第三章 MPLS技术

张喆
通信与信息工程学院

第三章 MPLS技术

- 3.1 MPLS的标记交换原理及LDP协议(回顾)
- 3.2 MPLS中的流量工程
- 3.3 MPLS VPN
- 3.4 VPLS
- 3.6 GMPLS
- 3.6 高速路由器的设计

本章要点

- 掌握标记分发协议的工作原理
- 掌握VPN
- 了解MPLS的流量工程
- 高速路由器的设计

网络层关注的是可达性问题。
其中的每个协议都解决了一个子问题。

终端如何获取IP地址?

DHCP

Debugging

ICMP

ping

traceroute

终端如何同外网的其他终端通信?

相同网络内的终端如何通信?

Broadcast

ARP

Carleton University

IPv6

Gateways

NAT

Routing protocols

OSPF, RIP, BGP

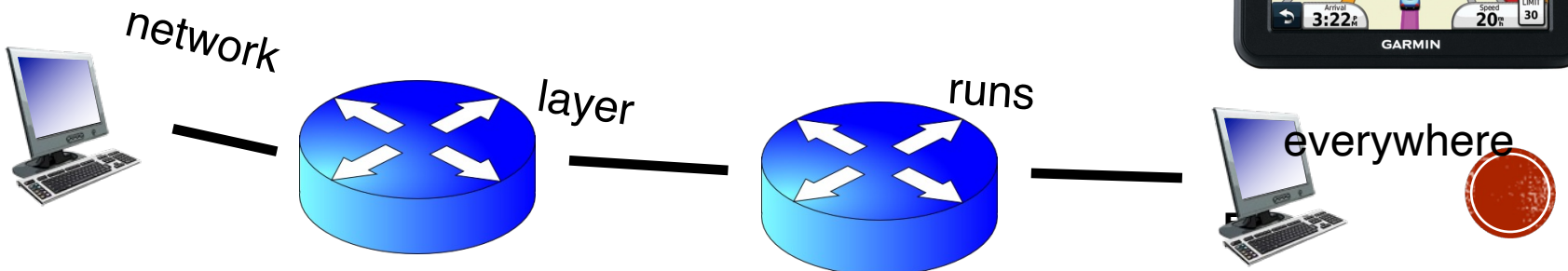
PRINCETON

网络层关键功能

比如: 公路旅行

- 转发 (data plane): 将数据包从路由器的输入移动到相应的路由器输出
- 转发: 每一个岔路口该怎么走

- 路由 (control plane):确定数据包从源到目的地所采取的路径
- 路由: 从出发地到目的地整体规划



Routing is a fundamental problem in networking.

如何设计一个用于在互联网上导航的“高德地图”？

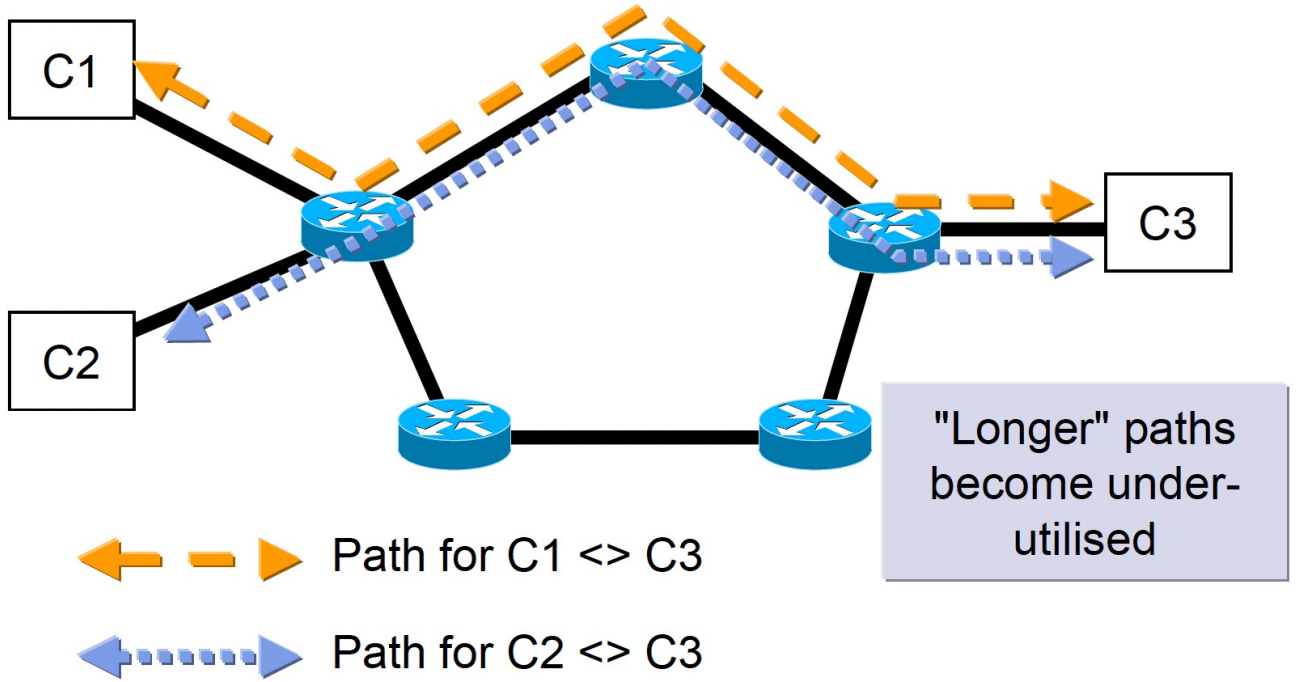
路由

传统路由的缺点

- 性能：
  - 路由表增大->处理时延增大
    - 每条转发决定需要大概1000条机器指令
  - 只支持最短路径算法
  - 基于数据包而非数据流
  - 最长前缀匹配难以转移到硅片(silicon)上。
    - 人们期望构建线速(wire-speed)路由器。
  - 难以实现QoS架构与服务

经典的 “FISH PROBLEM”

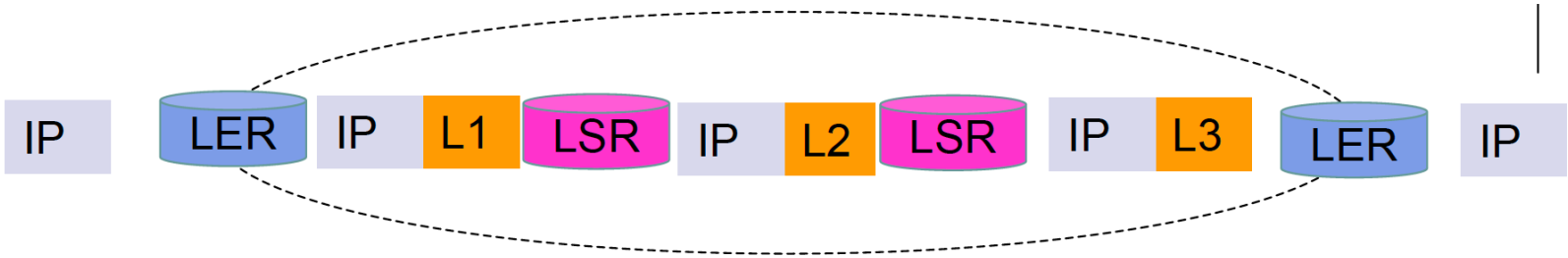
- Routing Protocols Create A Single "Shortest Path"



SOME TERMINOLOGY…

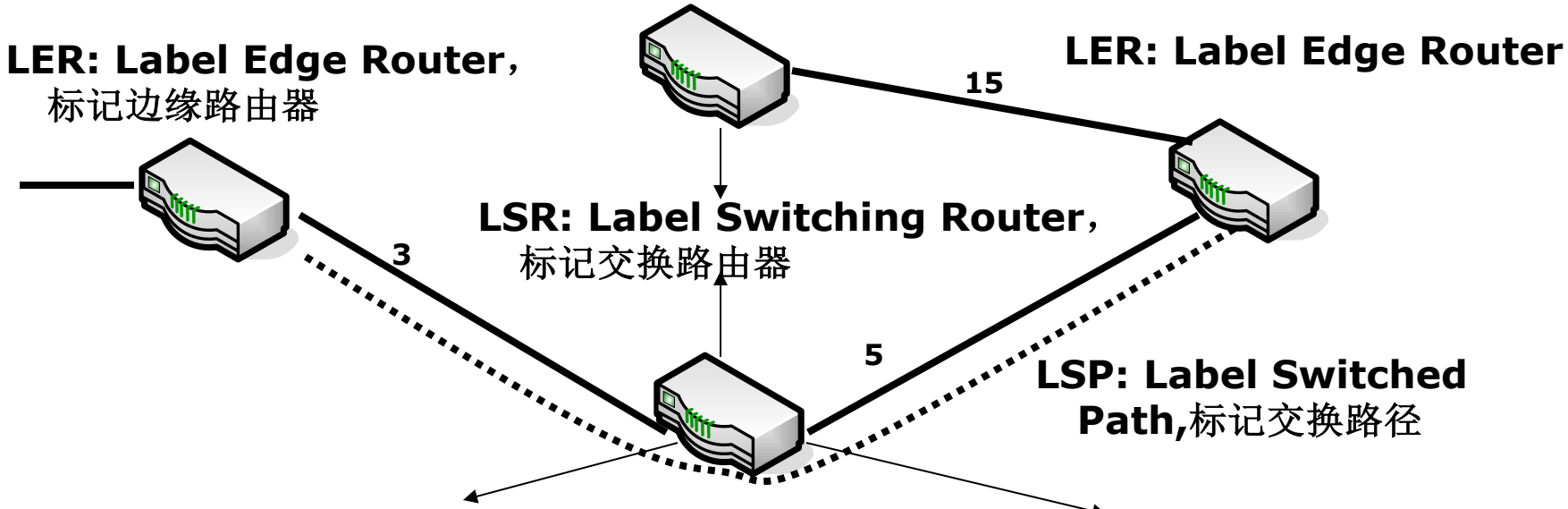
- Network Engineering:
  - Put the bandwidth where the traffic is
    - Physical cable deployment
    - Virtual connection provisioning
- Traffic Engineering:
  - Put the traffic where the bandwidth is
    - On-line or off-line optimization of routes
    - Implies the ability to diversify routes

MPLS概念



- *Multiprotocol Label Switching (MPLS): one of TE-aware routing protocols*
- 一组使MPLS网络得以实现的协议。
  - 数据包由边缘路由器（执行最长前缀匹配）分配标签(label)。
  - 数据包使用标签交换沿着MPLS网络中的*Label-Switched Path*（LSP）进行转发。
  - LSP可以通过多个第二层链路进行创建。
    - 支持ATM、以太网、PPP、帧中继等多种链路类型。
  - LSP可以支持多个第三层协议，例如IPv4、IPv6以及其他协议。

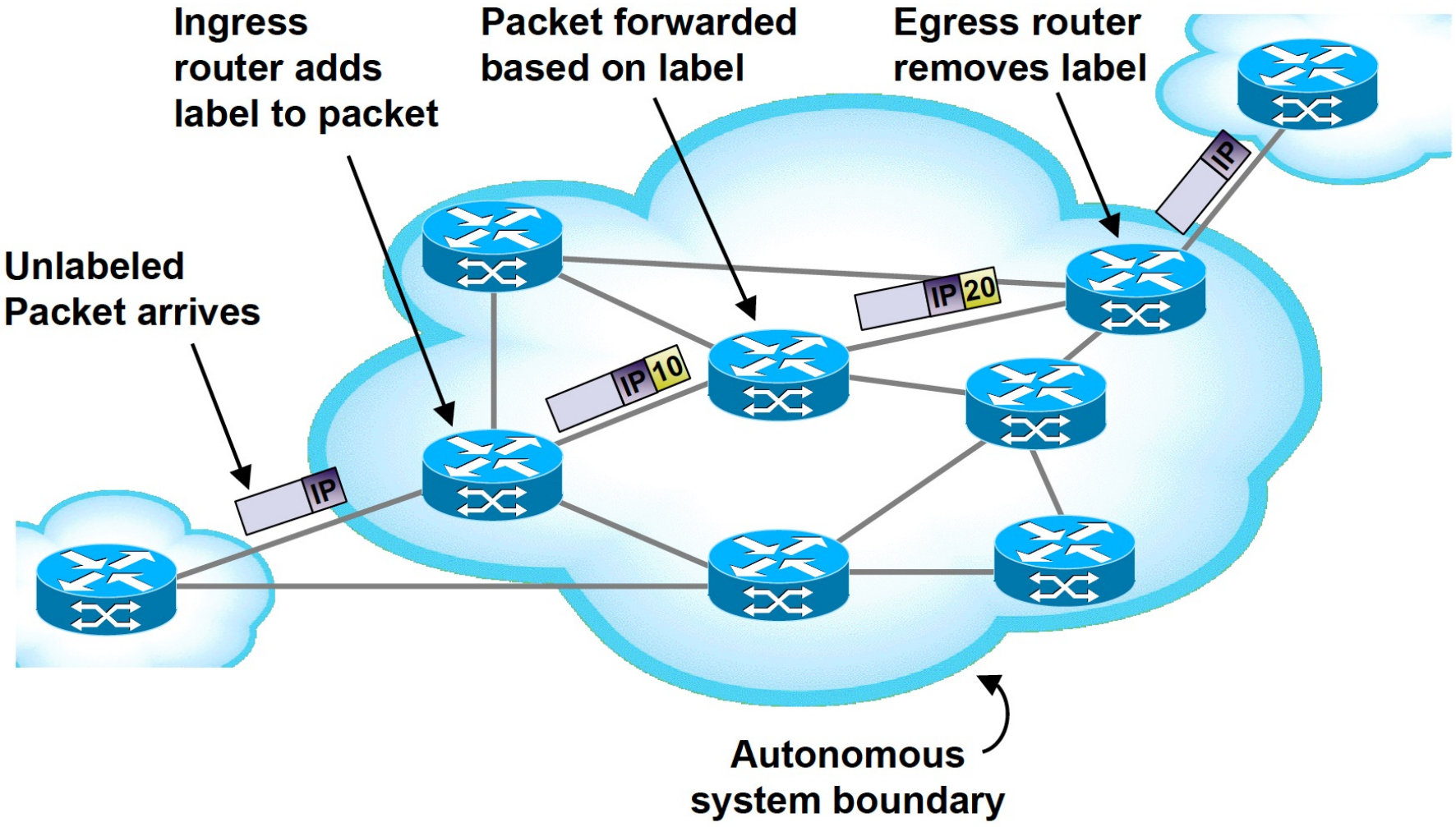
MPLS概念

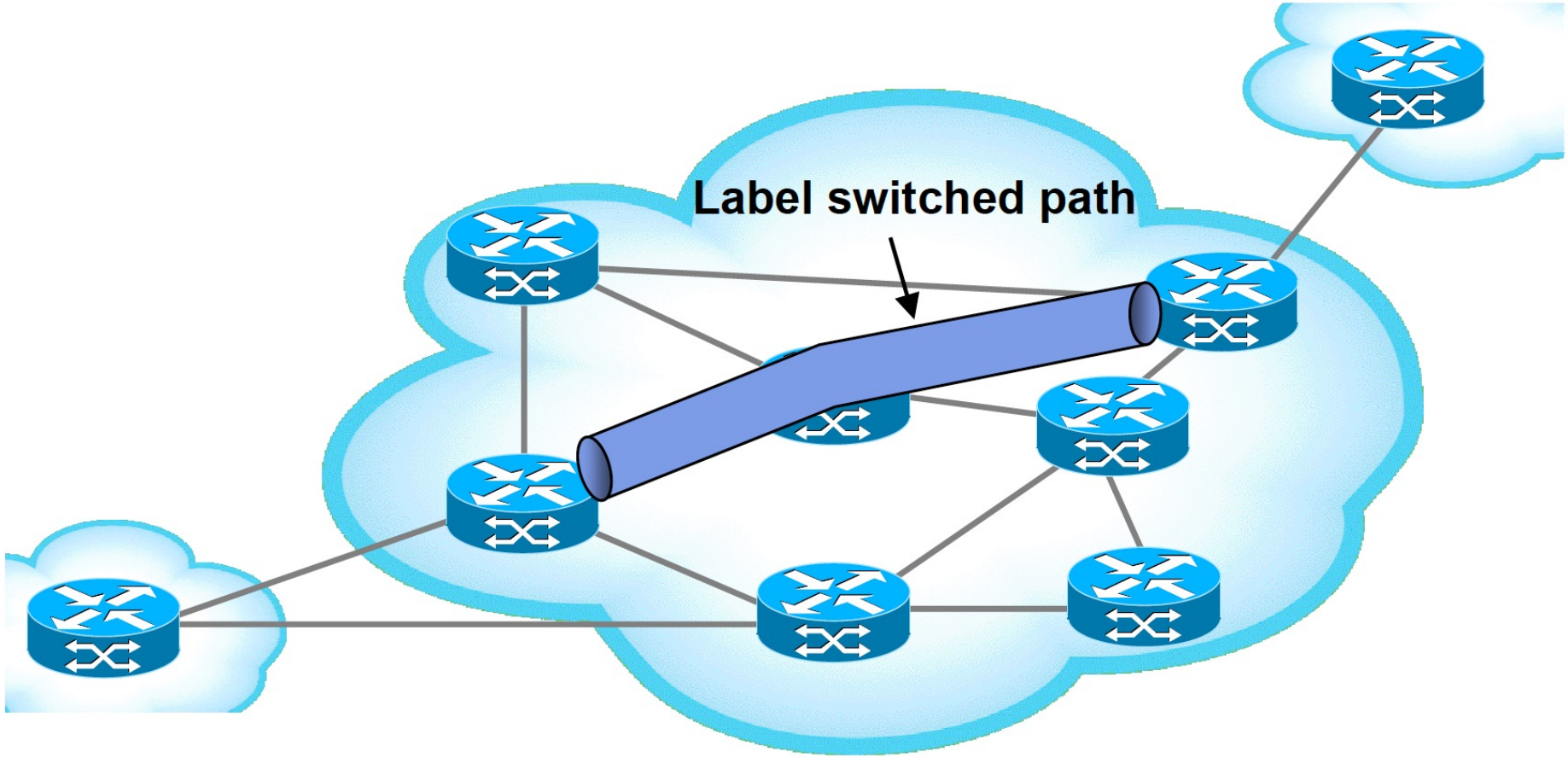


目的网段	输入端口	输入标记	输出端口	输出标记
10.10	1	5	2	3
20.10	3	15	2	23

LFIB：Label Forwarding Information Base
标记转发信息表

MPLS工作原理


Ingress
router adds
label to packet
Packet forwarded
based on label
Egress router
removes label
Unlabeled
Packet arrives
IP
IP 10
IP 20
IP
Autonomous
system boundary

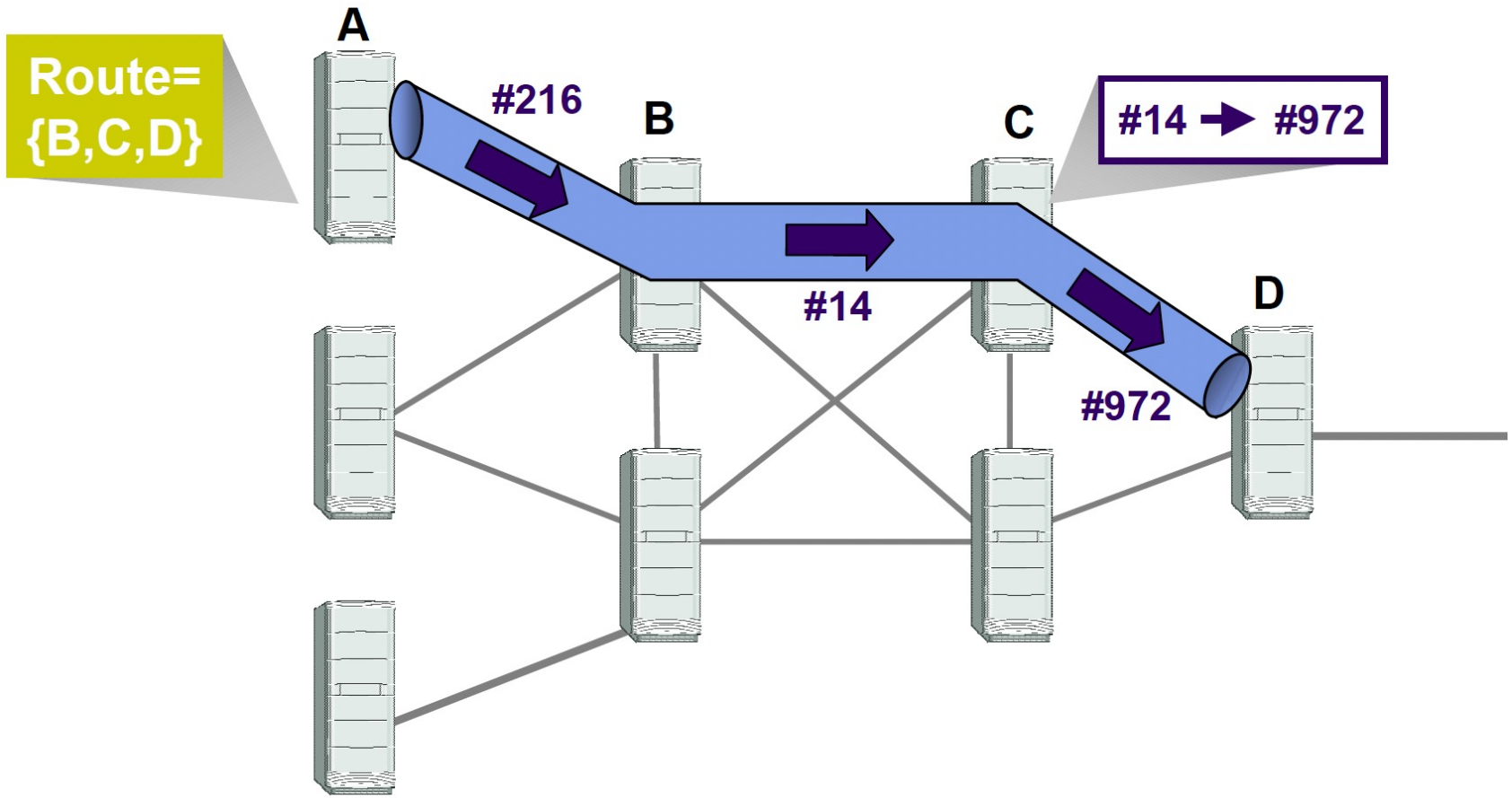


- Label Switched Path is like a pipe or tunnel
- While traveling on a label switched path, forwarding is based on the label only, not on destination IP address in packet

标签交换路径（LABEL SWITCHED PATHS, LSPS）

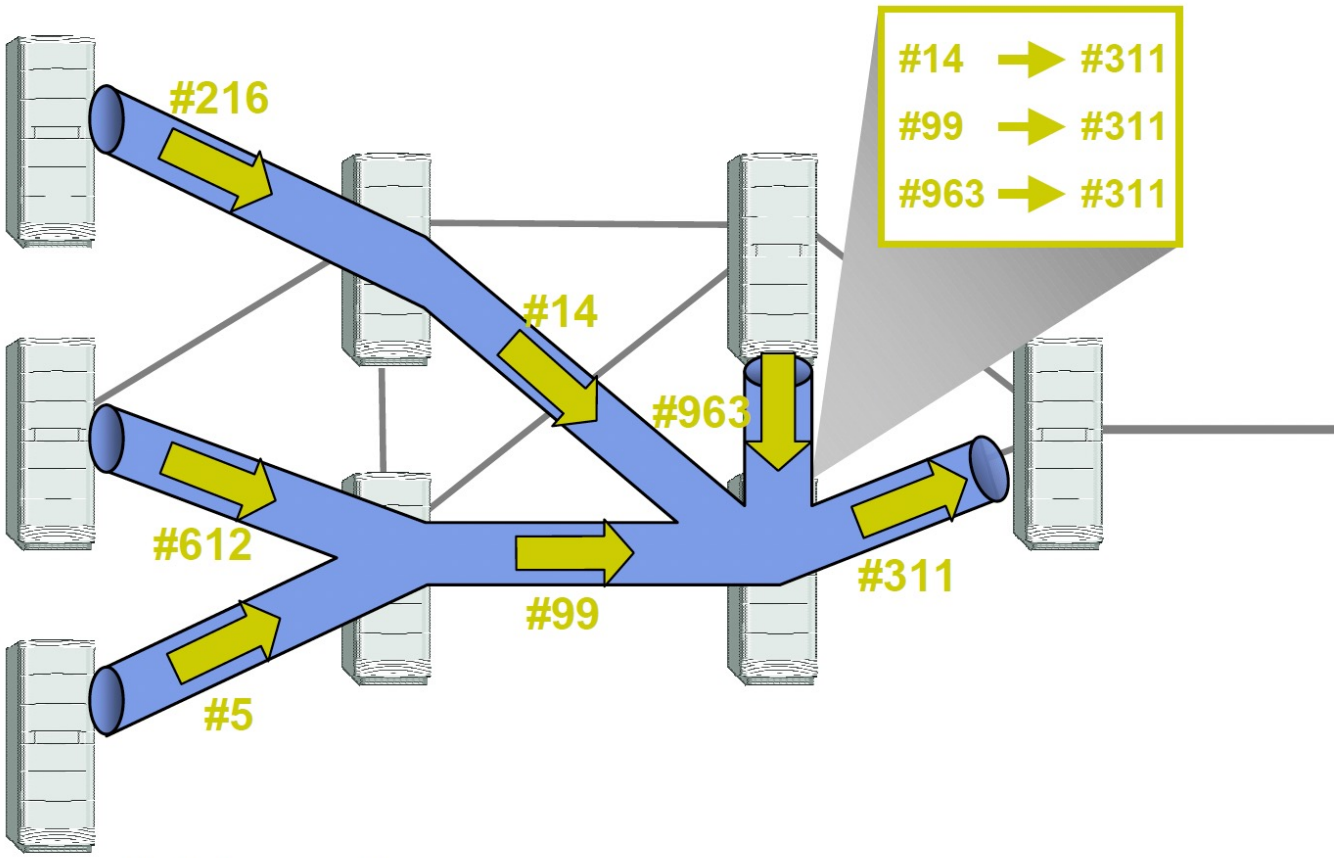
- LSPs:
  - 通常被称为“隧道”
  - 始终是单向的
  - 为方便起见，可以分类为以下两种类型：
    - **点对点（Point-to-point），或**
    - **合并（Merging）**

POINT-TO-POINT LSP



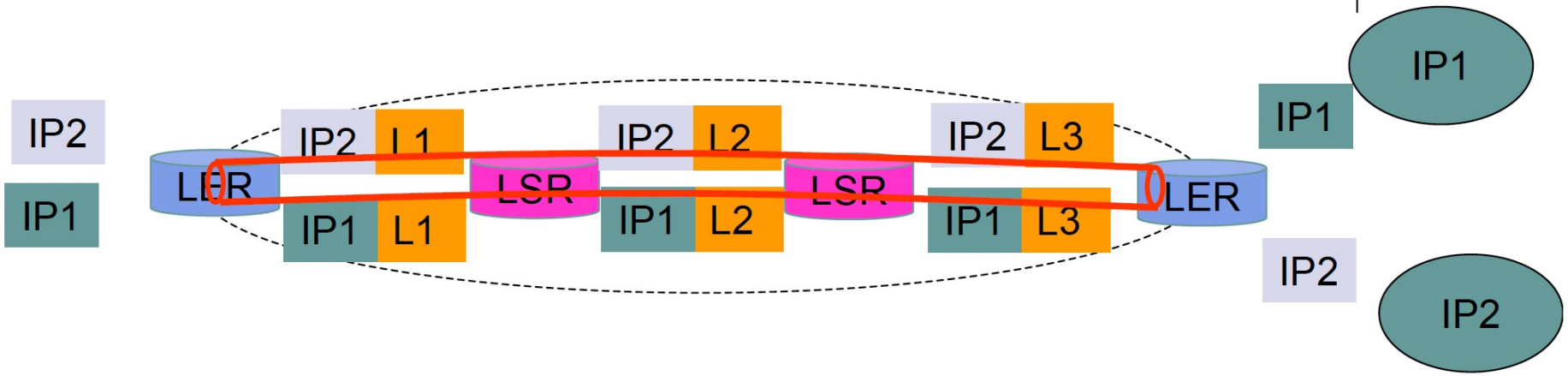
- LSP follows route chosen when LSP is set up.

MERGING LSP



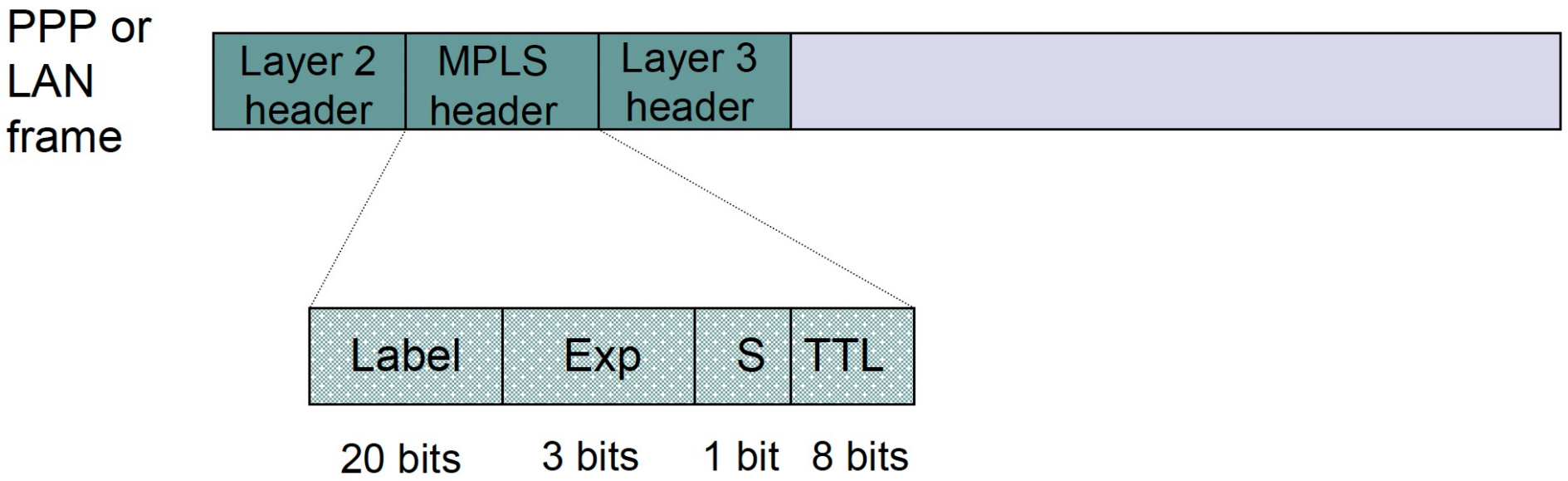
- LSP forms a “sink tree”
- The branches of the LSP always follows the same route as normal IP forwarding; that is, the *shortest path*

Forwarding Equivalence Class



- *FEC:* set of packets that are forwarded in the same manner
  - Over the same path, with the same forwarding treatment
  - Packets in an FEC have same next-hop router
  - Packets in same FEC may have different network layer header
  - Each FEC requires a *single entry* in the forwarding table
  - Coarse Granularity FEC: packets for all networks whose destination address matches a given address prefix
  - Fine Granularity FEC: packets that belong to a particular application running between a pair of computers

MPLS Labels

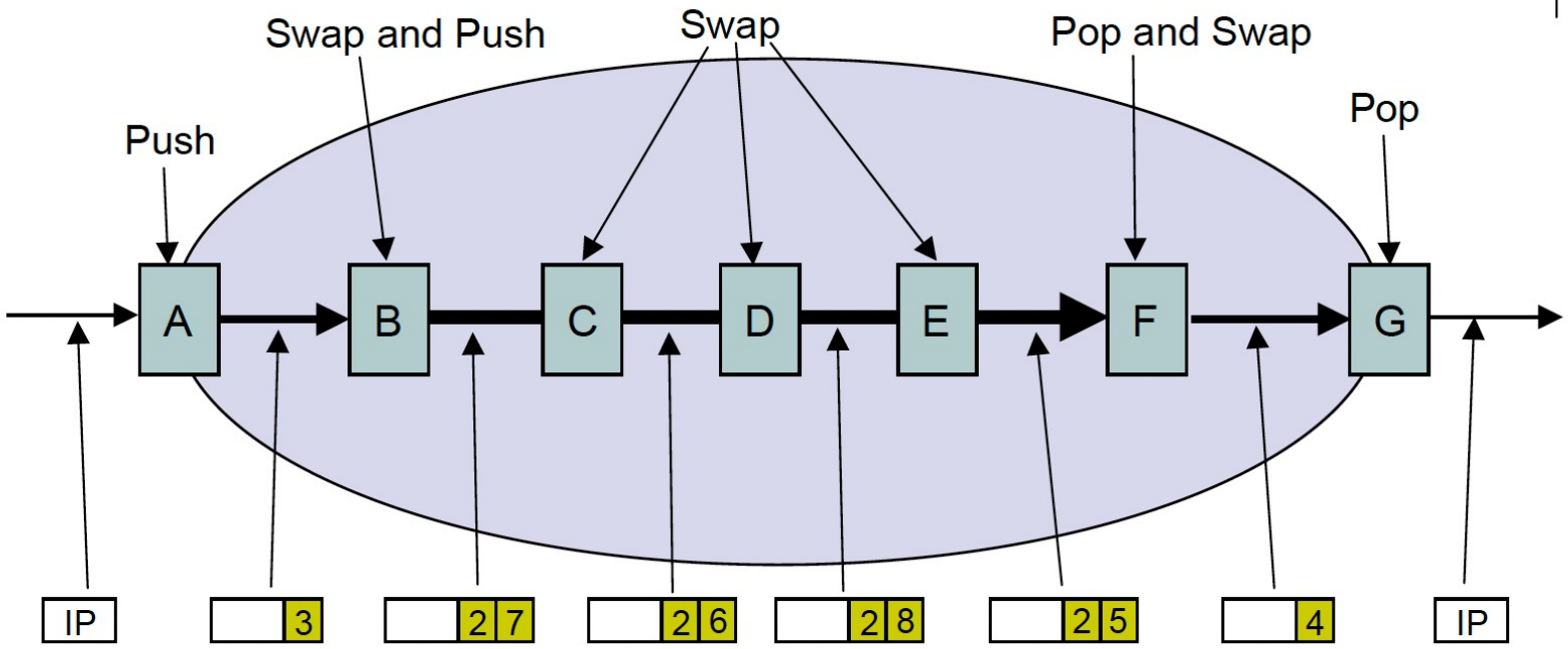


- *Shim header* between layer 2 & layer 3 header (32 bits)
  - 20-bit label + 1-bit hierarchical stack field + 8-bit TTL
  - 3-bit “experimental” field (can be used to specity 8 DiffServ PHBs)

标签堆栈(LABEL STACKING)

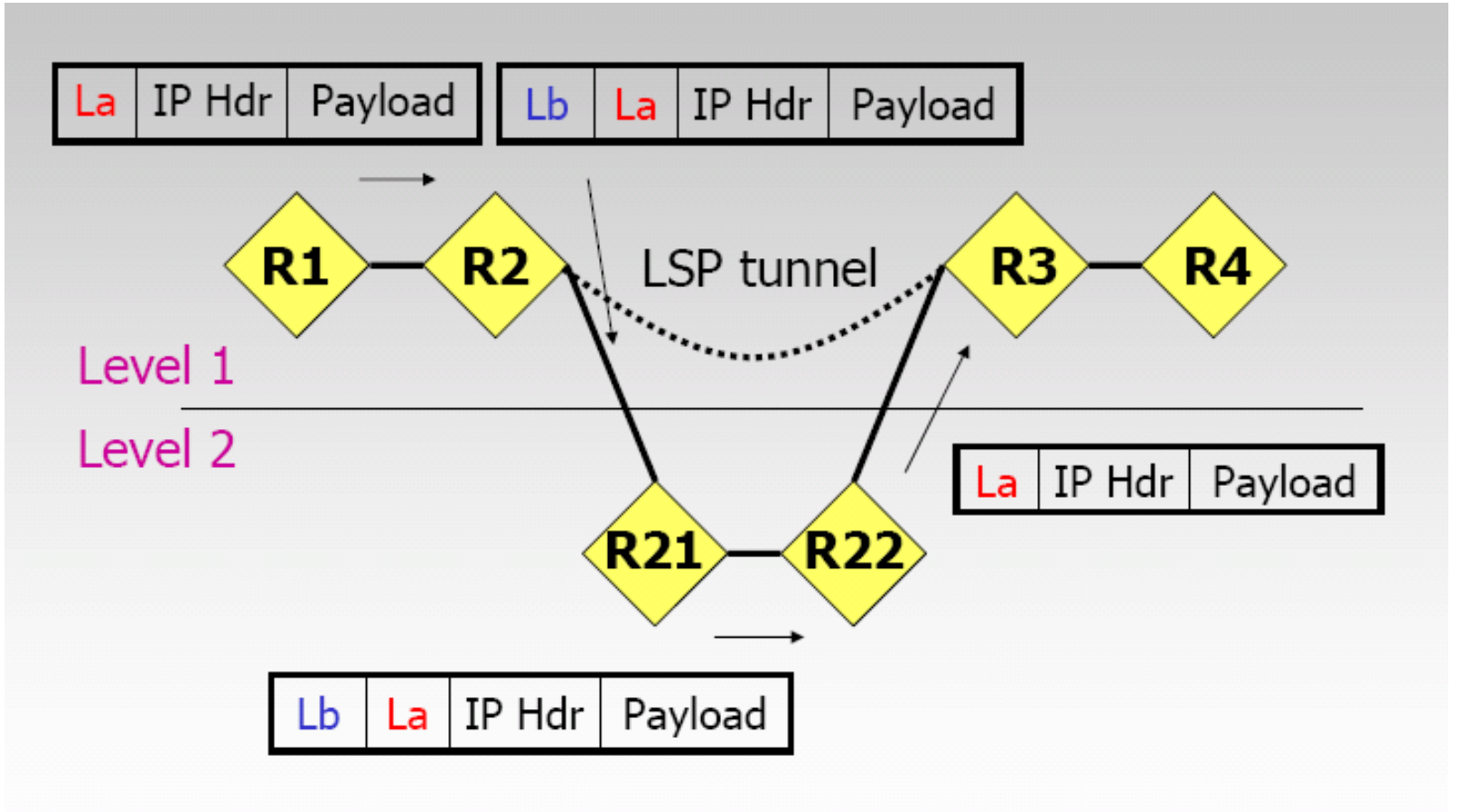
- 对于大型网络，只用一个标签进行路径标识显得“力不从心”。
- MPLS中分组**可以携带多个标签**，这些标签在分组中以“堆栈”的形式存在，对标签堆栈的操作按照“**后进先出**”的原则；

Label Stacking



- MPLS allows multiple labels to be stacked
  - Ingress LSR performs *label push* (S=1 in label)
  - Egress LSR performs *label pop*
  - Intermediate LSRs can perform additional pushes & pops (S=0 in label) to create tunnels
  - Above figure has tunnel between A & G; tunnel between B&F

标签堆栈


La
IP Hdr
Payload
Lb
La
IP Hdr
Payload
R1
R2
LSP tunnel
R3
R4
Level 1
Level 2
La
IP Hdr
Payload
R21
R22
Lb
La
IP Hdr
Payload

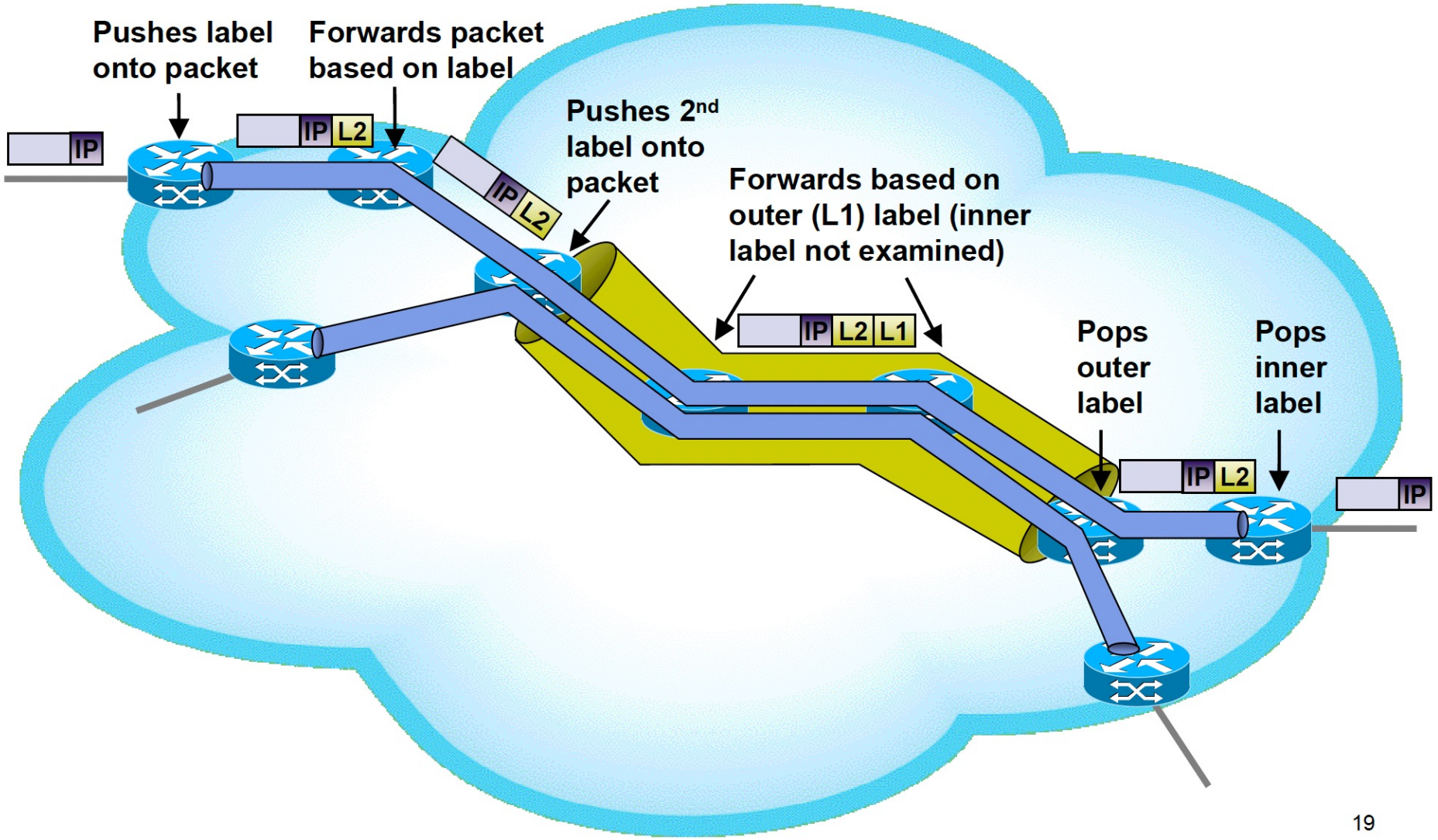
标签堆栈

- **决定如何转发分组的标签始终是栈顶标签**，事实上标签交换路由器并不考虑此标签堆栈有几层，对标签分组的处理在不同子网和不同层次的网络中相互独立；

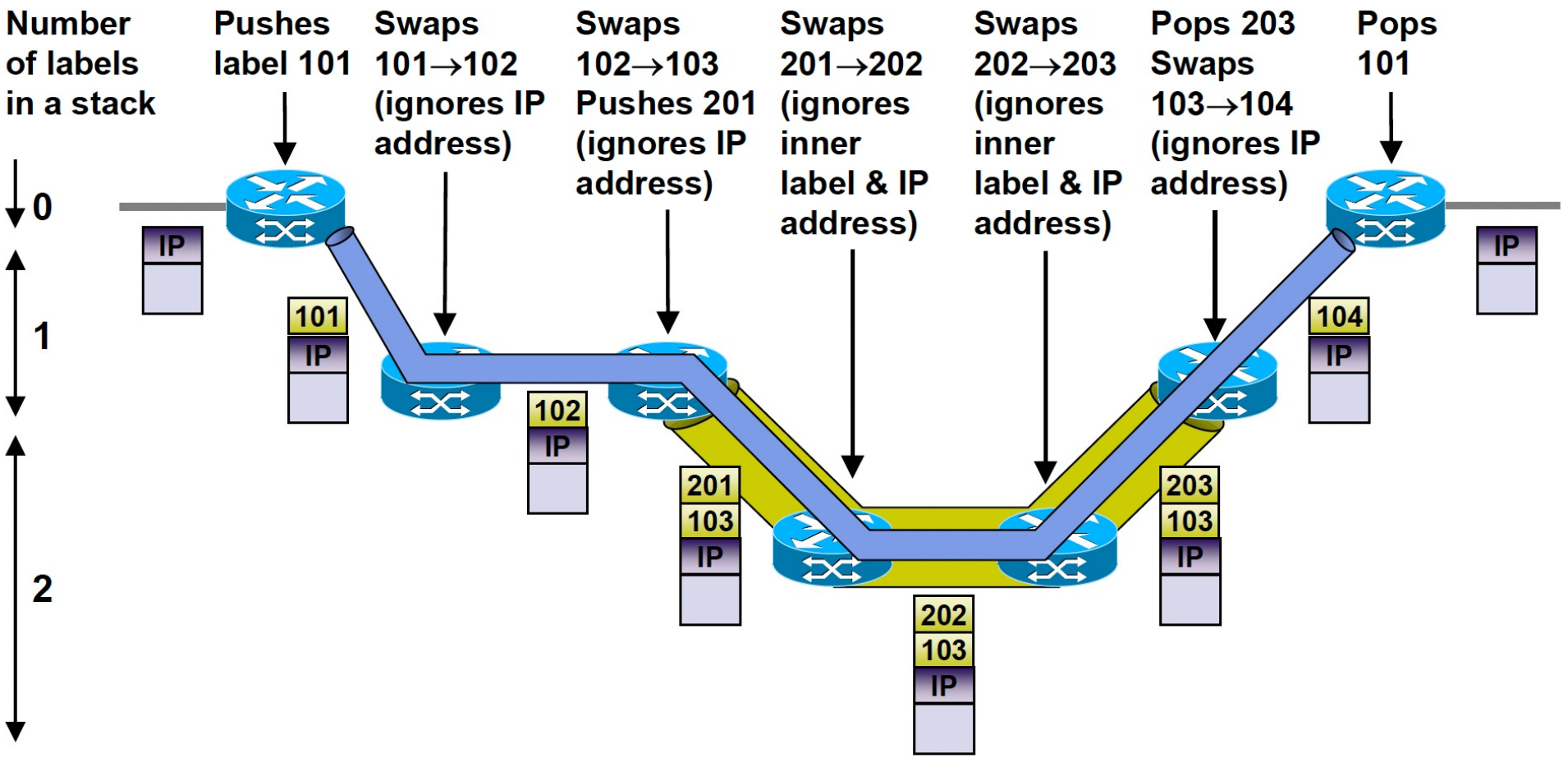
- MPLS中将标签堆栈的层数叫做标签堆栈的“深度”；

- 通过这种**隧道嵌套技术**可以支持任意庞大的网络；

Label Stacking Example



Label Stacking Details



MPLS的技术特点（1）

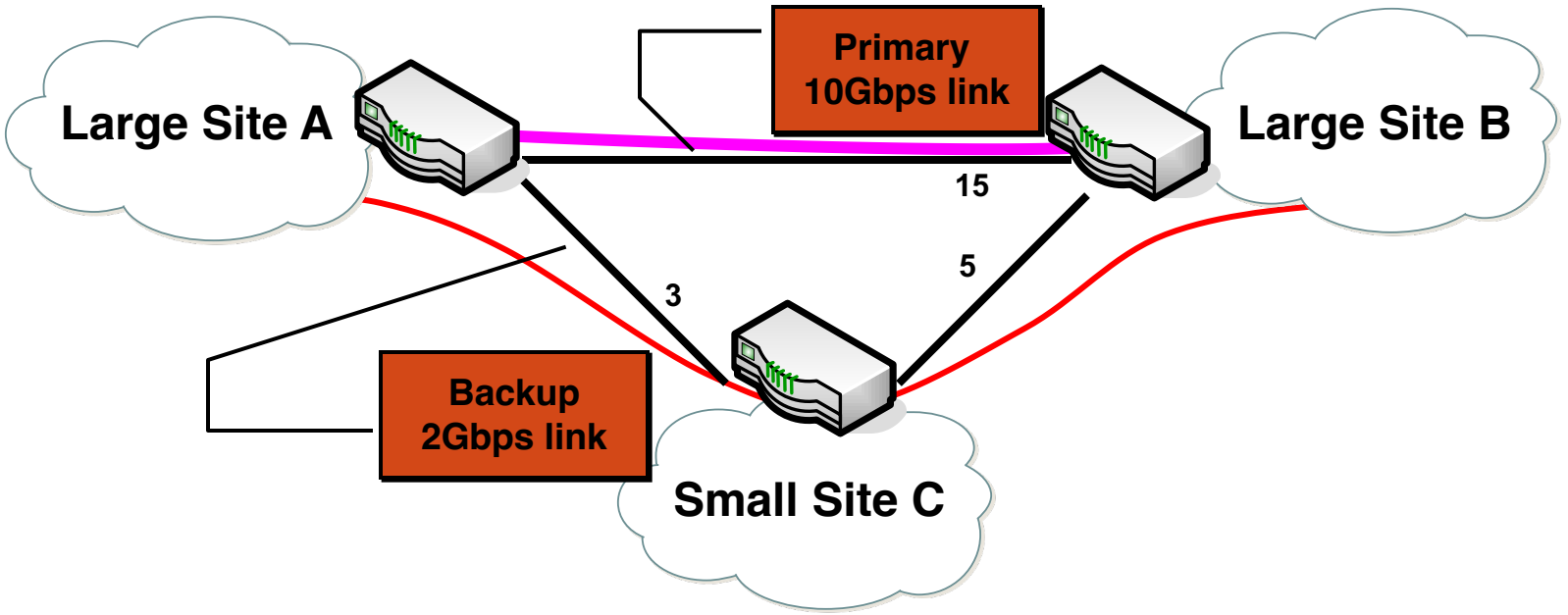
- **1 MPLS简化了分组的转发**
  - 基于定长短标签定完全匹配;
  - MPLS易制造高速路由器
- **2 MPLS支持有效的显式路由explicit routing**
  - 显式路由在网络**负荷调节**，保证QoS要求等方面起着重要作用;
  - 传统IP网络中，每个分组头都携带显式路由是不可能的;
  - MPLS只是在LSP建立时使用

MPLS的技术特点（2）

- **3 MPLS有利于实现流量工程（Traffic Engineering）**
  - TE:根据用户数据业务量及当前网络状态选择数据传输路径的过程，主要用来**平衡网络中的负荷**；

- **4 MPLS支持QoS选路**
  - QoS选路是指对特定的数据流，按其QoS要求来为它选择路由的方法。

MPLS中的流量工程



- 基于其它参数(QoS、源地址等)将分组分为不同的FEC，并且分配不同的标记，从而选择不同的LSP
- 实现不同路径上的负载均衡

MPLS的技术特点（3）

- **5 从IP分组到转发等价类的映射**

  - MPLS只需要在**其域的入口**进行一次从IP分组到FEC的映射，使得**IP分组到FEC的复杂转换得以简化**；

  - 注意：从**IP分组到服务等级的映射可能需要知道发送IP分组的用户**，从而才可能根据源地址、目的地址、输入接口或其他特征实现分组过滤，但某些信息只能在网络入口节点才能获得。

MPLS的技术特点（4）

- **6 MPLS支持多网络功能划分**
  - MPLS引入了标记粒度的概念，使其能分层地将处理功能划分给不同的网络单元，让靠近用户的网络边缘节点承担更多的工作；
  - 与此同时，**核心网络则尽可能地简单**。

MPLS的技术特点（5）

- **7 MPLS实现了用户不同服务级别要求的单一转发规范**；
- **8 MPLS提高了网络扩展性**：传统的IP与ATM结合是依靠中间层的翻译，这种方式带来了一系列的后果，如虚电路的“N的平方”问题，MPLS通过减少对等实体的数量、去掉路由器之间全网格状的n平方逻辑链路连接，（MPLS省略了把IP地址和路由映射到ATM交换表上的复杂性，）提高了可扩展性。

如何分发以及绑定标签呢？

- 标签分发协议
- 标签分发模式
- 标签分发控制方法

标签分发协议有哪些?

- **标签分发协议（LDP：Label Distribution Protocol ）**
- **RSVP**

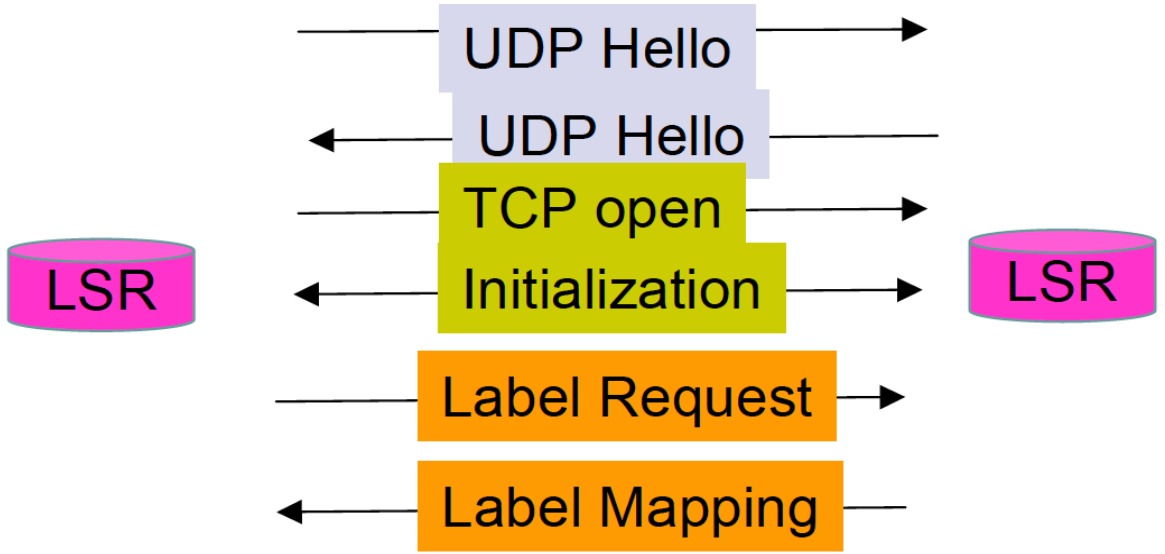
标签分发协议（LDP）的概念

- **标签分发协议（LDP：Label Distribution Protocol）:控制LSR之间交换标签与FEC绑定消息，协调LSR之间工作的一系列规程。**

- **在RFC 3036中详细定义。**

- **LDP功能：让LSR实现FEC与标签的绑定，并将这种绑定通知给相邻的LSR，使各LSR对收到的标签绑定达成共识。**

标签分发的工作原理与过程



- **Label Distribution Protocol (LDP), RFC 3036:**
  - Topology-driven assignment (routes specified by routing protocol)
  - Hello messages over UDP
  - TCP connection & negotiation (session parameters & label distribution option, label ranges, valid timers)
  - Message exchange (label request/mapping/withdraw)

RSVP

- RSVP同样可以用作标签分发
  - Simplex
    - Request resource: Sender->receiver
    - Sender sends PATH message to describe traffic flow
  - Receiver-oriented
    - Receivers initiate and maintain resource reservations
    - Receiver sends RESV message to reserve resource
  - Soft-state at intermediate routers
    - Reservation valid for specified duration
    - Released after timeout

标签的分发模式

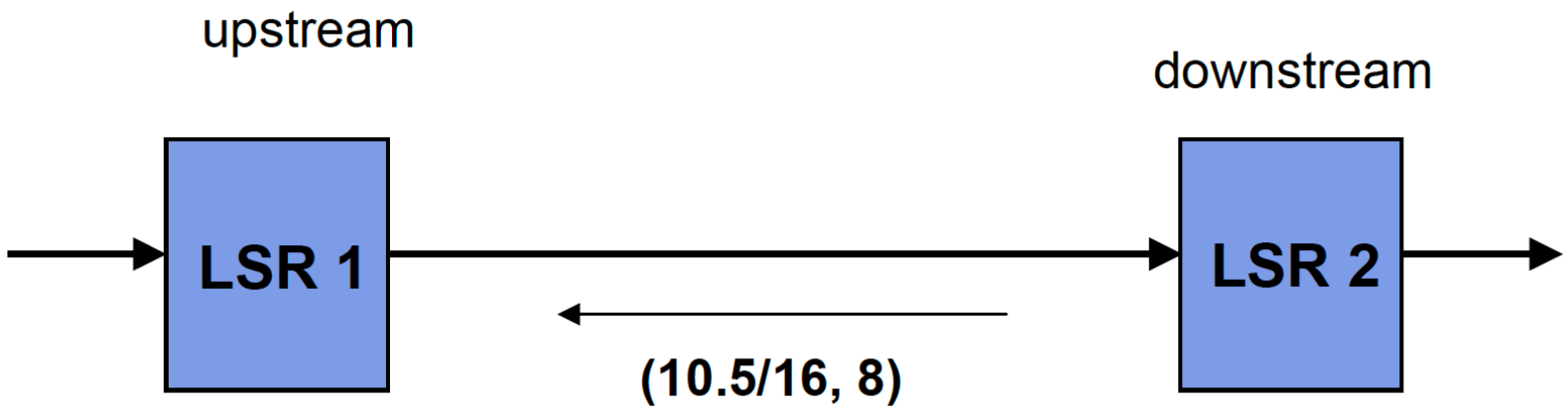
- Downstream Unsolicited (MPLS-BGP、LDP)

下游节点向上游节点主动公告标记

- Downstream-on-Demand (RSVP-TE)

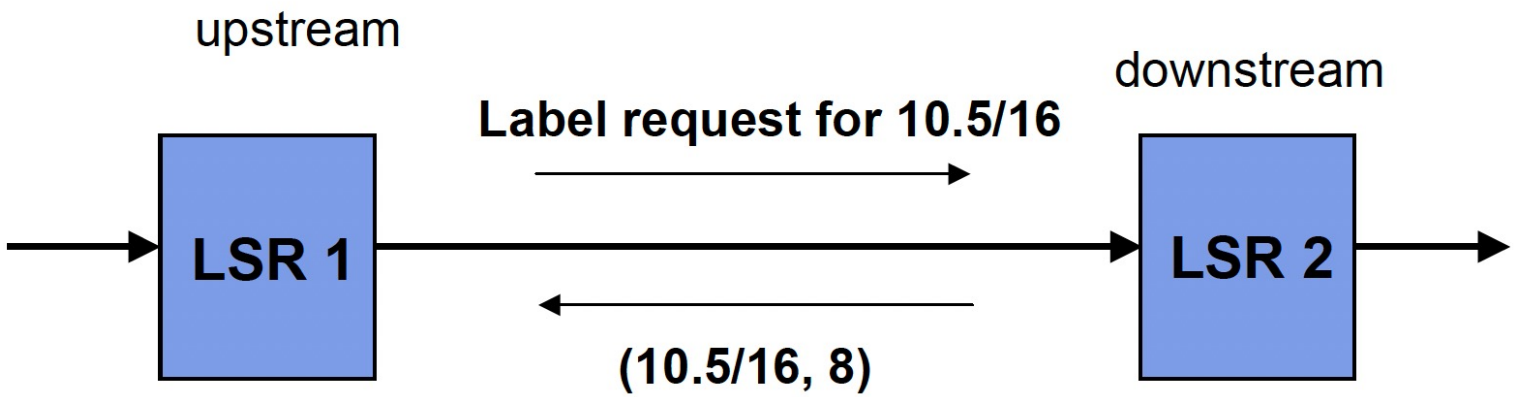
上游节点向下游节点请求标记

DOWNSTREAM UNSOLICITED MODE



- LSR2 becomes aware of a next hop for an FEC
- LSR2 creates a label for the FEC and forwards it to LSR1
- LSR1 can use this label if it finds that LSR2 is next-hop for that FEC

DOWNSTREAM-ON-DEMAND MODE



- LSR1 becomes aware LSR2 is next-hop in an FEC
- LSR1 requests a label from LSR2 for given FEC
- LSR2 checks that it has next-hop for FEC, responds with label

标签分发控制方法

- Independent label distribution control:
  - LSR可以独立完成FEC与label的绑定，并分发给其邻居

- Ordered label distribution control: LSR can distribute label if
  - It is an egress LSR
  - It has received FEC-label binding for that FEC from its next hop
  - 需要协同

MPLS整个过程

- Topology determination
- Path selection/creation
- Data forwarding

MPLS整个过程

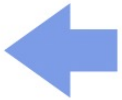
- Topology determination

TOPOLOGY DETERMINATION

- 建立在现有的链路状态路由协议之上: OSPF, IS-IS
- 添加TE扩展: OSPF-TE & IS-IS-TE.
  - Available bandwidth/resource information

MPLS整个过程

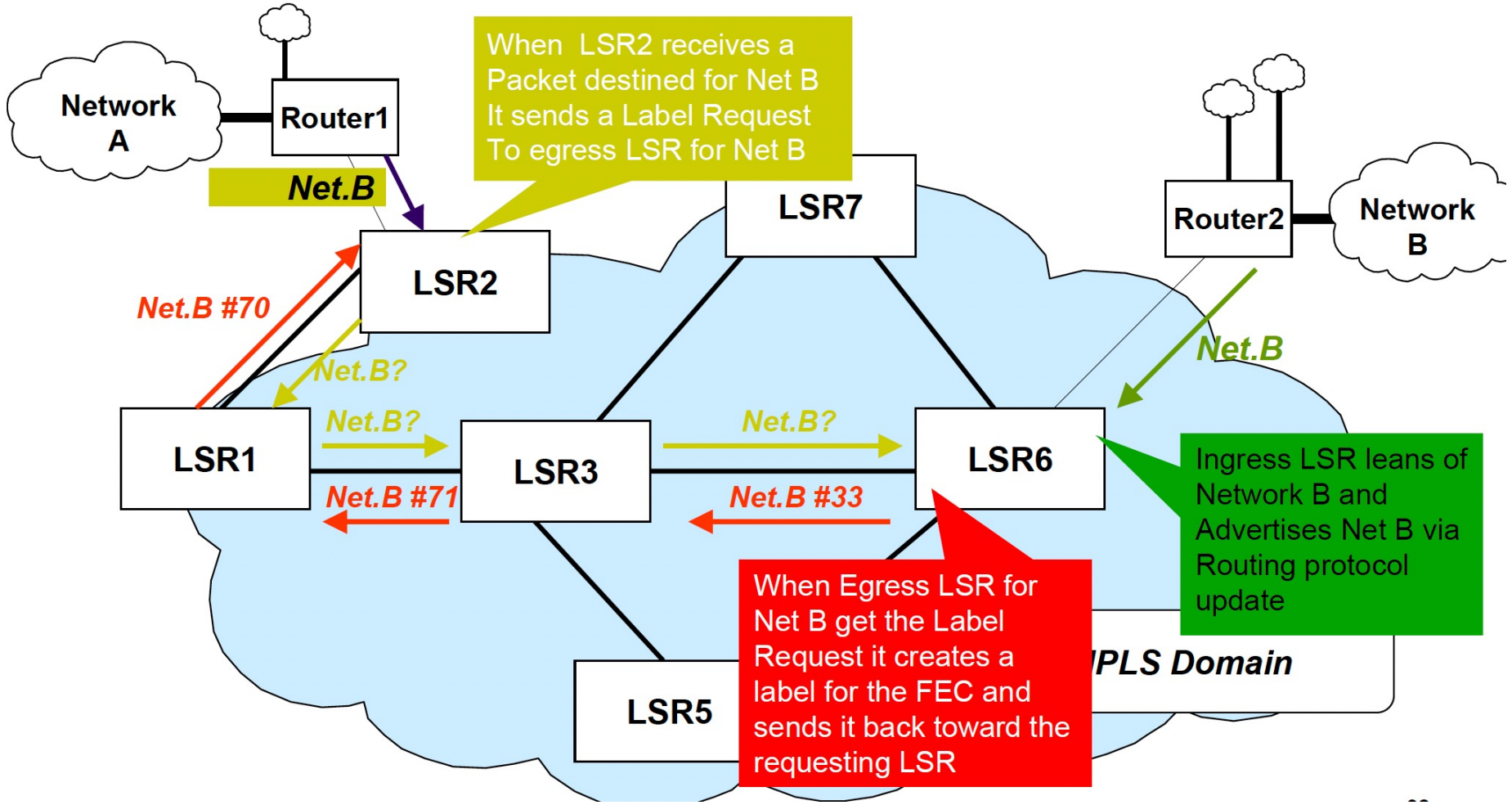
- Path selection/creation



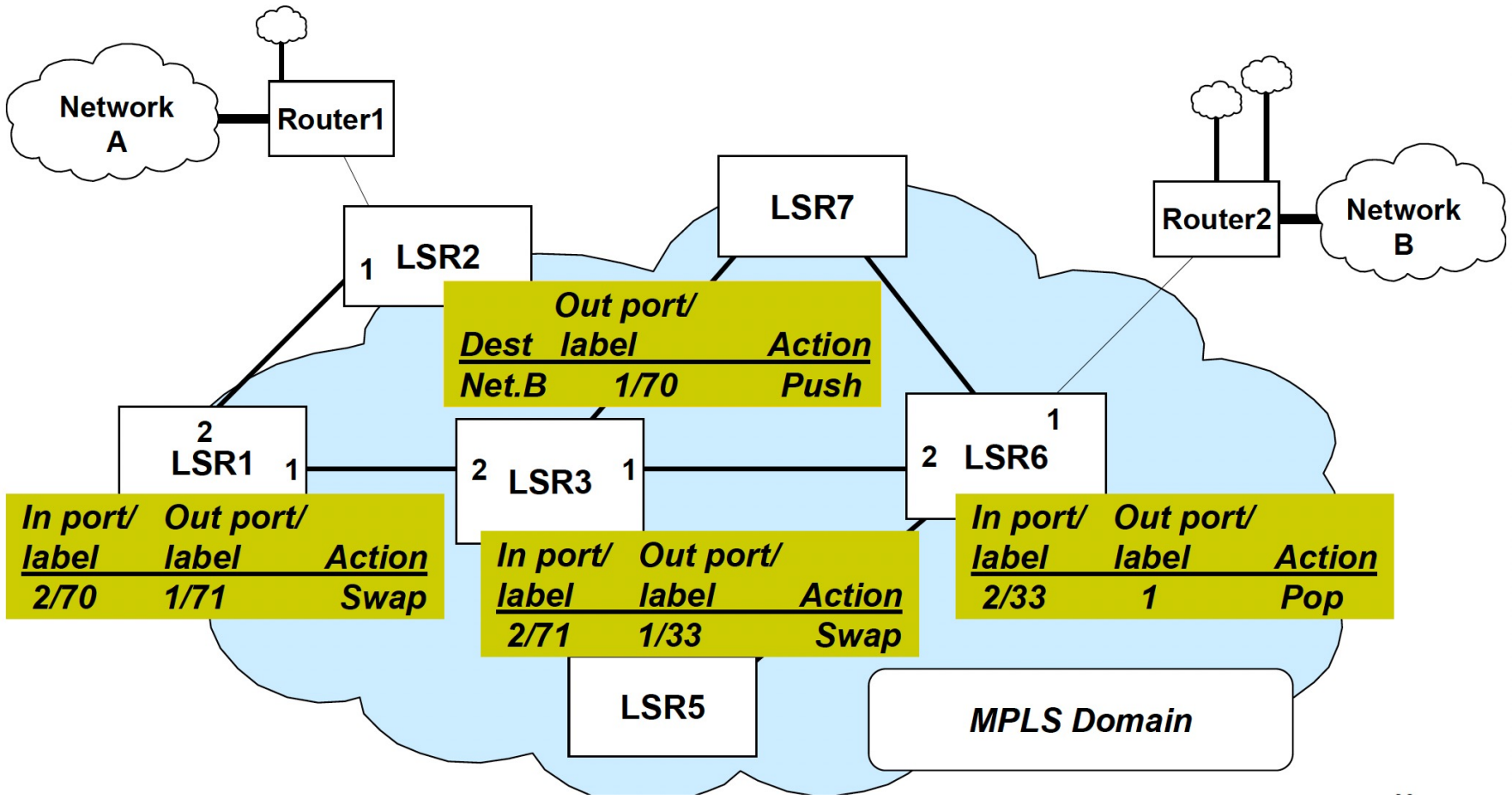
PATH SELECTION

- 两种方法：
  - Data Driven（逐跳路由）
    - Path is determined as label request messages progress inside the network
    - Using IP routing protocols for routing request messages
    - Sometimes referred to as connectionless MPLS
  - Explicit Route（显式路由）
    - Path is determined by the source route
    - Connection-oriented MPLS

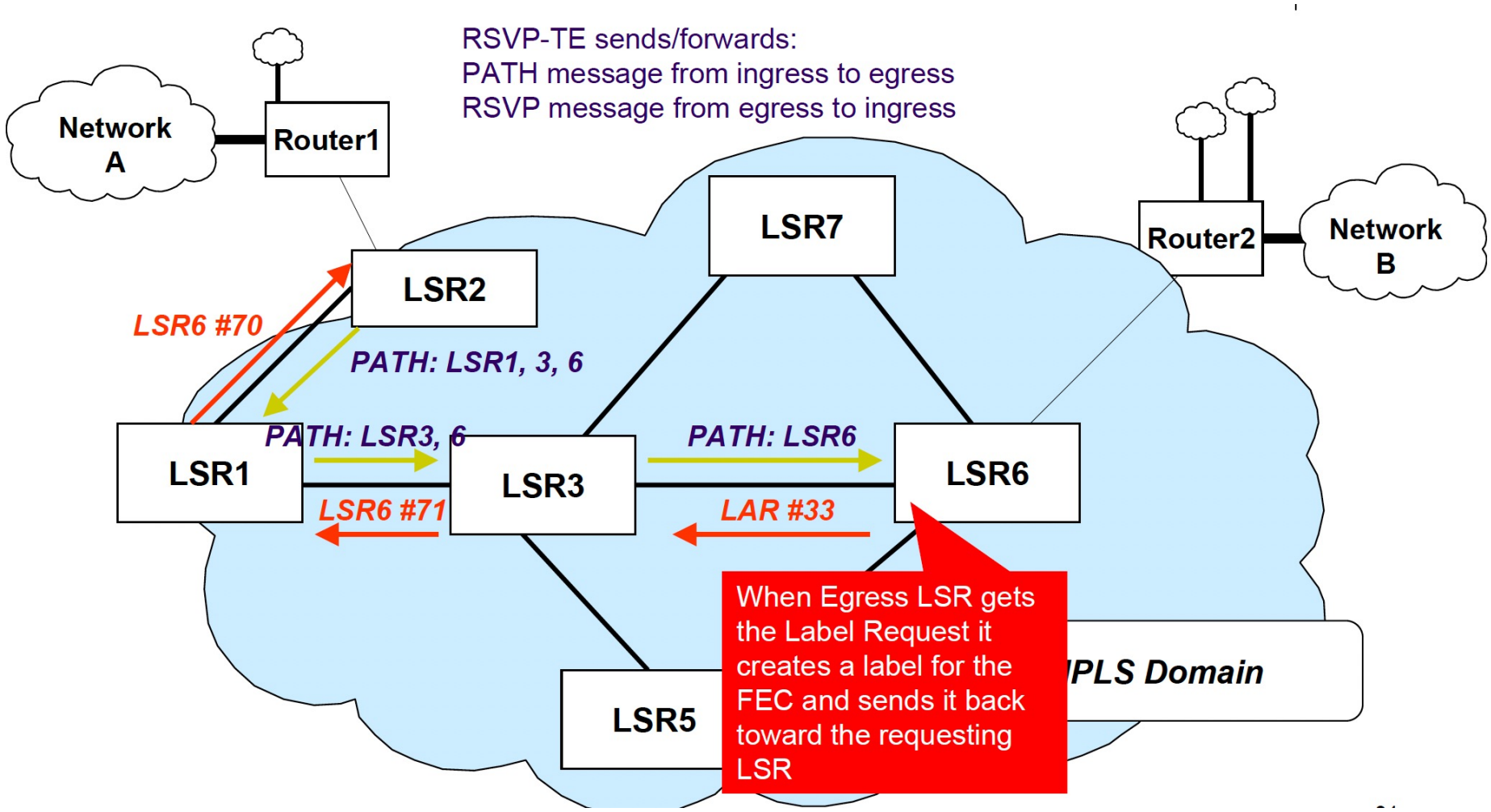
DOWNSTREAM-ON-DEMAND DATA DRIVEN



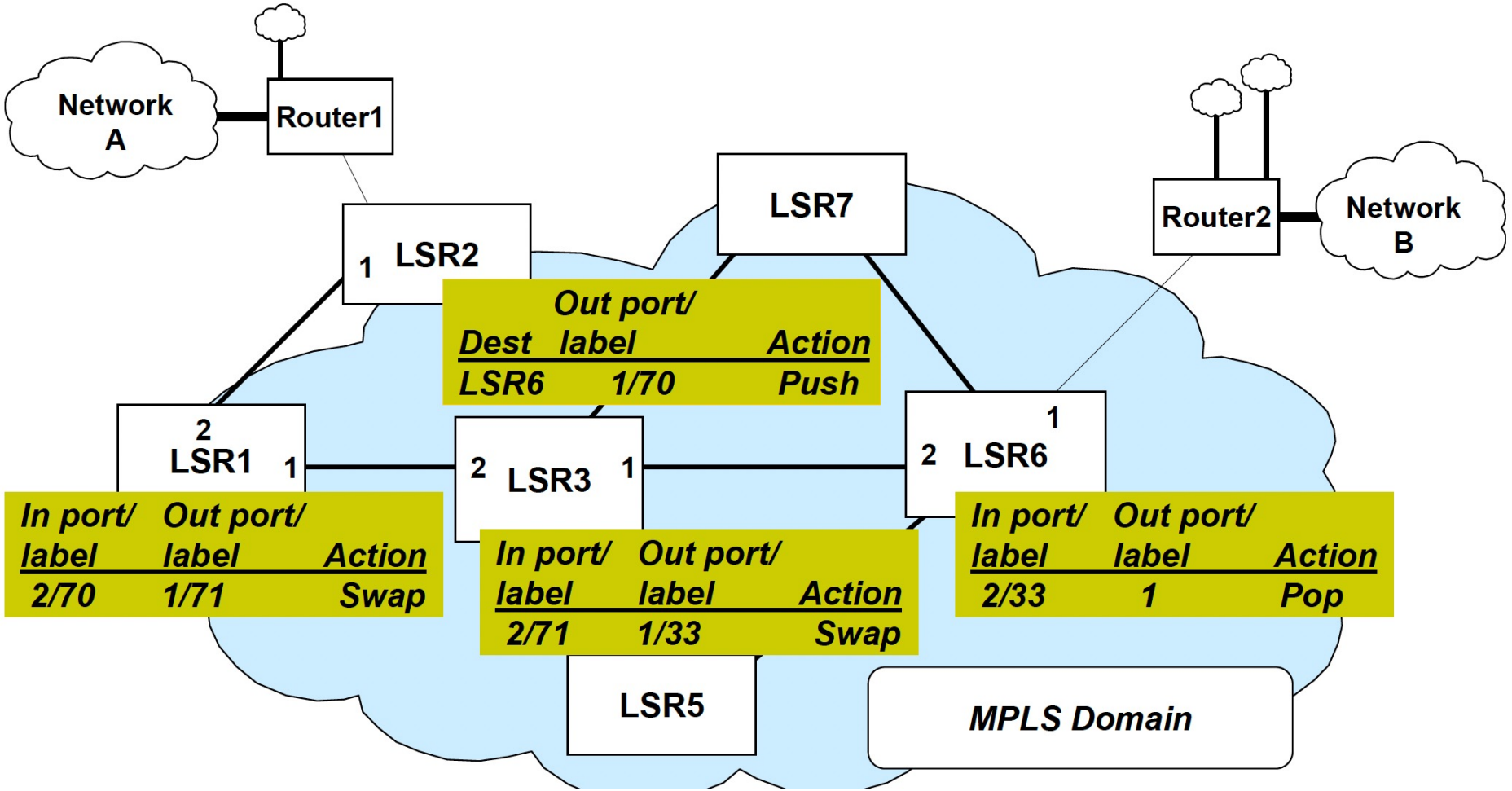
PATH-CREATED



DOWNSTREAM-ON-DEMAND EXPLICIT ROUTE



PATH-CREATED



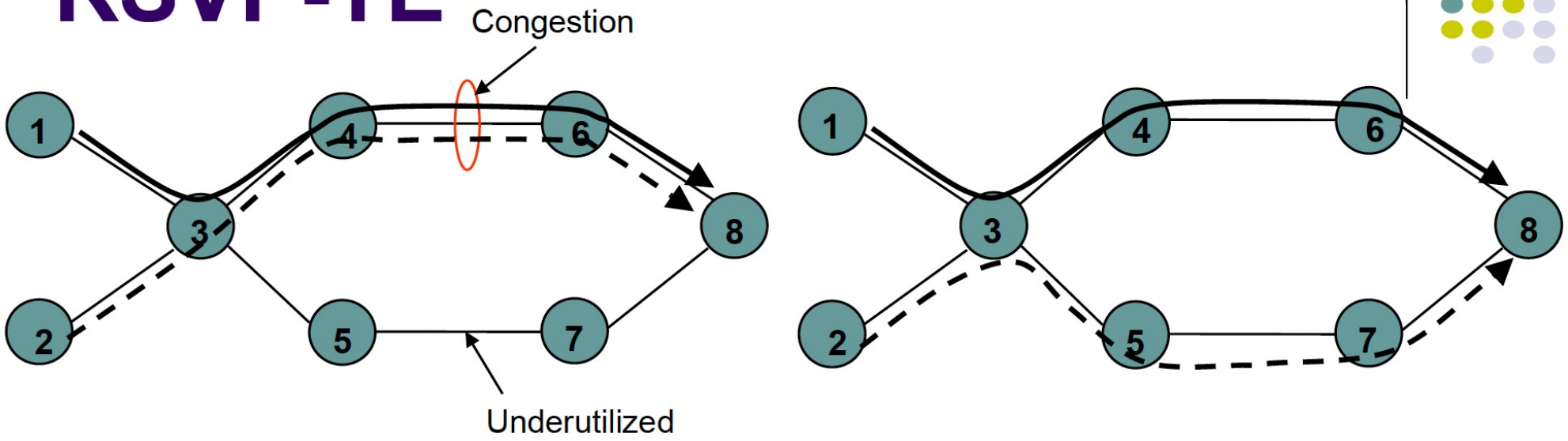
两种LSP建立方式的比较

- **逐跳路由**：实现简单，利用传统路由协议（如OSPF、IS-IS）以及现有设备中的路由功能，但对于故障路径的恢复有赖于路由协议的汇聚时间，并且**不具备流量工程能力。**
- **显式路由**：根据各种约束参数来计算路径，可以赋予不同LSP以不同的服务等级，可以为故障的LSP进行快速重路由，适于**实现流量工程与QoS业务**。

PATH SELECTION

- 传统标签分发协议，如LDP，RSVP不支持TE
  - 需要扩展：
    - LDP->CR-LDP (Constrained-based LDP)
    - RSVP->RSVP-TE
- 路径选择：需要考虑约束条件：如带宽，时延，…
  - SPF->CSPF (Constrain-based SPF)
- 现有路由协议需要扩展：
  - OSPF->OSPF-TE
  - ISIS->ISIS-TE

RSVP-TE

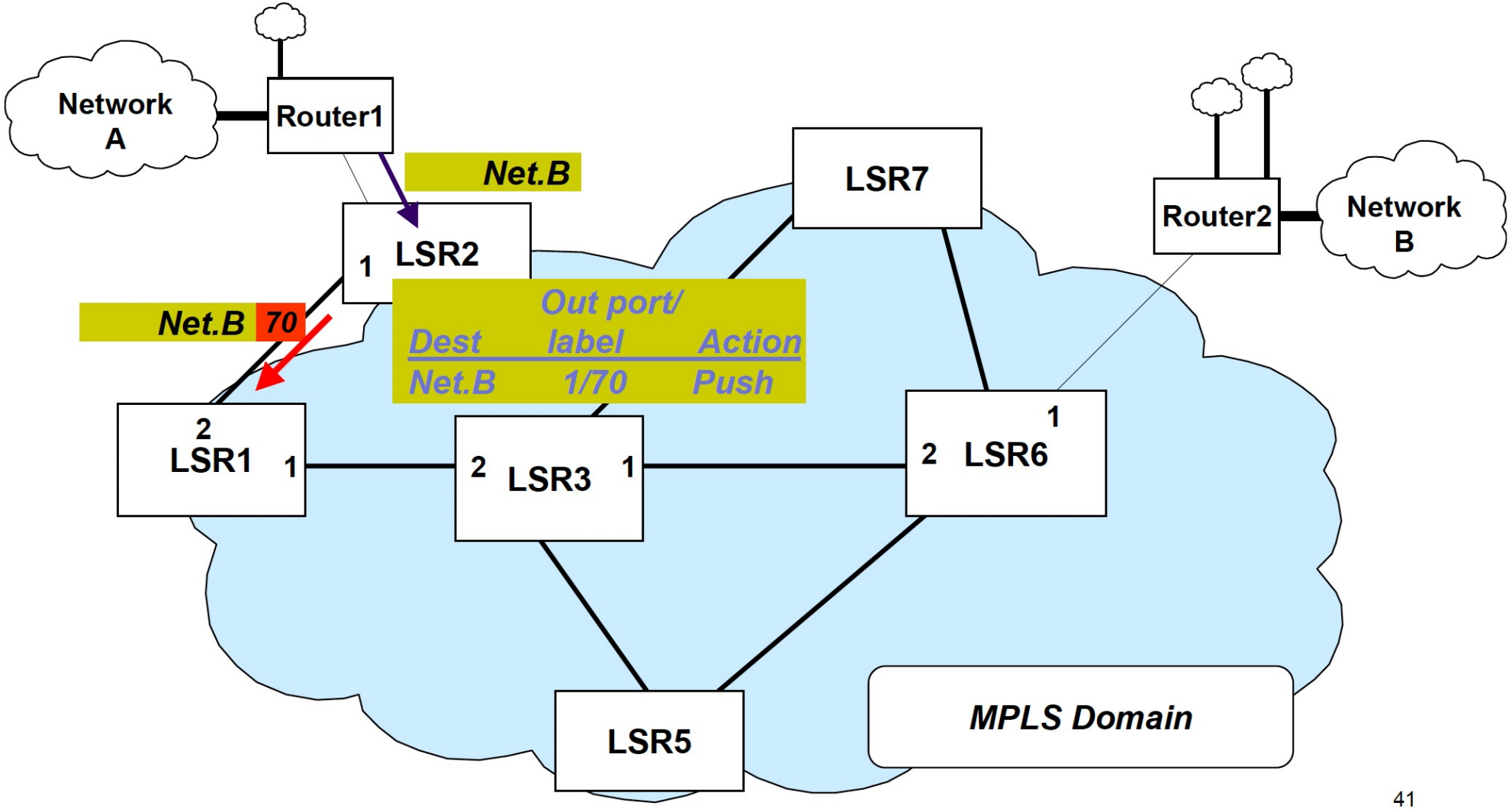


- Extensions to RSVP for *traffic-engineered LSPs*
  - Request-driven label distribution to create explicit route LSPs
  - Single node (usually ingress) determines route
  - Enables traffic engineering

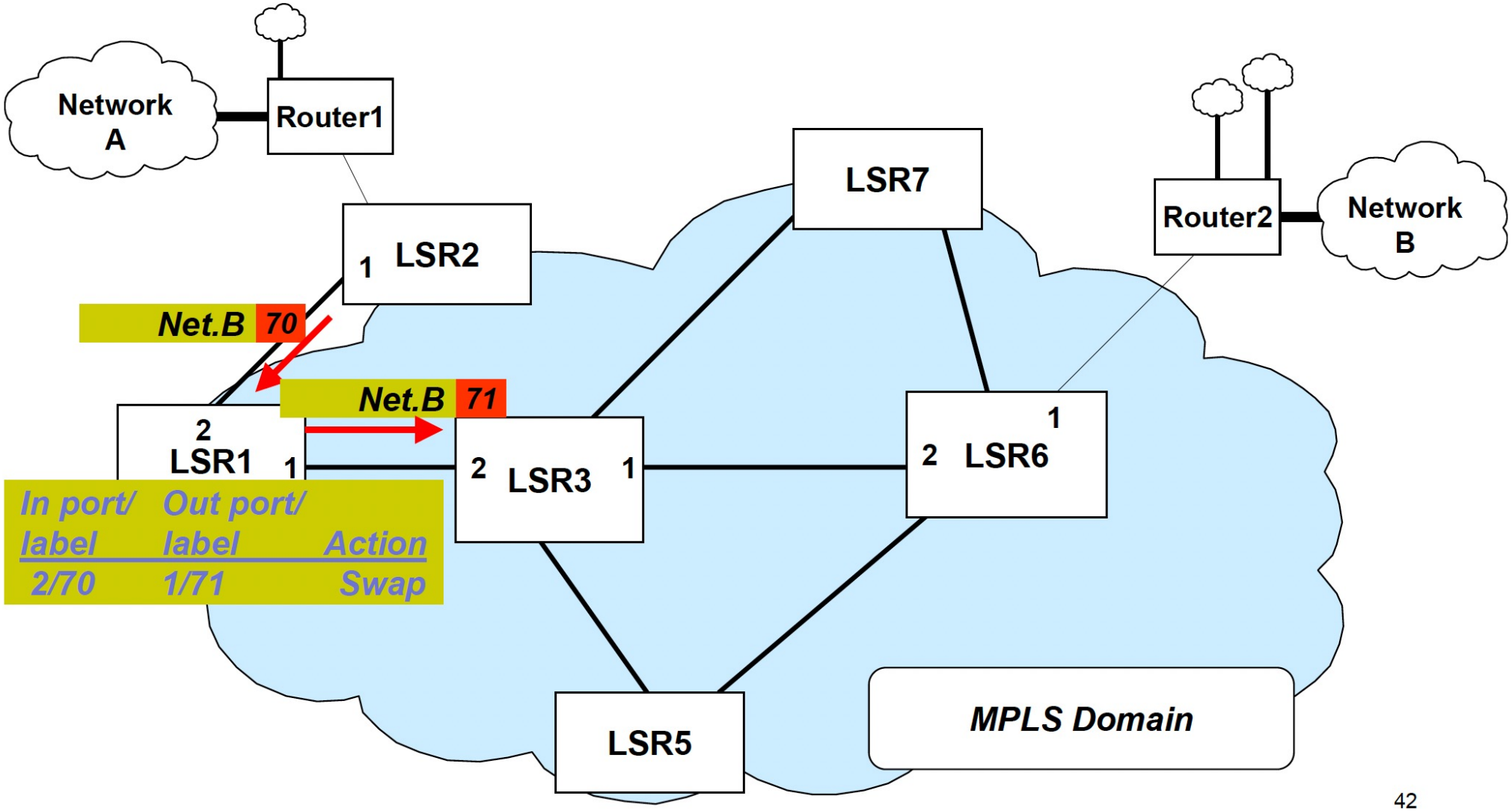
MPLS整个过程

- Data forwarding

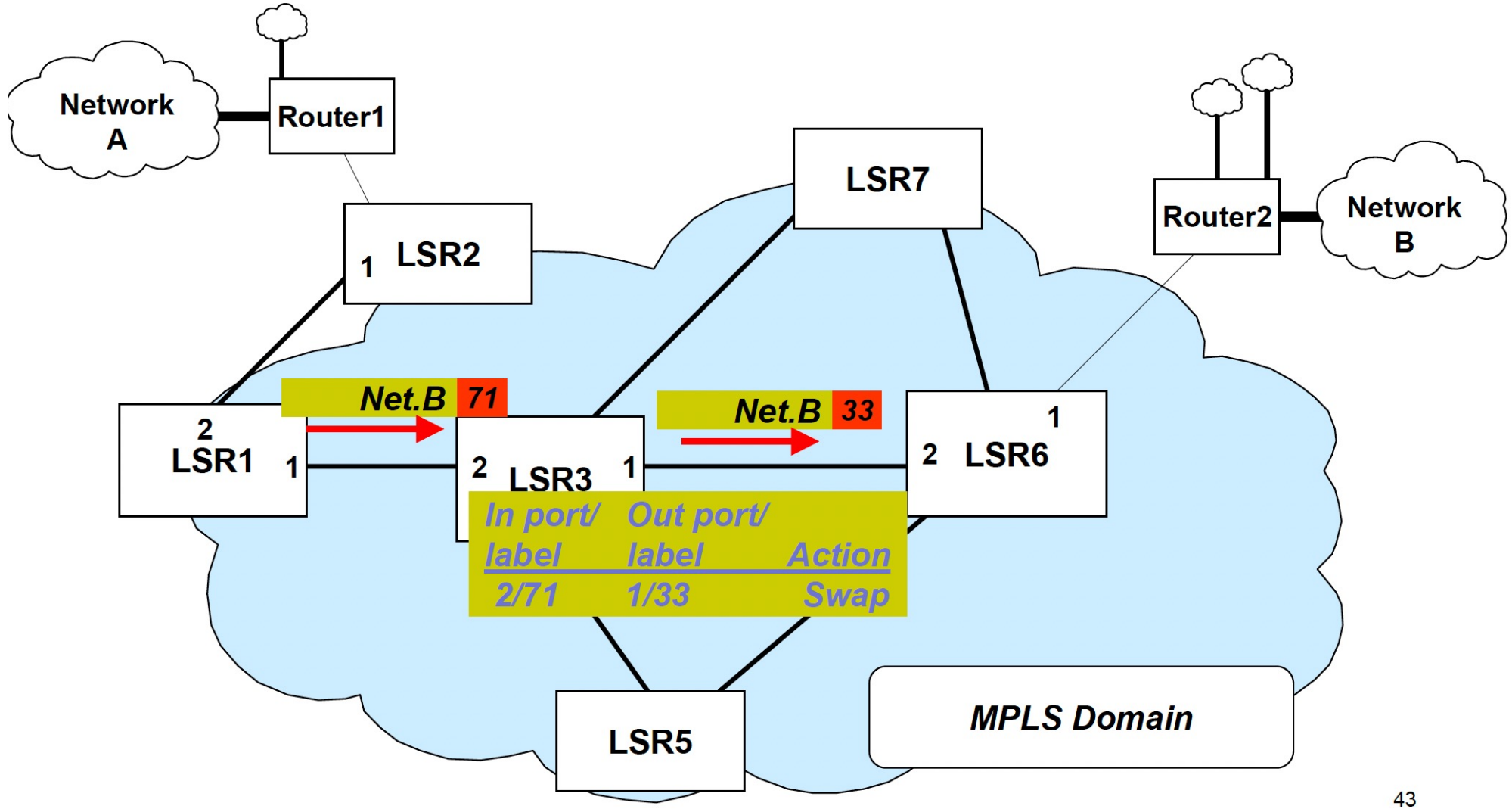
Data Forwarding – *Unlabelled packet to Ingress*



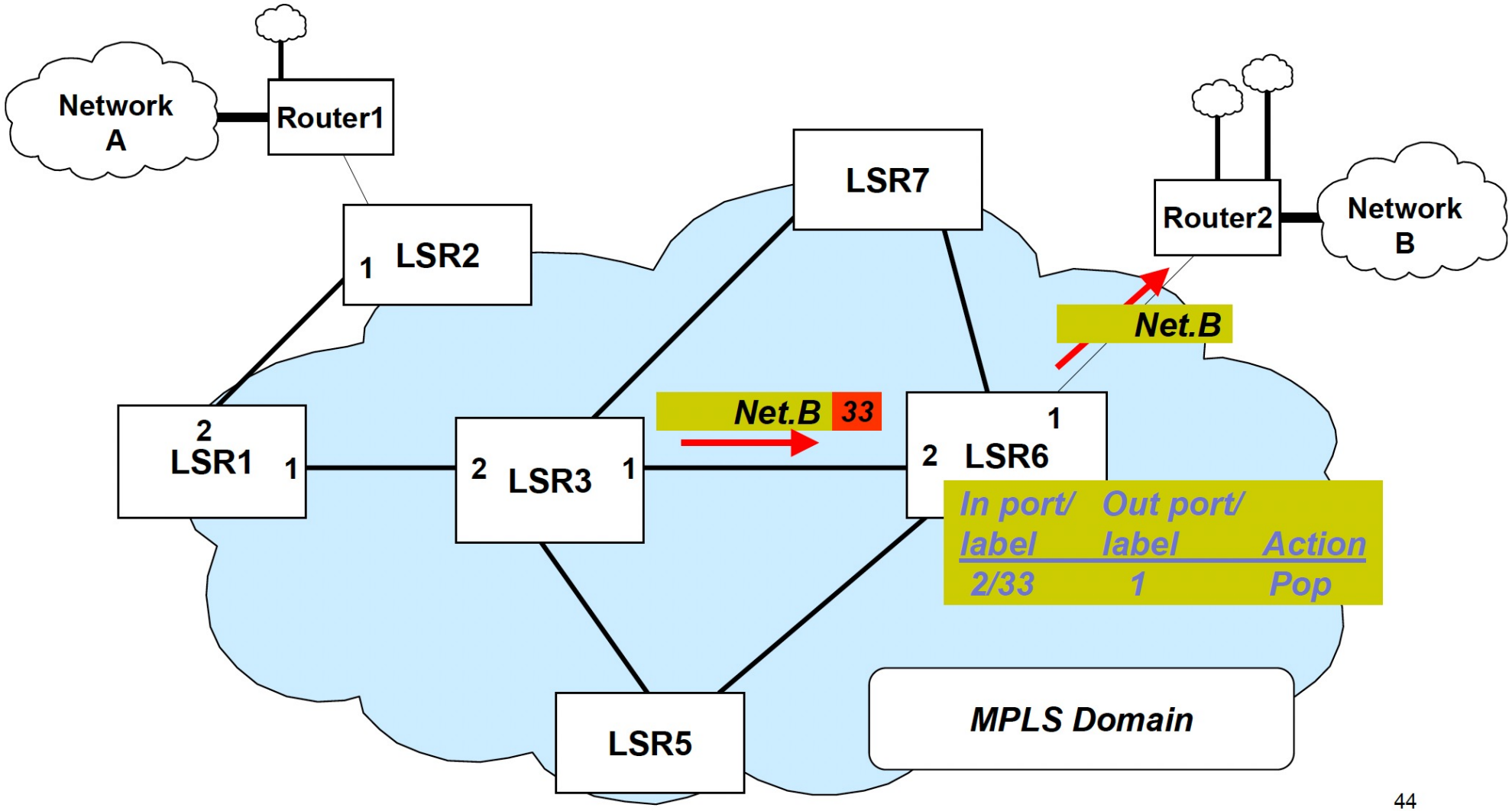
Data Forwarding – *LSR1 – LSR3*



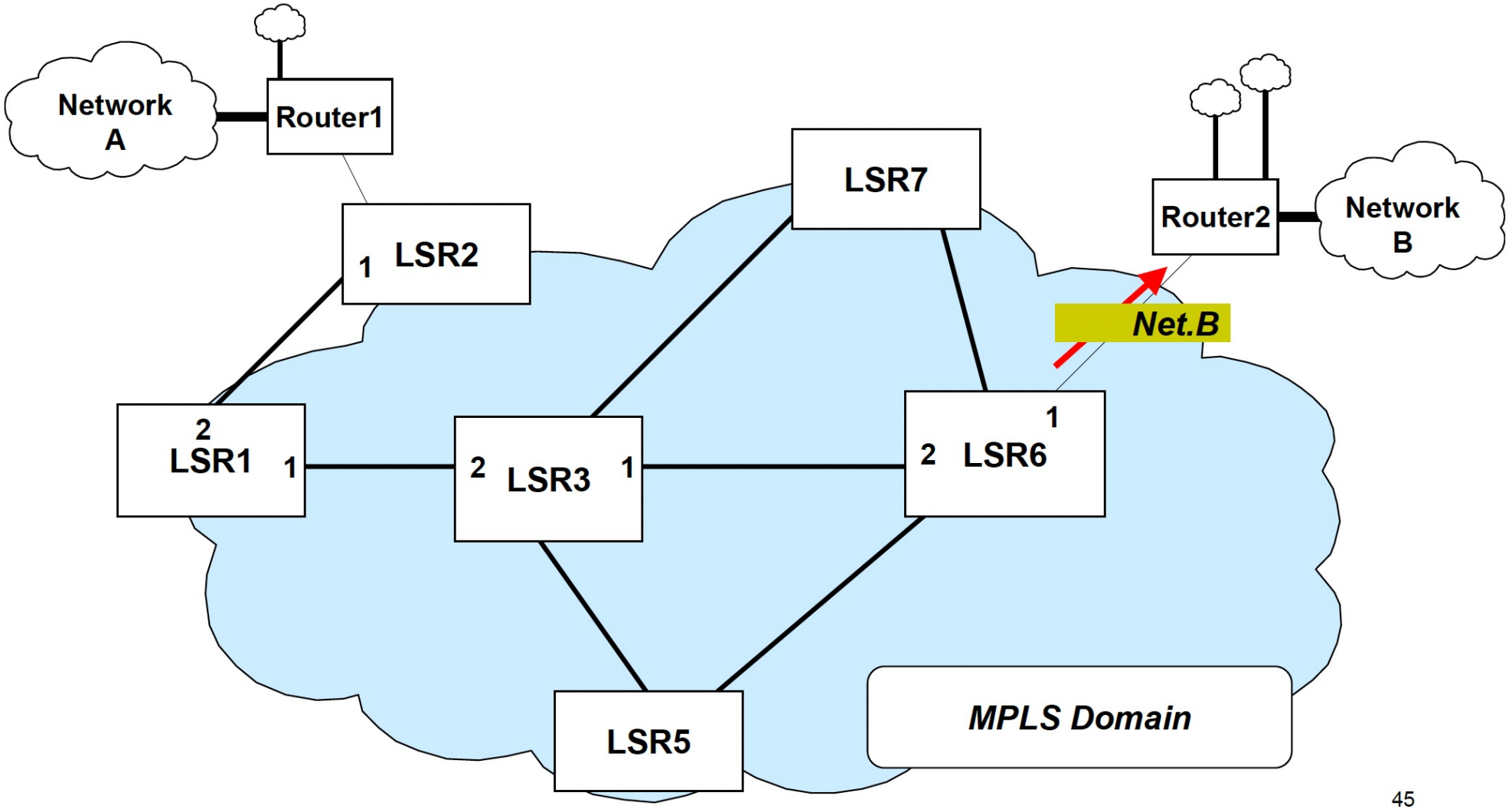
Data Forwarding – *LSR3 – LSR6*



Data Forwarding – *LSR6 – Egress Router*



Data Forwarding – *Unlabelled packet delivered*



效率问题

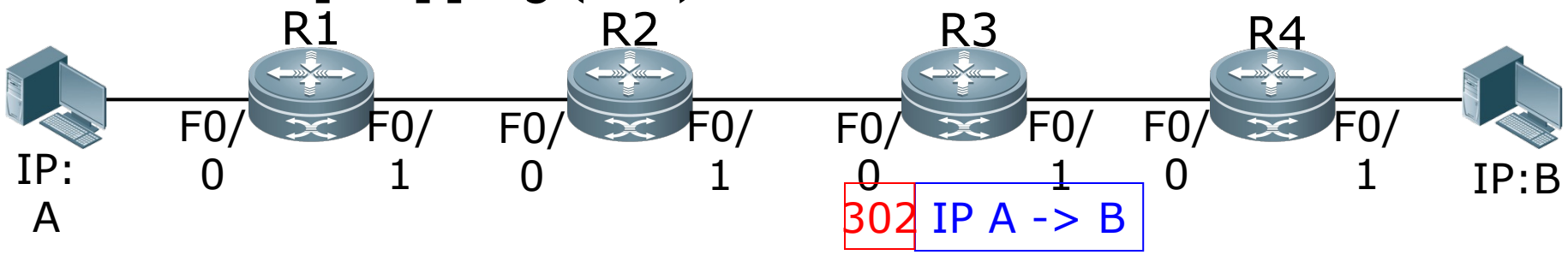
- 对于出口LER，先查找LFIB表，发现要将标签弹出，于是它将标签弹出，弹出后发现是个IP报文，于是又去查FIB表，最终将这个IP数据包转发出去。进行了两次查找。这降低了转发效率。

- 标签可以在（倒数第二跳）上弹出，出口LER只需查找FIB表将收到的IP报文进行转发

倒数第二跳弹出机制（PENULTIMATE HOP POPPING）

- R1、R2和R3如何将A->B的报文送到目的地?
  - ⇨POP动作：R3收到标签报文后，按照MPLS转发表中相关表项对报文中的标签进行弹出

  - ⇨Penultimate Hop Popping (PHP):倒数第二跳弹出

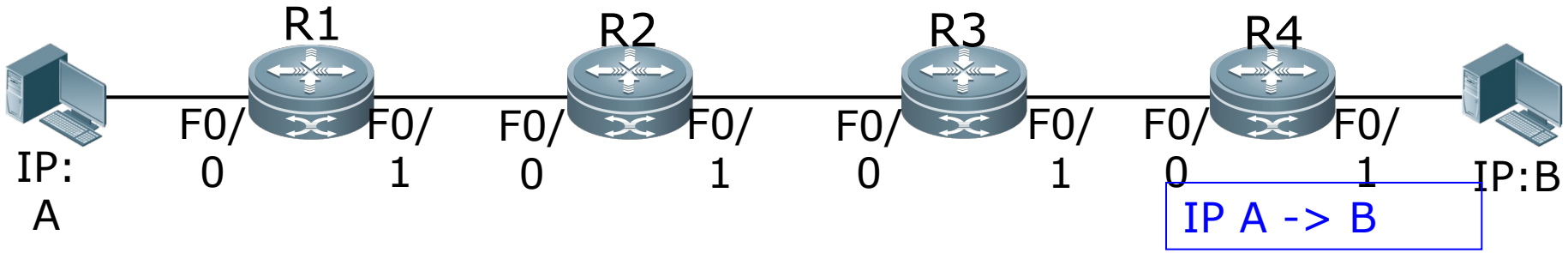


in标签	out标签	前缀	下一跳接口
301	201	A	F0/1
302	POP	B	F0/0

倒数第二跳弹出机制

- R1、R2和R3如何将A->B的报文送到目的地?
  - ⇨POP动作：R3收到标签报文后，按照MPLS转发表中相关表项对报文中的标签进行弹出

  - ⇨PHP:倒数第二跳弹出



R4查找路由表,将报文从相应接口转发出去

Thank You