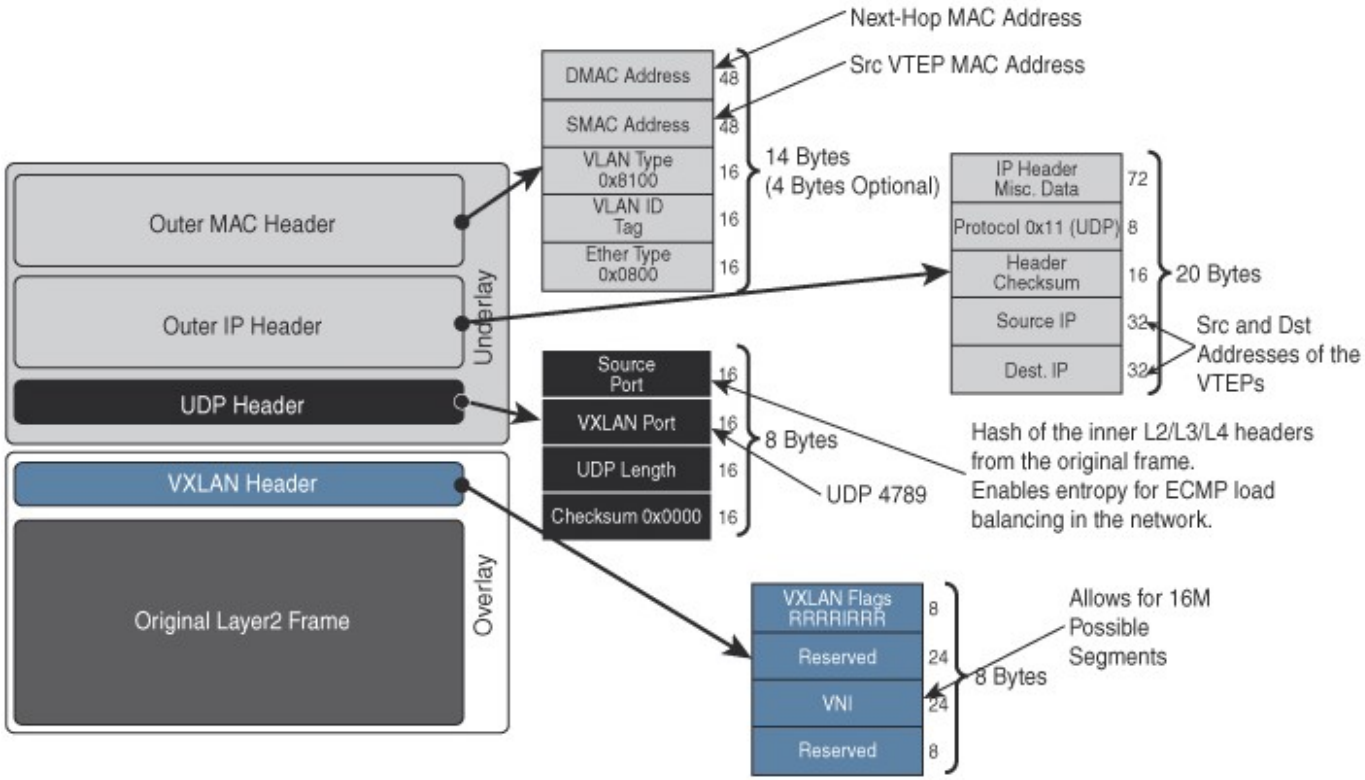
Virtual Extensible LAN (VXLAN)

- Encapsulates OSI Layer 2 Ethernet frames within Layer 4 UDP/IP datagrams .
  - Default port 4789.
- VLAN may be additionally identified by a **VNI field** with 24 bits.
  - 802.1Q tag only has 12 bits.
- The original inner 802.1Q header of the Layer 2 Ethernet frame is removed and mapped to a VNI to complete the VXLAN header.





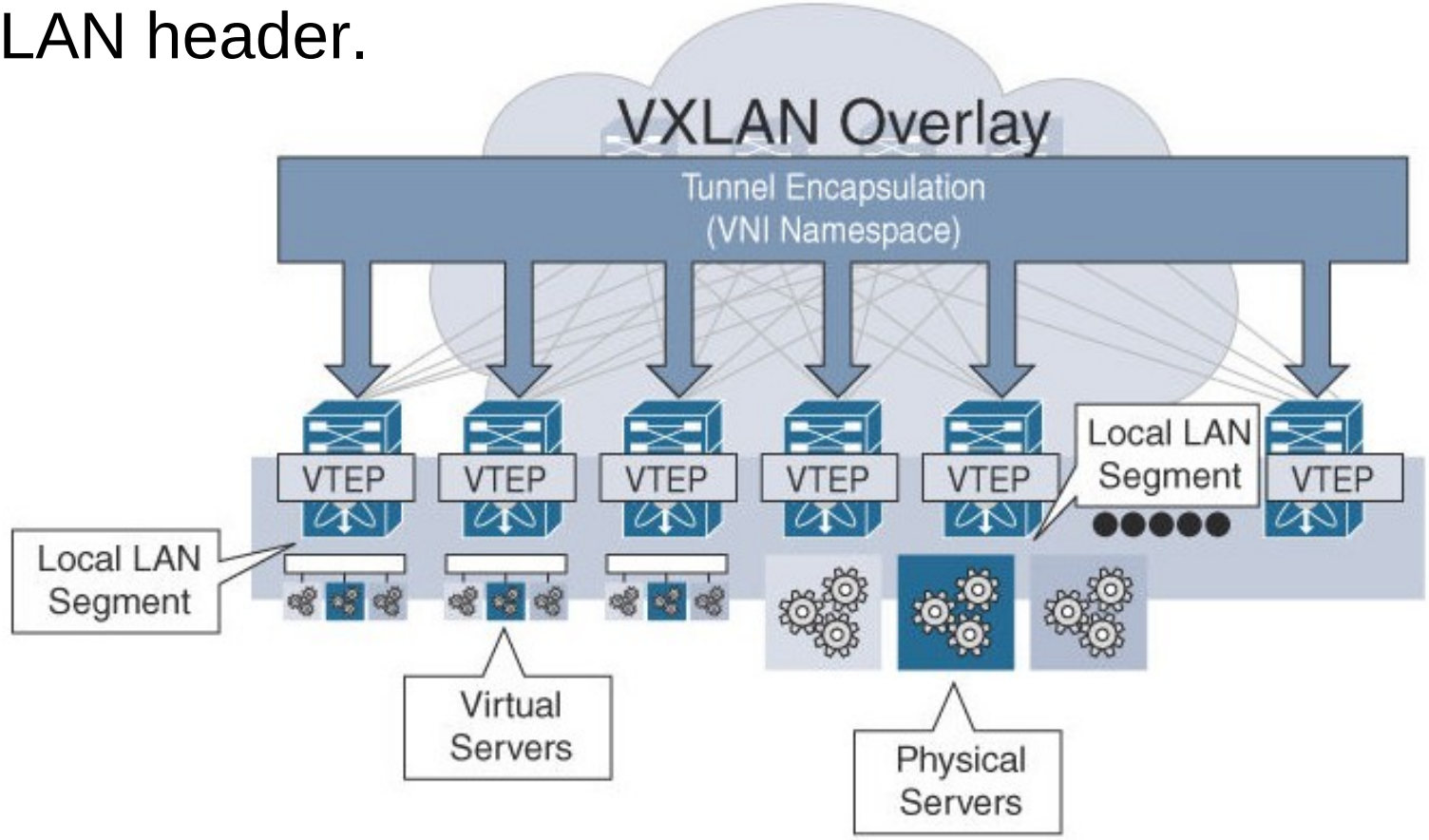
VXLAN Header/Packet




```
▸ Ethernet II, Src: ca:01:25:92:00:08 (ca:01:25:92:00:08), Dst: 0c:32:45:6d:00:00 (0c:32:45:6d:00:00)
▸ Internet Protocol Version 4, Src: 192.0.0.3, Dst: 192.0.0.1
▸ User Datagram Protocol, Src Port: 56255, Dst Port: 8472
▾ Virtual eXtensible Local Area Network
  ▸ Flags: 0x0800, VXLAN Network ID (VNI)
    Group Policy ID: 0
    VXLAN Network Identifier (VNI): 101
    Reserved: 0
▸ Ethernet II, Src: 0c:88:63:63:00:01 (0c:88:63:63:00:01), Dst: Private_66:68:00 (00:50:79:66:68:00)
▸ Internet Protocol Version 4, Src: 10.1.3.100, Dst: 10.1.1.100
▸ Internet Control Message Protocol
```

VTEP (VXLAN Tunnel Endpoint)

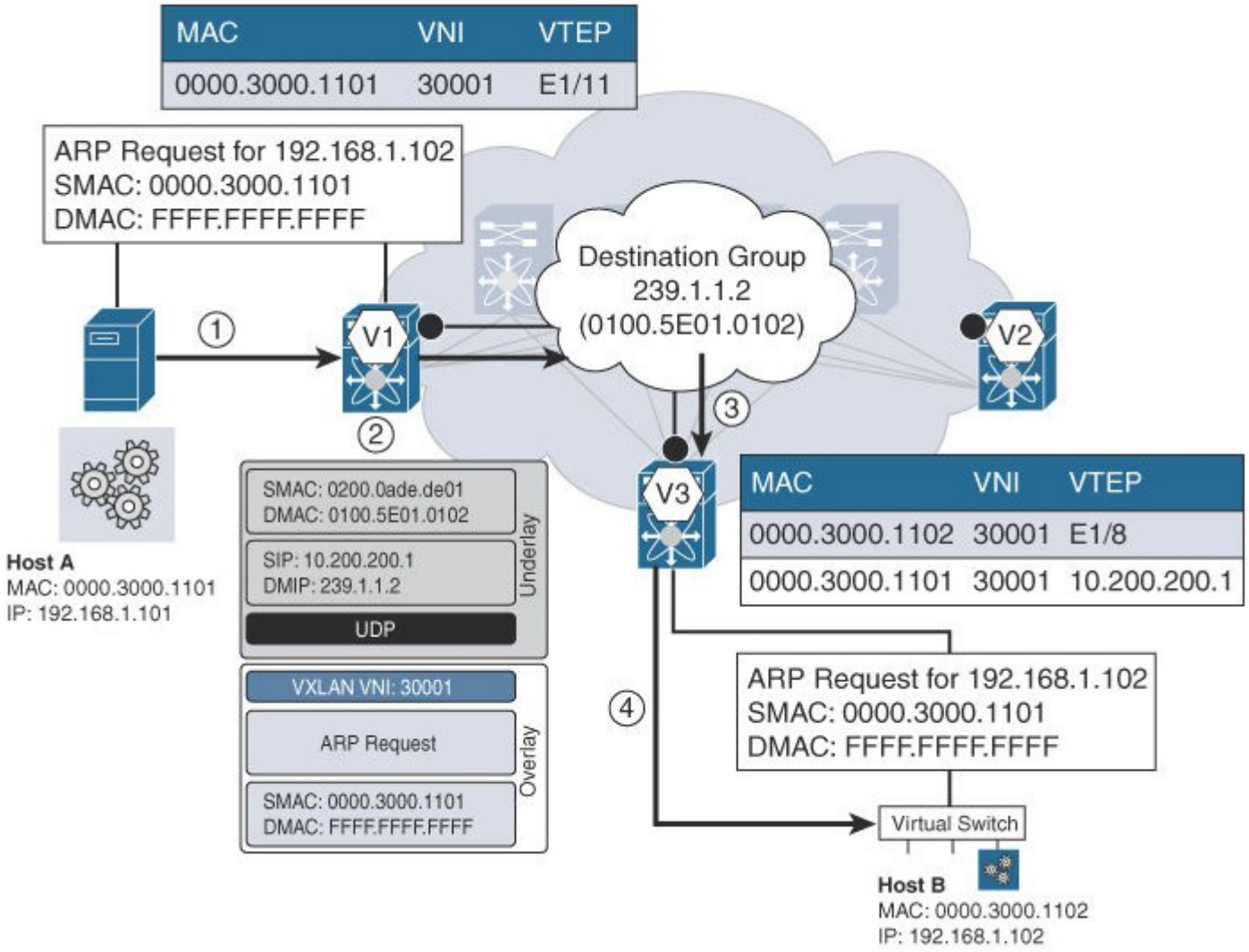
- The edge devices in a VXLAN network have the VXLAN Tunnel Endpoints (VTEP).
- Are responsible for encapsulation and decapsulation of the VXLAN header.



VTEP: VXLAN Tunnel Endpoint
VNI/VNID: VXLAN Network Identifier

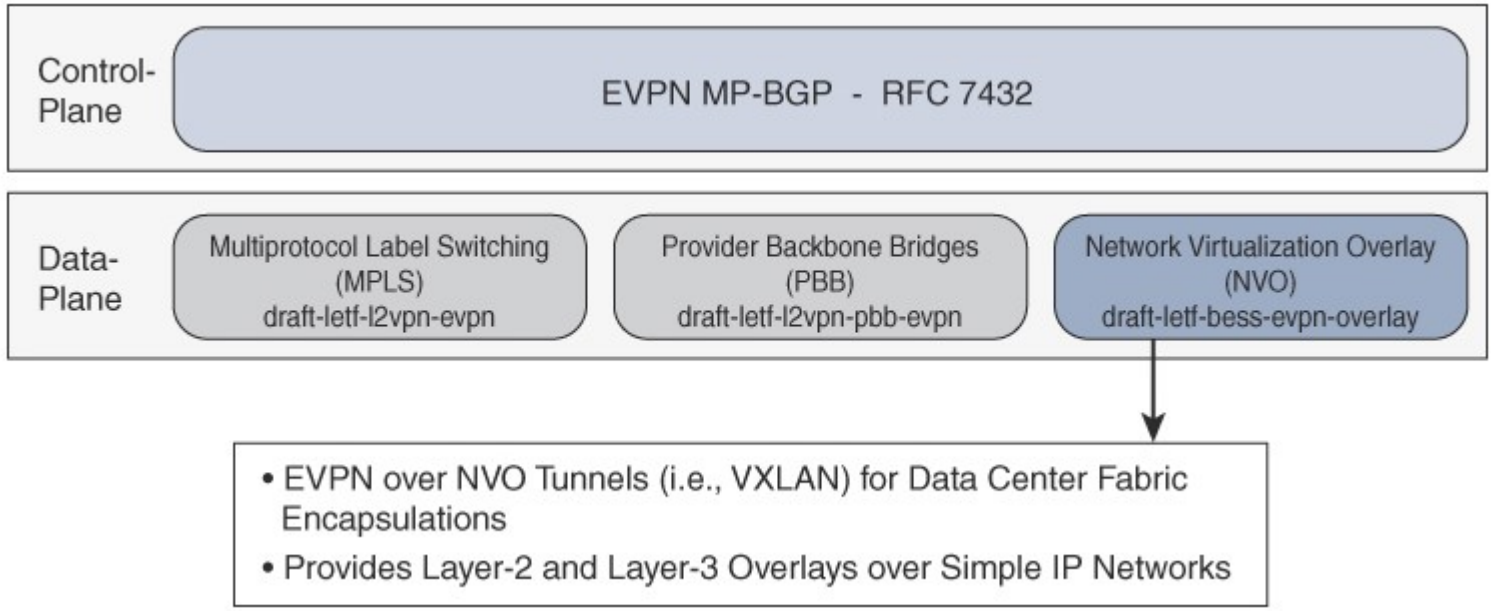
VXLAN Flood and Learn

- The multidestination traffic is flooded over the VXLAN between VTEPs.
  - To learn about the host MACs located behind the VTEPs so that subsequent traffic can be unicast.
  - This is referred to as an F&L mechanism.
- A native F&L based approach is far from optimal since the broadcast domain for a VXLAN now spans Layer 3 boundaries.



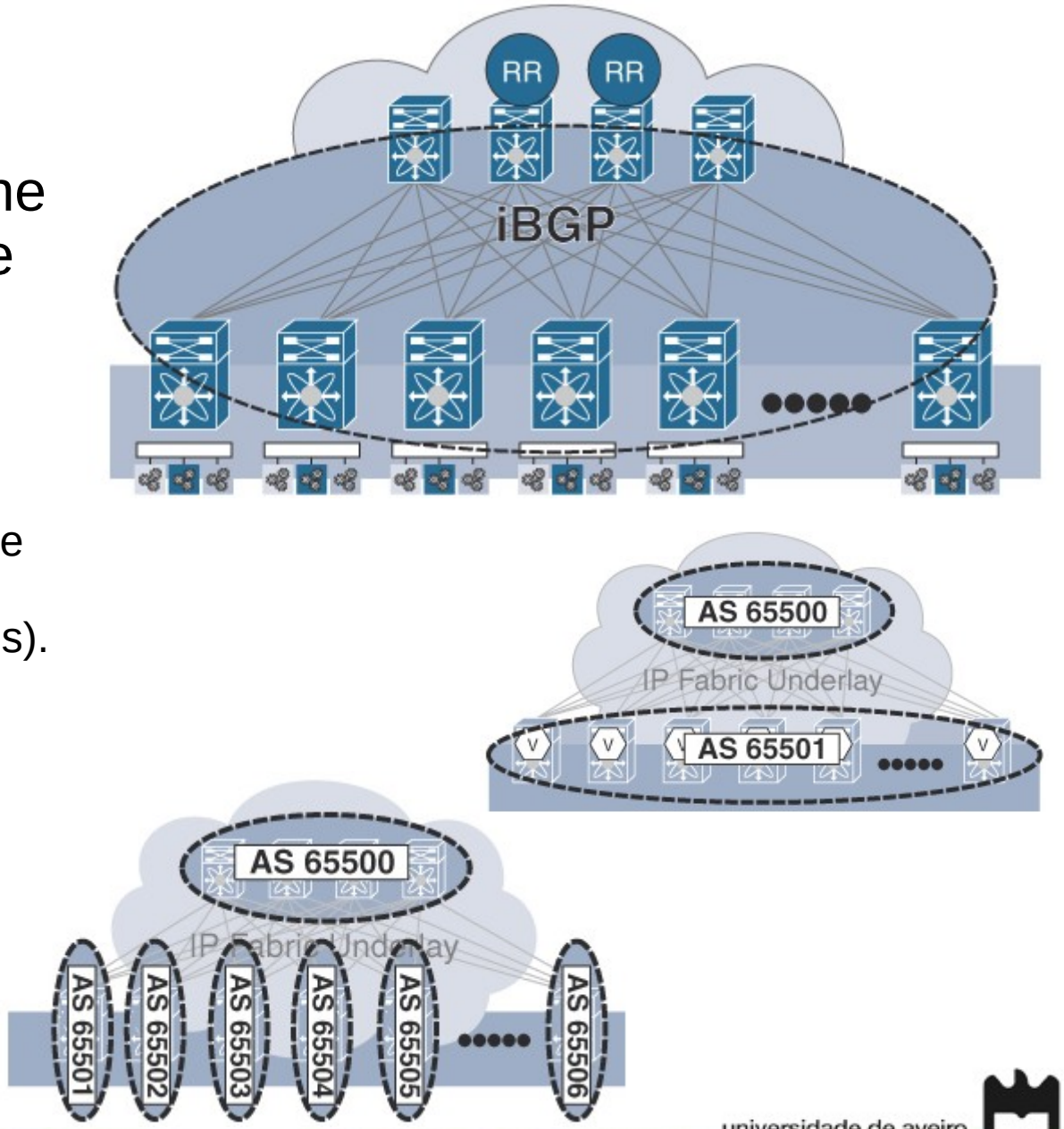
EVPN MP-BGP

- To mitigate the VXLAN Flood and Learn problem, it was introduced the concept of Ethernet VPN with MP-BGP, provided by the Address family L2VPN EVPN.
- The Address family L2VPN EVPN provides a method to transport VPN-aware Layer 2 (MAC) and Layer 3 (IP) information across a single MP-BGP peering session.
- The EVPN MP-BGP RFC allow for multiple data plane transport: MPLS, PBB and NVO.
  - A possible (and more common) EVPN over NVO solution for datacenters is VXLAN.



BGP EVPN wih VXLAN

- BGP is used to announce and learn remote VTEP addresses.
- VXLAN is used to transport to the specific remote VTEP where the destination device is.
- BGP relations can be:
  - Only internal BGP.
    - To avoid a full BGP mesh, Route Reflectors should be used (usually all or some of the spines).
  - External BGP relations between private AS.
    - Leaf in a single private AS.
    - Each Leaf is a private AS.



EVPN Route types

- Route Type-2
  - Defines the MAC/IP advertisement route.
  - Responsible for the distribution of MAC and IP address reachability information.
- Route Type-3
  - Called “inclusive multicast Ethernet tag route”.
  - Used to create the a distribution list for unknown unicast, multicast and broadcast packets (ingress replication).
  - Provides a way to replicate multidestination traffic in a unicast.

RD (8 Octets)
ESI (10 Octets)
Ethernet Tag ID (4 Octets)
MAC Address Length (1 Octet)
MAC Address (6 Octets)
IP Address Length (1 Octet)
IP Address (0, 4, or 16 Octets)
MPLS Label1 (3 Octets)
MPLS Label2 (0 or 3 Octets)

RD (8 Octets)
ESI (10 Octets)
Ethernet Tag ID (4 Octets)
IP Address Length (1 Octet)
Originating Router's IP Address (4 or 16 Octets)

EVPN Route Type-2

```
Border Gateway Protocol - UPDATE Message
  Marker: ffffffffffffffffffffffffffffffff
  Length: 144
  Type: UPDATE Message (2)
  Withdrawn Routes Length: 0
  Total Path Attribute Length: 121
 Path attributes
  Path Attribute - MP_REACH_NLRI
   Flags: 0x90, Optional, Extended-Length, Non-transitive, Complete
    Type Code: MP_REACH_NLRI (14)
    Length: 83
    Address family identifier (AFI): Layer-2 VPN (25)
    Subsequent address family identifier (SAFI): EVPN (70)
   Next hop: 192.0.0.3
    Number of Subnetwork points of attachment (SNPA): 0
   Network Layer Reachability Information (NLRI)
    EVPN NLRI: MAC Advertisement Route
      Route Type: MAC Advertisement Route (2)
      Length: 33
      Route Distinguisher: 0001c00000030003 (192.0.0.3:3)
     ESI: 00:00:00:00:00:00:00:00:00:00
      Ethernet Tag ID: 0
      MAC Address Length: 48
      MAC Address: Private_66:68:02 (00:50:79:66:68:02)
      IP Address Length: 0
     IP Address: NOT INCLUDED
      VNI: 101
    EVPN NLRI: MAC Advertisement Route
      Route Type: MAC Advertisement Route (2)
      Length: 37
      Route Distinguisher: 0001c00000030003 (192.0.0.3:3)
     ESI: 00:00:00:00:00:00:00:00:00:00
      Ethernet Tag ID: 0
      MAC Address Length: 48
      MAC Address: Private_66:68:02 (00:50:79:66:68:02)
      IP Address Length: 32
      IPv4 address: 10.1.3.100
      VNI: 101
  Path Attribute - ORIGIN: IGP
  Path Attribute - AS_PATH: empty
  Path Attribute - LOCAL_PREF: 100
  Path Attribute - EXTENDED_COMMUNITIES
   Flags: 0xc0, Optional, Transitive, Complete
    Type Code: EXTENDED_COMMUNITIES (16)
    Length: 16
   Carried extended communities: (2 communities)
    Encapsulation: VXLAN Encapsulation [Transitive Opaque]
    Route Target: 100:101 [Transitive 2-Octet AS-Specific]
```

- Announces a MAC address and respective IP address of a remote device.
  - And respective next-hop.
- EXTENDED_COMMUNITY attribute is uses to announce the type of encapsulation and the route target.
- Sent when Leaf device learns a new MAC address.

EVPN Route Type-3

```
▾ Border Gateway Protocol - UPDATE Message
    Marker: ffffffffffffffffffffffffffffffff
    Length: 122
    Type: UPDATE Message (2)
    Withdrawn Routes Length: 0
    Total Path Attribute Length: 99
  ▾ Path attributes
    ▾ Path Attribute - MP_REACH_NLRI
      ▸ Flags: 0x90, Optional, Extended-Length, Non-transitive, Complete
        Type Code: MP_REACH_NLRI (14)
        Length: 28
        Address family identifier (AFI): Layer-2 VPN (25)
        Subsequent address family identifier (SAFI): EVPN (70)
      ▸ Next hop: 192.0.0.2
        Number of Subnetwork points of attachment (SNPA): 0
      ▾ Network Layer Reachability Information (NLRI)
        ▾ EVPN NLRI: Inclusive Multicast Route
            Route Type: Inclusive Multicast Route (3)
            Length: 17
            Route Distinguisher: 0001c00000020002 (192.0.0.2:2)
            Ethernet Tag ID: 0
            IP Address Length: 32
            IPv4 address: 192.0.0.2
    ▸ Path Attribute - ORIGIN: IGP
    ▸ Path Attribute - AS_PATH: empty
    ▸ Path Attribute - MULTI_EXIT_DISC: 0
    ▸ Path Attribute - LOCAL_PREF: 100
    ▸ Path Attribute - ORIGINATOR_ID: 192.0.0.2
    ▸ Path Attribute - CLUSTER_LIST: 192.0.0.1
    ▸ Path Attribute - EXTENDED_COMMUNITIES
    ▾ Path Attribute - PMSI_TUNNEL_ATTRIBUTE
      ▸ Flags: 0xc0, Optional, Transitive, Complete
        Type Code: PMSI_TUNNEL_ATTRIBUTE (22)
        Length: 9
        Flags: 0
        Tunnel Type: Ingress Replication (6)
        VNI: 102
      ▸ Tunnel ID: tunnel end point -> 192.0.0.2
```

- Defines the next hop for unknown unicast, multicast and broadcast.
- Must also carry a Provider Multicast Service Interface (PMSI) Tunnel attribute.
  - Defines tunnel type.
  - For EVPN with VXLAN the tunnel type is “Ingress Replication”.
- Sent when a new Leaf (BGP peer) is added.

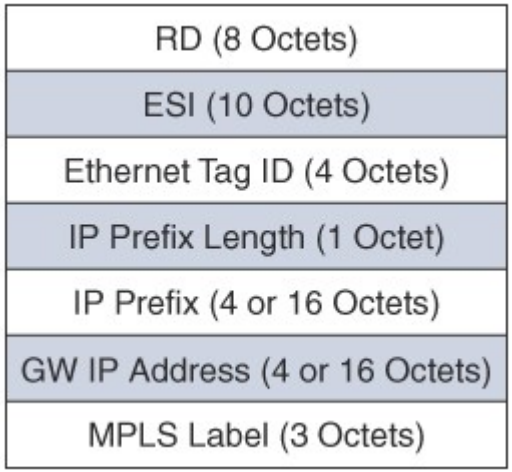
BGP Route Table – L2VPN EVPN

```
# show bgp l2vpn evpn
BGP table version is 1, local router ID is 192.0.0.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete
EVPN type-1 prefix: [1]:[EthTag]:[ESI]:[IPlen]:[VTEP-IP]:[Frag-id]
EVPN type-2 prefix: [2]:[EthTag]:[MAClen]:[MAC]:[IPlen]:[IP]
EVPN type-3 prefix: [3]:[EthTag]:[IPlen]:[OrigIP]
EVPN type-4 prefix: [4]:[ESI]:[IPlen]:[OrigIP]
EVPN type-5 prefix: [5]:[EthTag]:[IPlen]:[IP]

   Network          Next Hop            Metric LocPrf Weight Path
Route Distinguisher: 192.0.0.1:2
*> [3]:[0]:[32]:[192.0.0.1]
                    192.0.0.1                              32768 i
                    ET:8 RT:100:102
...
*>i[2]:[0]:[48]:[00:50:79:66:68:01]
                    192.0.0.2                100       0 i
                    RT:100:101 ET:8
...
*>i[2]:[0]:[48]:[00:50:79:66:68:02]:[32]:[10.1.3.100]
                    192.0.0.3                100       0 i
                    RT:100:101 ET:8
...
*>i[3]:[0]:[32]:[192.0.0.3]
                    192.0.0.3                100       0 i
                    RT:100:101 ET:8
```

Layer 3 VPN over EVPN with VXLAN

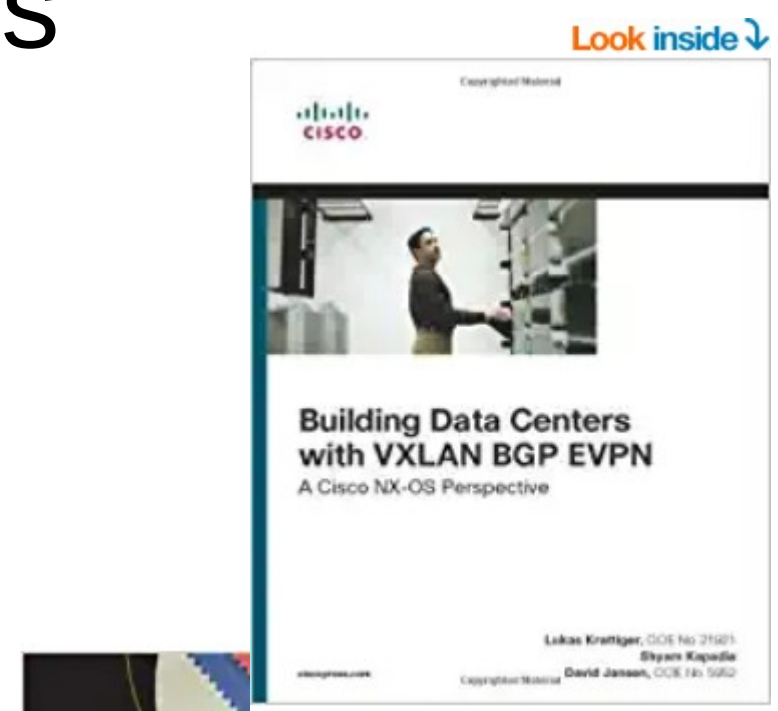
- As an alternative Layer3 VPN to MPLS VPN, it is possible to create a Layer3 VPN over an EVPN with VXLAN.
  - Using announcements of Route Type-5.
- Route Type-5
  - Announces IP network prefixes.

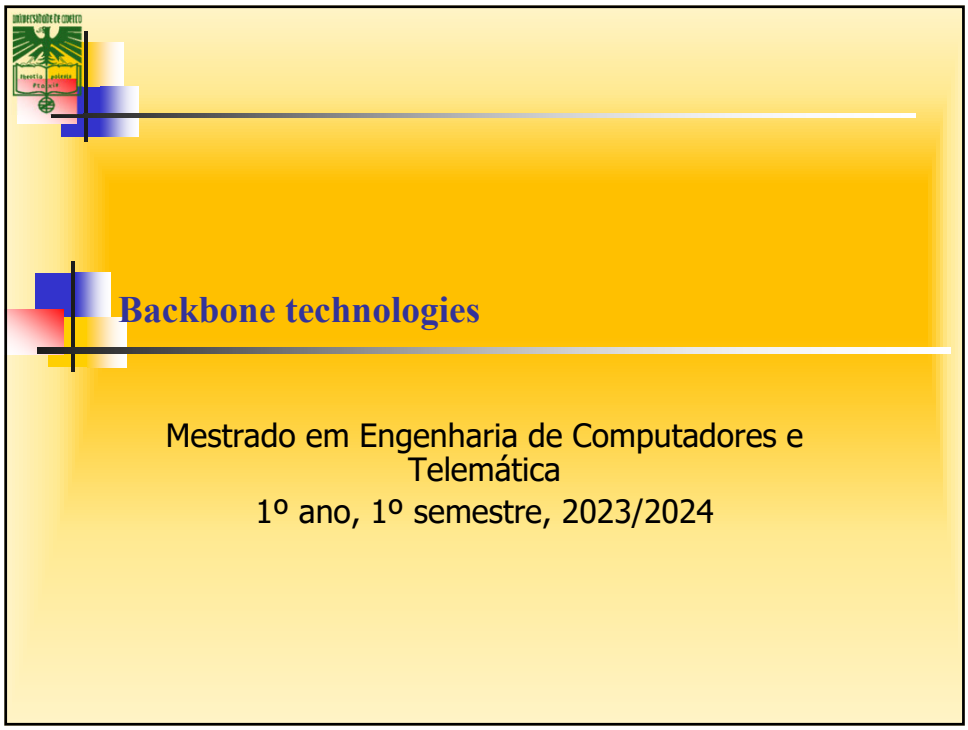



```
▾ Border Gateway Protocol - UPDATE Message
    Marker: ffffffffffffffffffffffffffffffff
    Length: 121
    Type: UPDATE Message (2)
    Withdrawn Routes Length: 0
    Total Path Attribute Length: 98
  ▾ Path attributes
    ▾ Path Attribute - MP_REACH_NLRI
      ▸ Flags: 0x90, Optional, Extended-Length, Non-transitive, Complete
        Type Code: MP_REACH_NLRI (14)
        Length: 45
        Address family identifier (AFI): Layer-2 VPN (25)
        Subsequent address family identifier (SAFI): EVPN (70)
      ▸ Next hop: 192.0.0.1
        Number of Subnetwork points of attachment (SNPA): 0
      ▾ Network Layer Reachability Information (NLRI)
        ▾ EVPN NLRI: IP Prefix route
            Route Type: IP Prefix route (5)
            Length: 34
            Route Distinguisher: 00010a0101010004 (10.1.1.1:4)
          ▸ ESI: 00:00:00:00:00:00:00:00:00:00
            Ethernet Tag ID: 0
            IP prefix length: 16
            IPv4 address: 10.1.0.0
            IPv4 Gateway address: 0.0.0.0
            VNI: 101
    ▸ Path Attribute - ORIGIN: INCOMPLETE
    ▸ Path Attribute - AS_PATH: empty
    ▸ Path Attribute - MULTI_EXIT_DISC: 0
    ▸ Path Attribute - LOCAL_PREF: 100
    ▾ Path Attribute - EXTENDED_COMMUNITIES
      ▸ Flags: 0xc0, Optional, Transitive, Complete
        Type Code: EXTENDED_COMMUNITIES (16)
        Length: 24
      ▾ Carried extended communities: (3 communities)
        ▸ Encapsulation: VXLAN Encapsulation [Transitive Opaque]
        ▸ Route Target: 100:101 [Transitive 2-Octet AS-Specific]
        ▸ EVPN Router's MAC: Router's MAC: 0c:32:45:6d:00:01 [Transitive EVPN]
```

References

- Building Data Centers with VXLAN BGP EVPN: A Cisco NX-OS Perspective (Networking Technology), 1st Edition, David Jansen, Lukas Krattiger, Shyam Kapadia, Cisco Press (March 31, 2017), ISBN-13: 978-1587144677.
- The Fast-Track Guide to VXLAN BGP EVPN Fabrics: Implement Today's Multi-Tenant Software-Defined Networks, 1st Edition, Rene Cardona, Apress (May 19, 2021), ISBN-13:978-1484269299.





1

Traffic Engineering (TE)

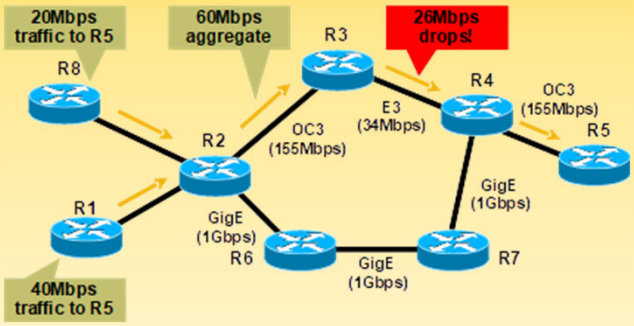
- Network Engineering
  - Build your network to carry your predicted traffic!
  - Traffic patterns are impossible to predict!
  - Routing is based on the destination and does not allow to take the maximum possible advantage of the network resources.
  - IP source routing (using options field of IP header) is not usable in practice due to security reasons.
- Traffic Engineering
  - Manipulate your traffic path to fit your network!
    - Can be done with routing protocol costs (difficult deployment), or MPLS.
    - With RIP or OSPF or ANY OTHER IGP it is not possible to condition multiple traffic flows.
  - Increase efficiency of bandwidth resources.
    - Prevent over-utilized (congested) links whilst other links are under-utilized.
  - Ensure the most desirable/appropriate path for some/all traffic.
    - Override the shortest path selected by the routing protocols.

2

2

Example – avoiding congestion

- On IP networks, *IntServ* and *DiffServ* are "routing independent architectures", retain the issues from routing
- IP network routing is based on the destination and does not allow to take the maximum possible advantage of the network resources
  - Shortest path will lead to congestion, even with available resources in the core
  - With **RIP or OSPF or ANY OTHER IGP** it is not possible to condition both flows.

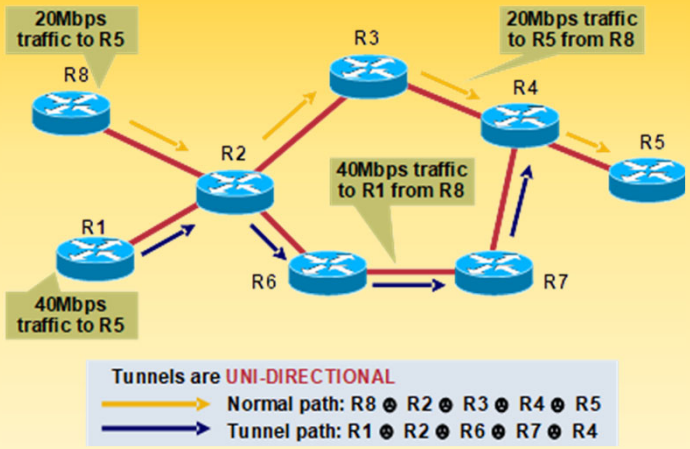


3

3

Using TE to solve… (source based routing)

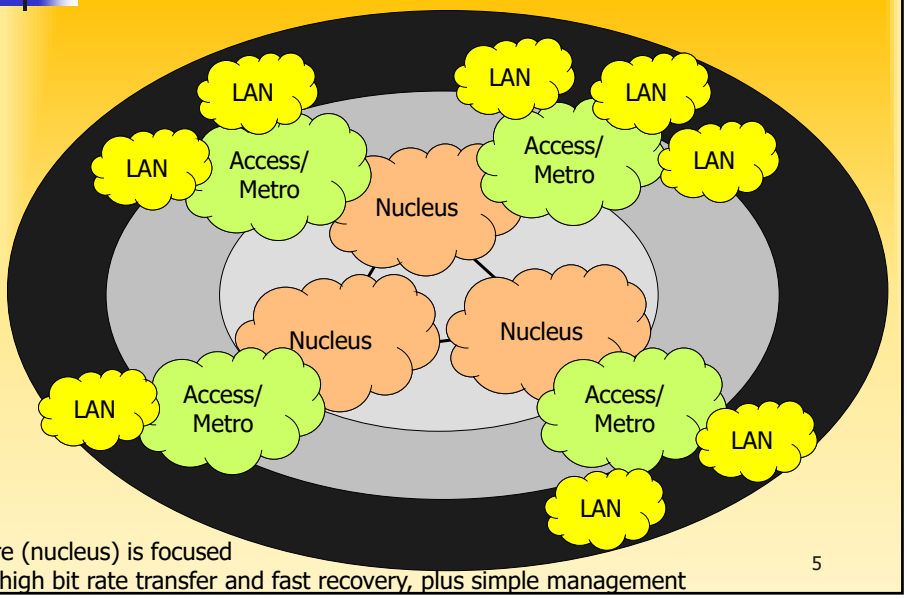
- Tunnels (virtual entities) explore all capacity
  - Packets will transport, from their source, a list of routers' addresses that define their path to the destination (*Options* field of the IP datagram header)



4

4

Network architecture of a public operator

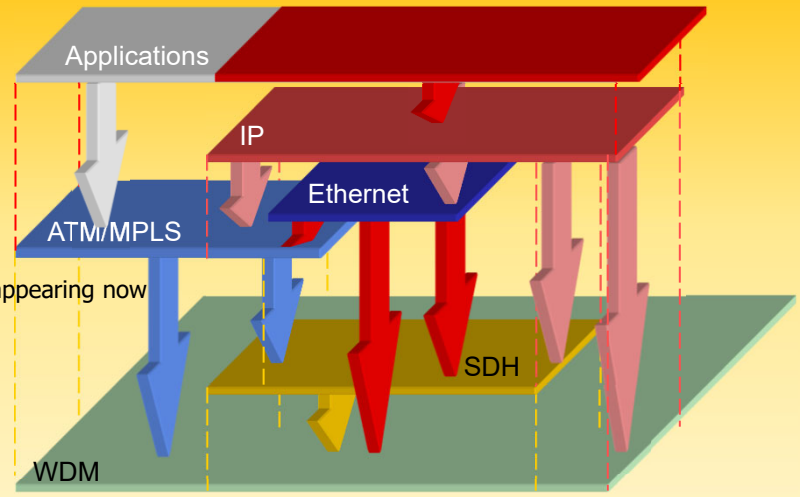


Core (nucleus) is focused
on high bit rate transfer and fast recovery, plus simple management

5

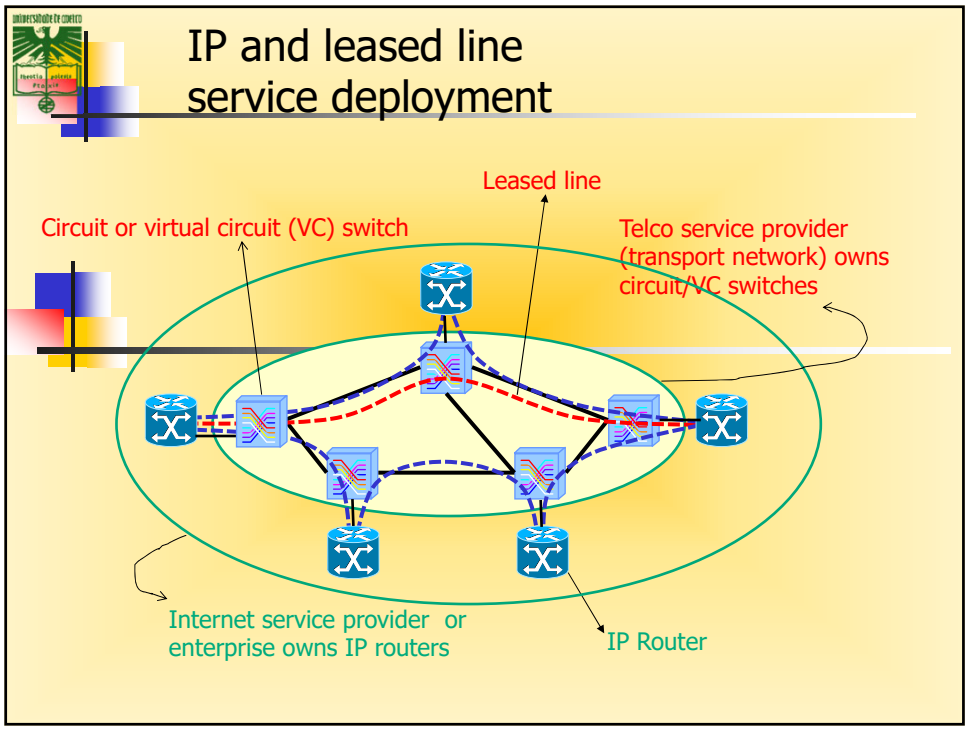
5

IP Transport Network as a Overlay

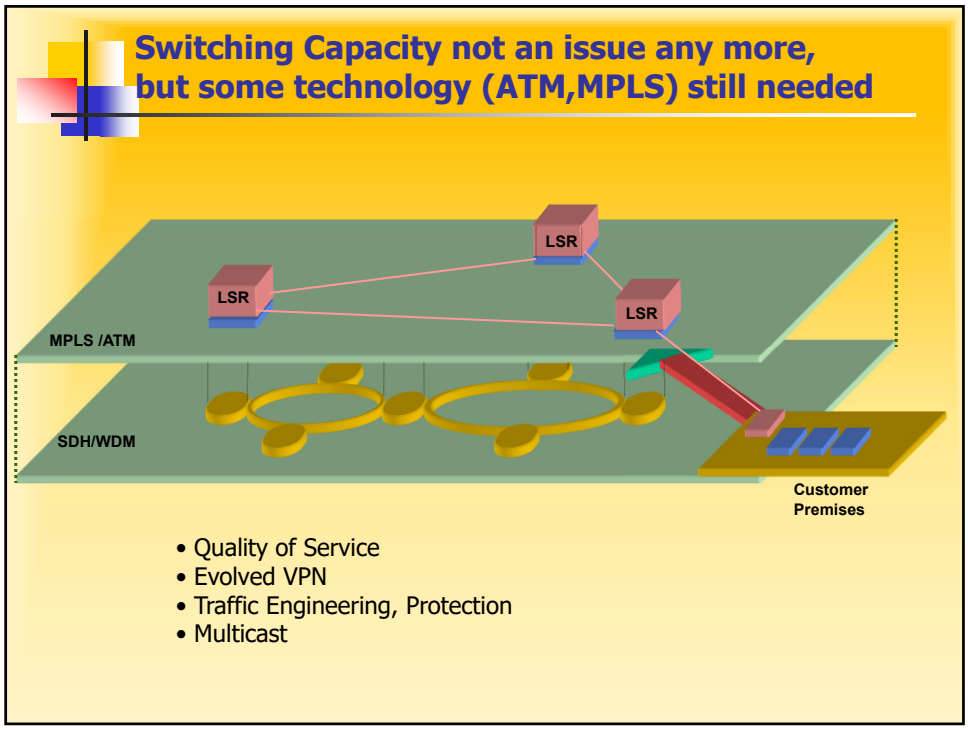


ATM is disappearing now

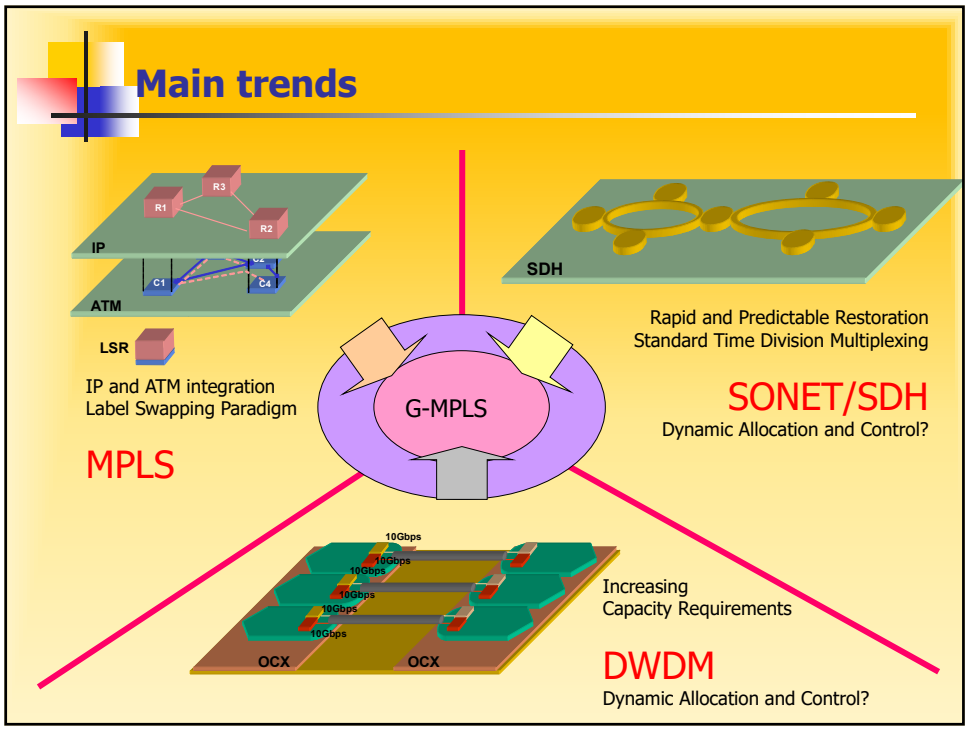
12


IP and leased line service deployment
Leased line
Circuit or virtual circuit (VC) switch
Telco service provider (transport network) owns circuit/VC switches
Internet service provider or enterprise owns IP routers
IP Router

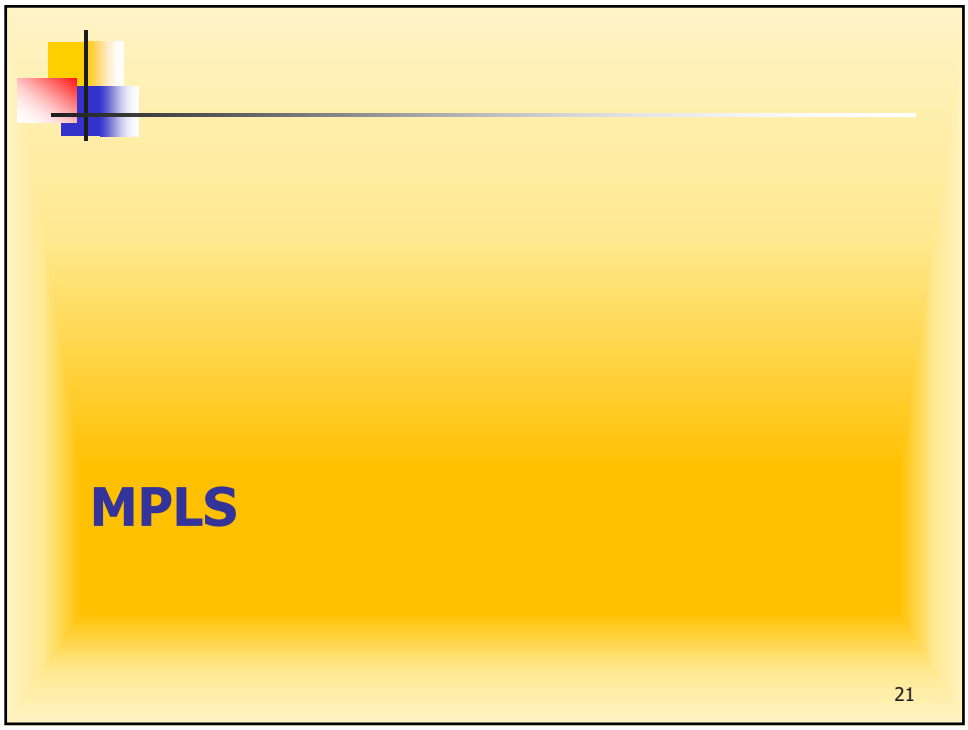
13


Switching Capacity not an issue any more, but some technology (ATM,MPLS) still needed
LSR
LSR
LSR
MPLS /ATM
SDH/WDM
Customer Premises
• Quality of Service
• Evolved VPN
• Traffic Engineering, Protection
• Multicast

14


Main trends
R1
R2
R3
IP
ATM
C1
C4
LSR
IP and ATM integration
Label Swapping Paradigm
MPLS
G-MPLS
SDH
Rapid and Predictable Restoration
Standard Time Division Multiplexing
SONET/SDH
Dynamic Allocation and Control?
10Gbps
OCX
OCX
Increasing
Capacity Requirements
DWDM
Dynamic Allocation and Control?

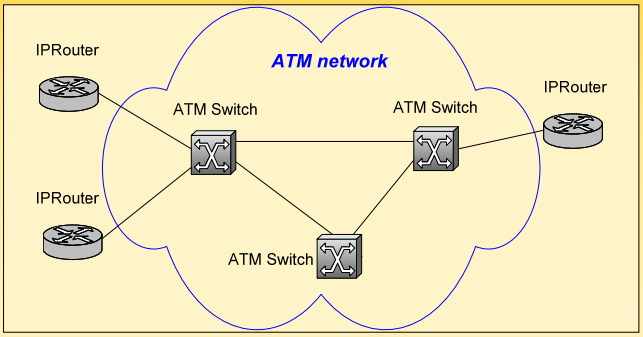
20


MPLS
21

21

IP networks over ATM

- IP routers are interconnected by an ATM network
- Connections between IP routers are implemented through virtual circuits (VCCs) or virtual paths (VPCs) on the ATM network
- It is necessary to manage two protocol layers
  (ATM is not available anymore on new networks)

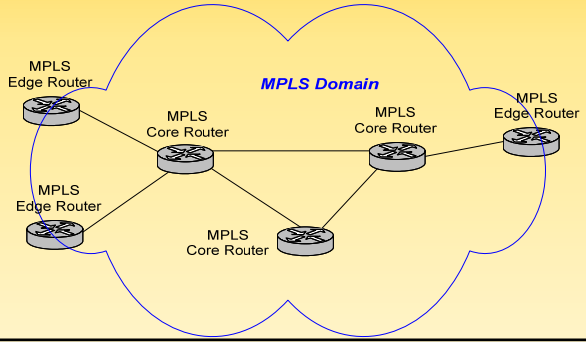


22

22

MPLS networks

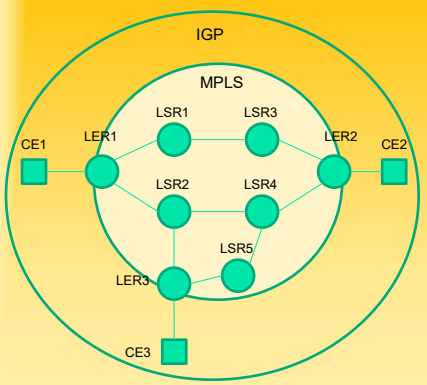
- Packets are labeled at the source with the label of the first hop
- Routers route packets based on their labels, just like ATM does with the VCI and VPI fields
- Advantages
  - Simplification of the packet routing process on routers
  - Traffic engineering capability equivalent to ATM
  - Simplification of the network management (a single protocol layer)

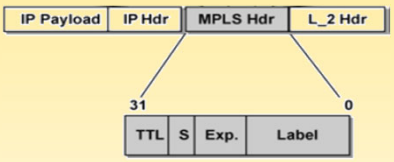


23

23

IP/MPLS





- **FEC – Forwarding Equivalence Class**
  - Identifies group of MPLS packets that are handled in the same way;
- **LABEL**
  - Local identifier based on the concept of FEC;
- **LSR – Label Switching Router**
  - LSRs in the same domain form an MPLS domain, using label forwarding;
- **LER – Label Edge Router**
  - LERs interacts with devices outside the domain, using IP technologies;

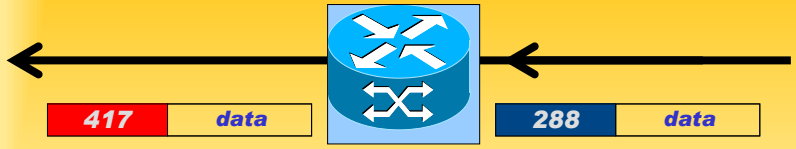
24

IP/MPLS

- **LSP – Label Switched Path**
  - Path along which FEC travels through an MPLS network;
- **LDP – Label Distribution Protocol**
  - The protocol used by MPLS for control that classifies FECs, distributes labels, and establish and maintains LSPs;
- **LFIB – Label Forwarding Information Base**
  - Forwarding table created by routing protocols and labels;
- **LSP tunneling**
  - Explicit route LSP between two LSRs that are not directly connected (indicated for TE);
- **Multi-level label stack**
  - Label stack carried by MPLS packets organized as LIFO (last-in first-out) stack;

25

Forwarding via Label Swapping

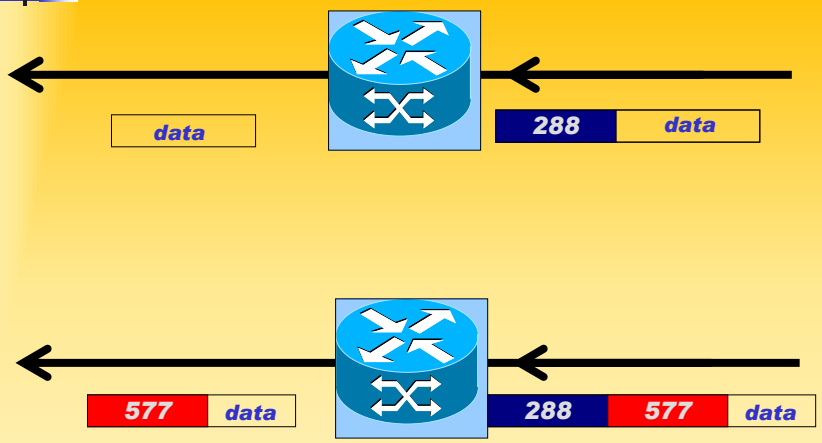


Labels are short, fixed-length values.

26

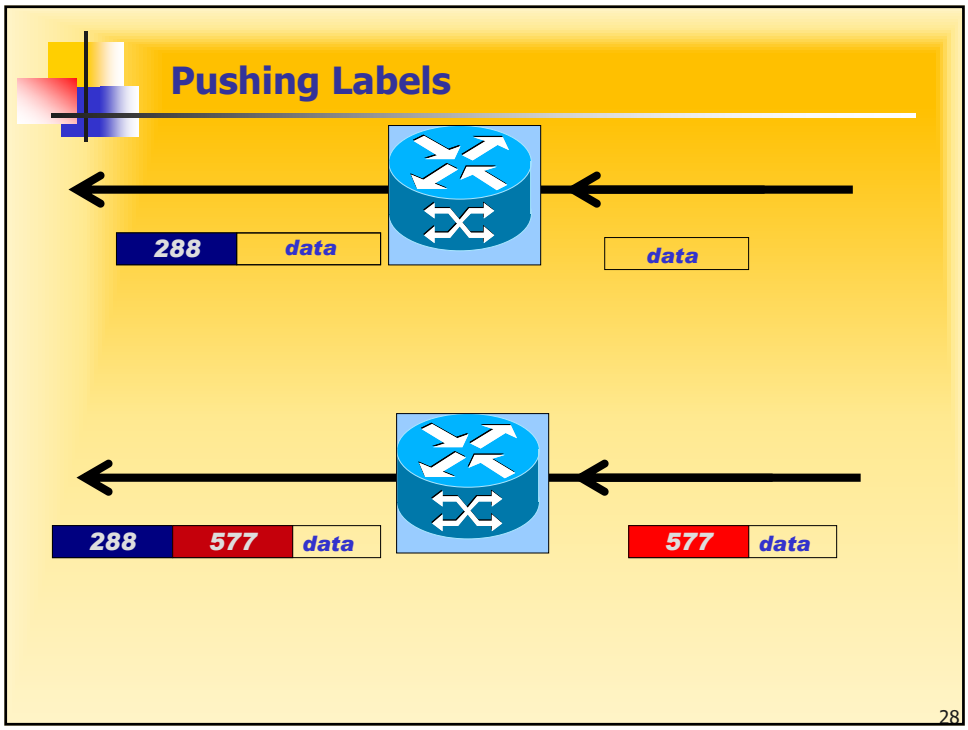
26

Popping Labels

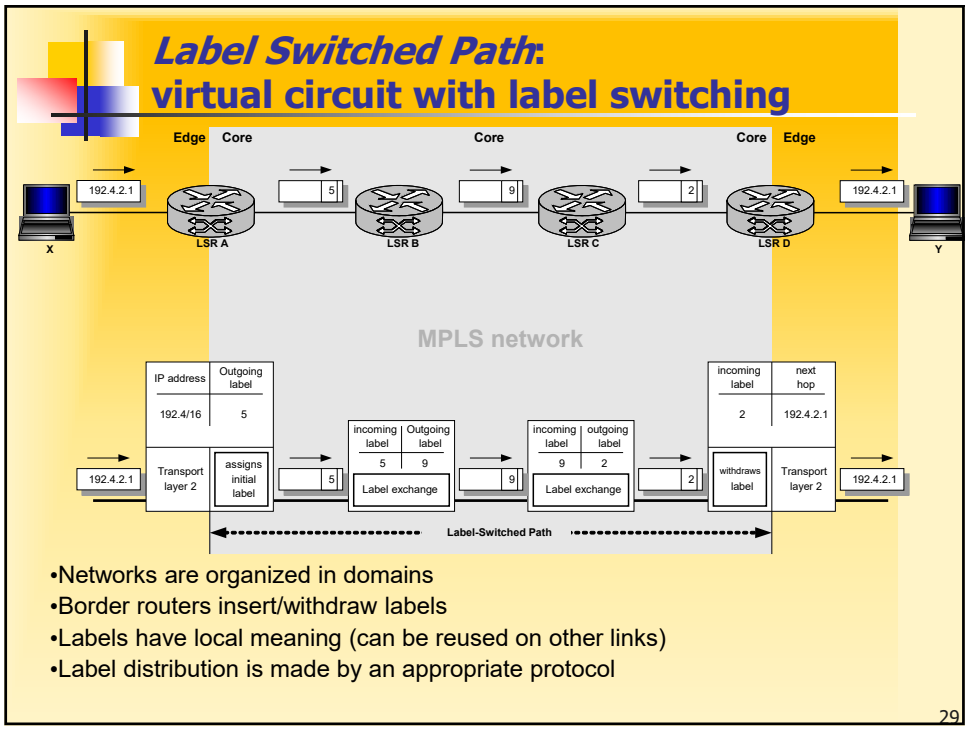


27

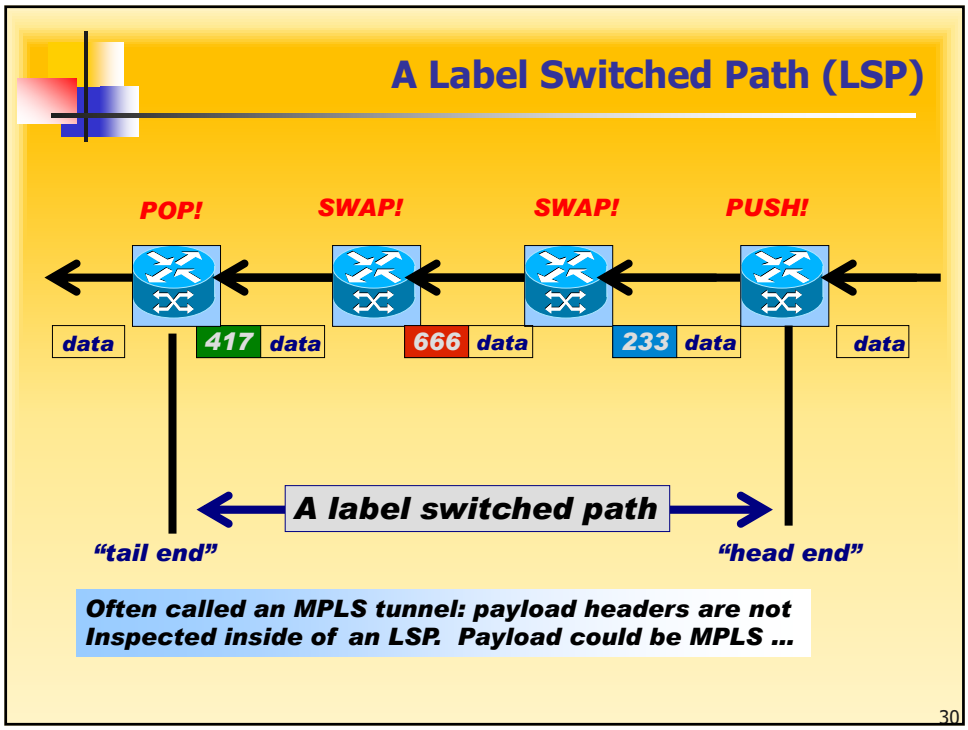
27


Pushing Labels
288
data
data
288
577
data
577
data
28

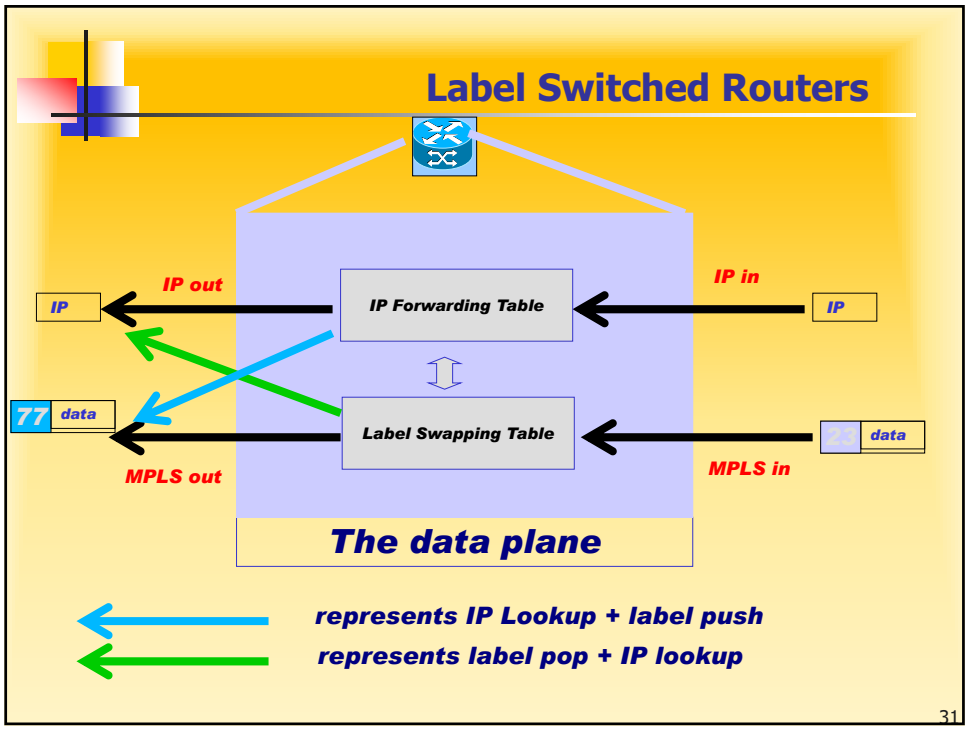
28


Label Switched Path:
virtual circuit with label switching
Edge
Core
Core
Core
Edge
192.4.2.1
5
9
2
192.4.2.1
X
LSR A
LSR B
LSR C
LSR D
Y
MPLS network
IP address
Outgoing label
192.4/16
5
incoming label
Outgoing label
5
9
incoming label
outgoing label
9
2
incoming label
next hop
2
192.4.2.1
192.4.2.1
Transport layer 2
assigns initial label
5
Label exchange
9
Label exchange
2
withdraws label
Transport layer 2
192.4.2.1
Label-Switched Path
•Networks are organized in domains
•Border routers insert/withdraw labels
•Labels have local meaning (can be reused on other links)
•Label distribution is made by an appropriate protocol
29

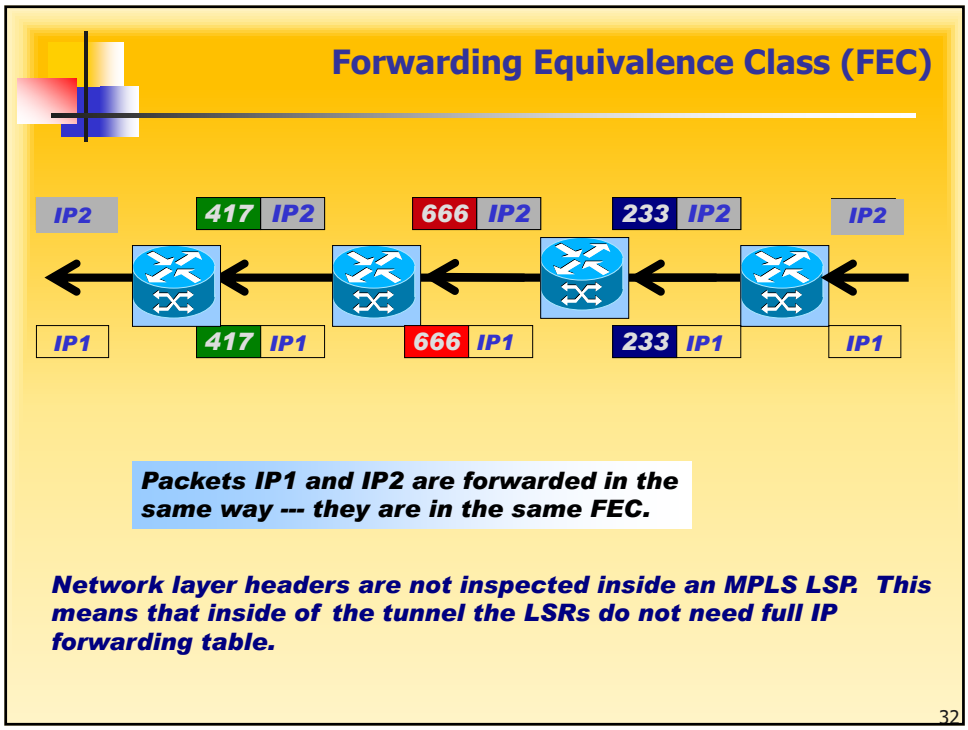
29


A Label Switched Path (LSP)
POP!
SWAP!
SWAP!
PUSH!
data
417 data
666 data
233 data
data
A label switched path
"tail end"
"head end"
Often called an MPLS tunnel: payload headers are not Inspected inside of an LSP. Payload could be MPLS ...
30

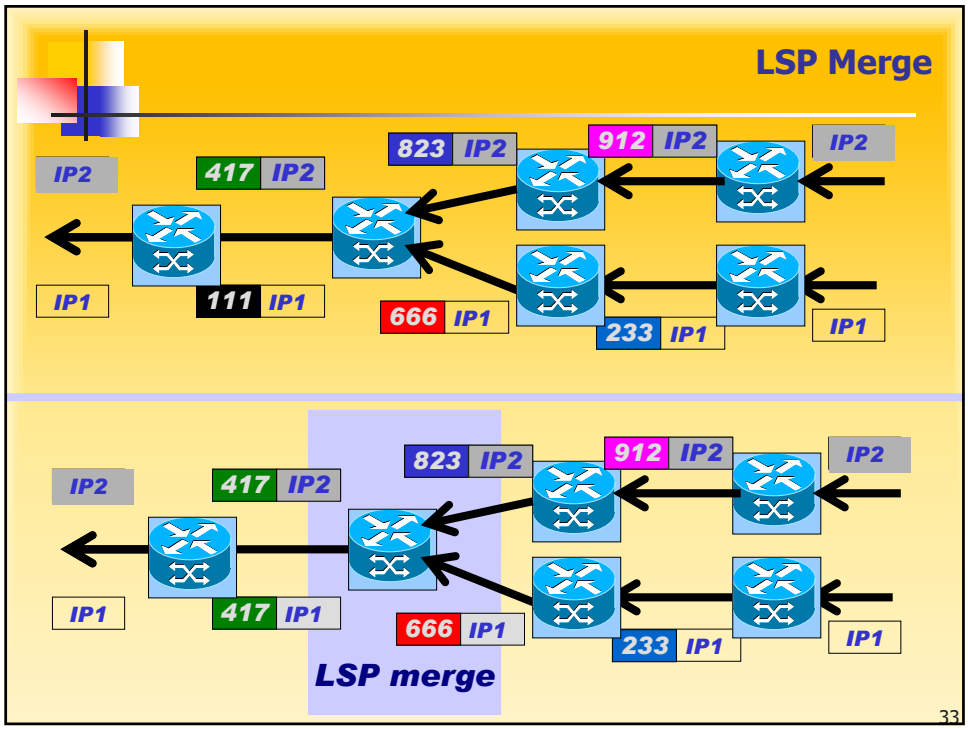
30


Label Switched Routers
IP out
IP in
IP
IP Forwarding Table
IP
77 data
Label Swapping Table
23 data
MPLS out
MPLS in
The data plane
represents IP Lookup + label push
represents label pop + IP lookup
31

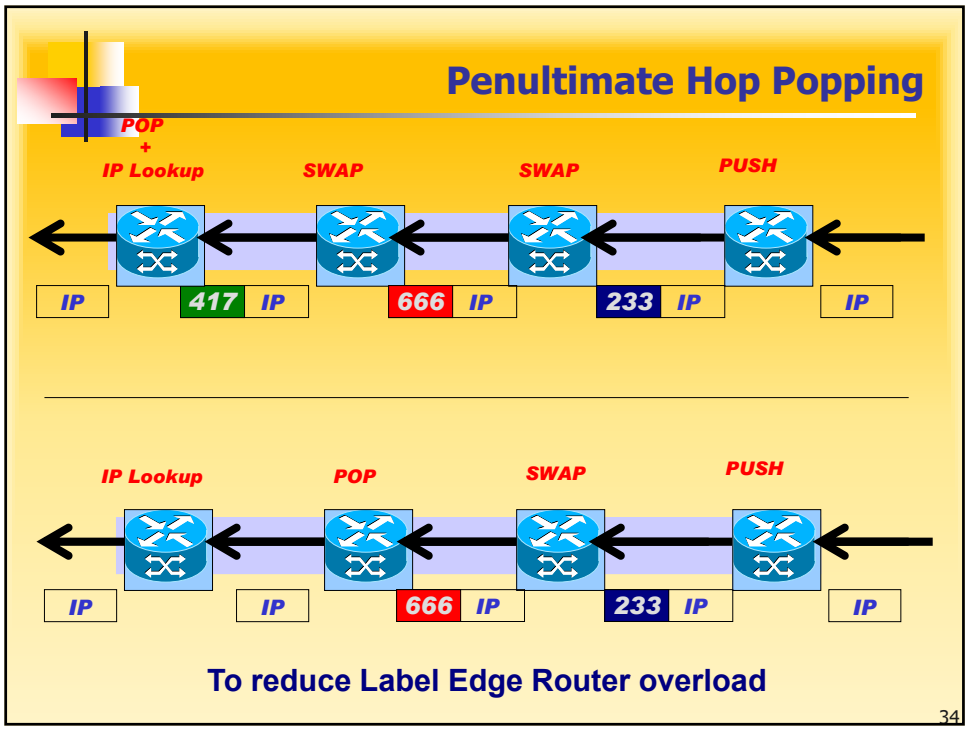
31


Forwarding Equivalence Class (FEC)
IP2
417 IP2
666 IP2
233 IP2
IP2
IP1
417 IP1
666 IP1
233 IP1
IP1
Packets IP1 and IP2 are forwarded in the same way --- they are in the same FEC.
Network layer headers are not inspected inside an MPLS LSP. This means that inside of the tunnel the LSRs do not need full IP forwarding table.
32

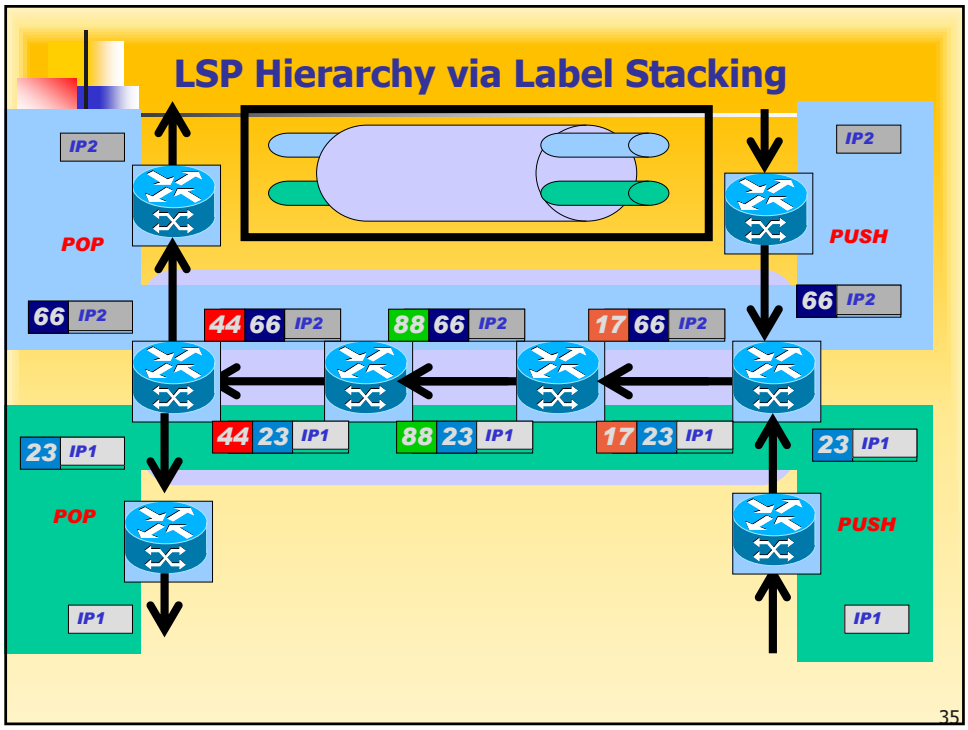
32


LSP Merge
IP2
417 IP2
823 IP2
912 IP2
IP2
IP1
111 IP1
666 IP1
233 IP1
IP1
IP2
417 IP2
823 IP2
912 IP2
IP2
IP1
417 IP1
666 IP1
233 IP1
IP1
LSP merge
33

33

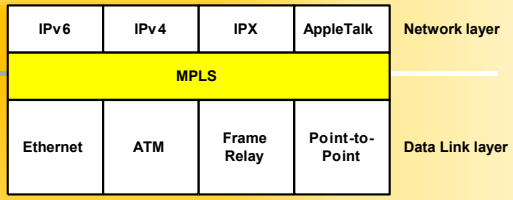

Penultimate Hop Popping
POP
+
IP Lookup
SWAP
SWAP
PUSH
IP
417 IP
666 IP
233 IP
IP
IP Lookup
POP
SWAP
PUSH
IP
IP
666 IP
233 IP
IP
To reduce Label Edge Router overload
34

34

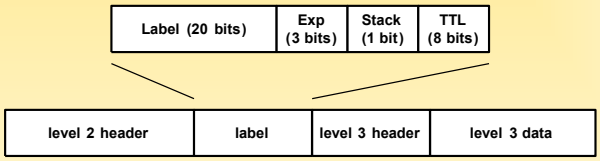

LSP Hierarchy via Label Stacking
IP2
POP
66 IP2
44 66 IP2
88 66 IP2
17 66 IP2
IP2
PUSH
66 IP2
23 IP1
44 23 IP1
88 23 IP1
17 23 IP1
23 IP1
POP
IP1
PUSH
IP1
35

35

Labels



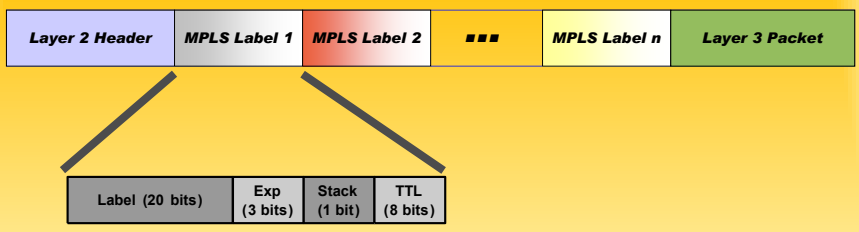
- On some Data Link (level 2) technologies, label is given by the appropriate fields of their header
  - ATM technology : VPI (Virtual Path ID) and VCI (Virtual Channel ID) fields
  - Frame Relay technology: DLCI (Data Link Connection Identifier) field
- On other Data Link technologies (Point-to-Point, Ethernet), the label is inserted between layer 2 and layer 3 headers



36

36

Generic MPLS Encapsulation



RFC 3032. MPLS Label Stack Encoding

- ***Label:*** ***Label Value, 20 bits***
- ***Exp:*** ***Experimental, 3 bits***
- ***S:*** ***Bottom of Stack, 1 bit***
- ***TTL:*** ***Time to Live, 8 bits***

37

37

IP/MPLS Network Establishment

- **Discovery link and topology**
  - IP routing table is built
    - LSRs and LERs use routing protocols to discover network topology eg. OSPF, ISIS, (BGP);
    - CEs advertise their addresses using routing protocols into MPLS cloud;
  - Forwarding Information Base (FIB) is built, initially without label information
- **Label Assignment**
  - FECs creation
    - LSRs classify with the same FEC all packets handled on the same way;
  - Allocate Labels
    - Every LSR allocates locally labels for every destination in the IP routing table (LIB and LFIB setup);

38

Label Distribution Protocols

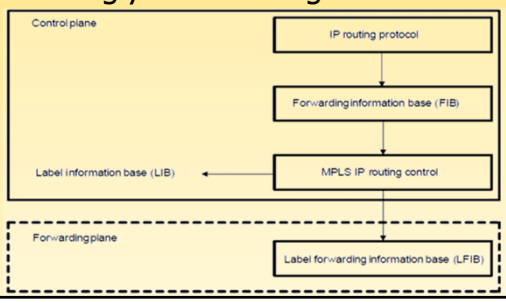
- Unconstrained routing
  - Label Distribution Protocol (LDP).
  - Path is chosen based on IGP shortest path.
- Constrained routing
  - Constrained by explicit path definition and/or performance requirements (e.g., available bandwidth).
  - Resource Reservation Protocol with Traffic Engineering (RSVP-TE).
    - Evolution of RSVP to support traffic engineering and label distribution.
  - Constrained based Routing LDP (CR-LDP).
    - Evolution of LDP to support constrained routing.
    - Deprecated!
- MPLS VPN scope
  - MP-BGP using address family VPN IPv4 and family specific MP_REACH_NLRI attribute.

39

39

IP/MPLS Network Establishment

- **Label distribution operation and LSP Establishment**
  - Discovery
    - Basic Discovery – LSRs send LDP link Hellos UDP (multicast) for directly connected peers.
    - Extended Discovery – LSRs send LDP targeted Hellos UDP for a specific (remote) IP peer.
  - Session Establish and Maintenance
    - TCP session is established and it is maintained through periodically Keep-Alive messages
  - LFIBs are established accordingly with routing and Label tables.



40

Label Distribution Protocol (LDP)

- Dynamic distribution of label binding information.
- LSR discovery.
- Reliable transport with TCP.
- Incremental maintenance of label swapping tables (only deltas are exchanged).
- Designed to be extensible with Type-Length-Value (TLV) coding of messages.
- Modes of behavior that are negotiated during session initialization
  - Label distribution control (ordered or independent).
  - Label retention (liberal or conservative).
  - Label advertisement (unsolicited or on-demand).

41

41

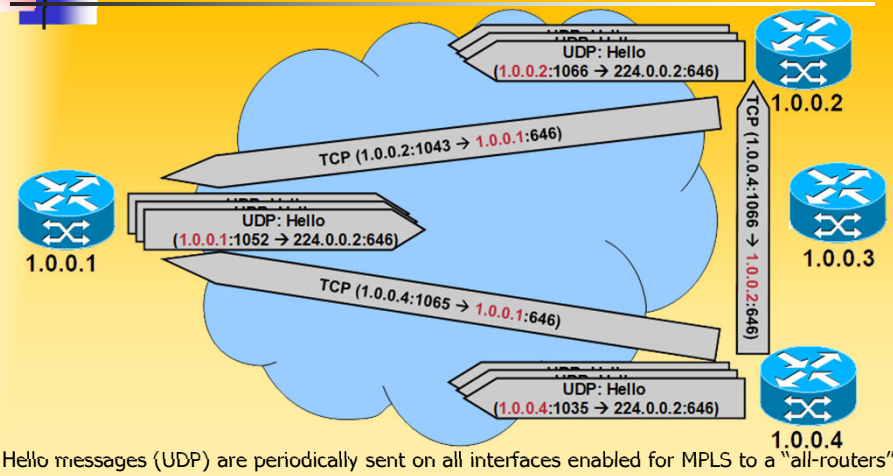
LDP Messages

- Discovery messages
  - Announce and maintain the presence of an LSR in a network.
  - **Hello Messages** (UDP) sent to "all-routers" multicast address.
  - Once neighbor is discovered, a LDP session is established over TCP.
- Session messages
  - Establish (**Initialization Message**) and maintain (**KeepAlive Message**) sessions between LDP peers.
- Advertisement messages
  - When a new LDP session is initialized and before sending label information an LSR advertises its interface addresses with one or more **Address Messages**.
  - An LSR withdraw previously advertised interface addresses with **Address Withdraw Messages**.
  - Create, change, and delete label mappings for FECs.
    - **Label Mapping, Label Request, Label Abort Request, Label Withdraw,** and **Label Release Messages.**
- Notification messages
  - Provide advisory information and to signal error information.

42

42

LDP Neighbour Discovery



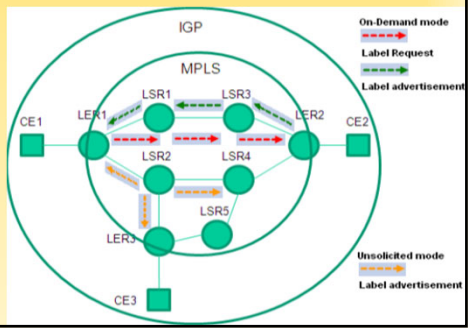
- Hello messages (UDP) are periodically sent on all interfaces enabled for MPLS to a "all-routers" multicast address (224.0.0.2).
- If there is another router on that interface it will respond by trying to establish a LDP/TCP session with the source of the hello messages.
- Both TCP and UDP messages use well-known LDP port number 646
- LDP Session is started by the router with higher IP address.

43

43

IP/MPLS Network Establishment

- LSP Establishment and Maintenance
  - Downstream On-demand mode– Upstream LSR sends a label request message (with FEC description) to its Downstream:
    - Ordered mode – a LSR only sends label (response) to its Upstream when it receives the label from its Downstream;
    - Independent mode – a LSR sends the label (response) when receives any request label;
  - Downstream Unsolicited mode – a Downstream LSR advertises label binding information to its Upstream LSR unsolicited after session to be established, without request ;
  - Label retention mode
    - Conservative mode
      - LSRs keeps only the labels from next hops
      - Indicated for limited label space
    - Liberal mode
      - LSRs keeps any labels, even if those are not from next hops
      - Indicated for quick adaptation of route changes

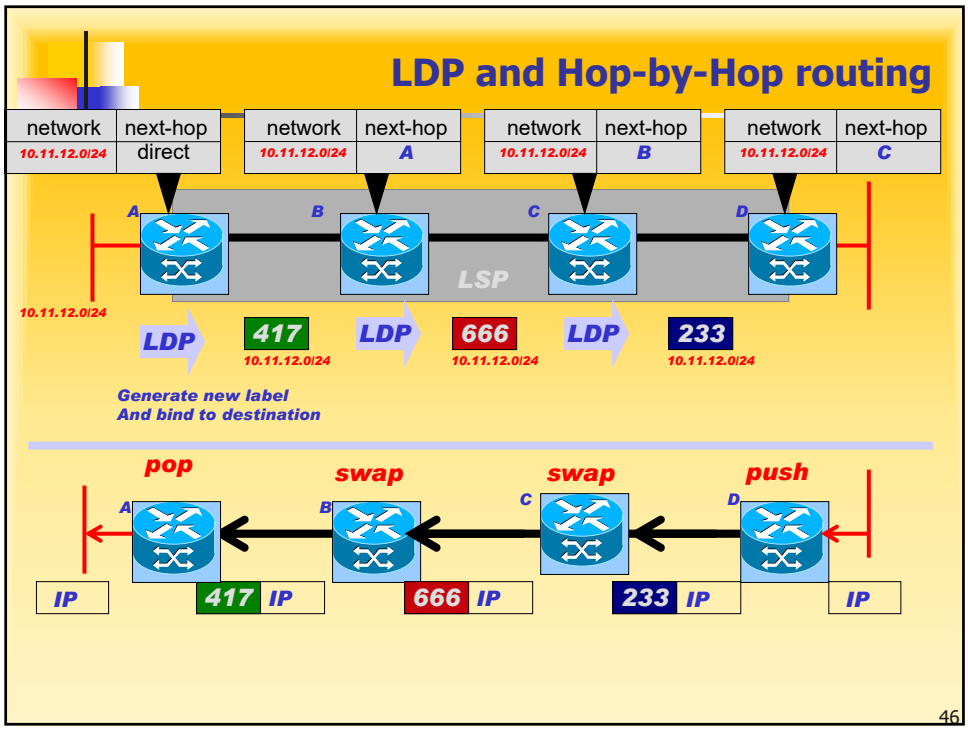


44

Multiprotocol Label Switching (MPLS) – TE usage

MSC Engenharia de Computadores e Telemática
1º ano, 1º semestre, 2023/2024

45



46

46

Constrained based Routing

- A cost is associated to each link
- Each link has a further set of attributes that represent performance metrics

The routing objective is to determine the lowest cost path that does not violate the restrictions that were assigned

- Restrictions can be associated to a set of performance characteristics, like for example, bandwidth, delay, priority, etc.
  - For the bandwidth case, the restriction that is imposed to the routing algorithm is that the path must have, on each connection it traverses, a bandwidth higher than a certain threshold.
  - In this case, the connection attribute used is the available bandwidth.

47

47

Constraint Based Routing?

Basic components

1. Specify path constraints
2. Extend topology database to include resource and constraint information
3. Find paths that do not violate constraints and optimize some metric
4. Signal to reserve resources along path
5. Set up LSP along path (with explicit route)
6. Map ingress traffic to the appropriate LSPs

Problem here: OSPF areas hide information for scalability. So these extensions work best only within an area...

Extend Link State Protocols (IS-IS, OSPF)

Extend RSVP or LDP or both!

Note: (3) could be offline (e.g. network operational center – NOC - systems) or online (perhaps an extension to OSPF)

Problem here: what is the "correct" resource model for IP services?

48

48

Resource Reservation + Label Distribution

Two emerging/competing/dueling approaches:

Add label distribution and explicit routes to a resource reservation protocol

RSVP-TE = **RSVP**

CR-LDP = **LDP**

Add explicit routes and resource reservation to a label distribution protocol

RSVP-TE: Extensions to RSVP for LSP Tunnels
draft-ietf-mpls-rsvp-lsp-tunnel-08.txt

Constraint-Based LSP Setup using LDP
draft-ietf-mpls-cr-ldp-05.txt

As of February 2003, the IETF MPLS working group deprecated CR-LDP and decided to focus purely on RSVP-TE.

RFC 3468: The Multiprotocol Label Switching (MPLS) Working Group decision on MPLS signalling protocols

49

49

LSPs establishing protocols

- RSVP-TE (*Resource Reservation Protocol – Traffic Engineering*)
  - Extension of the RSVP protocol
- CR-LDP (*Constrained based Routing – Label Distribution Protocol*)
  - Extension of the LDP protocol, deprecated

- Both protocols enable:
  - The specification of a route to a LSP
  - To chose the labels on each link of the route
  - To make resources reservation for the LSP
- Routes are previously determined:
  - By management (Traffic engineering), in a NOC
  - By a *Constrained based Routing* type protocol

50

50

RSVP-TE vs. CR-LPD

RSVP-TE	***CR-LDP***
• ***Soft state periodically refreshed***	• ***State maintained incrementally***
• ***IntServ QoS model***	• ***New QoS model derived from ATM models***

And the QoS model determines the additional information attached to links and nodes and distributed with extended link state protocols…

And what about that other Internet QoS model, diffserv?

51

51

Recall....
ReSerVation Protocol (RSVP)

- ReSerVation Protocol (RSVP) was developed to communicate resource needs between hosts and network devices
  - Associated to the Intserv QoS model
- RSVP allows:
  - The source to describe the characteristics of the IP packets flow.
  - Destinations to describe the reservation they want.
  - Routers to know how to process the packets flow in order to fulfil the requested reservation.
- Encapsulated on IP (protocol type = 46 (0x2E))
- Signalling is based on PATH and RESV messages.
  - PATH announces the traffic characteristics at the sender.
  - RESV achieves reservations that were initiated by the receivers.
  - If the reservation is not possible, a RESV ERR message is sent.
- The routers reservation states have to be periodically refreshed (soft states).

54

54

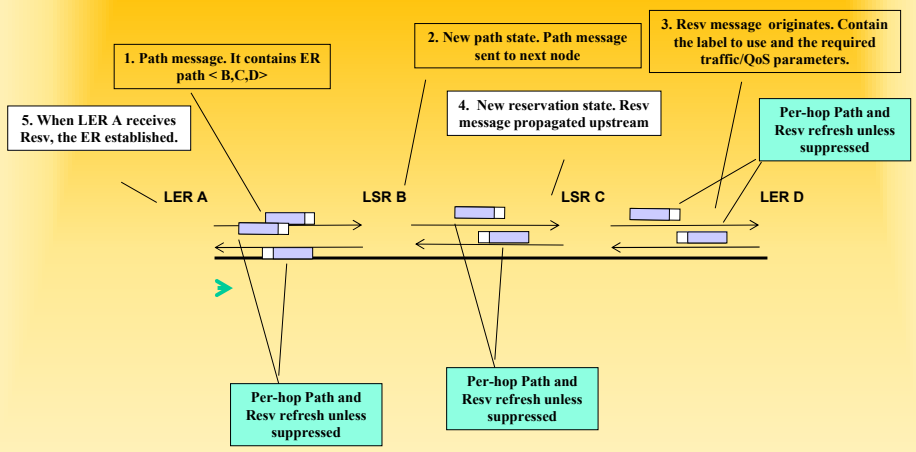
Resource Reservation Protocol with Traffic Engineering (RSVP-TE)

- Evolution of RSVP
  - RFC 3209: RSVP-TE: Extensions to RSVP for LSP Tunnels. (12/2001)
  - RFC 5151: Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions. (2/2008)
- To map traffic flows onto the physical network topology through label switched paths, resource and constraint network information are required
  - Provided by Extend Link State Protocols (IS-IS or OSPF with TE extensions).
    - RFC 3630: Traffic Engineering (TE) Extensions to OSPF Version 2. (9/2003)
    - RFC 5305: IS-IS Extensions for Traffic Engineering. (10/2008)

55

55

Establishment of a LSP with the RSVP-TE protocol



56

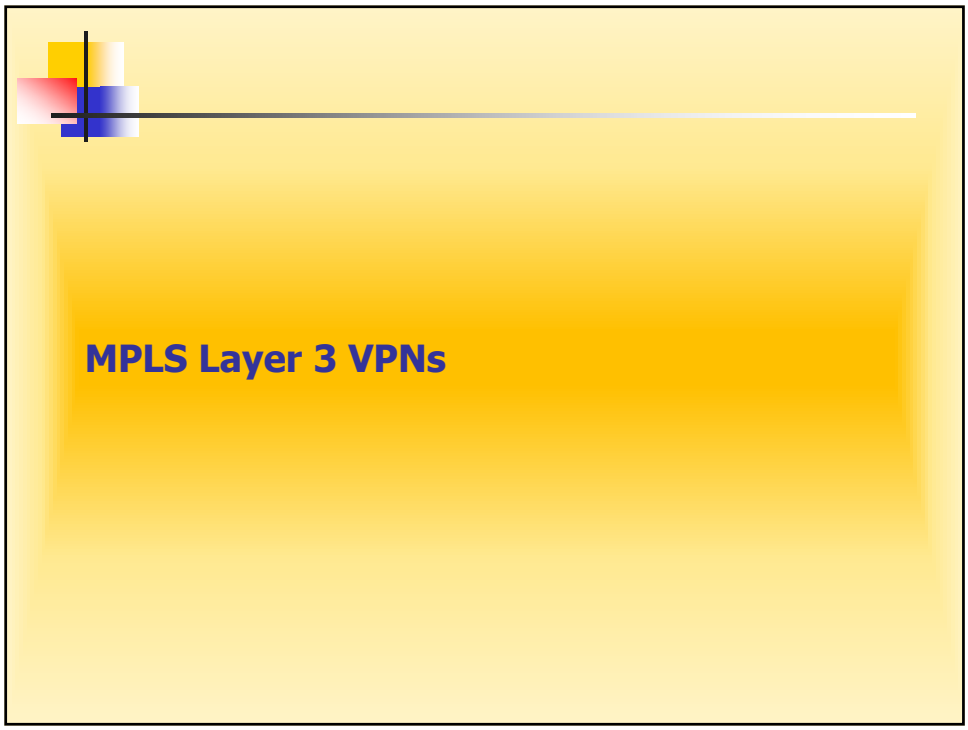
56

REVIEW: MPLS - Major Drivers

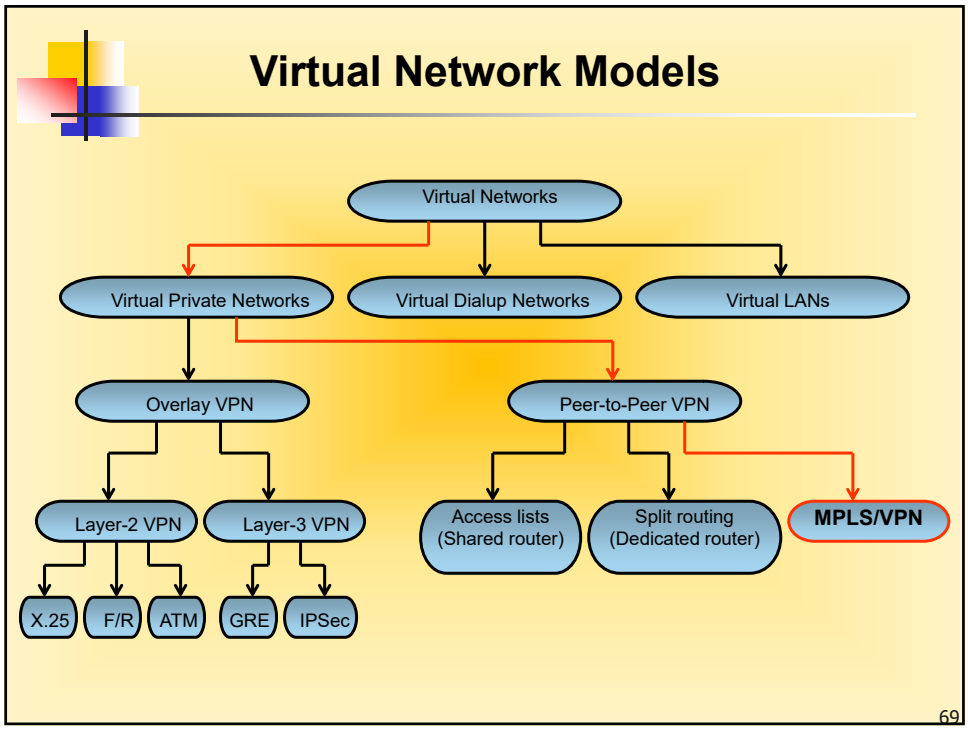
- Provide IP VPN Services *"the leased line"*
  - Scalable IP VPN service – Build once and sell many
  - Managed Central Services – Building value added services and offering them across VPNs
- Managing traffic on the network using MPLS Traffic Engineering
  - Providing tighter SLA/QoS (Guaranteed BW Services)
  - Protecting bandwidth - Bandwidth Protection Services
- Integrating Layer 2 & Layer 3 Infrastructure
  - Layer 2 services such as ATM (or Frame Relay) over MPLS
  - Mimic layer 2 services over a highly scalable layer 3 infrastructure

57

57


MPLS Layer 3 VPNs

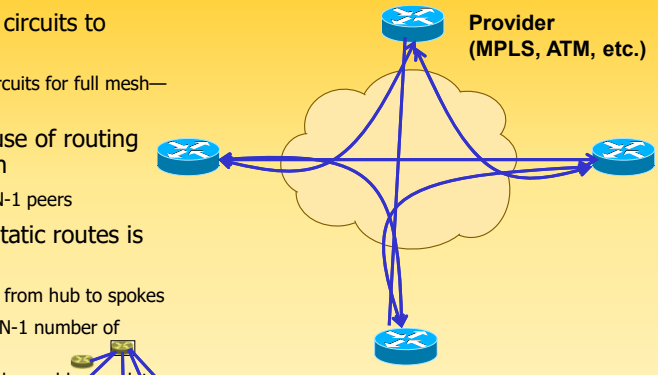
68

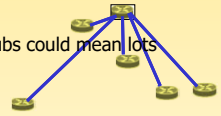

Virtual Network Models
Virtual Networks
Virtual Private Networks
Virtual Dialup Networks
Virtual LANs
Overlay VPN
Peer-to-Peer VPN
Layer-2 VPN
Layer-3 VPN
Access lists (Shared router)
Split routing (Dedicated router)
MPLS/VPN
X.25
F/R
ATM
GRE
IPSec
69

69

Overlay Network

- Provider sells a circuit service
  - customer wants to deploy its own network over those services
- Customers purchases circuits to connect sites, runs IP
  - N sites, (N*(N-1))/2 circuits for full mesh—expensive
- scalability issue because of routing peers in mesh approach
  - N sites, each site has N-1 peers
- Hub and spoke with static routes is simpler,
  - still buying N-1 circuits from hub to spokes
  - suffers from the same N-1 number of routing peers
  - Spokes distant from hubs could mean lots of long-haul circuits





70

70

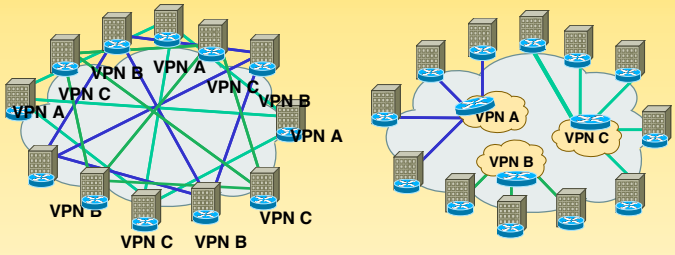
IP/MPLS Applications

- **MPLS-based VPNs**
  - MPLS L3VPN – VPN-IP over MPLS
    - VPN is a secure IP-based network between geographically dispersed sites that can communicate securely over a shared backbone;
    - MPLS VPNs provides the capability to deploy and administer scalable Layer 3 VPN backbone services to business customers
  - MPLS L2VPN – Any Technology (AT) or Transport over MPLS (e.g.: EoMPLS)
    - AT over MPLS transport Layer2 packets over MPLS network;
    - Allow the use of MPLS network to provide connectivity between customer sites with existing Layer2 networks;
  - MPLS-TE
    - Extends existing IP protocols and makes use of MPLS forwarding capabilities to provide TE
    - Brings explicit routing capabilities to MPLS networks

72

MPLS L3 VPNs using BGP (RFC2547)

- End user perspective
  - Virtual Private IP service
  - **Simple routing – just point default to provider**
  - Full site-site connectivity without the usual drawbacks (routing complexity, scaling, configuration, cost)
- Major benefit for provider – scalability

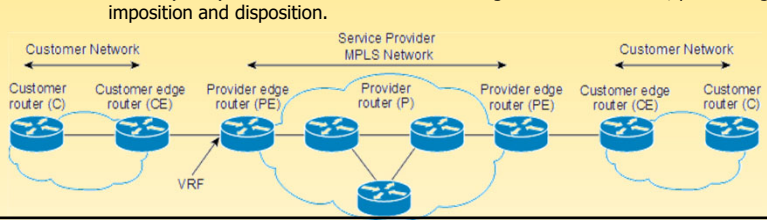


73

73

MPLS VPN Terminology

- Customer router (C) is connected only to other customer devices.
- Customer Edge (CE) router peers at Layer 3 to the Provider Edge (PE).
  - The PE-CE Interface runs either a dynamic routing protocol (eBGP, RIPv2, EIGRP, or OSPF) or has static routing (Static, Connected).
- Provider (P) router, resides in the core of the provider network.
  - Participates in the control plane for customer prefixes. The P router is also referred to as a Label Switch Router (LSR), in reference to its primary role in the core of the network, performing label switching/swapping of MPLS traffic.
- Provider Edge (PE) router, sits at the edge of the MPLS SP network.
  - In an MPLS VPN context, separate VRF routing tables are allocated for each user group.
  - Contains a global routing table for routes in the core SP infrastructure.
  - The PE is sometimes referred to as a Label Edge Router (LER) or Edge Label Switch Router (ELSR) in reference to its role at the edge of the MPLS cloud, performing label imposition and disposition.



74

74

VPN Routing and Forwarding Instance (VRF)

- PE routers maintain separate routing tables.
  - Virtual Routing and Forwarding (VRF) instance is separate from the global routing table that exists on PE routers

– Global routing table
  - Contains all PE and P routes (perhaps BGP)
  - Populated by the VPN backbone IGP

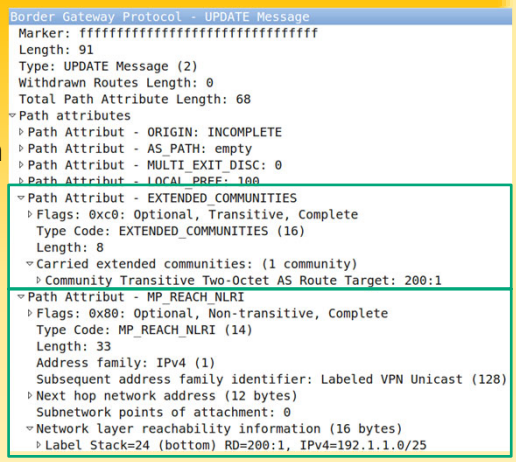
– VRF (VPN routing and forwarding)
  - Routing and forwarding table associated with one or more directly connected sites (CE routers)
  - VRF is associated with any type of interface, whether logical or physical (e.g. sub/virtual/tunnel)
  - Interfaces may share the same VRF if the connected sites share the same routing information
  - Routes are injected into the VRF from the CE-PE routing protocols for that VRF and any MP-BGP announcements that match the defined VRF.

75

Carrying VPN Routes in BGP

- VRFs by themselves aren't all that useful
  - Need some way to get the VRF routing information off the PE and to other PEs
  - This is done with BGP

- Additions to MP-BGP to Carry MPLS-VPN Info
  - RD: Route Distinguisher
  - RT: Route Target
  - VPNv4 address family
  - MPLS Label



76

76

Terminology, 1/2

- RR—Route Reflector
  - A router (usually not involved in packet forwarding) that distributes BGP routes within a provider's network
- PE—Provider Edge router
  - The interface between the customer and the MPLS-VPN network; only PEs (and maybe RRs) know anything about MPLS-VPN routes
- P—Provider router
  - A router in the core of the MPLS-VPN network, speaks LDP/RSVP but not necessarily VPNv4
- CE—Customer Edge router
  - The customer router which connects to the PE; does not know anything about labels, only IP (most of the time)
- LDP—Label Distribution Protocol
  - Distributes labels with a provider's network that mirror the IGP, one way to get from one PE to another
- LSP—Label Switched Path
  - The chain of labels that are swapped at each hop to get from one PE to another

77

77

Terminology, 2/2

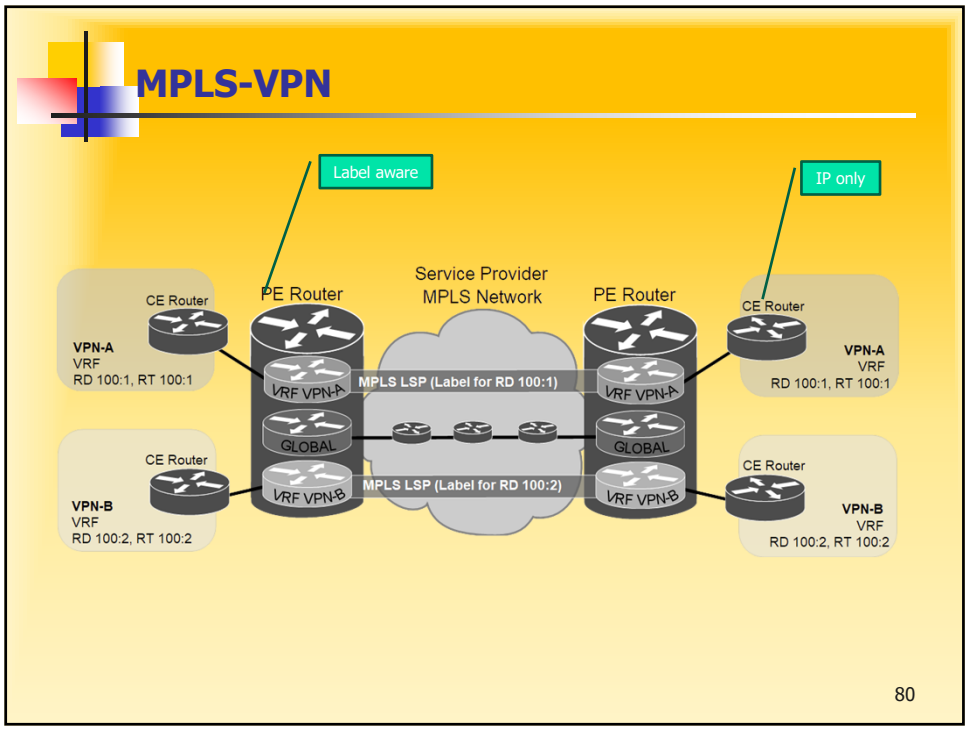
- VPN—Virtual Private Network
  - A network deployed on top of another network, where the two networks are separate and never communicate
- VRF—Virtual Routing and Forwarding instance
  - Mechanism in IOS used to build per-interface route and forwarding information bases (RIB and FIB)
  - VRF exports and imports one or more RT (route targets).
- VPNv4
  - Address family used in BGP to carry MPLS-VPN routes
- RD
  - Route Distinguisher, used to uniquely identify the same network/mask from different VRFs (i.e., 10.0.0.0/8 from VPN A and 10.0.0.0/8 from VPN B)
  - objective: make routes unique, hide routes from different customers
- RT
  - Route Target, used to control import and export policies, to build arbitrary VPN topologies for customers
  - exported RTs can be carried in BGP

Example:

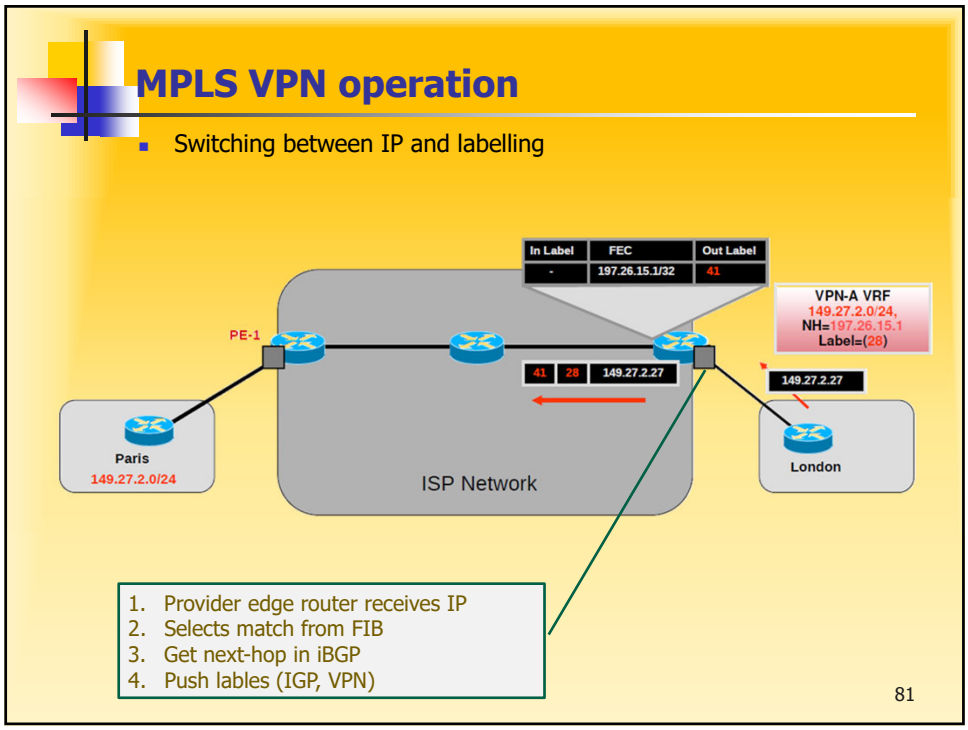
```
ip vrf VPN-A
rd 100:1
route-target export 100:1
route-target import 100:1
```

78

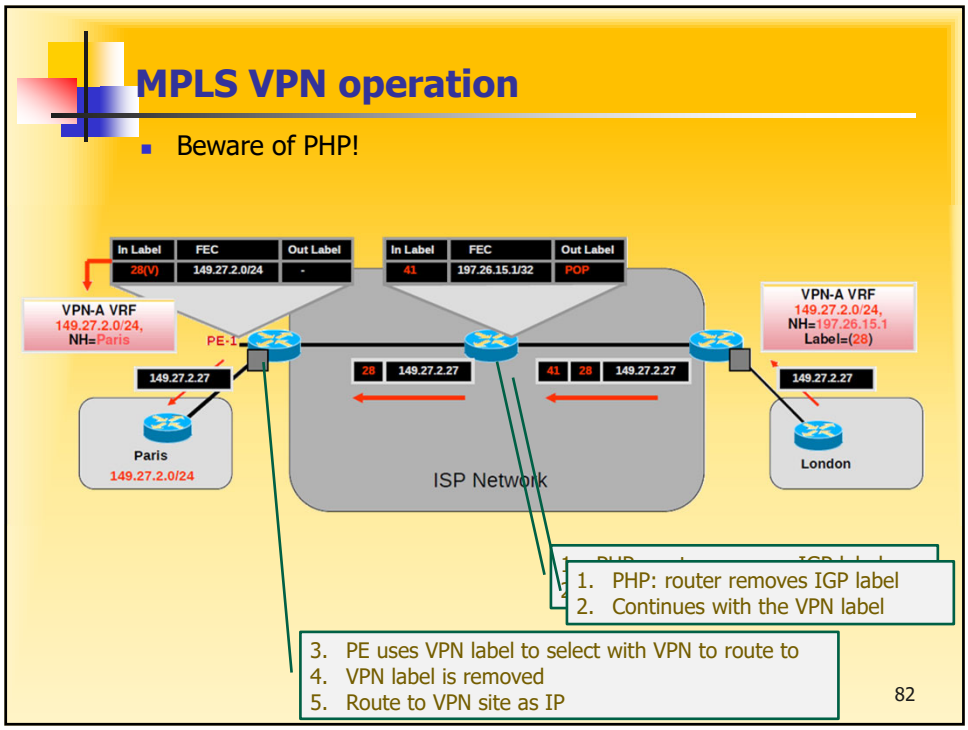
78


MPLS-VPN
Label aware
IP only
Service Provider
MPLS Network
CE Router
PE Router
PE Router
CE Router
VPN-A
VRF
RD 100:1, RT 100:1
VRF VPN-A
MPLS LSP (Label for RD 100:1)
GLOBAL
MPLS LSP (Label for RD 100:2)
VRF VPN-B
CE Router
VPN-B
VRF
RD 100:2, RT 100:2
80

80


MPLS VPN operation
Switching between IP and labelling
In Label
FEC
Out Label
-
197.26.15.1/32
41
VPN-A VRF
149.27.2.0/24,
NH=197.26.15.1
Label=(28)
PE-1
41
28
149.27.2.27
149.27.2.27
Paris
149.27.2.0/24
ISP Network
London
1. Provider edge router receives IP
2. Selects match from FIB
3. Get next-hop in iBGP
4. Push lables (IGP, VPN)
81

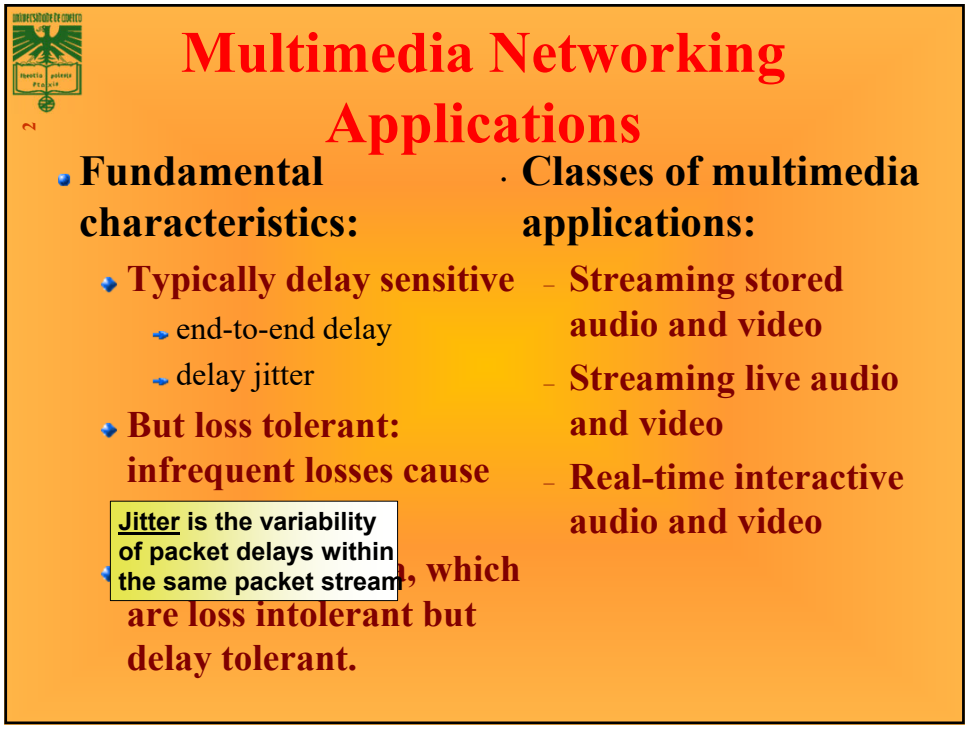
81


MPLS VPN operation
Beware of PHP!
In Label
FEC
Out Label
28(V)
149.27.2.0/24
-
In Label
FEC
Out Label
41
197.26.15.1/32
POP
VPN-A VRF
149.27.2.0/24,
NH=Paris
PE-1
VPN-A VRF
149.27.2.0/24,
NH=197.26.15.1
Label=(28)
149.27.2.27
28
149.27.2.27
41
28
149.27.2.27
149.27.2.27
Paris
149.27.2.0/24
ISP Network
London
1. PHP: router removes IGP label
2. Continues with the VPN label
3. PE uses VPN label to select with VPN to route to
4. VPN label is removed
5. Route to VPN site as IP
82

82

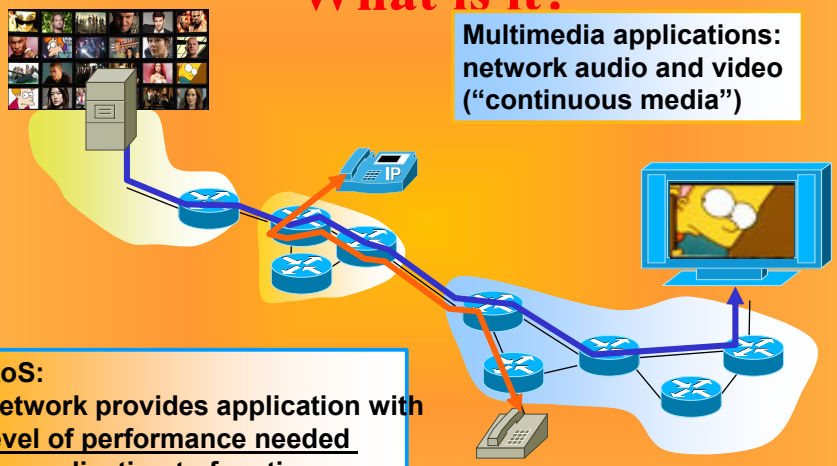

Bringing it together
Multimedia in IP
(Web view)

1


Multimedia Networking Applications
Fundamental characteristics:
Typically delay sensitive
end-to-end delay
delay jitter
But loss tolerant: infrequent losses cause
Jitter is the variability of packet delays within the same packet stream
, which are loss intolerant but delay tolerant.
Classes of multimedia applications:
Streaming stored audio and video
Streaming live audio and video
Real-time interactive audio and video

2

Multimedia, Quality of Service: What is it?



Multimedia applications: network audio and video ("continuous media")

QoS:
Network provides application with level of performance needed for application to function.

3

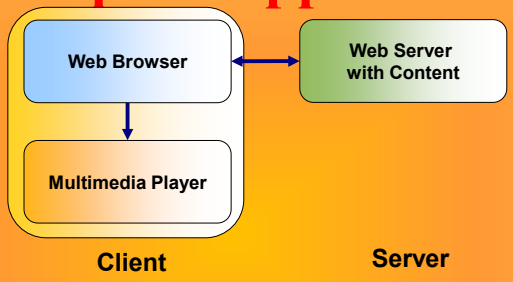
Internet Multimedia Support

- Integrated services philosophy.
  - Requires dedicated links/channels with QoS requirements.
- Differentiated services philosophy.
  - Fewer changes to Internet infrastructure.
- Best effort.
  - No major changes.
  - More bandwidth when needed.
  - Application-level control and distribution.

Would require QoS
Only possible in private networks or operator infrastructure

4

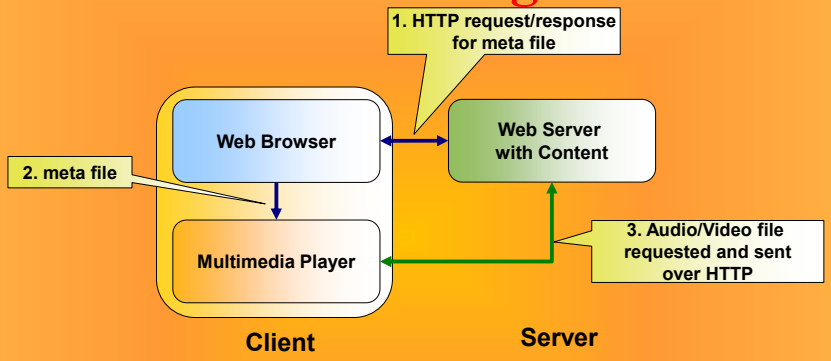
Internet Multimedia: Simplest Approach



- Audio or video stored in file.
- Files transferred as HTTP object (or using P2P).
  - Received in entirety at client as a file.
  - Then passed to default player in client.
- Audio&video is not streamed!
- No "pipelining", long delays until playout!

5

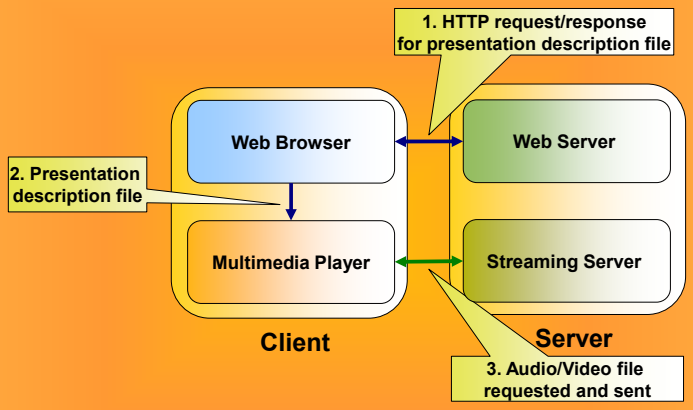
Internet Multimedia: Web Streaming



- Browser GETs metafile.
  - Content negotiation may happen.
- Browser launches player, passing metafile
- Player contacts server.
- Server streams audio/video to player.

6

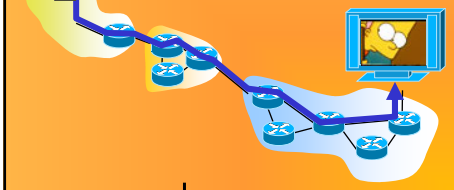
Streaming from a streaming server



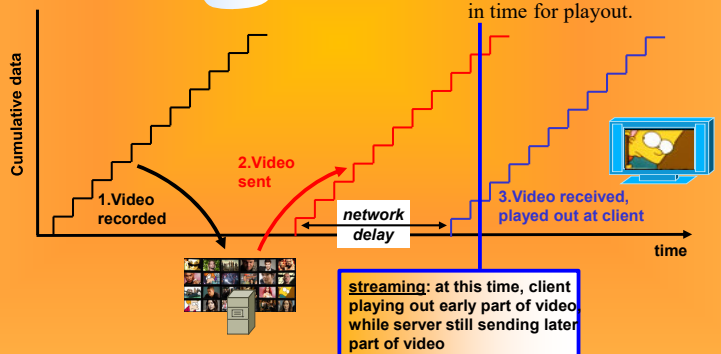
- This architecture allows for non-HTTP protocol between server and media player.
- Can use UDP or TCP transport.

7

Web Streaming: Stored Multimedia

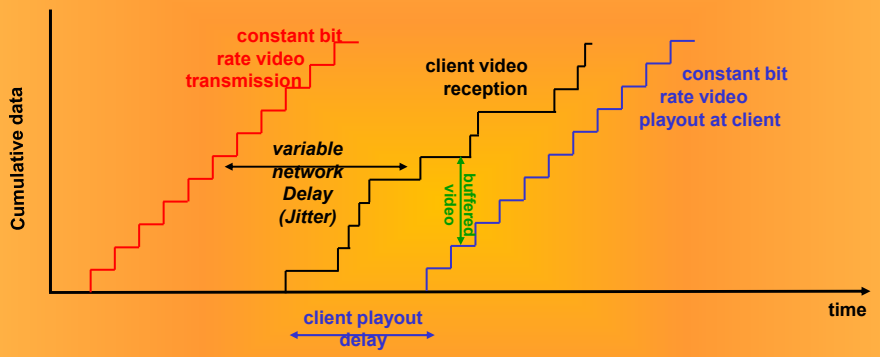


- Streaming:
  - Media stored at source is transmitted to client.
  - Client visualization begins before all data has arrived.
    - Timing constraint for transmitted data: in time for playout.



8

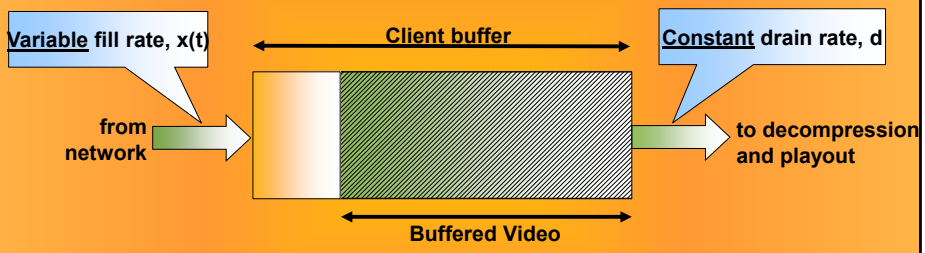
Streaming Multimedia: Client Buffering



- Client-side buffering, playout delay compensate for network-added delay, delay jitter.
  - Size of buffer/playout delay are parameters that can be adjusted dynamically
  - For VoIP, delay between 2 packets is about 20ms (plus *jitter*).

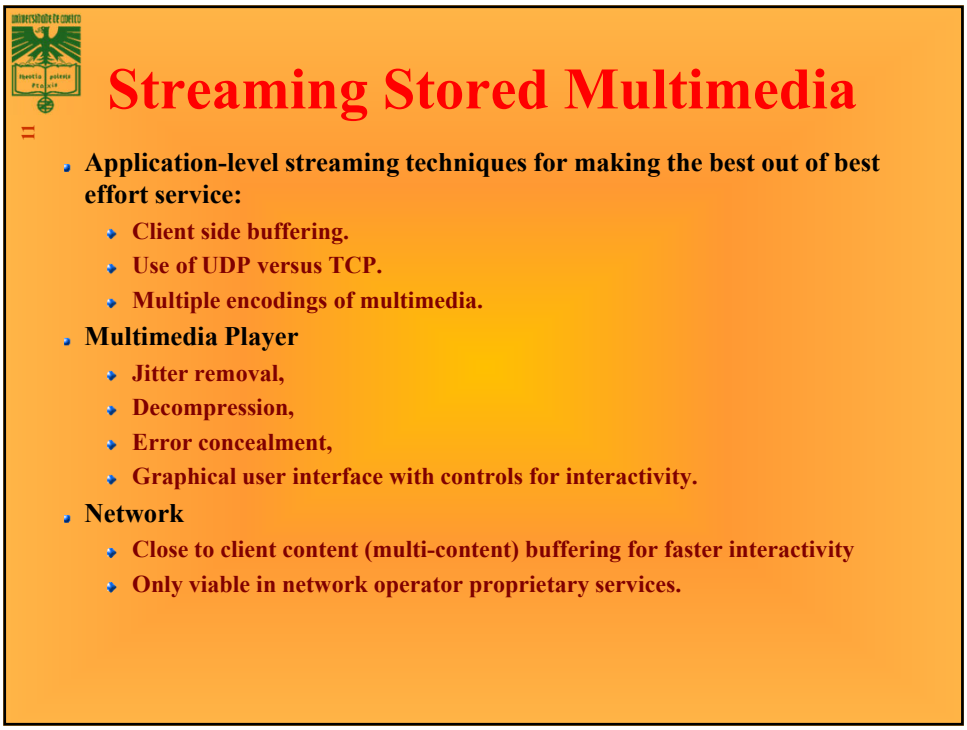
9

Streaming Multimedia: Client Buffering

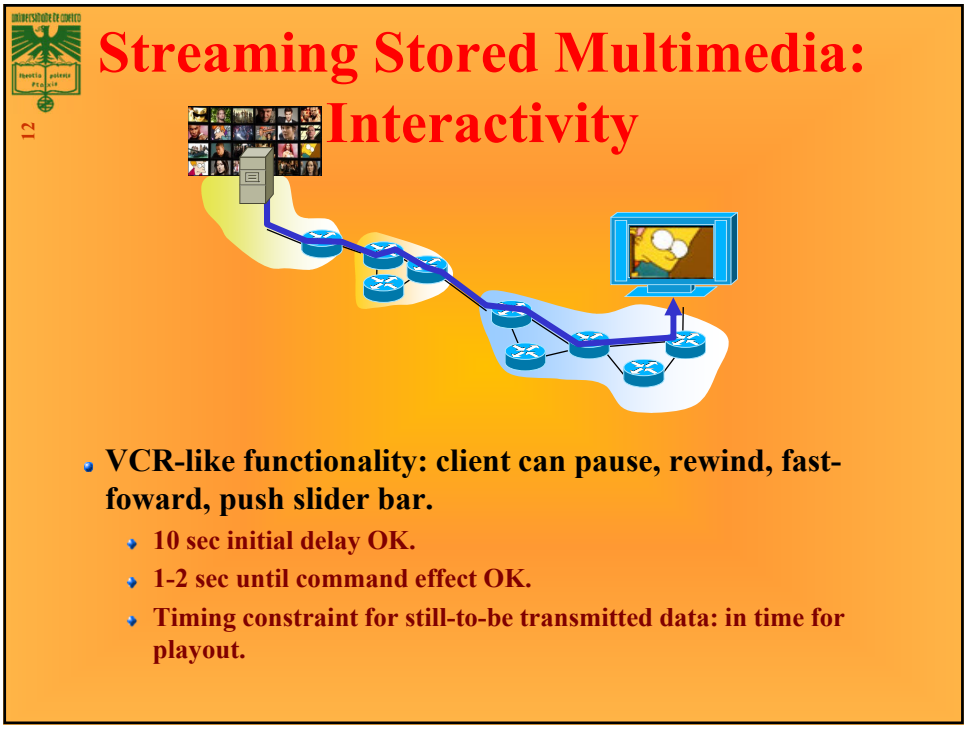


- Client-side buffering, playout delay compensate for network-added delay, delay jitter.
  - Size of buffer/playout delay are parameters that can be adjusted dynamically

10


11
Streaming Stored Multimedia
Application-level streaming techniques for making the best out of best effort service:
Client side buffering.
Use of UDP versus TCP.
Multiple encodings of multimedia.
Multimedia Player
Jitter removal,
Decompression,
Error concealment,
Graphical user interface with controls for interactivity.
Network
Close to client content (multi-content) buffering for faster interactivity
Only viable in network operator proprietary services.

11


12
Streaming Stored Multimedia: Interactivity
VCR-like functionality: client can pause, rewind, fast-foward, push slider bar.
10 sec initial delay OK.
1-2 sec until command effect OK.
Timing constraint for still-to-be transmitted data: in time for playout.

12

13

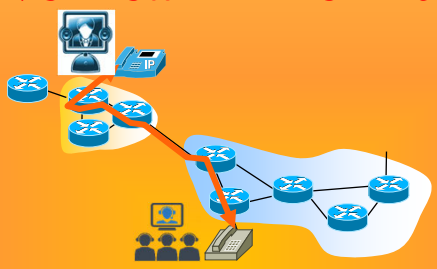
Streaming Live Multimedia

- Examples:
  - Internet TV/radio show.
  - Live sporting event.
- Streaming
  - Playback buffer.
  - Playback can lag tens of seconds after transmission.
  - Still have timing constraint.
- Interactivity
  - Fast forward impossible.
  - Rewind, pause possible!

13

14

Interactive Real-Time Multimedia



- Applications:
  - IP telephony, video conference, online-game multimedia actions, distributed interactive worlds.
- End-end delay requirements:
  - Audio: < 150 msec good, < 400 msec OK
    - Includes application-level (packetization) and network delays.
    - Higher delays noticeable, impair interactivity.
- Requires session initialization
  - Advertise its IP address, port number, encoding algorithms, required contents, available contents

14

15

UDP Streaming vs. TCP Streaming

- **UDP**
  - **Server sends at rate appropriate for client .**
    - Often send rate = encoding rate = constant rate.
    - Then, fill rate = constant rate - packet loss.
  - **Short playout delay (2-5 seconds) to compensate for network delay jitter.**
  - **Error recover: time permitting.**
- **TCP**
  - **Send at maximum possible rate under TCP.**
  - **Fill rate fluctuates due to TCP congestion control.**
  - **Larger playout delay: smooth TCP delivery rate.**
  - **HTTP/TCP passes more easily through firewalls.**

15

16

HTTP/TCP Streaming

- **Multiple versions with distinct/complementary characteristics are generated for the same content**
  - **With different bitrates, resolutions, frame rates.**
- **Each version is divided into time segments.**
  - **e.g., two seconds.**
- **Each segment is provided on a web server and can be retrieved through standard HTTP GET requests.**
- **Examples of protocols:**
  - **MPEG's Dynamic Adaptive Streaming over HTTP (DASH).**
    - Standard ISO/IEC 23009-1. YouTube's default.
  - **Adobe HTTP Dynamic Streaming (HDS).**
  - **Apple HTTP Live Streaming (HLS).**
  - **Microsoft Smooth Streaming (MSS).**

16

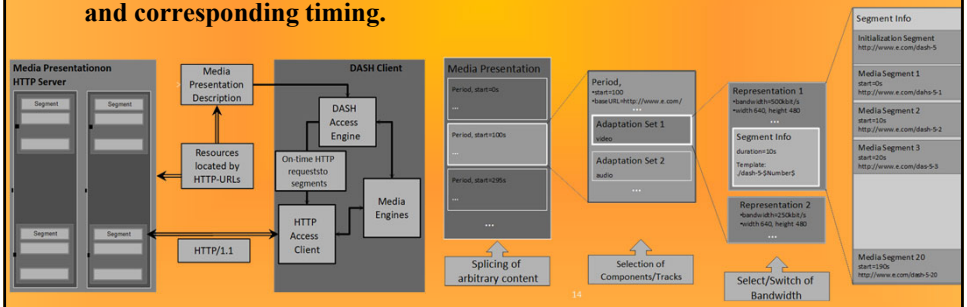
User Control of Streaming Media: RTSP

- RTSP (Real Time Streaming Protocol): RFC 2326
  - Client-server application layer protocol.
  - For user to control display: rewind, fast forward, pause, resume, repositioning, etc…
- Does not define how audio/video is encapsulated for streaming over network.
- Does not restrict how streamed media is transported.
  - Can be transported over UDP or TCP.
- Does not specify how the media player buffers audio/video.
- RTSP messages are also sent out-of-band:
  - RTSP control messages use different port numbers than the media stream: out-of-band
    - Port 554
  - The media stream is considered "in-band"

17

Dynamic Adaptive Streaming over HTTP (DASH)

- Developed to be an Open Standard Delivery Format.
  - MPEG DASH ISO/IEC 23009-1.
- Video streaming solution where pieces of video streams/files are requested with HTTP and spliced together by the client.
  - Client entirely controls delivery.
- Media Presentation Description (MPD) describes accessible Segments and corresponding timing.



22

23

WebRTC

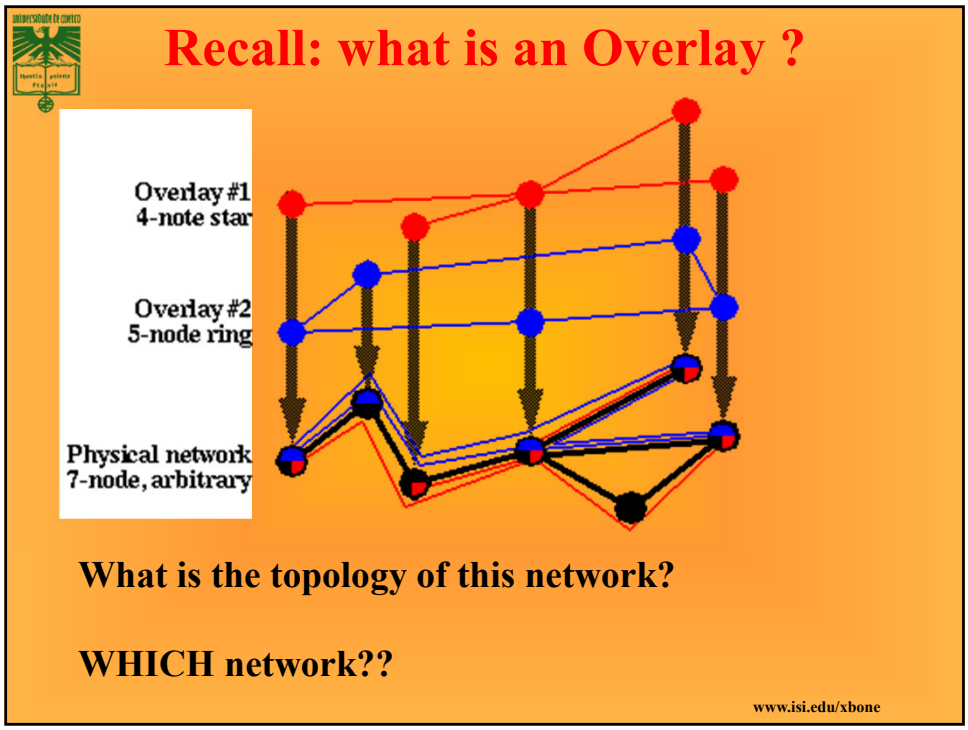
- Peer-to-peer connections.
  - An instance allows an application to establish peer-to-peer communications with another instance in another browser, or to another endpoint implementing the required protocols.
- RTP Media.
  - Allow a web application to send and receive media stream over a peer-to-peer connection (discussed in a minute)
- Peer-to-peer Data
  - Allows a web application to send and receive generic application data over a peer-to-peer connection.
- Peer-to-peer DTMF.

23

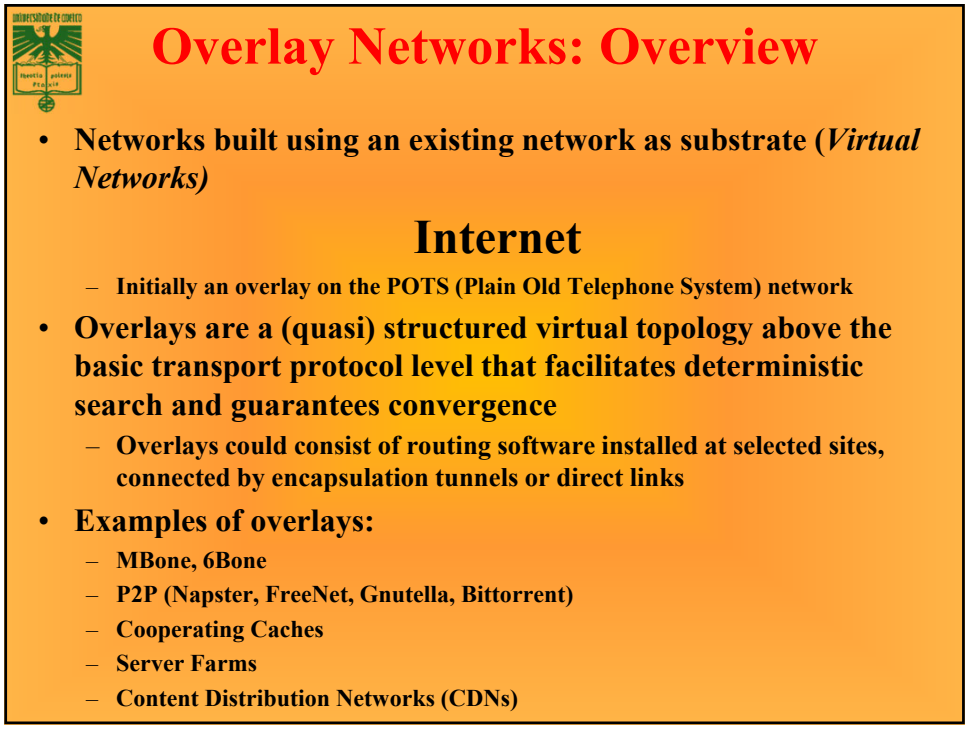
CDNs

Everyone in the same network ?

24


Recall: what is an Overlay ?
Overlay #1
4-note star
Overlay #2
5-node ring
Physical network
7-node, arbitrary
What is the topology of this network?
WHICH network??
www.isi.edu/xbone

26


Overlay Networks: Overview
• Networks built using an existing network as substrate (*Virtual Networks)*
Internet
– Initially an overlay on the POTS (Plain Old Telephone System) network
• Overlays are a (quasi) structured virtual topology above the basic transport protocol level that facilitates deterministic search and guarantees convergence
– Overlays could consist of routing software installed at selected sites, connected by encapsulation tunnels or direct links
• Examples of overlays:
– MBone, 6Bone
– P2P (Napster, FreeNet, Gnutella, Bittorrent)
– Cooperating Caches
– Server Farms
– Content Distribution Networks (CDNs)

27

Content Distribution Networks

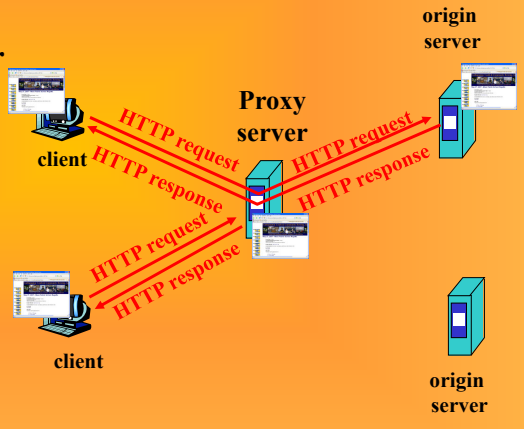
Client-Server and distribution models
Caching and load balancing

28

(recall FR): Web caches (proxy server)

Goal: satisfy client request without involving origin server

- user sets browser: Web accesses via proxy server
- browser sends all HTTP requests to proxy
  - object in cache: cache returns object
  - else proxy requests object from origin server, then returns object to client



30

More about Web caching

- Proxy server acts as both client and server
- typically proxy server is installed by ISP (university, company, residential ISP)

Why Web caching?

- reduce response time for client request
- reduce traffic on an institution's access link.

31

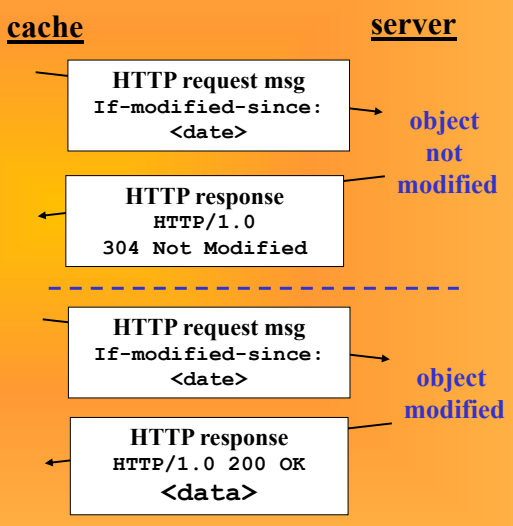
Conditional GET

- **Goal:** don't send object if cache has up-to-date cached version
- cache: specify date of cached copy in HTTP request

```
If-modified-since:
  <date>
```

- server: response contains no object if cached copy is up-to-date:

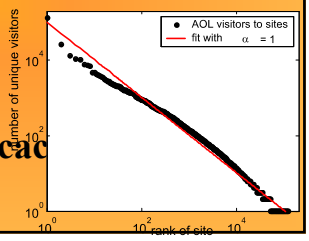
```
HTTP/1.0 304 Not
  Modified
```



36

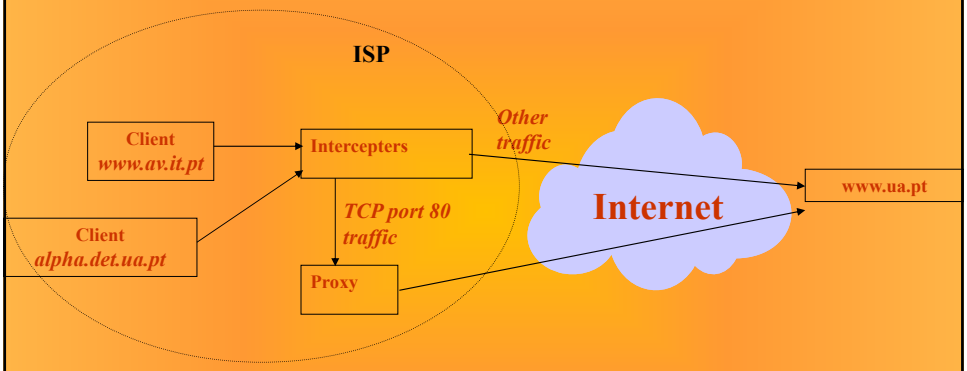
Optimizing performance

- **Where to cache content?**
  - **Popularity of Web objects is Zipf-like**
    - a few elements that score *very* high (the left tail in the diagrams)
    - a medium number of elements with middle-of-the-road scores (the middle part of the diagram)
    - a huge number of elements that score very low (the right tail in the diagram)
  - **Small number of sites cover large fraction of requests**
- **Given this observation, how should cac replacement work?**



37

Potential content network structure: Caching Proxies

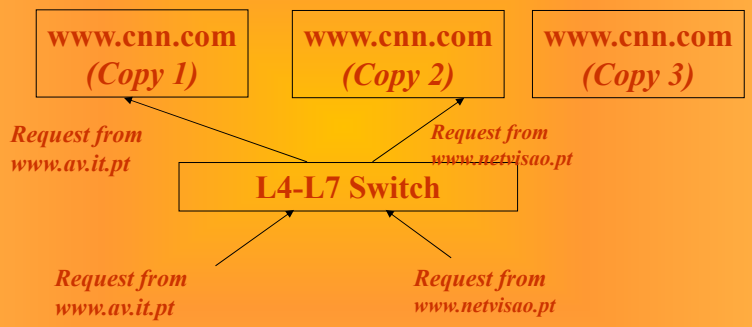


- **Mostly motivated by ISP business interests – reduction in bandwidth consumption of ISP from the Internet**
  - **Reduced network traffic**
  - **Reduced user perceived latency**

41

Potential content network structure : Server Farms

Simple solution to the content distribution problem: deploy a large group of servers on site

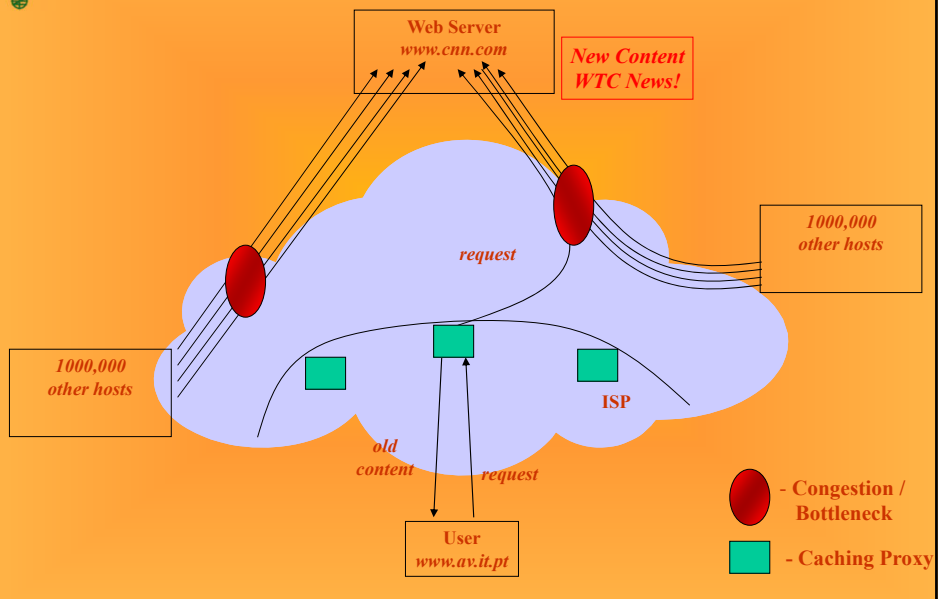


- **Arbitrate client requests to servers using an "intelligent" L4-L7 switch**
  - **Quite used today**

42

Flash Crowds:

Consider, On September 11, 2001



43

Why Not Web-only approaches for content networks?

- Integrating file caching in proxies
  - Optimized for 10KB objects
  - 10GB = 1.000.000 x 10KB
- Memory pressure
  - Disk access is 1000 times slower
  - Working sets do not fit in memory
- Waste of resources
  - More servers needed
  - Provisioning is a must

Problems with *Server farms and Caching proxies*

- Server farms do nothing about problems due to network congestion, or to improve latency issues due to the network
- Caching proxies serve only their clients, not all users on the Internet
- Content providers *(say, Web servers)* cannot rely on existence and *correct* implementation of caching proxies
- Accounting issues with caching proxies.
  For instance, *www.cnn.com* needs to know the number of hits to the webpage for advertisements displayed on the webpage

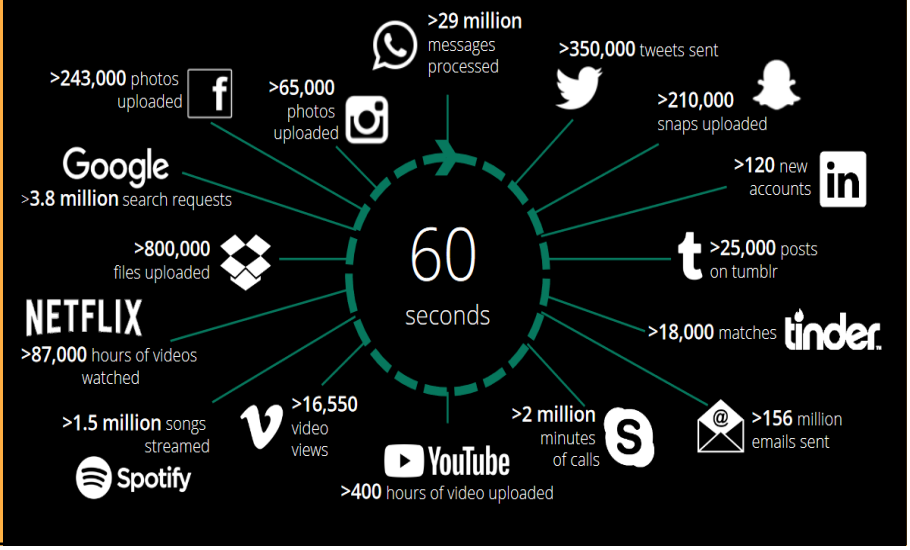
CDNs

47

So much happened in our digitalized world in 2017 – and we have the numbers behind it

Things that happened online in 2017 within 60 seconds

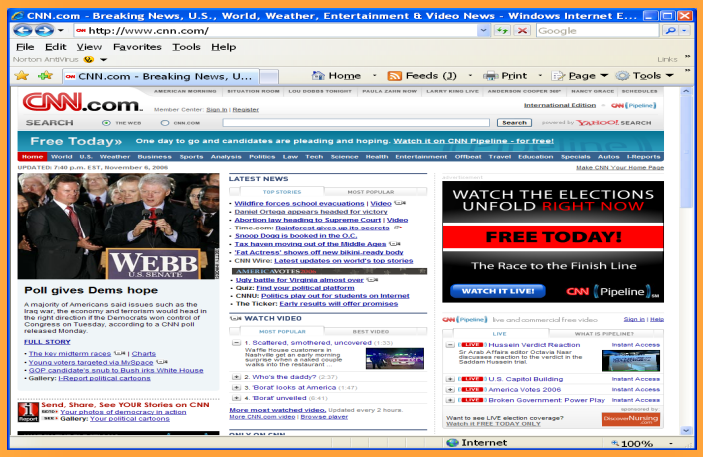
Lots of multimedia content!



49

50

CDNs Target environment?



Most Web files are small (1KB ~ 100KB)

(initially….)

51

Motivation

- **IP based networks**
- **Web based applications have become the norm for corporate internal networks and many business-to-business interactions**
- **Large acceptance and explosive growth**
  - **Serious performance problems**
  - **Degraded user experience**

  For a large set of applications, including VIDEO access
- **Improving the performance of networked applications**
  - **Use many sites at different points within the network**
    - Stand alone servers
    - Routers

CDNs basics

- **What is a CDN?**
  - **A network of servers delivering content on behalf of an origin site**
    - A number of CDN companies well established now
      - E.g. Akamai, Digital Island, Speedera, CDN77, Cloudfare, Stackpat
    - Many companies are exploring CDNs
      - Avoid congested portions of the Internet
- **Consist of**
  - **Edge servers deployed at several ISP (Internet Service Provider) access locations and network exchange points**
- **Large-file service with no custom client, no custom server, no prepositioning**
- **Improve the response time of an Internet site**
  - **Offloading the delivery of bandwidth-intensive objects, such as images and video clips**
- **Intelligent Internet infrastructure that improves the performance and scalability of distributed applications by moving the bulk of their *computation* to servers located at the edge of the network**
  - **Applications are logically split into two components**
    - Executed at an edge server close to the user
    - Executed on a traditional application server

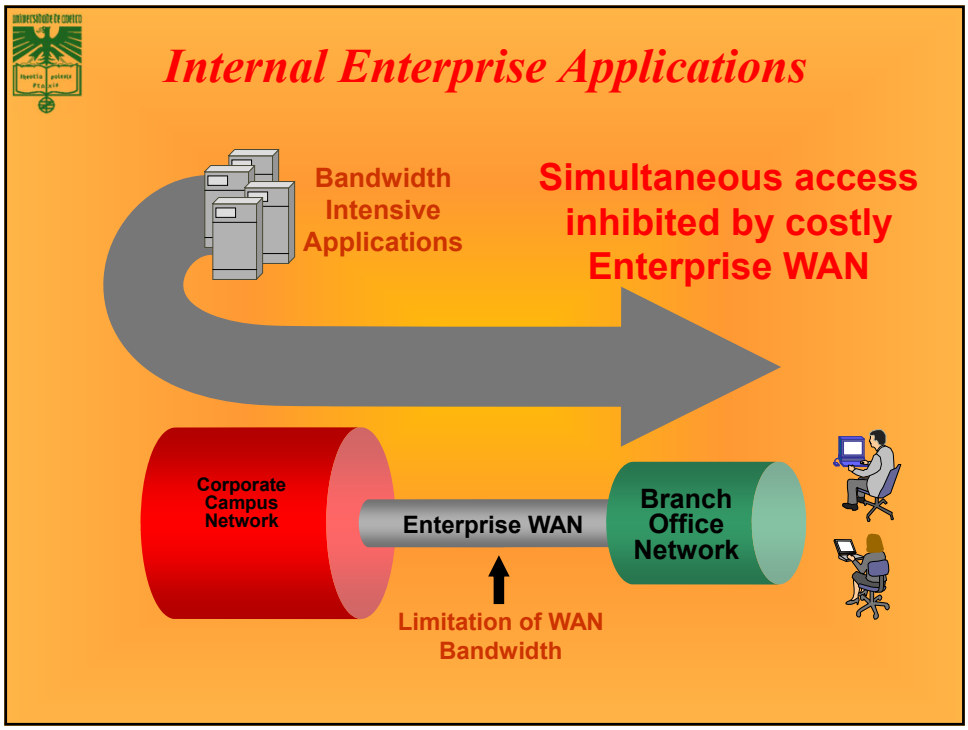
52

53

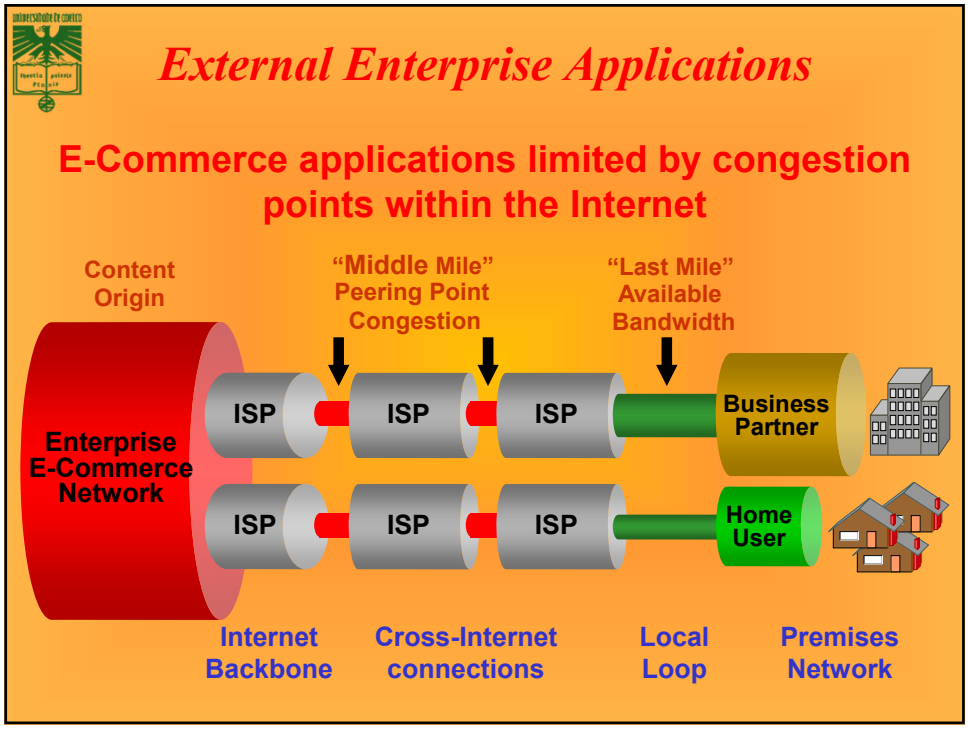
CDN Generations

- **First generation (early 90ies)**
  - **Accelerate the performance of web sites**
  - **Support increasing volumes of traffic**
    - Key disruption event: 9/11
    - Akamai technologies created
- **Second generation (early 2000ies)**
  - **Support high volumes of multimedia traffic**
  - **Audio/video intensive networks**
    - All ISPs developed/used CDNs
- **Third generation (2010+)**
  - **Cloud computing**
    - Amazon cloud (2008)
  - **UGC (user generated content)**
  - **P2P and interactivity**
    - AT&T distributed data centers (2011)
  - **Mobile support, and device adapted content**

53

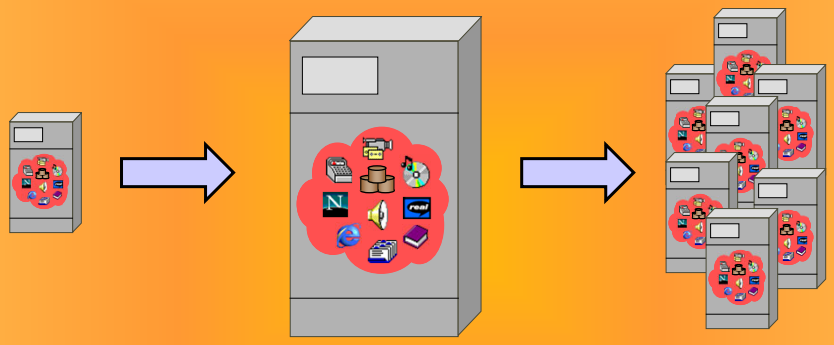

Internal Enterprise Applications
Bandwidth Intensive Applications
Simultaneous access inhibited by costly Enterprise WAN
Corporate Campus Network
Enterprise WAN
Branch Office Network
Limitation of WAN Bandwidth

54


External Enterprise Applications
E-Commerce applications limited by congestion points within the Internet
Content Origin
"Middle Mile" Peering Point Congestion
"Last Mile" Available Bandwidth
Enterprise E-Commerce Network
ISP
ISP
ISP
Business Partner
ISP
ISP
ISP
Home User
Internet Backbone
Cross-Internet connections
Local Loop
Premises Network

55

Content Scaling



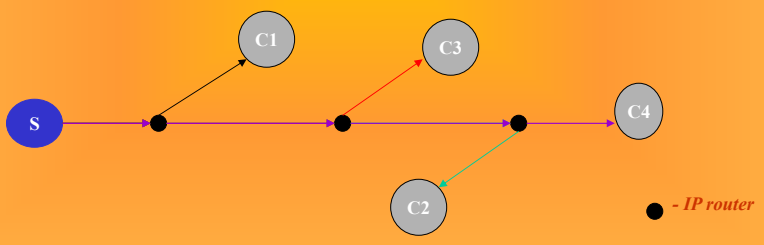
- Need to scale content to handle numerous clients
  - One can only scale 'vertically' to a point

Multiple servers/locations introduces new issues

56

Early Motivations for Content Networks (1st generation)

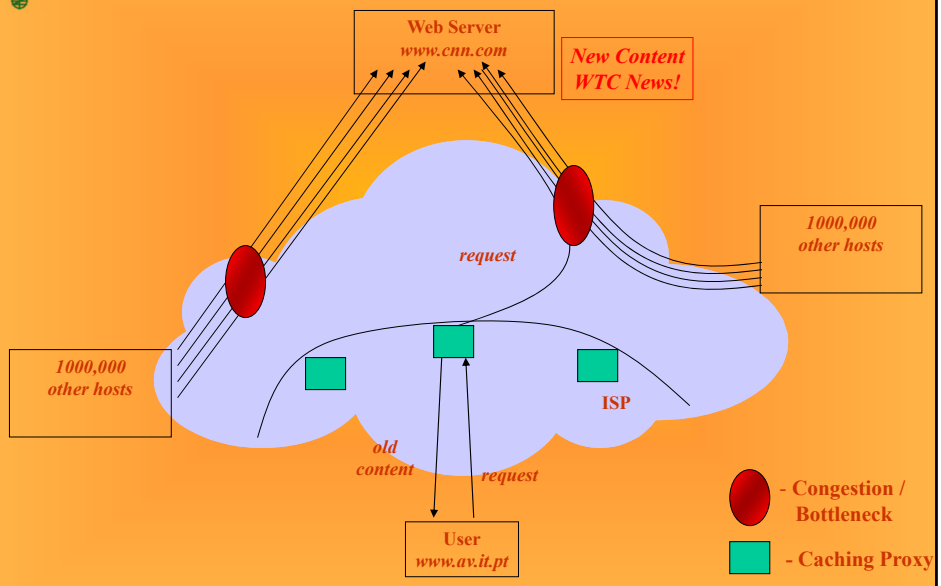
- More hops between client and Web server => more congestion!
- Same data flowing repeatedly over links between clients and Web server
- Origin server is bottleneck as number of users grows
- Flash Crowds *(for instance, Sept. 11)*
  - *The Content Distribution Problem:* Arrange a rendezvous between a content source at the origin server *(www…com)* and a content sink *(users)*



57

Flash Crowds:

Consider, On September 11, 2001



59

60

Flash crowd solution: CDNs..

What is a CDN?

A network of servers delivering content on behalf of an origin site

Large-file service with

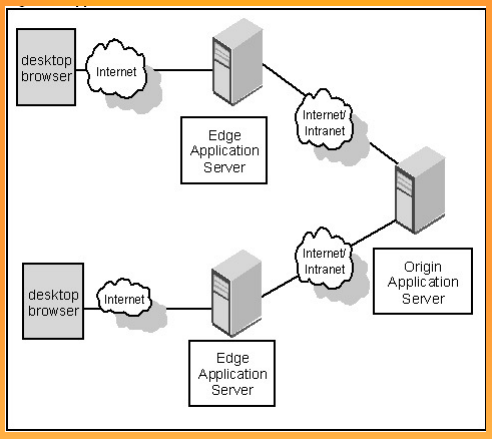
- **No custom client**
- **No custom server**
- **No prepositioning**
- **No rehosting**
- **No manual provisoning**

60

61

Model

- Application offload (1st generation concern)



61

62

Content distribution networks

- Client attempts to access the main server site for an application
- It is redirected to one of the other sites
- Each site caches information
  - Avoid going to the main server to get the information/application
- Access a closelly located site
  - Avoid congestion on the path to the main server
- Set of sites used to improve the performance of web-based applications collectivelly
  - Content distribution network

62

Inside a CDN

- **Servers are deployed in clusters for reliability**
  - **Some may be offline**
    - Could be due to failure
    - Also could be "suspended" (e.g., to save power or for upgrade)
- **Could be multiple clusters per location (e.g., in multiple racks)**
- **Server locations**
  - **Well-connected points of presence (PoPs)**
  - **Inside of ISPs**

63

64

Advantages

- **Better scalability**
- **Higher availability**
- **Improved response time from a centrally managed solution**
- **Nodes constituting the distribution network are designed to be**
  - **Self-configuring**
  - **Self-managing**
  - **Self-diagnosing**
  - **Self-healing**

to ensure easy management and operational convenience

64

65

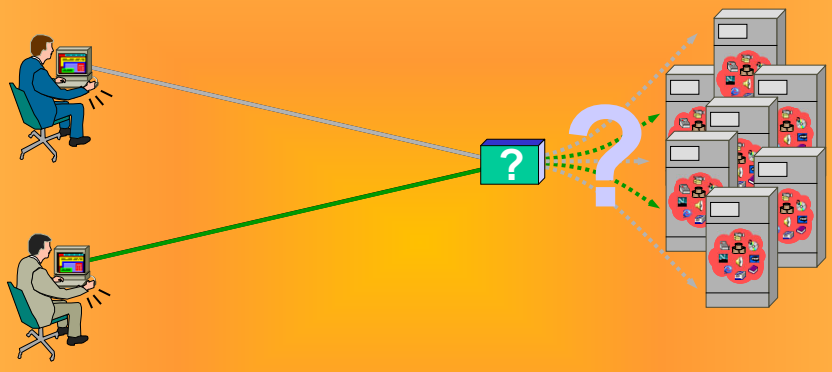
Challenges

- Keep consistency among the enterprise data hosted by the offloaded applications
- Share session state among edge and origin application servers
- Distribution, configuration, and management
- Develop programming models consistent with current industry standards such as J2EE
- Application security.

- There is active research into general frameworks to be used to support distributed applications, as well as prototyping the ideas for specific application instances

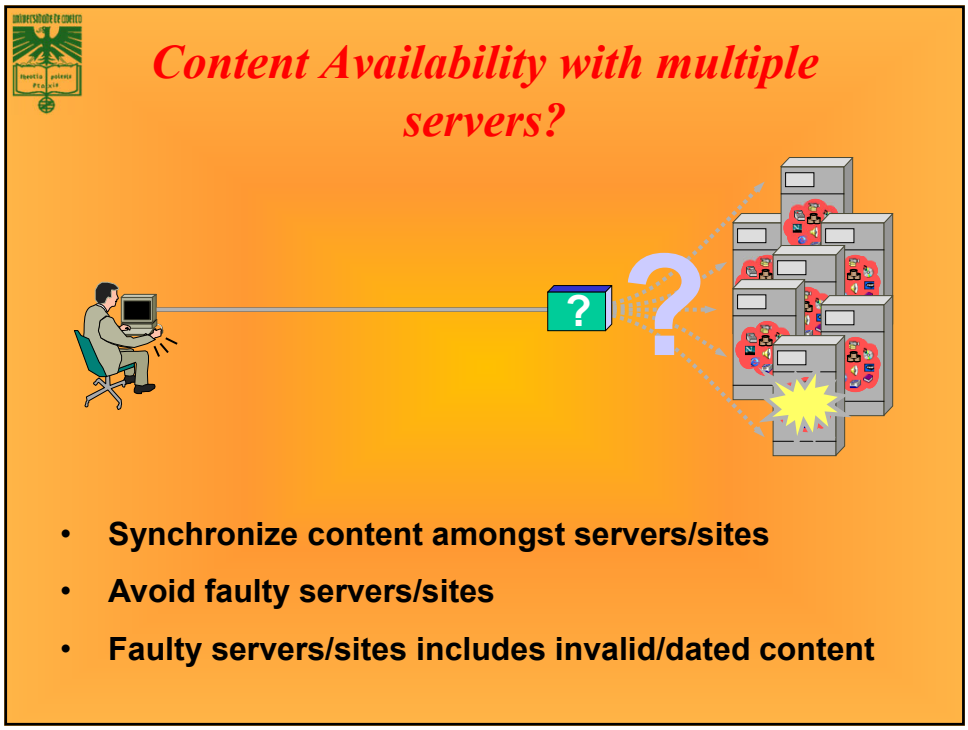
65

Load-Sharing Content

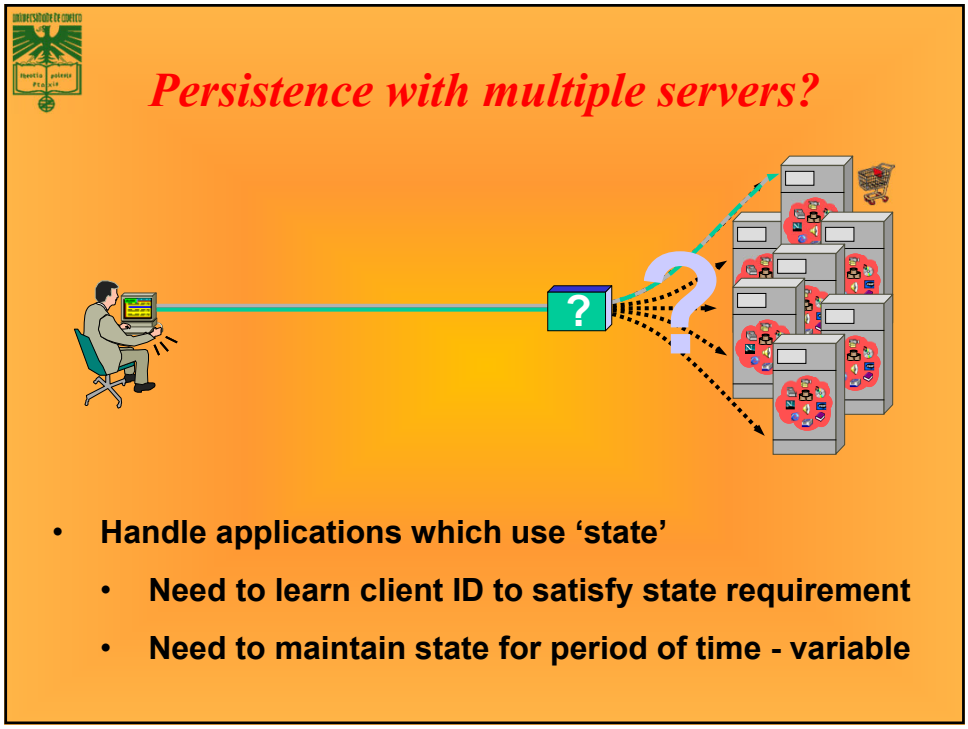


- Handle requests fairly amongst servers/sites
- Easily add servers/sites to content service
- Adjust connections based on server/site load

66


Content Availability with multiple servers?
?
?
• Synchronize content amongst servers/sites
• Avoid faulty servers/sites
• Faulty servers/sites includes invalid/dated content

67


Persistence with multiple servers?
?
?
• Handle applications which use 'state'
• Need to learn client ID to satisfy state requirement
• Need to maintain state for period of time - variable

68

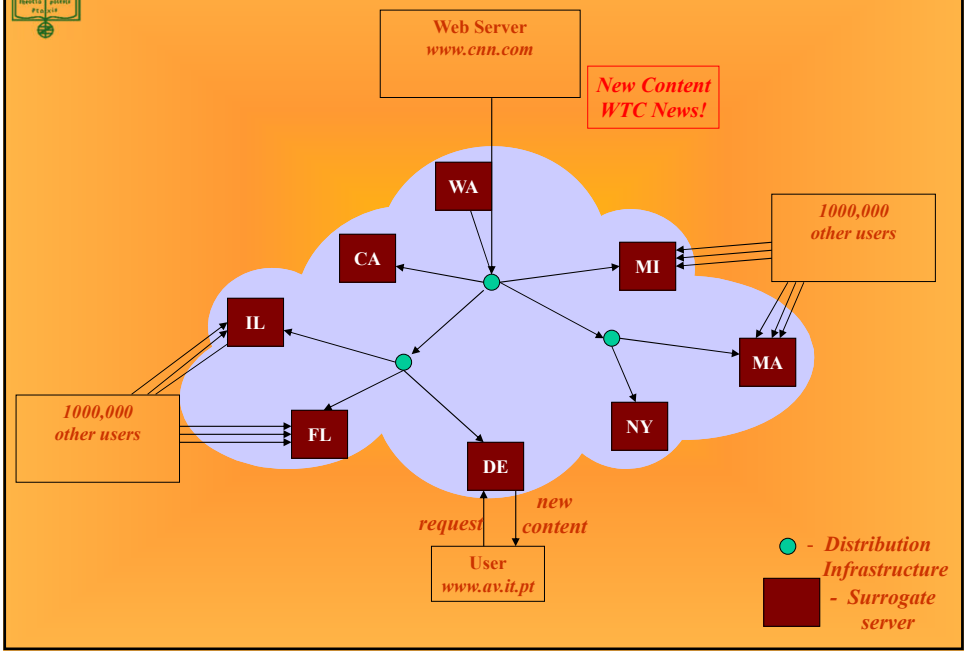
69

Outline

- Overall context
- Challenges
- Potential alternatives?
- **Architecture**

69

CDN, On September 11, 2001



70

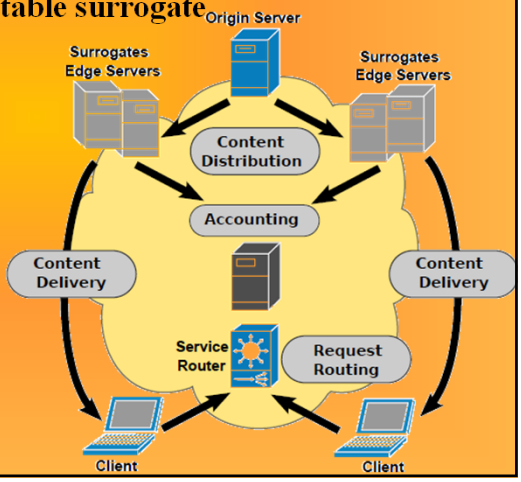
With CDNs

- Overlay network to distribute content from origin servers to users
  - Avoids large amounts of same data repeatedly traversing potentially congested links on the Internet
  - Reduces Web server load
  - Reduces user perceived latency
  - Tries to route around congested networks
- CDN is not a cache!
  - Caches are used by ISPs to reduce bandwidth consumption, CDNs are used by content providers to improve quality of service to end users
  - Caches are reactive, CDNs are proactive
  - Caching proxies cater to their users (web clients) and not to content providers (web servers), CDNs cater to the content providers (web servers) and clients
  - CDNs give control over the content to the content providers, caching proxies do not

71

CDN Components

- *Content Delivery Infrastructure:* Delivering content from producer to clients by surrogates
- *Request Routing Infrastructure:* Steering or directing content request from a client to a suitable surrogate
- *Distribution Infrastructure:* Moving or replicating content from content source (origin server, content provider) to surrogates
- *Accounting Infrastructure:* Logging and reporting of distribution and delivery activities



72

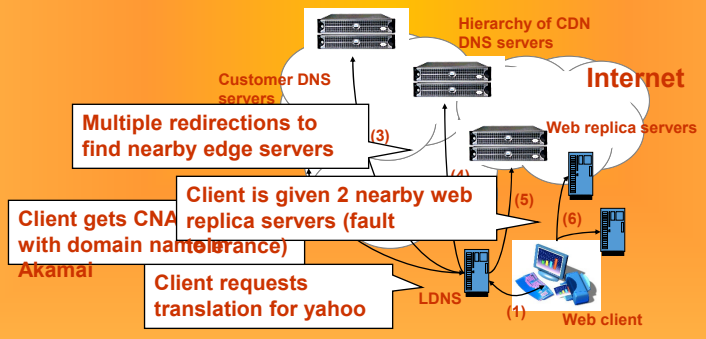
Mapping clients to servers

- CDNs need a way to send clients to the "best" server
  - The best server can change over time
  - And this depends on client location, network conditions, server load, …
  - What existing technology can we use for this?
- DNS-based redirection
  - Clients request www.foo.com
  - DNS server directs client to one or more IPs based on request IP
  - Use short TTL to limit the effect of caching

73

DNS Redirection

- Web client's request redirected to 'close' by server
  - Client gets web site's DNS CNAME entry with domain name in CDN network
  - Hierarchy of CDN's DNS servers direct client to 2 nearby servers



74

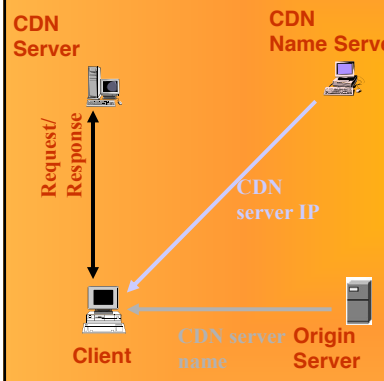
DNS Redirection Considerations

- **Advantages**
  - **Uses existing, scalable DNS infrastructure**
  - **URLs can stay essentially the same**
- **Limitations**
  - **DNS servers see only the DNS server IP**
    - Assumes that client and DNS server are close. Is this accurate?
  - **Content owner must give up control**
  - **Unicast addresses can limit reliability**

75

76

What other CDN techniques are being used?



- **DNS redirection (DR)**
  - **Full-site delivery**
  - **Partial-site delivery**
- **URL rewriting**
- **Hybrid scheme**
  - **URL rewriting + DNS redirection**
- **Manual hyperlink selection**
- **HTTP redirection**
- **Layer 4 switching**
- **Layer 7 switching**
- **Anycast**

76

77

Offloading a portal

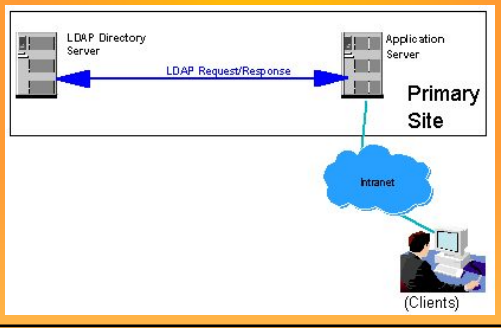
- Portal servers allow users to access content and applications from a single access point
  - Users can create persistent, customized views of applications and content chosen from the set of applications and content by the portal administrators
- Portal server pages are personalized
- Often include dynamic content
- Significant amount of computation required for page assembly
  - Application offload

77

78

Offloading an Enterprise directory

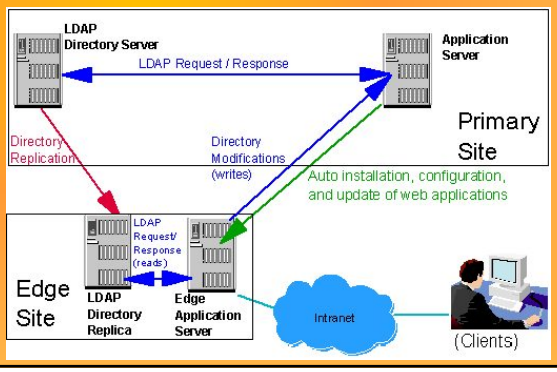
- E.g. a common e-Workplace tool
- The employee data is often stored in a central LDAP directory
  - Separate web-based application providing the interface to the directory



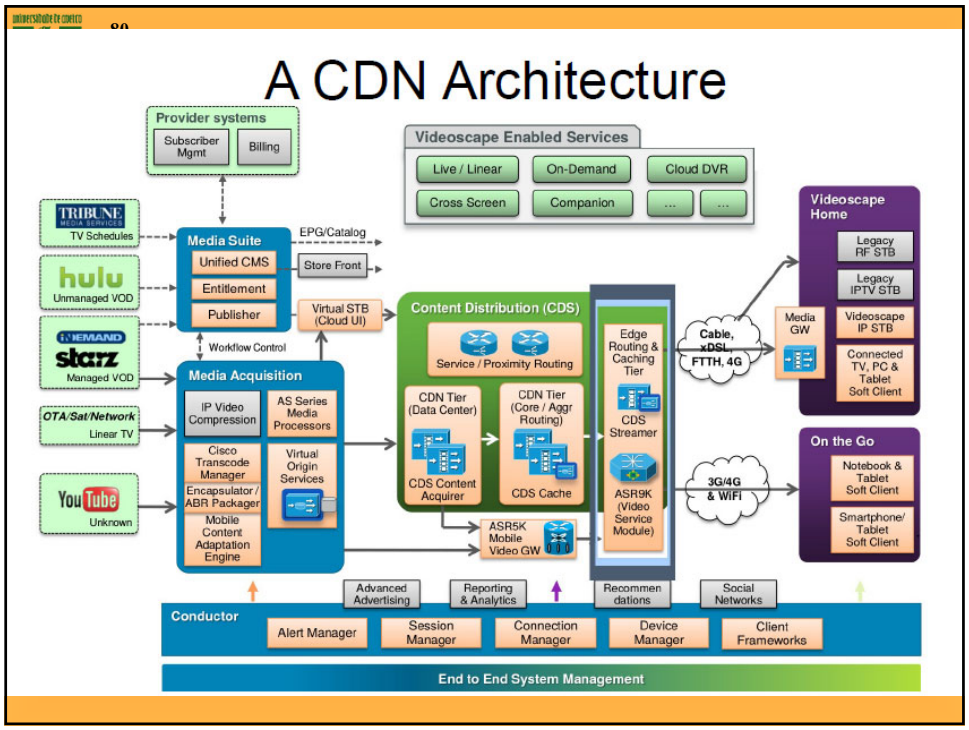
78

Offloading an Enterprise directory

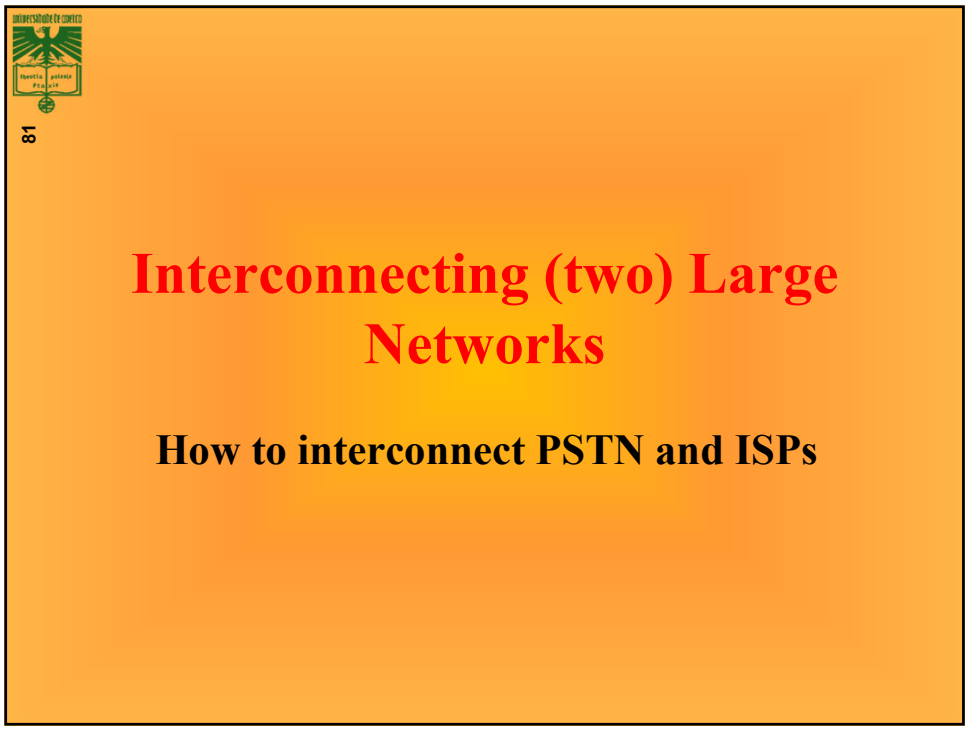
- **Centralized directory**
  - **Convenient to manage**
  - **Performance for clients accessing the directory from remote sites can be poor**
    - E.g. transcontinental network connections suffer from a long delay
- **Offloaded version of the application**



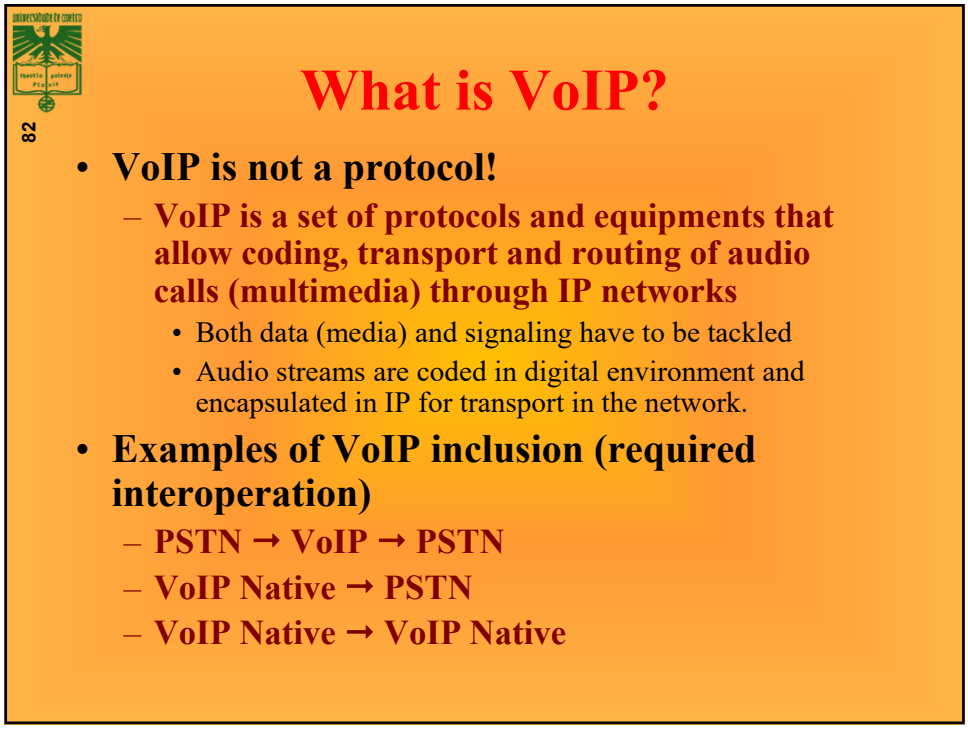
79



80


81
Interconnecting (two) Large Networks
How to interconnect PSTN and ISPs

81


82
What is VoIP?
• VoIP is not a protocol!
– VoIP is a set of protocols and equipments that allow coding, transport and routing of audio calls (multimedia) through IP networks
• Both data (media) and signaling have to be tackled
• Audio streams are coded in digital environment and encapsulated in IP for transport in the network.
• Examples of VoIP inclusion (required interoperation)
– PSTN → VoIP → PSTN
– VoIP Native → PSTN
– VoIP Native → VoIP Native

82

83

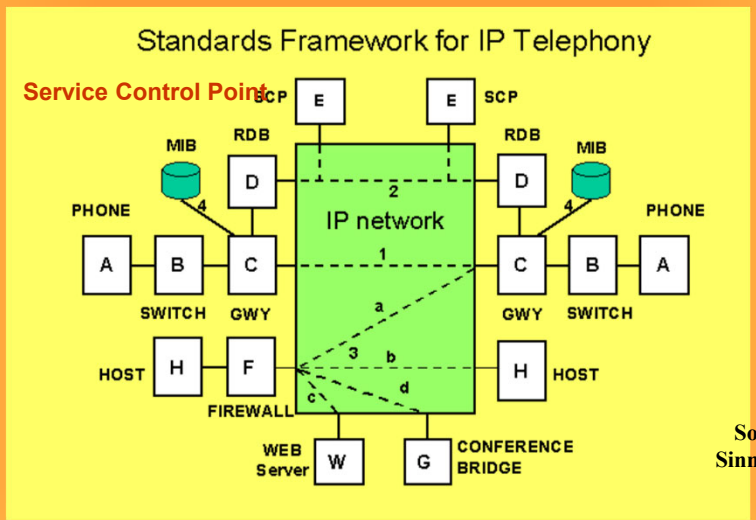
VoIP advantages

- **Cost reduction**
  - **Do not need to pay for PSTN circuits for call transport (user side) / consolidate infrastructure (provider side)**
  - **Bandwidth reduction**
    - Distributed nature of VoIP
    - Operation costs reduction – voice and data traffic both in the same network
- **'Open' standards and interoperability between operators**
  - **Does not depend on proprietary solutions**
- **Integration of voice and data networks**
  - **Considered as 'just another IP application'**
  - **Two major approaches: ITU-T (early on) and IETF (current)**
  - **As long as the quality is similar to the PSTN network, companies can easily invest in new services and applications**

83

84

Voice over IP Framework



Source: Sinnreich et al

Most challenges are associated with control plane.

84

85

Different levels of VoIP problem

1. **The transport level**
   - How to transport multimedia information.

   Covers also content, but mostly RTP (and associated protocols)
2. **The session control**
   - How to signal a VoIP session.

   Covers also application protocols, but we talk mostly about SIP and H.323 or RTSP (web)
3. **The gateway control**
   - How to signal interface entities between Internet and POTS.

   Mostly Megaco

85

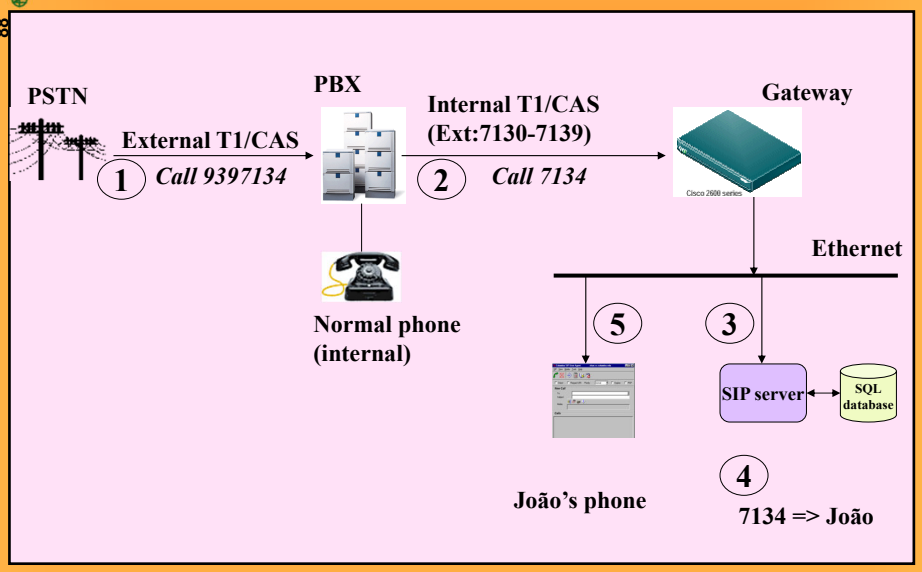
87

PSTN interoperation

- **Translate audio (PCMU/PCMA)**
- **Translate signalling (PRI/T1,ISUP)**
  - Different signalling
  - Advanced capabilities in SIP are not mapped in PSTN
- **Translate identifiers (phone numbers)**
- **Determine transition points**

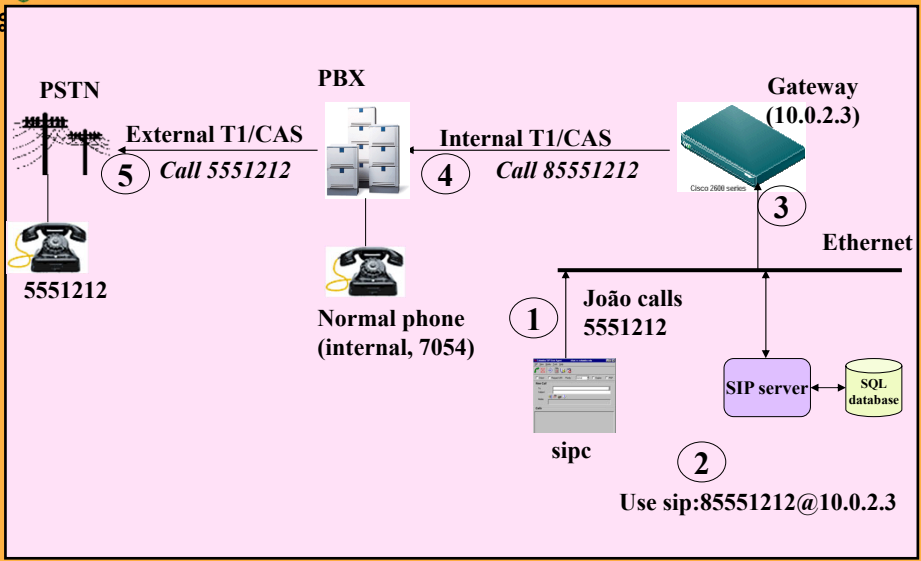
87

PSTN calls to IP



88

IP calls to PSTN



89

90

ISPs and PSTN

- **Having VoIP (specially voice) sessions connecting to old-style phone networks implies:**
  1. **Interconnecting voice signalling**
  2. **Interconnecting data (voice)**
     - Typically this is set by routing tables in both sides
  3. **Linking both inteconnection actions**
  4. **Selecting where to do each one of these**

90

91

What about REAL interoperation?

- **Signaling boxes between the data and circuit systems must be interconnected**
  - **Multiple interconnection points may exist**
- **Systems must select best interconnection points**
  - **This implies best routing solution**
    - And this is mixed routing – both in data and circuit systems
  - **Interoperation points may be different for the data and control planes**
- **Different types of boxes may exist (interoperation of data/control/both)**

91

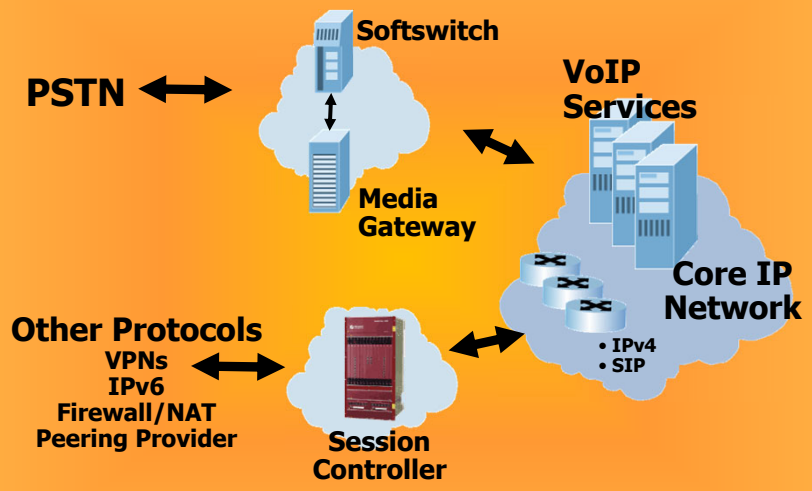
92

VoIP and PSTN Interoperability in Large Scalable Scenarios

- Requires an application programming interface and a corresponding protocol for controlling VoIP Gateways from external call control elements.
- Signaling must be inter-operable between PSTN and VoIP.
- Protocols:
  - Media Gateway Controller Protocol (MGCP) - RFC 2705
  - MGCP evolution/successor → H.248/Megaco (RFC 3015) → H.248.1/Gateway Control Protocol (RFC 3525)
    - These are control plane signaling only.
  - SIGTRAN (Signaling Transport) is the standard telephony protocol used to transport Signaling System 7 (SS7) signals over the Internet.
    - Stream Control Transmission Protocol (SCTP) – RFC 3286
      - Is an IP transport designed for transporting signaling information over an IP network.
      - Reliable transport protocol with support for framing of individual message boundaries.

92

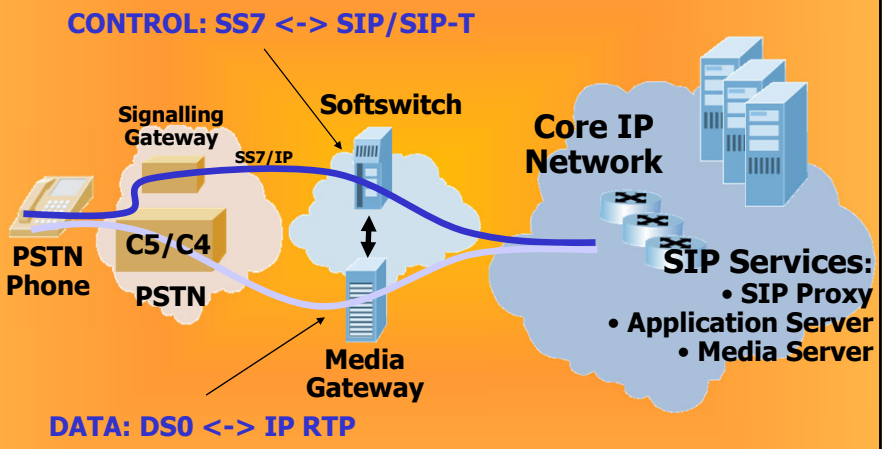
Network interoperation



Technologically addressed by media gateways and softswitches (PSTN ⇦⇨ IP) and session controllers (IP ⇦⇨ IP), and associated VoIP data plane protocols

93

PSTN/IP Interworking



Softswitches are associated with Media Gateways (often colocated in the same box), to bridge both signalling and data.

94

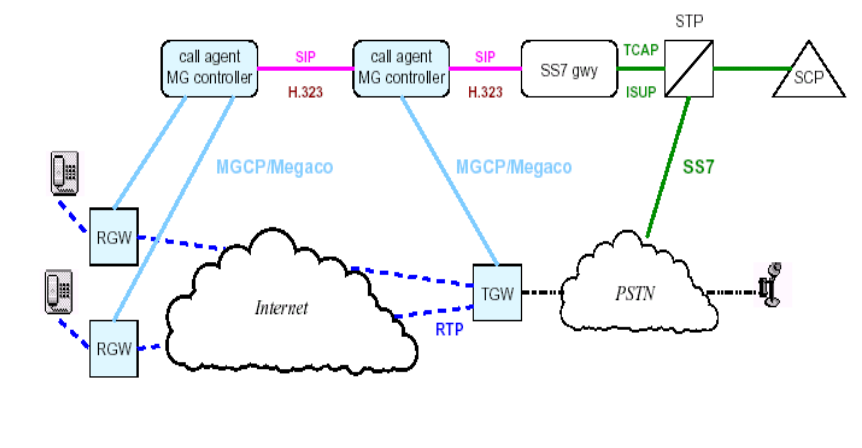
MGCP e Megaco

- Media Gateway Controller Protocol (RFC 2705)
- Controls phone Gateways resorting to *external* control elements, the media gateway controllers (MGC) a.k.a. call agents
  - Gateways: Eg: RGW (residential gateway): physical interconnection between VoIP networks and phone interfaces at homes
  - The call control "intelligence" is outside the gateways, and is controlled by external elements
  - master-slave philosophy
- Objective: scalable gateway infrastructure between PSTN and IP networks
- MGCP Successor: H.248/Megaco
- These are control plane signaling ONLY.

95

MGCP Architecture

Objective: VoIP large scalable implementations



RGW: Residential Gateway
TGW: Trunk Gateway

96

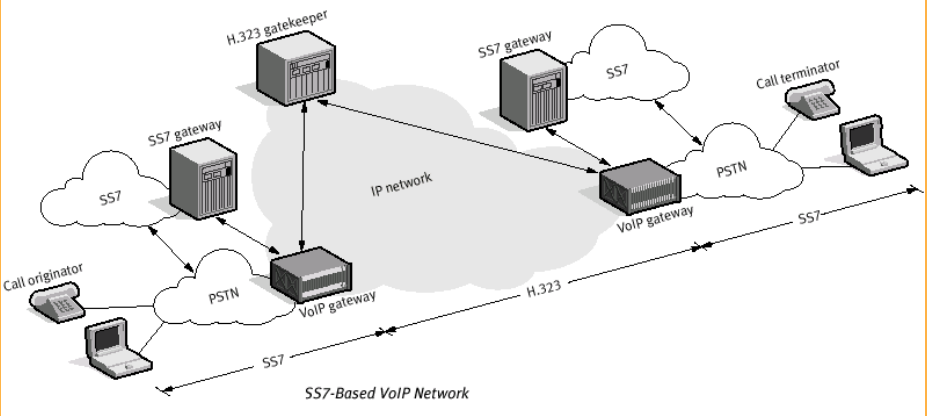
MGCP/H.248 Elements

97

- **Media Gateway Controller (MGC)**
  - **Controls the parts of the call state that pertain to connection control for media channels in a MG.**
- **Media Gateway (MG: RGW/TGW)**
  - **Converts media provided in one type of network to the format required in another type of network.**
  - **MG could terminate bearer channels from a switched circuit network (e.g., DS0s) and media streams from a packet network (e.g., RTP streams in an IP network).**
- **Signaling Gateway (SG)**
  - **Responsible for transferring signaling messages (e.g., SS7 messages) to different protocols and transports.**
    - Signaling Transport (SIGTRAN)
    - e.g., SS7 to SIGTRAN (SCTP/IP).

97

VoIP and SS7



98

Operator Networks and the FMBC

The arrival of common services

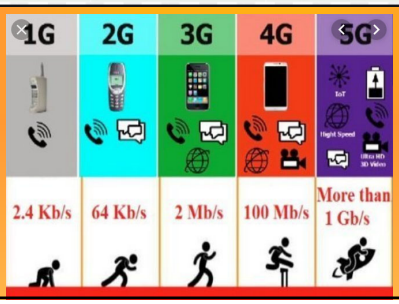
100

Rationale for the class

- Relate with the evolution of mobile networks
- Perceive the technologies underlying the integration of services, coming from the mobile networks, and integrating multimedia communications
- Understand the interworking of signal and data between networks
- Follow up the development of the FMC and FMBC concepts, as evolution of IP-based services

Wide Communication technologies: Cellular

Comparison	2G	3G	4G	5G
Introduced in year	1993	2001	2009	2018
Technology	GSM	WCDMA	LTE, WiMAX	MIMO, mm Waves
Access system	TDMA, CDMA	CDMA	CDMA	OFDM, BDMA
Switching type	Circuit switching for voice and packet switching for data	Packet switching except for air interference	Packet switching	Packet switching
Internet service	Narrowband	Broadband	Ultra broadband	Wireless World Wide Web
Bandwidth	25 MHz	25 MHz	100 MHz	30 GHz to 300 GHz
Advantage	Multimedia features (SMS, MMS), internet access and SIM introduced	High security, international roaming	Speed, high speed handoffs, global mobility	Extremely high speeds, low latency
Applications	Voice calls, short messages	Video conferencing, mobile TV, GPS	High speed applications, mobile TV, wearable devices	High resolution video streaming, remote control of vehicles, robots, and medical procedures



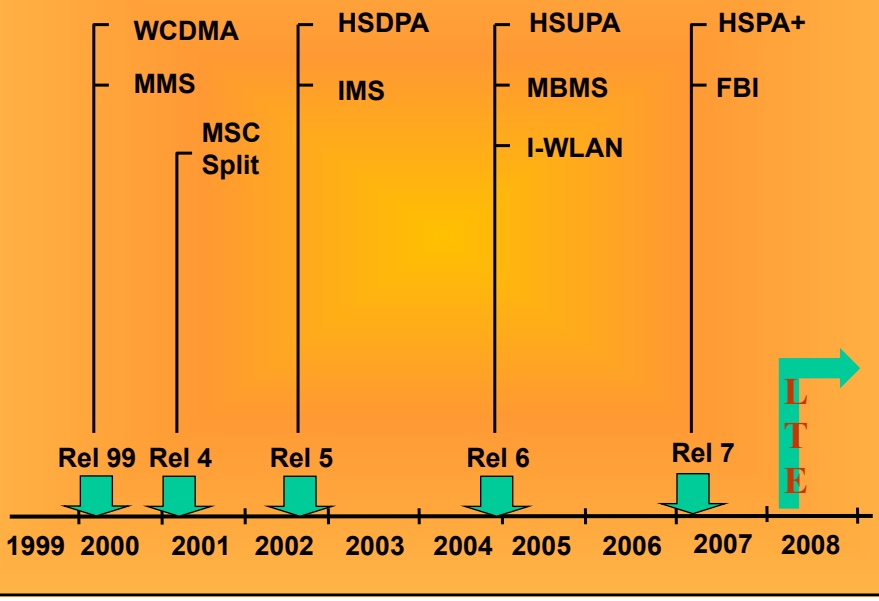
Early cellular systems

103

- 1G: analog systems (450-900 MHz)
  - Signalling: FSK
  - Share of medium: FDMA
  - NMT (Europe), AMPS (US)
- 2G: digital systems (900, 1800, 1900 MHz)
  - Share of medium : TDMA/CDMA
  - Circuit switching
  - GSM (Europe), IS-136 (US), PDC (Japan)
- 2.5G: extensions for packet switching
  - Digital: GSM → GPRS
  - Analog: AMPS → CDPD
- 3G: networks for data applications
  - High rates, data, Internet
  - Share of medium : TDMA/CDMA/CDMA
  - IMT-2000 (Europe: UMTS)

103

3GPP Release initial timeline: UMTS is phased



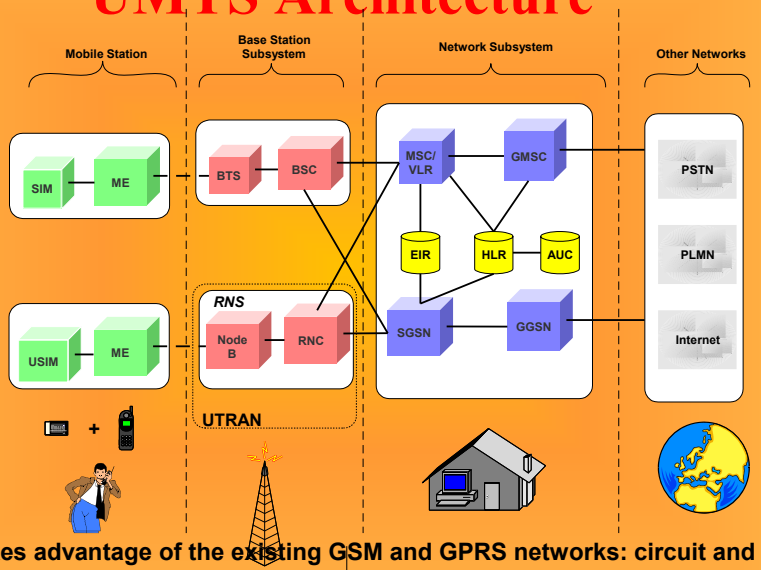
104

UMTS: first universal celular data system

- 3G system
- Oriented to generalized service diffusion and its future users trends
  - Combines cellular, wireless, paging, etc. functions
- "multimedia everywhere"
- Developed as an evolution path of 2.5G systems
  - Progressive evolution (GPRS-EDGE-UMTS)
- Direction towards IP networking, and relying of data services

105

UMTS Architecture

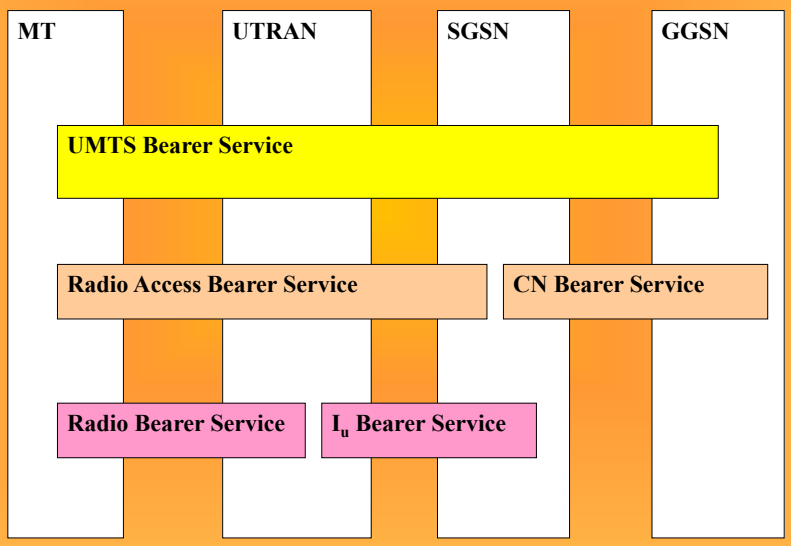


UMTS takes advantage of the existing GSM and GPRS networks: circuit and data paths
The main difference comes with the new radio interface.

106

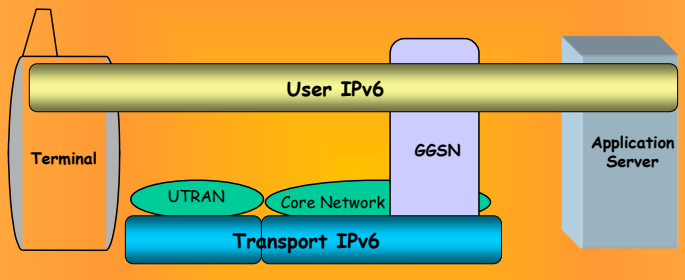
108

UMTS global architecture

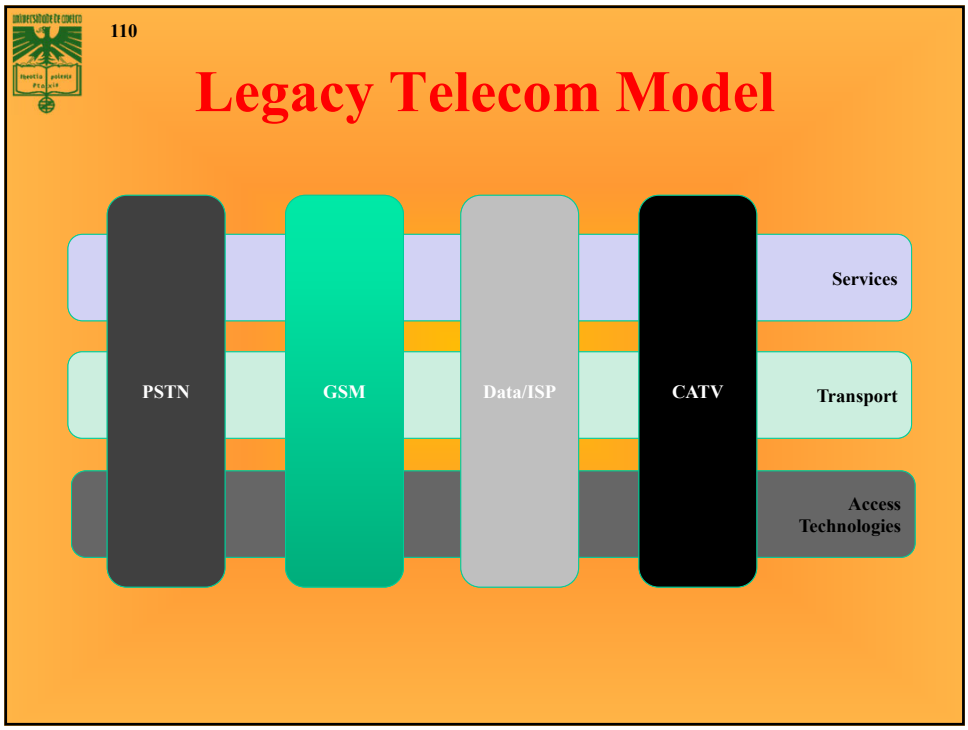


109

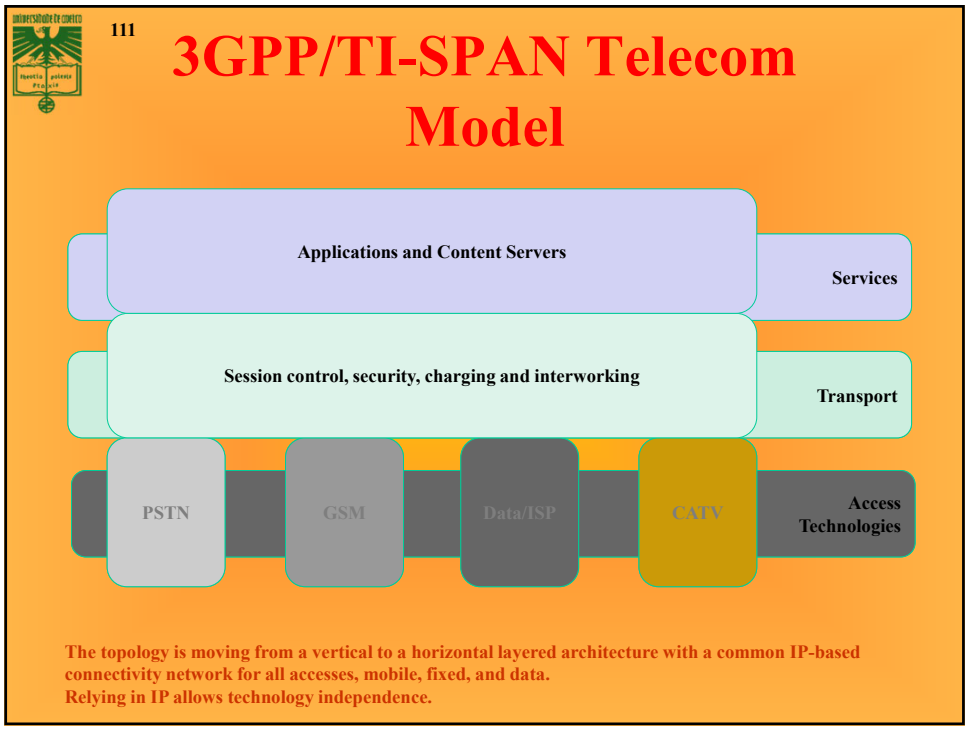
User plane vs transport plane



- User and transport planes are completely independent, i.e. the transport plane can run on a different IP version than the user plane
- UTRAN and Core Network transport can also run on different IP versions


110
Legacy Telecom Model
PSTN
GSM
Data/ISP
CATV
Services
Transport
Access Technologies

110

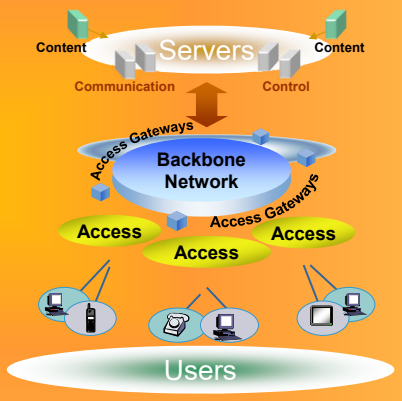

111
3GPP/TI-SPAN Telecom Model
Applications and Content Servers
Session control, security, charging and interworking
PSTN
GSM
Data/ISP
CATV
Services
Transport
Access Technologies
The topology is moving from a vertical to a horizontal layered architecture with a common IP-based connectivity network for all accesses, mobile, fixed, and data.
Relying in IP allows technology independence.

111

112

UMTS from release 5 on: IMS: IP Multimedia Subsystem

- **Same Core network**
- **Same User on different accesses**
- **Same Services**
- **Can use WLAN, ADSL, LAN, UTRAN (GPRS) etc. as accesses in ONE system**
- **Can have several devices and move between them**
- Addresses circuit-oriented session concepts!



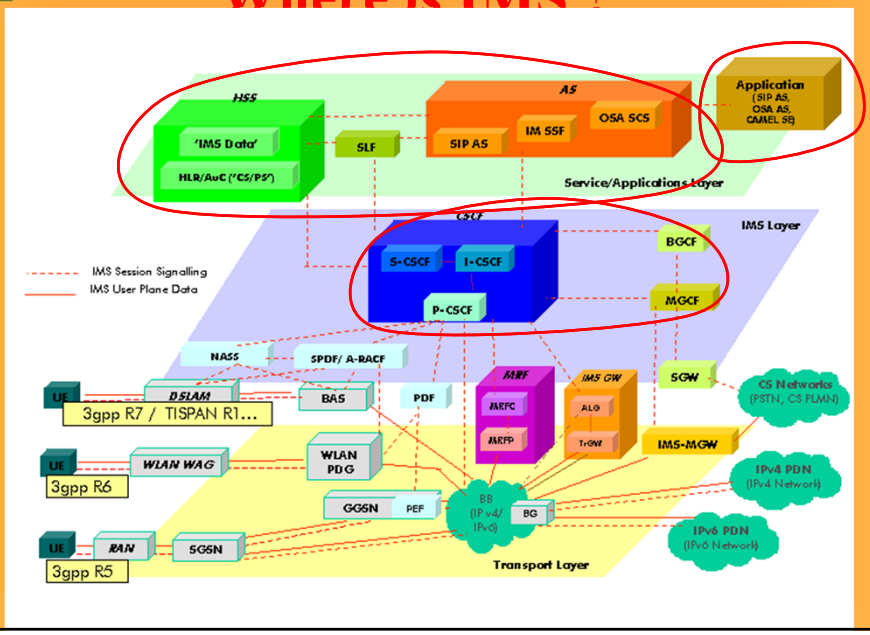
112

IP Multimedia Subsystem (IMS)

- **New IP-based mobile core network for 3G evolution**
- **Uses 3GPP variant of SIP & other IP protocols**
- **"Intelligent Network" over IP?**
- **New services drive IMS deployment**
  - **Push-to-Talk, FMC, IP Centrex**
- **PTT (PoC) & UMA FMC specs already turned over to 3GPP**
- **Developed by 3GPP for GSM-to-3G evolution**
  - **Defined in release 5; fully specified in release 6**

114

Where is IMS ?



x-CSCF

- **Serving - CSCF**
  - Controls the user's SIP Session
  - very few per domain
  - Located in the home domain
  - Is a SIP registrar (and proxy)
- **Proxy-CSCF**
  - IMS contact point for the user's SIP signaling
  - Several in a domain
  - Located in the visited domain
  - Terminals must know this proxy (e.g. DHCP used)
  - Compresses and decompresses SIP messages
  - Secures SIP messages
  - Assures correctness of SIP messages
- **Interrogating – CSCF**
  - domain's contact point for inter-domain SIP signaling
  - one or more per domain
  - In case there are more than one S-CSCFs in the domain, locates which S-CSCF is serving a user

118

SIP Protocol

- **Defined in IETF RFC 3261**
  - **"… an application-layer control (signaling) protocol for creating, modifying, and terminating sessions with one or more participants. These sessions include Internet telephone calls, multimedia distribution, and multimedia conferences."**
- **In IMS, SIP is modified to include extra functionality and support a specific set of functions only**
  - **SIP is to the Internet what SS#7 is to telephony**
- **At the core of IMS there are several SIP proxies:**
  - **I-CSCF, S-CSCF, P-CSCF**
  - **The Call Session Control function (CSCF) is the heart of the IMS architecture**
  - **The main functions of the CSCF:**
    - provide session control for terminals and applications using the IMS network
    - secure routing of the SIP messages,
    - subsequent monitoring of the SIP sessions and communicating with the policy architecture to support media authorization.
    - responsibility for interacting with the HSS.

118

119

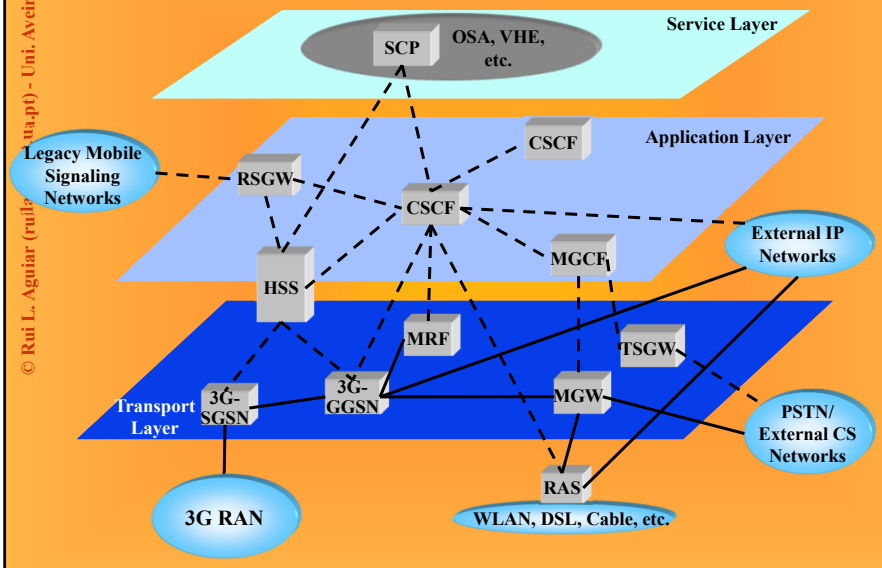
IMS Identity and User Profiles

- **IMS uses SIP identity: SIP URIs**
  - **e.g. sip:ruilaa@ua.pt**
  - **Opposed to phone numbers**
  - **A user is uniquely identified in the HSS by his IMPI (Private User Identity).**
  - **IMPI is a unique global identity defined by the Home operator**
  - **used only in the process of registration**
- **to establish communication with a user IMPU (Public User Identity) is necessary.**
  - **Every user has one or more IMPUs.**
  - **Each IMPI can have several IMPUs**
  - **Users can classify their public identities: business, family, friends, …**
    - E.g. sip:ruilaa@ua.pt, sip:steve.jobs@left.apple.com

119

120

UMTS Layered Approach for the PS services



120

121

Implications: what can be achieved with this add-on

- Any Device
- Any Access Technology
- Any Where

ALWAYS BEST CONNECTED

- One Network, multiple access technologies
- Common Session Control
- Generic Application Servers
- Single set of services that apply network wide
- Consistent user experience
- Operational efficiency
- New services/applications

121

Fixed Mobile (Broadcast) Convergence - FM(B)C

- One customer service
  - Handles mobile and fixed calls
  - Any network — mobile, WiFi, Broadband Cable…
  - Avoid mobile charging when in-building
- Single (customer) number with common suite of services
  - One voice mailbox, one phone directory…
  - Mobile, fixed, conference room
- New services? Irrespective of location, access technology or terminal device
  - Potentially gradative provision

Slide 122

122

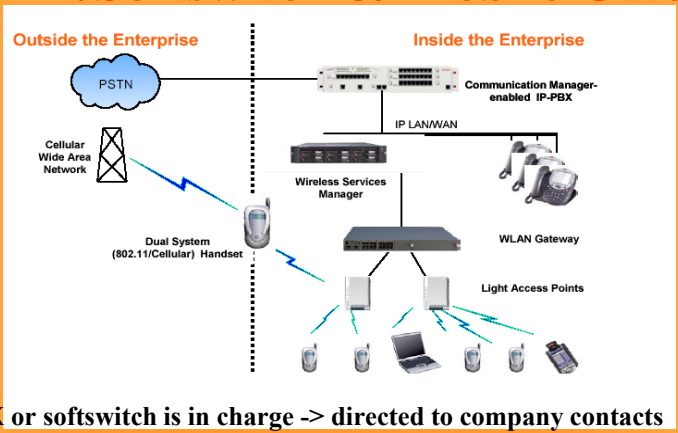
Implementing the FMBC concepts

- IP Multimedia Subsystem (IMS)
  - 3G vision of future IP-based mobile communications
- IP-PBX or softswitch with mobile network interface
  - Centered in company internal communications
- Wireless "fixed" line services
  - New (not FMBC) in developing nations, mobile, no handoffs
- Unlicensed Mobile Access (UMA)
  - voice & data services over WiFi

Slide 127

127

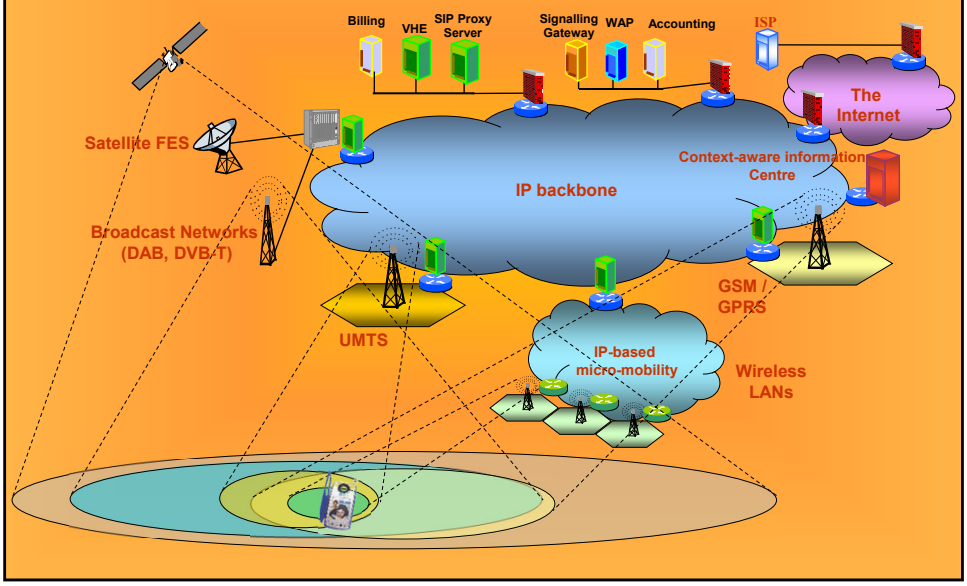
IP-PBX/Softswitch & Mobile Gateway



- IP-PBX or softswitch is in charge -> directed to company contacts
- Service hands off to mobile network when out of WLAN range
- Mobile network only used when necessary
  - "Tunnel" through the mobile network
  - Voice or text messages generated when necessary to connect enterprise calls to remote users

128

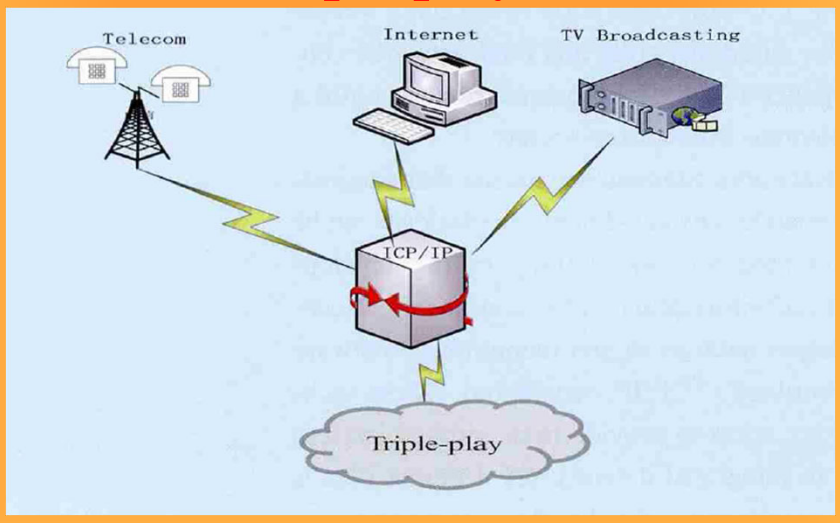
Inter-Working



130

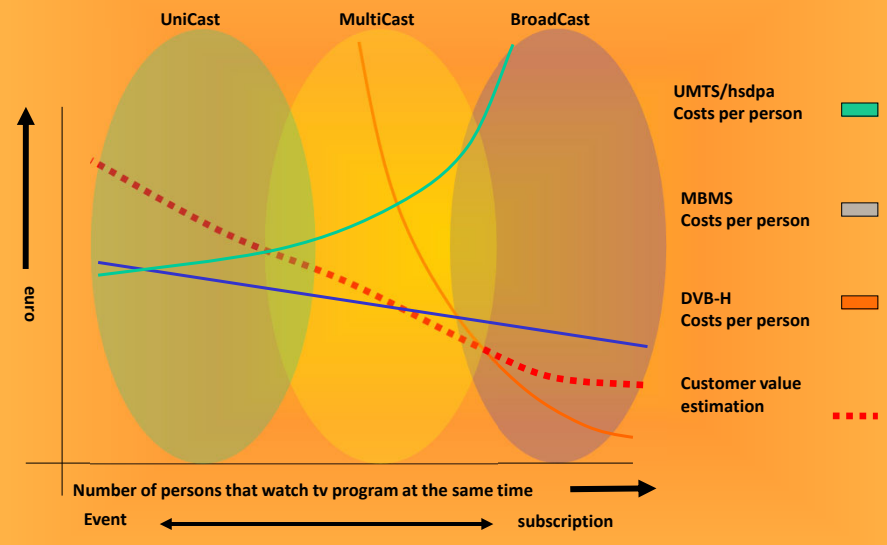
131

New service (FBC model): triple play



131

The broadcast problem: when to use it

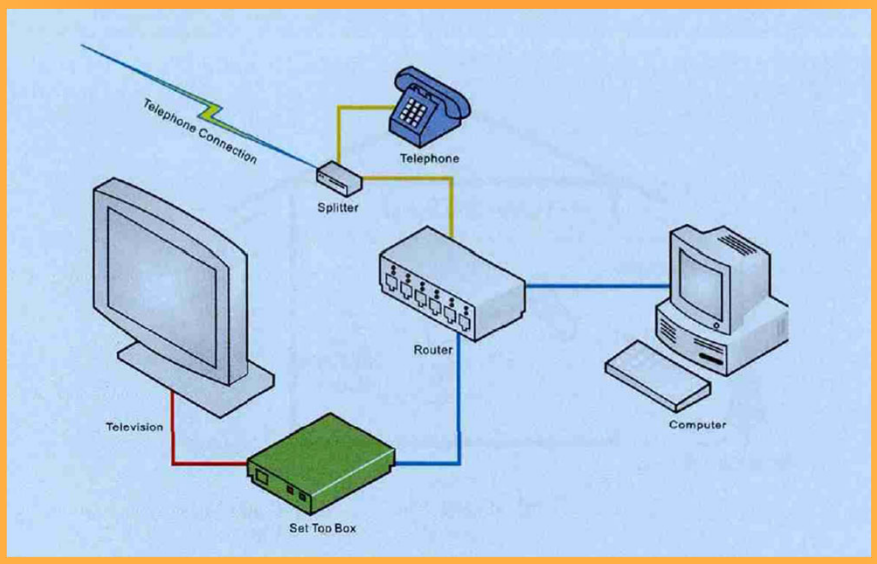


Delivery cost vs. value for customer mobile bearers

132

132

New methods of TV delivery



133

Potential of IP services for TV: Arriving of the multi- screen

PC watcher

TV broadcast

Digital TV packages

Digital video recorder

View programs that you missed

Video on demand

Personal TV services

DVB-H

UMTS

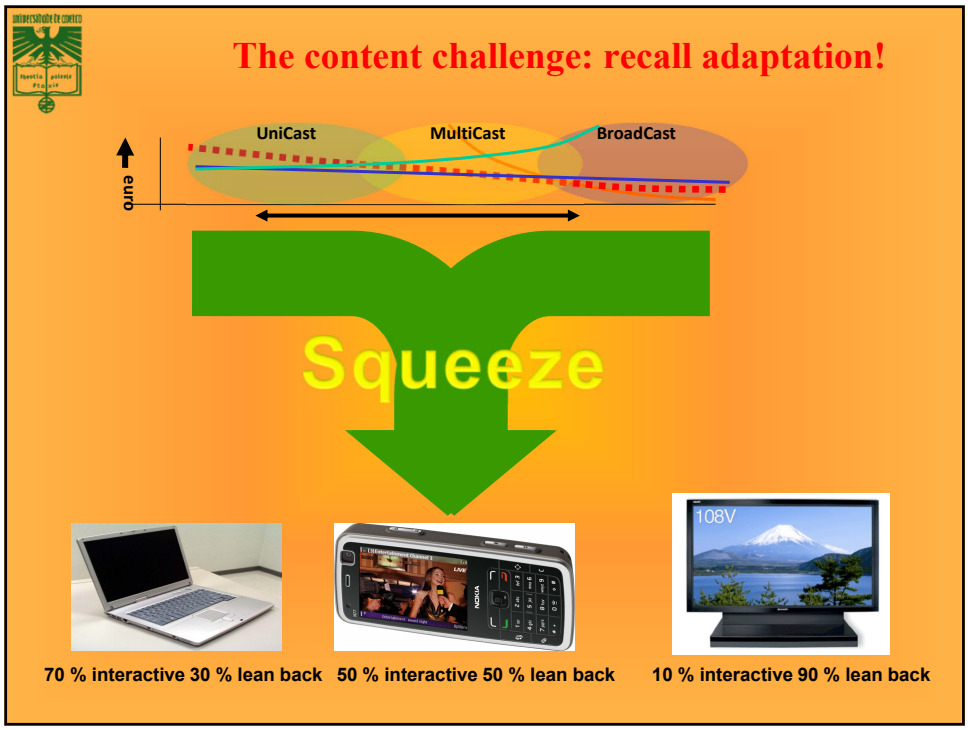
Wifi

Interactive

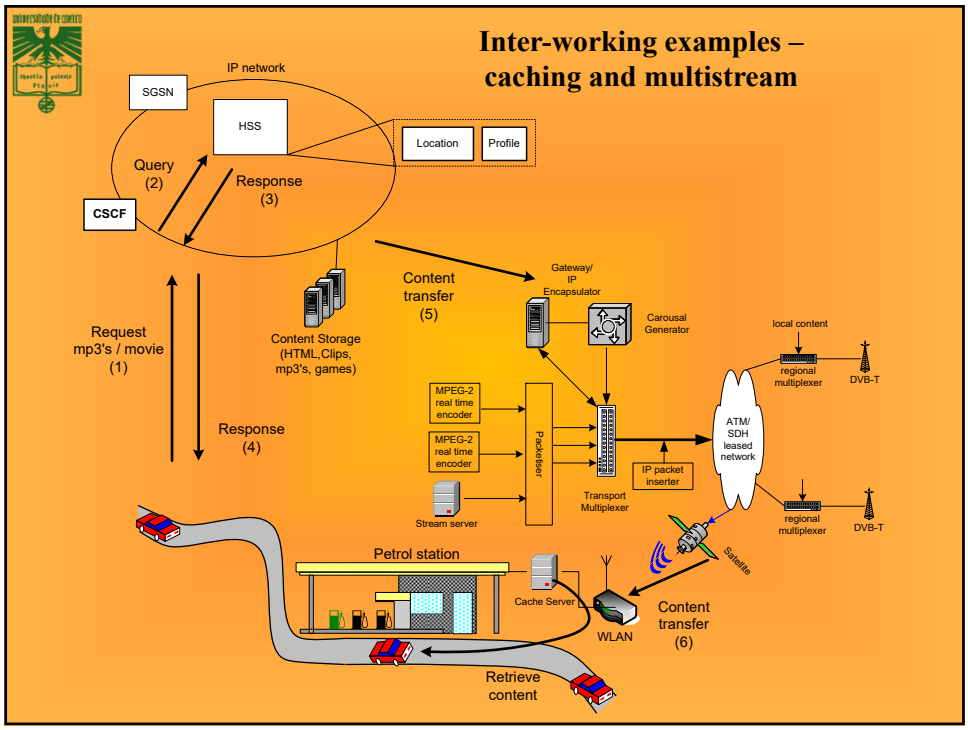
DVB-T IP-TV

interactive TV

134


The content challenge: recall adaptation!
UniCast
MultiCast
BroadCast
euro
Squeeze
108V
70 % interactive 30 % lean back
50 % interactive 50 % lean back
10 % interactive 90 % lean back

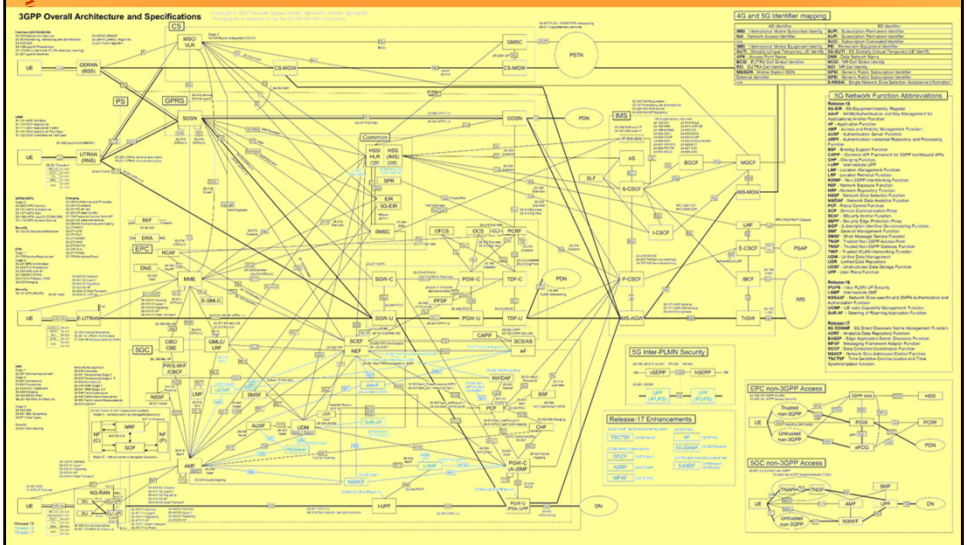
135


Inter-working examples – caching and multistream
IP network
SGSN
HSS
Location
Profile
Query (2)
Response (3)
CSCF
Content transfer (5)
Gateway/ IP Encapsulator
Carousal Generator
local content
regional multiplexer
DVB-T
Request mp3's / movie (1)
Content Storage (HTML,Clips, mp3's, games)
Response (4)
MPEG-2 real time encoder
MPEG-2 real time encoder
Packetiser
Stream server
Transport Multiplexer
IP packet inserter
ATM/ SDH leased network
regional multiplexer
DVB-T
Satellite
Petrol station
Cache Server
WLAN
Content transfer (6)
Retrieve content

136

Example: current mobile network

(without IP/MPLS/CDN/DWDM…)



Conclusions

- Overall large comercial networks are very complex
  - "Transport"/"ISP" is Multidimensional
  - Multistakeholder
  - Very large
  - Limited by costs
- No overall architecture vision easy to grasp exist, not even in comercial operators
  - Different departments handle different aspects