IP QoS技术

www.huawei.com

HUAWEI

前 言

在传统的IP网络中，所有的报文都被无区别的等同对待，对报文传送的可靠性、传送延迟等性能不提供任何保证。

随着IP网络上新应用的不断出现，对IP网络的服务质量也提出了新的要求，例如VoIP等实时业务就对报文的传输延迟提出了较高要求，如果报文传送延时太长，用户将不能接受（相对而言，E-Mail和FTP业务对时间延迟并不敏感）。为了支持具有不同服务需求的语音、视频以及数据等业务，要求网络能够区分出不同的通信，进而为之提供相应的服务。传统IP网络的尽力服务不可能识别和区分出网络中的各种通信类别，而具备通信类别的区分能力正是为不同的通信提供不同服务的前提，所以说传统网络的尽力服务模式已不能满足应用的需要。

QoS技术的出现便致力于解决这个问题。

培训目标

学完本课程后，您应该能：

- 理解IP QoS的各种技术原理。
- 理解IP QoS技术在高端产品中的实现原理。
- 掌握IP QoS在高端产品中的配置。

目 录

HUAWEI

IP QoS的业务需求

传统的IP网络

主要承载数据业务，采用尽力传送（Best Effort）的方式，服务质量显得无关紧要。

当前的IP网络

近年来，随着以IP技术为核心的Internet的飞速发展，以及各种新业务的出现（VOIP、VPN、ERP等），IP网络已由一个单纯的数据网络转变为具有商业价值的承载网，因此IP网络必须为其所承载的每一类业务提供相应的服务质量。

IP QoS的概念

QoS：Quality of Service，即服务质量

IP QoS 是指IP网络的一种能力，即在跨越多种底层网络技术（MP、FR、ATM、Ethernet、SDH、MPLS等）的IP网络上，为特定的业务提供其所需要的服务。服务质量包括：

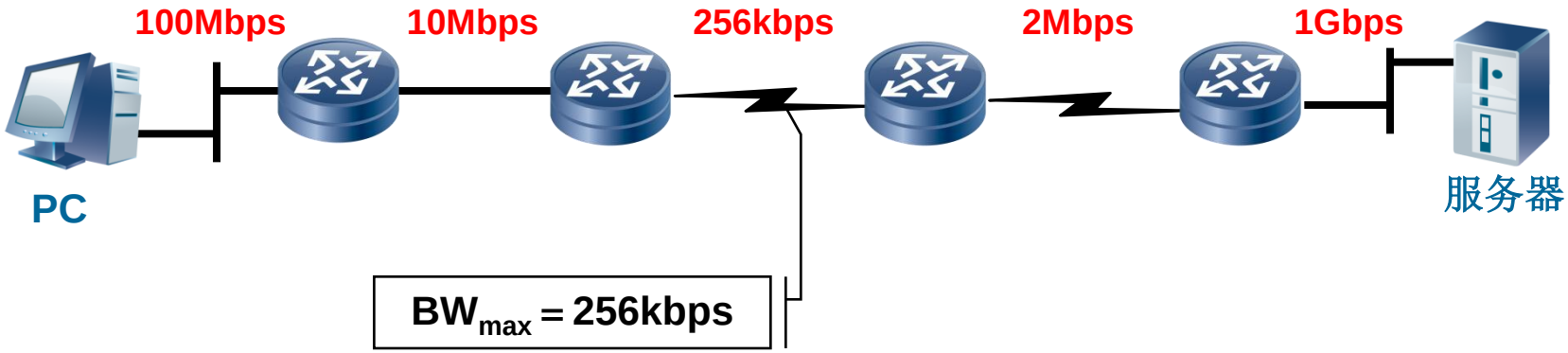
- 传输的带宽
- 传输的时延和抖动
- 数据的丢包率

网络中存在资源竞争，就存在对服务质量的要求

提高某类业务的服务质量同时也会损害其它业务的服务质量

带宽

BWmax = Min(100M,10M,256k,2M,1G) = 256kbps

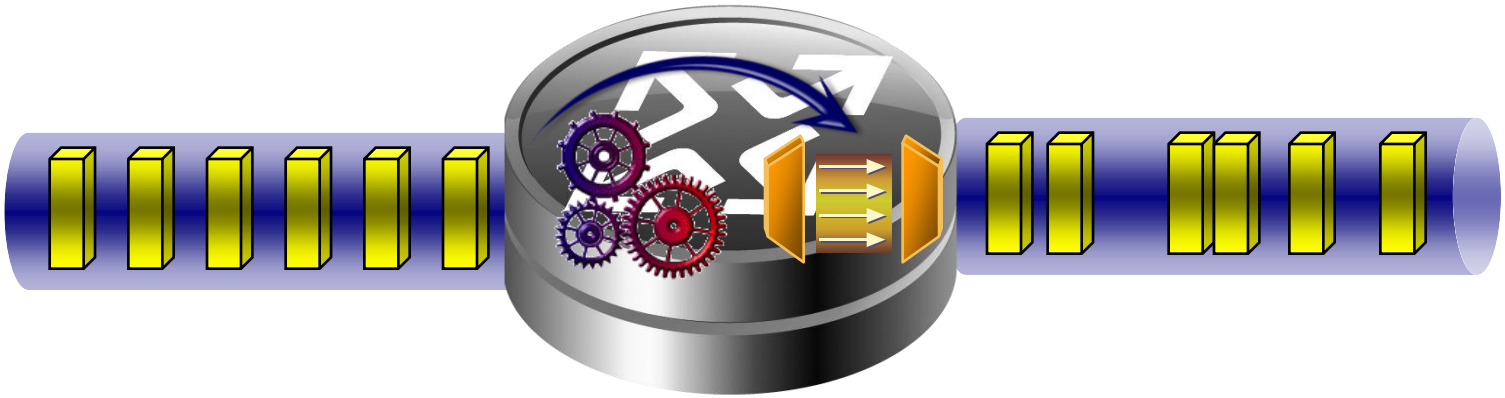


木桶理论：最大带宽BW_{max}等于数据传输路径上的最小带宽。

端到端时延

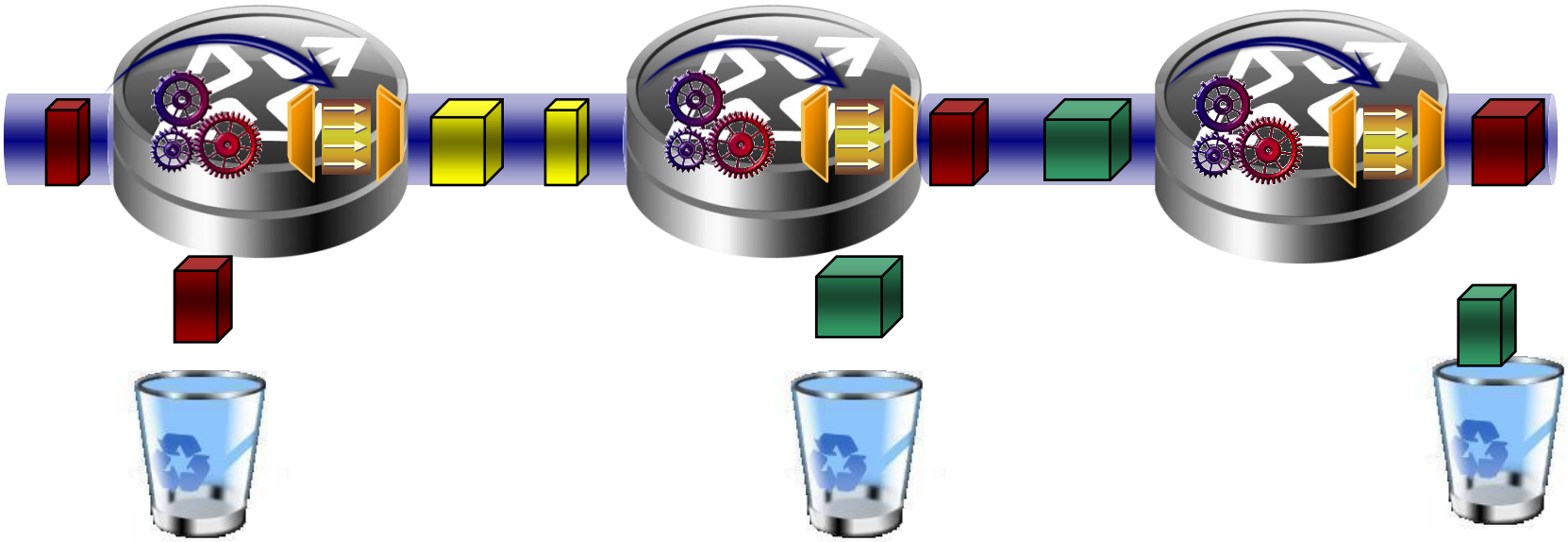
端到端时延等于路径上所有传输时延、处理时延与队列时延之和。

抖动



抖动是因为每个包的端到端时延不相等造成的。

丢包



丢包可能在传输过程的每一个环节发生。

QoS服务模型

QoS 根据网络质量和用户需求，通过不同的服务模型为用户提供服务。通常QoS 提供以下三种服务模型：

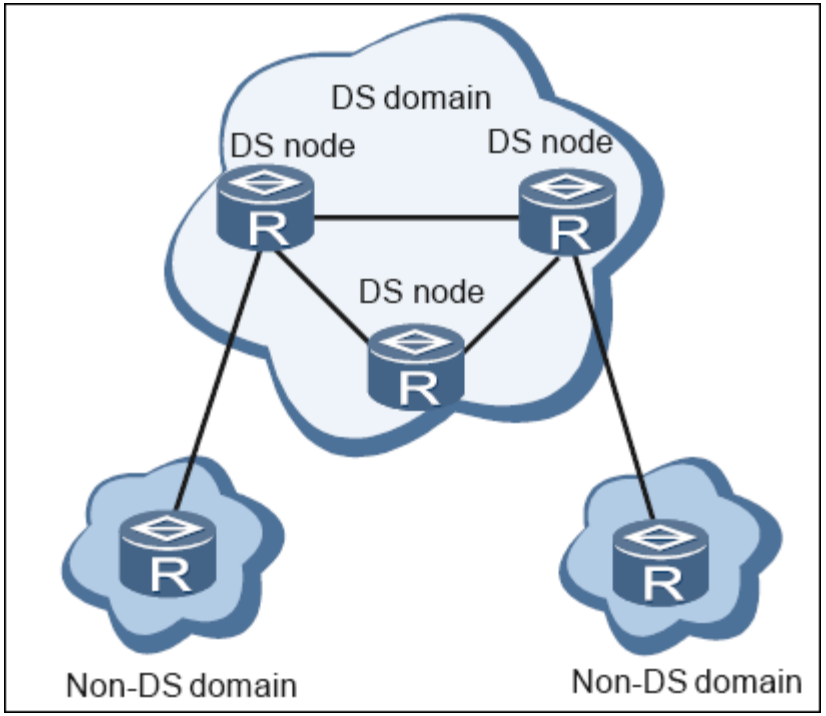
- Best-Effort Service模型：是最简单的服务模型。应用程序可以在任何时候，发出任意数量的报文，而且不需要事先获得批准，也不需要通知网络。

- Integrated Service模型：是一个综合服务模型，它可以满足多种QoS 需求。这种服务模型在发送报文前，需要向网络申请特定的服务。

- Differentiated Service模型：是通过设置报文头部的QoS 参数信息，来告知网络节点它的QoS 需求。报文传播路径上的各个路由器都可以通过对 报文头的分析来获知报文的服务需求类别。

IP 网络中的Diff-Serv 模型

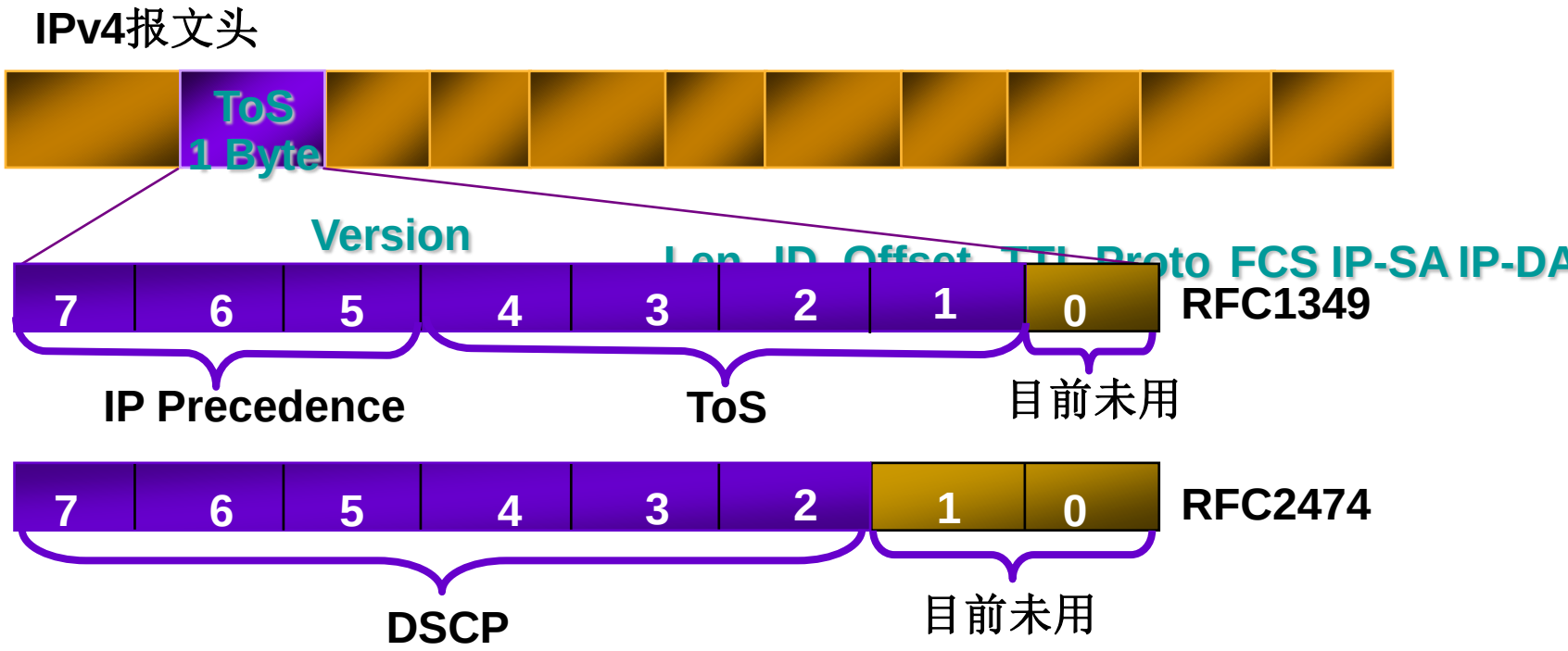
一般来讲，在提供IP网络的QoS时，为了适应不同规模的网络，在IP骨干网往往需要采用DiffServ体系结构。Diff-Serv 网络结构示意图：



DiffServ模型的几个重要概念

DSCP（DiffServ Code Point）

IP 优先级



HUAWEI

DiffServ模型的几个重要概念（续）

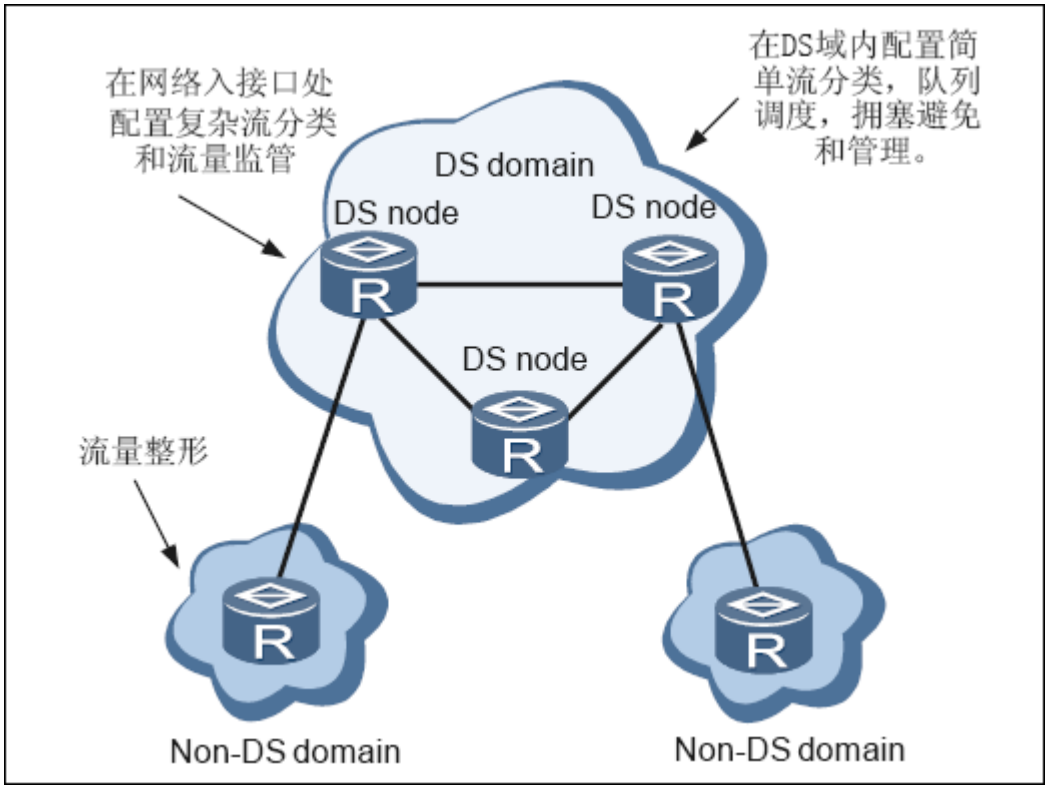
PHB（Per-Hop Behaviors），PHB是DS节点作用于数据流的行为。网络管理员可以配置DSCP到PHB的映射关系。如果DS节点接收到一个报文，检查其DSCP，发现未定义到PHB的映射，则DS节点将选择采用缺省PHB（即Best-Effort，DSCP=000000）进行转发处理。每个DS节点必须支持该缺省PHB。

PHB的分类，IETF DiffServ工作组目前定义了四种PHB：

- Default PHB
- Class-Selector PHB
- Expedited Forwarding PHB
- Assured Forwarding PHB

QoS 实现的相关技术

在Diff-Serv 模型中，常见QoS 特性应用：



问 题

什么是QoS?

QoS包括哪些方面?

常见的QoS服务模型有哪些?

IPV4报文中，DSCP、TOS和IP Precedence的关系?

总 结

完成本节的学习后，您应该掌握以下几点：

QoS的基本概念；

QoS的常见服务模型；

DiffServ服务模型相关概念。

目 录

流量分类和标记

流量分类及标记是部署QoS 的基础

可以根据ACL、以及报文自身信息对流量进行分类

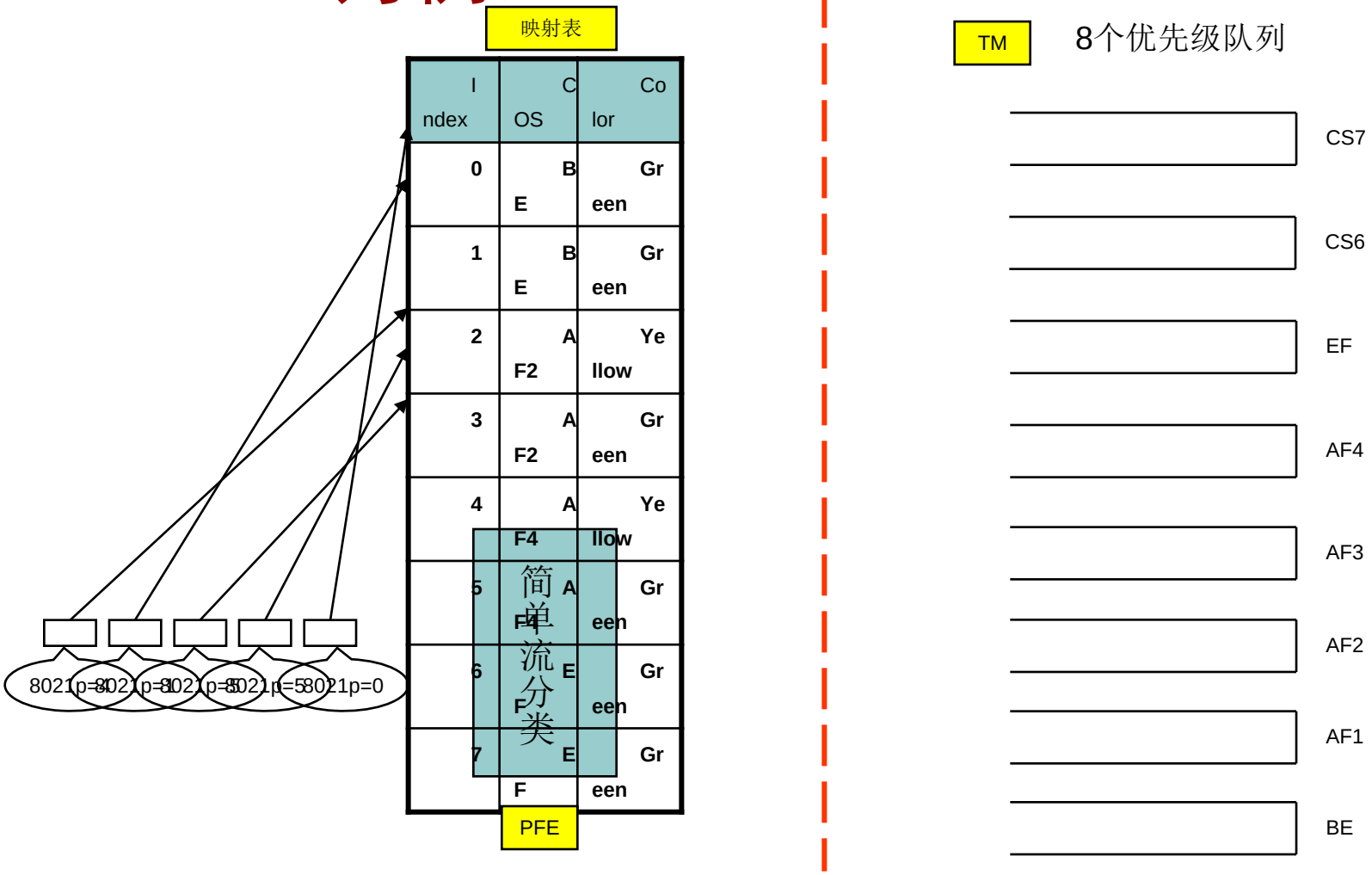
可以基于DSCP、IP Precedence、802.1P、MPLS EXP等信息对报文进行标记

流量分类

流量分类是按照一定的规则识别符合某类特征的报文，特征不同的报文享受到的服务不同。按照分类规则参考信息的不同，流量分类可以分为简单流分类和复杂流分类。

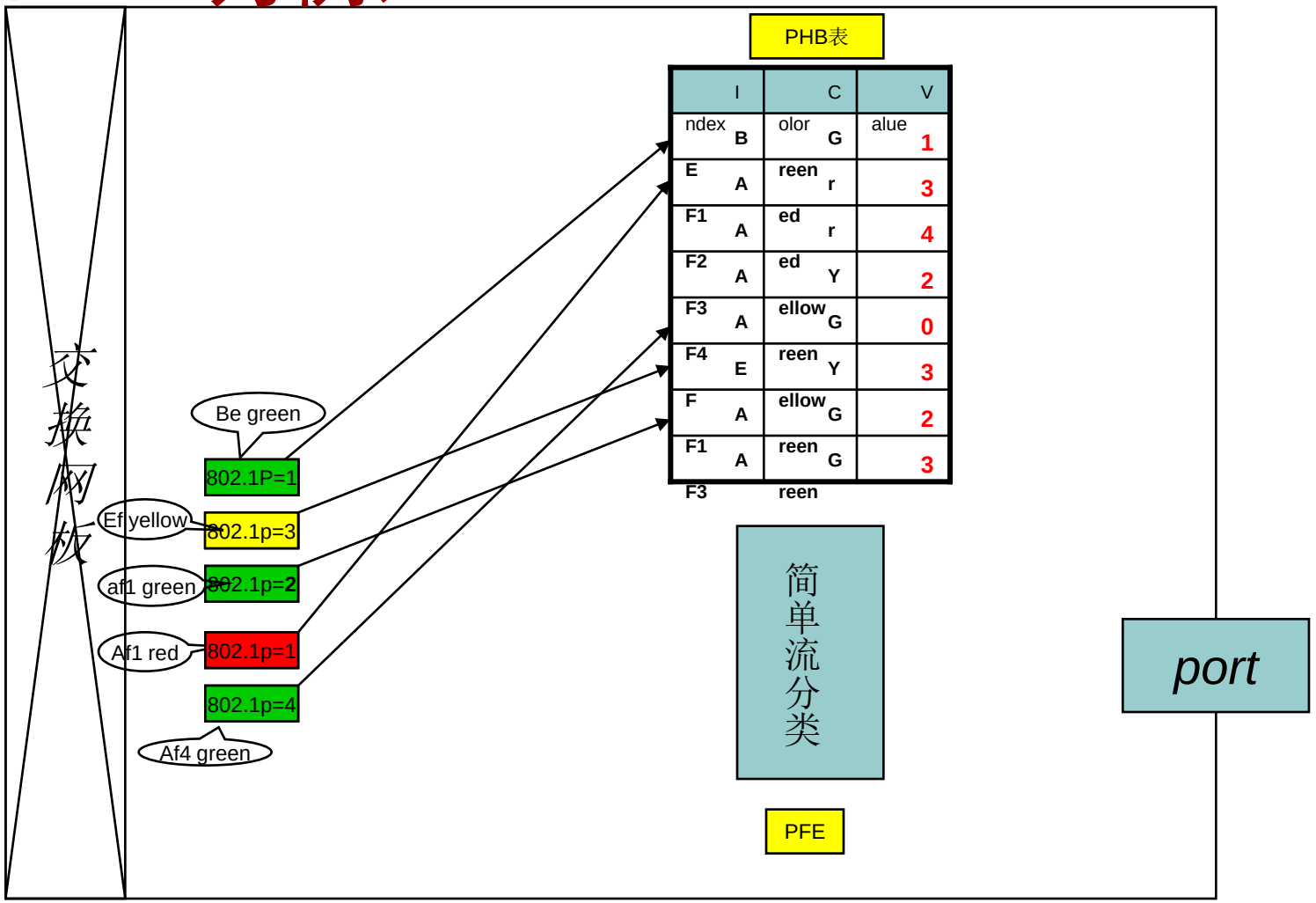
- 简单流分类是指采用简单的规则，如IP 报文头中的DSCP/IP-PRE值，MPLS报文的EXP域值，Vlan报文头中的802.1P 值对报文进行粗略的分类，以识别出具有不同优先级或服务等级特征的流量。
- 复杂流分类是指采用复杂的规则，如综合链路层、网络层、传输层信息（例如源MAC 地址、目的MAC 地址、源IP 地址、目的IP 地址、用户组号、协议类型或应用程序的TCP/UDP 端口号等）对报文进行精细的分类。通常在Diff-Serv 域的边界路由器上对流量进行复杂流分类。

简单流分类和标记（流量入方向以802.1P为例）



HUAWEI

简单流分类和标记（流量出方向以802.1P为例）



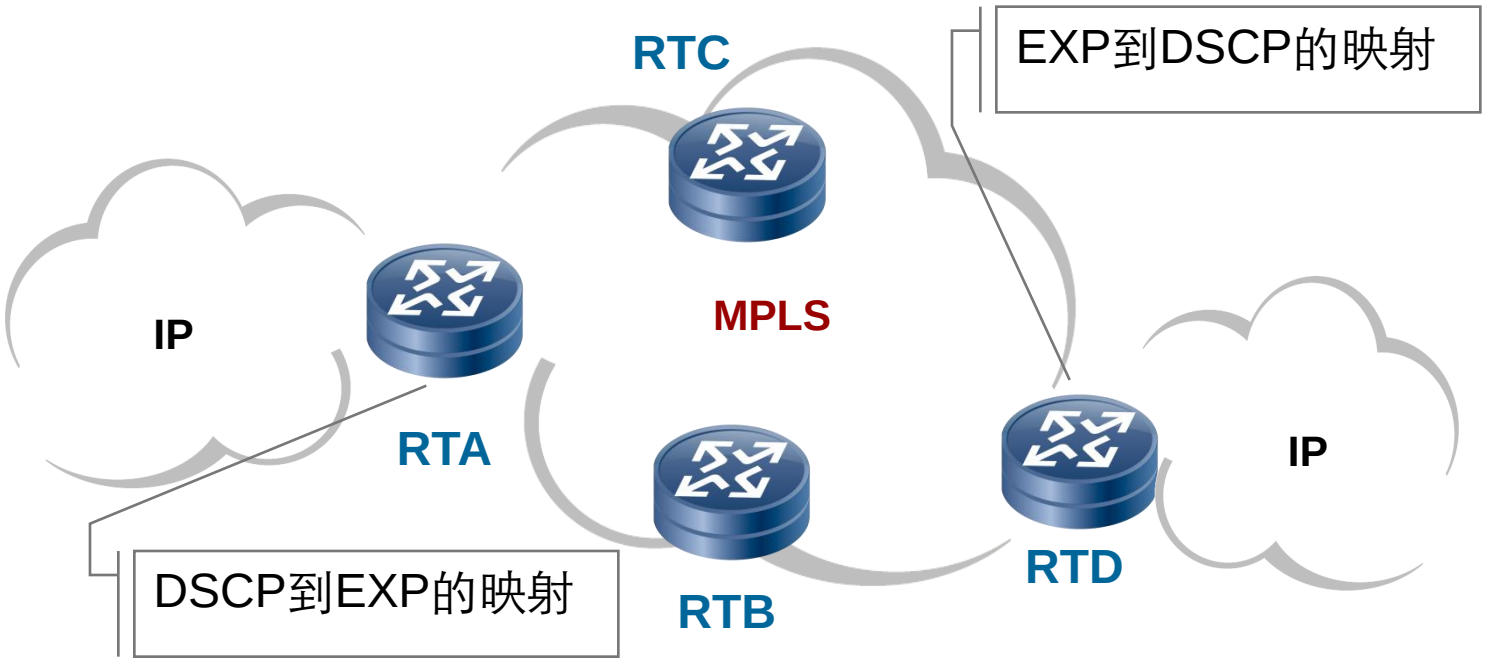
简单流分类与标记在产品中的实现

华为路由器产品支持配置8个DS域（定义见注释）。

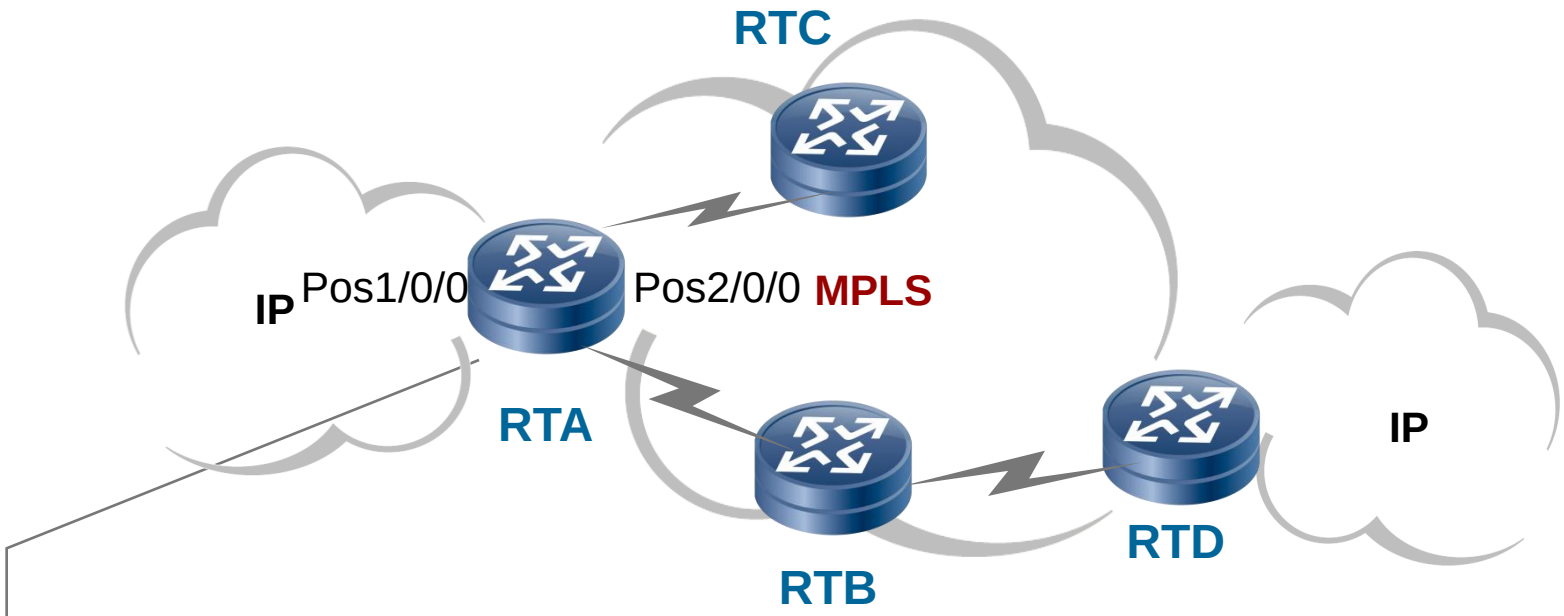
上行简单流分类，根据IP DSCP、MPLS EXP或802.1P将报文分为八种业务类型（CS7、CS6、EF、AF4—AF1、BE）、三种颜色（green、yellow、red），从而区分不同的业务（如，语音、视频、数据等）。在拥塞管理、队列调度时，不同业务进入不同的队列，得到差异化的调度。例如语音可以进入高优先级的PQ队列，保证低延时。上行若不做简单流分类，报文业务类型都为BE。

下行简单流分类，根据内部业务类型（CS7、CS6、EF、AF4—AF1、BE）、三种颜色（green、yellow、red），重新设置报文的IP DSCP、MPLS EXP或802.1P，实现了重标记的功能，重新标记IP DSCP、MPLS EXP或802.1P。下行未配置简单流分类时，IP DSCP、MPLS EXP或802.1P不做改变。

简单流分类应用场景举例

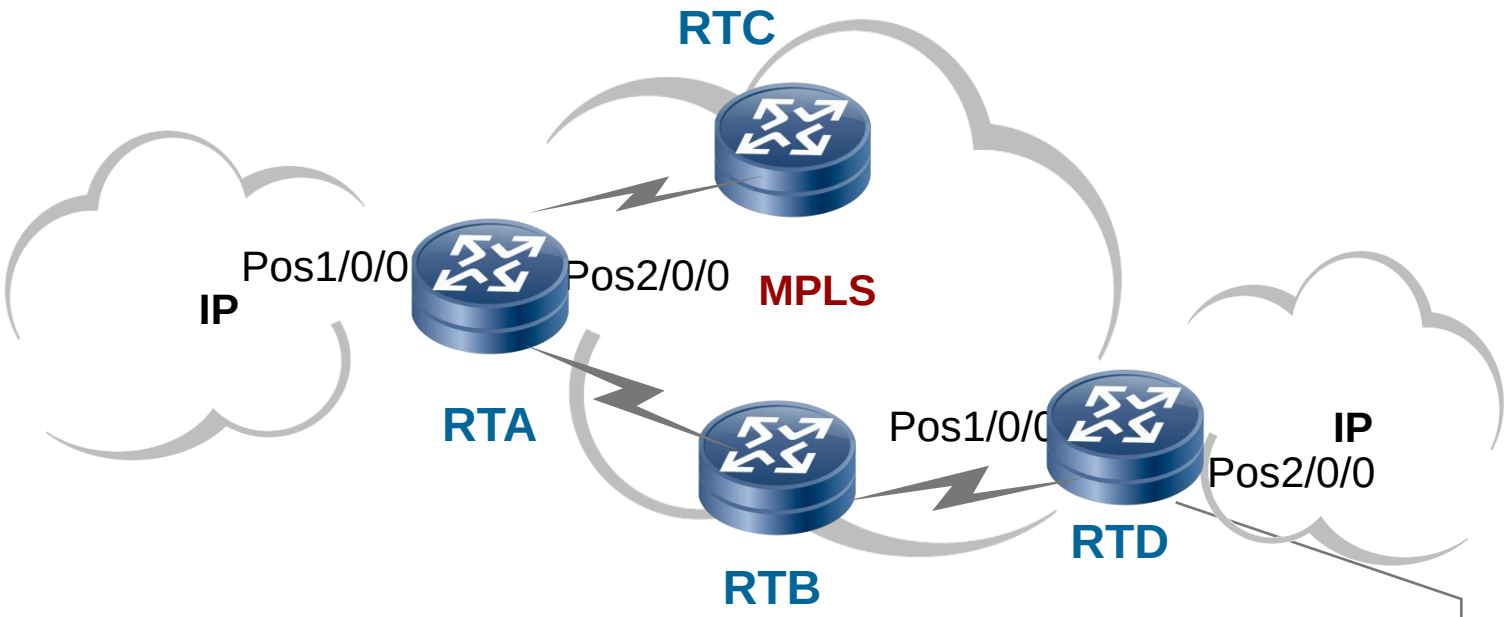


简单流分类的配置



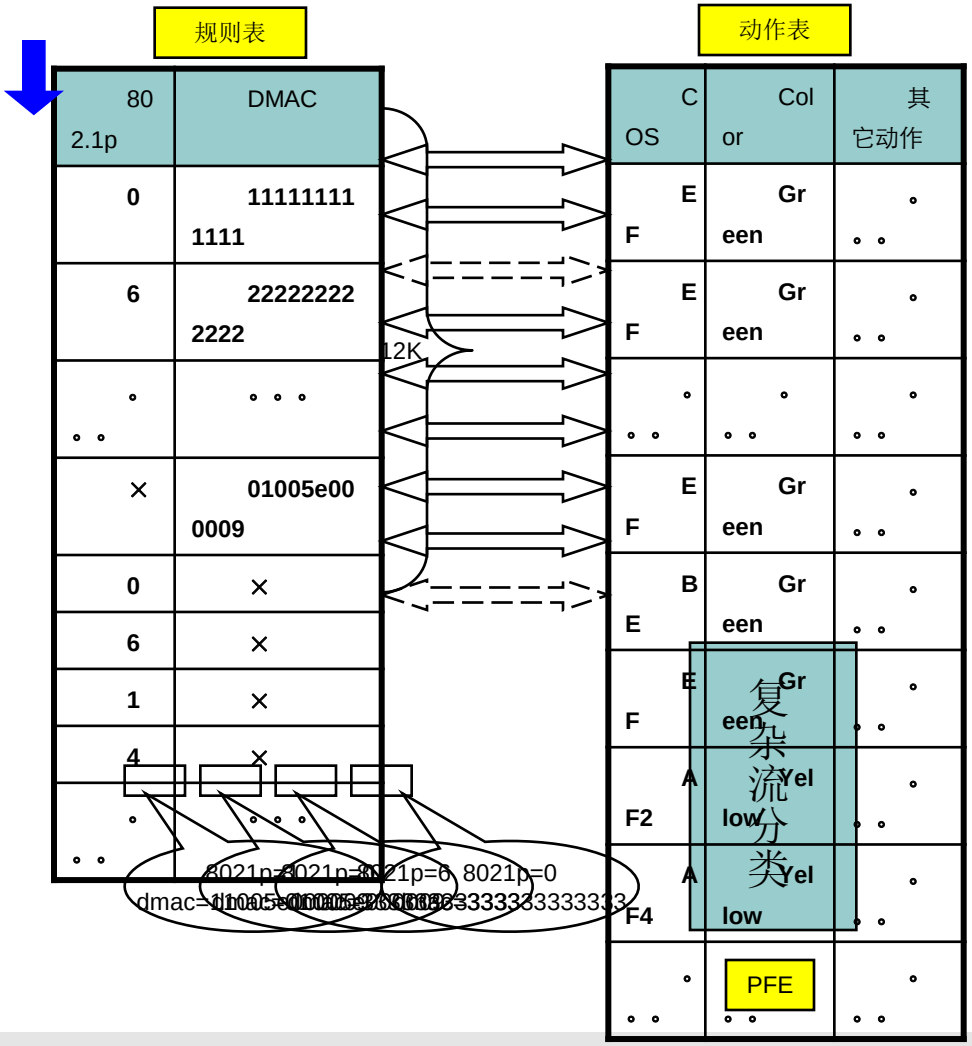

```
[RTA]diffserv domain A
[RTA-dsdomain-A]ip-dscp-inbound 18 phb af4 green
[RTA-dsdomain-A]mpls-exp-outbound af4 green map 5
[RTA]interface pos 1/0/0
[RTA-Pos1/0/0]trust upstream A
[RTA]interface pos 2/0/0
[RTA-Pos2/0/0]trust upstream A
```

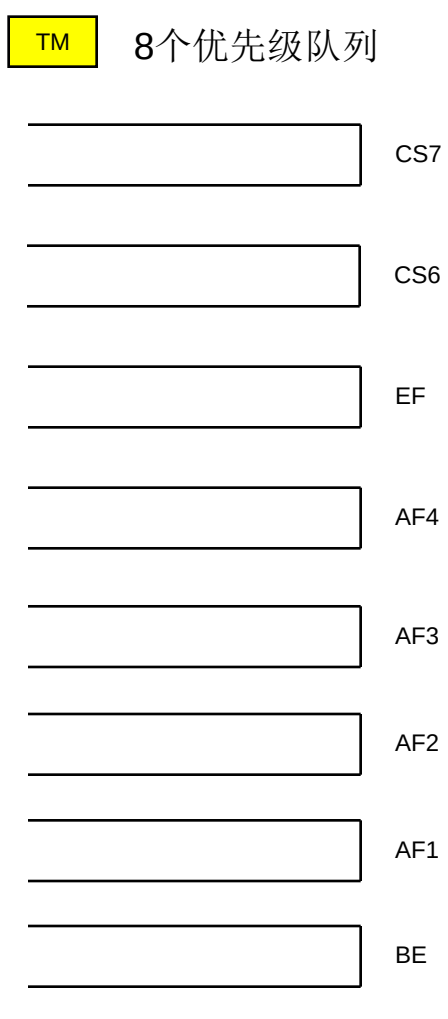
简单流分类的配置（续）




```
[RTD]diffserv domain B
[RTD-dsdomain-B]mpls-exp-inbound 5 phb af4 green
[RTD-dsdomain-B]ip-dscp-outbound af4 green map 18
[RTD]interface pos 1/0/0
[RTD-Pos1/0/0]trust upstream B
[RTDinterface pos 2/0/0
[RTD-Pos2/0/0]trust upstream B
```

复杂流分类





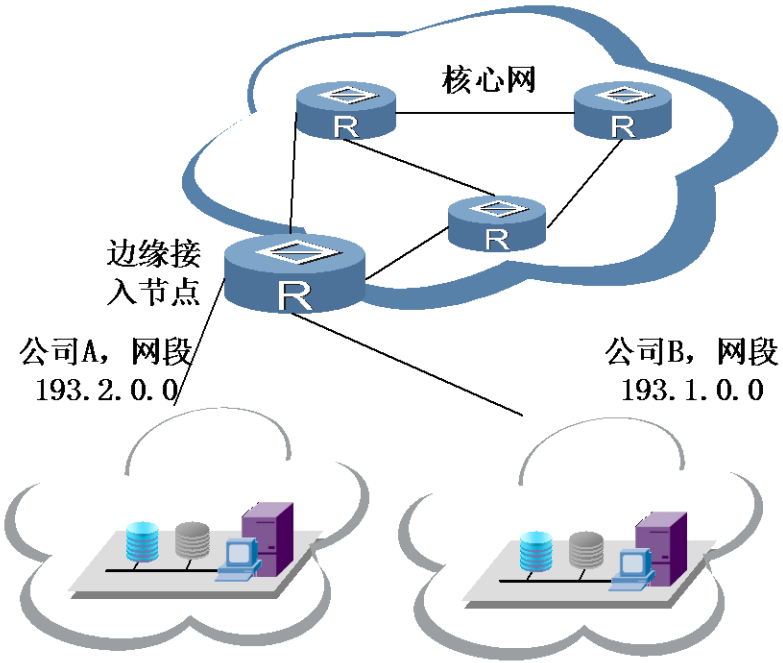
复杂流分类在产品中的实现

在实现复杂流分类时分为两个部分：规则部分和动作部分。

复杂流分类是报文匹配规则与符合规则后报文应采用的动作的结合。当处理报文时，根据报文中用来分类的字段信息组成关键字，查找规则表；如果报文能匹配上规则部分，则根据查找结果确定该规则对应的动作表，确定该报文应该执行何种动作。如果报文没有匹配上任何一条规则，则报文不做分类按普通报文正常转发。

ACL（Access Control List）：访问控制列表。用于复杂流分类的规则部分。

复杂流分类应用场景举例



如上图所示，假设A公司购买的带宽为200M，B公司购买的带宽为400M。为了实现带宽保证，可以在边缘接入节点上配置复杂流分类，根据IP地址区分A，B公司，然后执行不同的流量监管。

问 题

什么是流量分类?

流量分类包括哪些分类方法?

简单流分类是依据报文的什么信息来分类?

复杂流分类是依据报文的什么信息来分类?

总 结

完成本节的学习后，您应该掌握以下几点：

分类与标记的基本概念；

复杂流分类、简单流分类；

两种分类技术的实现与应用。

目 录

HUAWEI

流量监管介绍

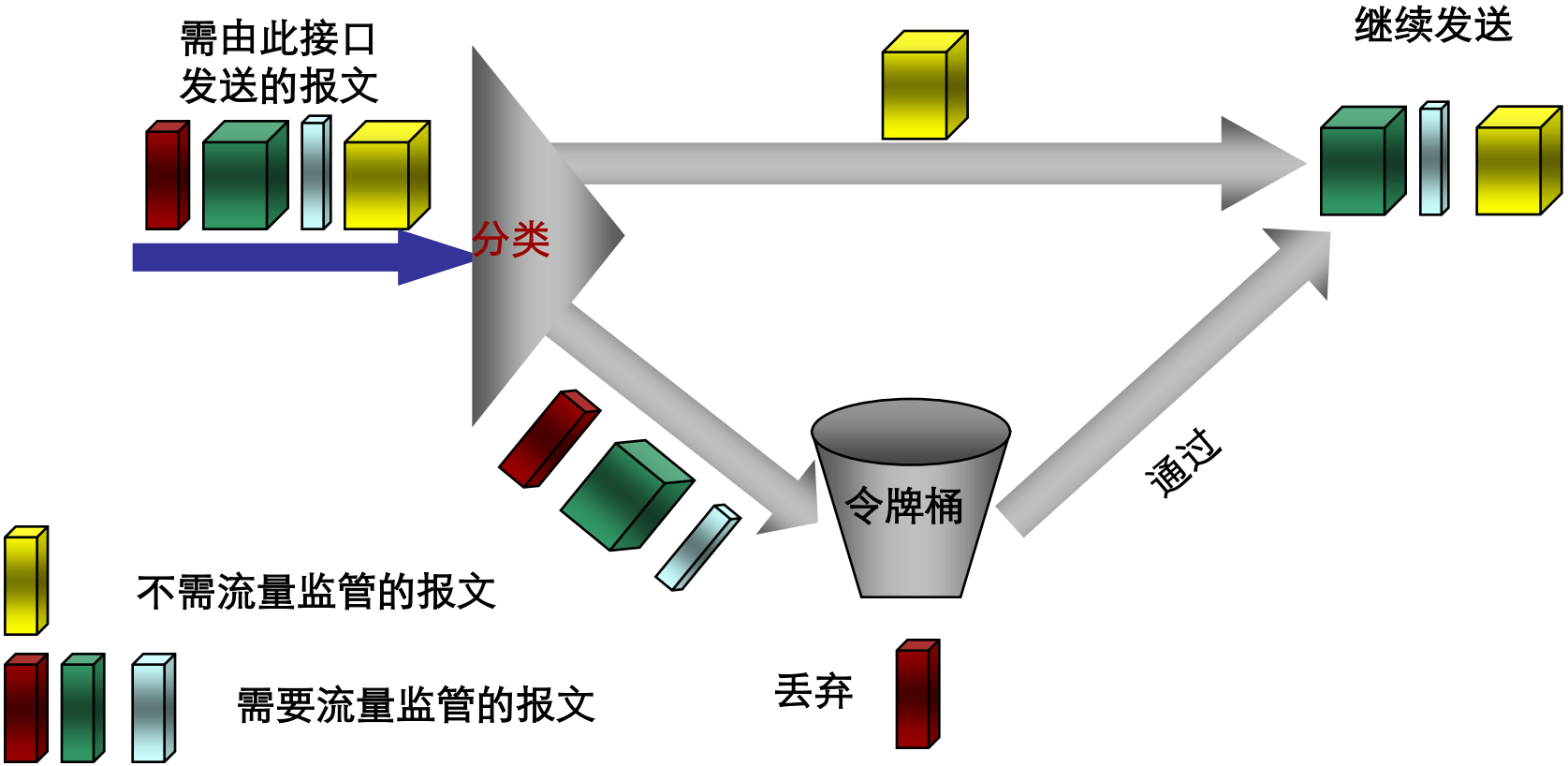
流量监管（Traffic-policing）是一种在入接口或出接口应用的对进入路由器的某流量进行限制的流量管理技术。

对于 ISP 来说，对用户送入网络中的流量进行控制是十分必要的。对于企业网，对某些应用的流量进行控制也是一个有力的控制网络状况的工具。

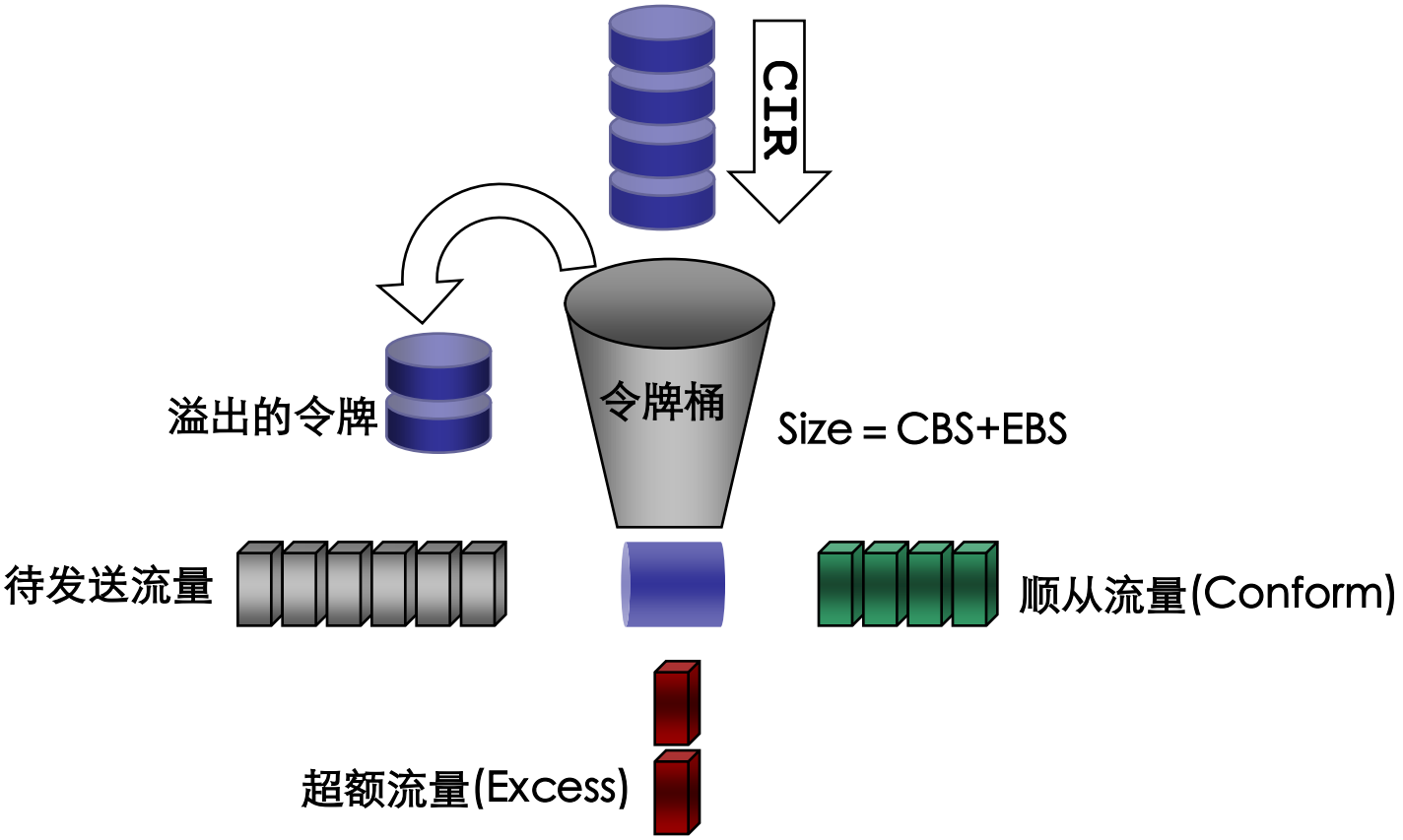
流量监管的典型应用是监督进入网络的某一流量的规格，把它限制在一个合理的范围之内，或者对超出的部分流量进行“惩罚”，以保护网络资源和运营商的利益。在报文满足一定的条件时，如某个连接的报文流量过大，流量监管就可以对该报文采取不同的处理动作，例如丢弃报文。

流量监管

网络管理者可以使用约定访问速度Car（Committed Access Rate）来对流量进行控制 。



Token Bucket：令牌桶



流量监管的具体实现

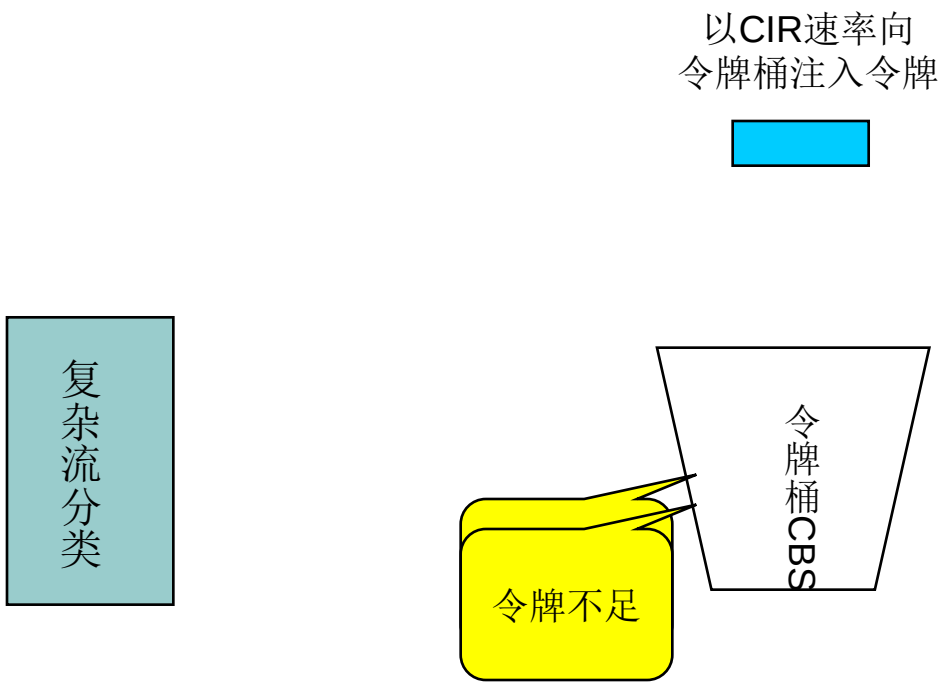
单桶单速率流量监管： 一个令牌桶，容量是CBS ，一个填充令牌的速率CIR。当有 B 字节的报文传过来的时候，根据桶的当前容量来对这个报文进行处理。

双桶单速率流量监控：两个令牌桶，一个的容量是CBS ，一个的容量是EBS ，一个填充令牌的速率CIR,两个令牌桶使用同一个填充速率。当有 B 字节的报文传过来的时候，根据两个桶的当前容量来对这个报文进行处理。

双桶双速率流量监控：两个令牌桶，一个的容量是CBS ，一个的容量是PBS 。这两个令牌桶分别使用两个填充令牌的速率，一个填充速率是CIR，一个填充速率是 PIR 。当有 B 字节的报文传过来得时候，根据两个桶的当前容量来对这个报文进行处理。

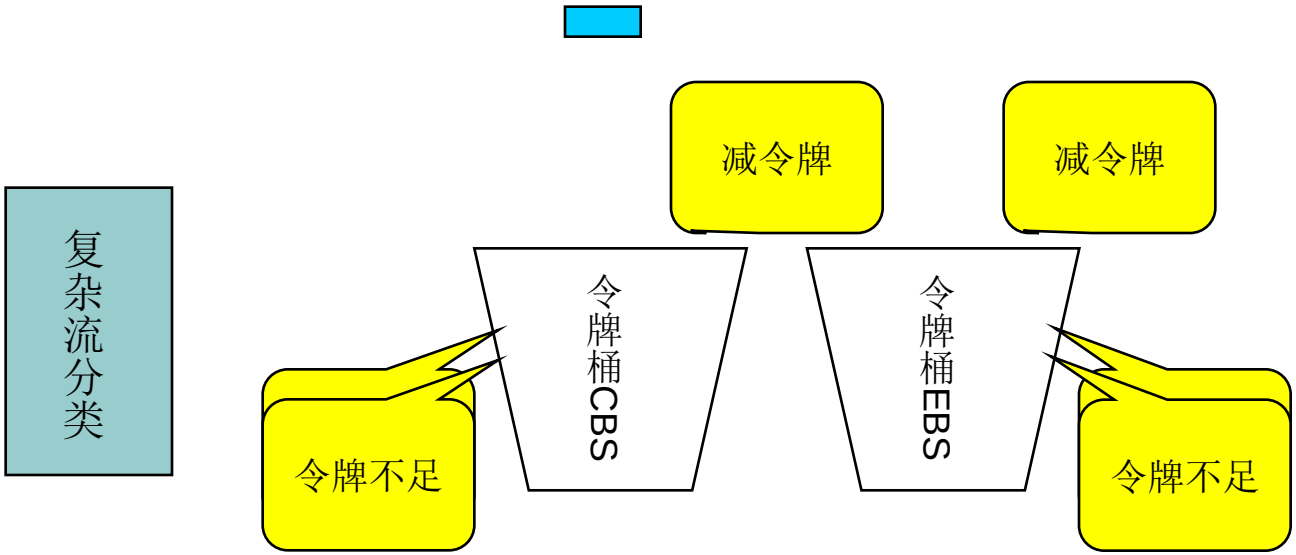
单桶单速流量监管

例：标记为green的报文通过，标记为red的报文丢弃。



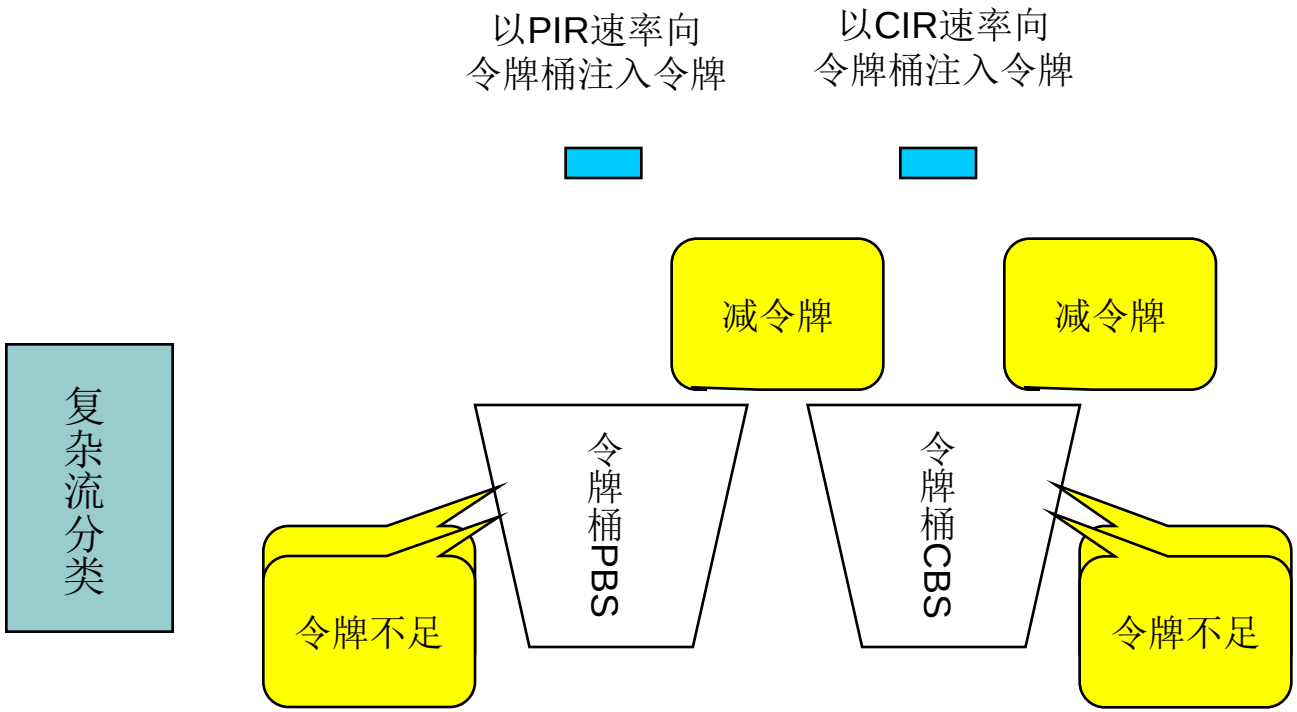
双桶单速流量监管

例：标记为green的报文通过，yellow报文通过，red的报文丢弃。
以CIR速率向令牌桶注入令牌，CBS令牌桶满后，再向EBS令牌桶填充

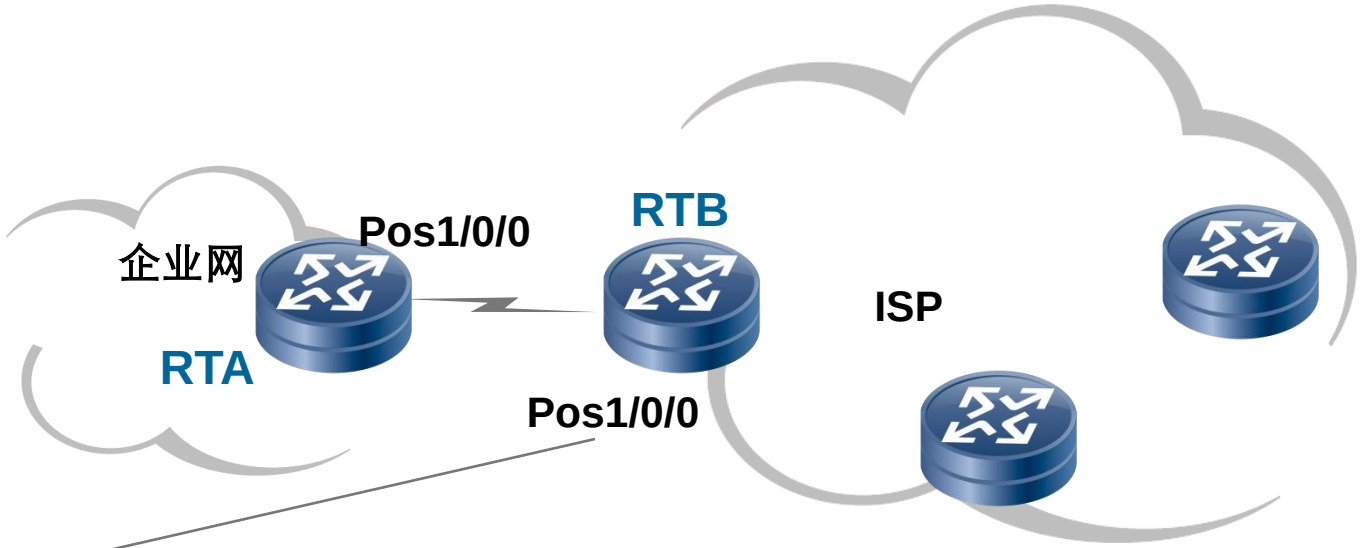


双桶双速流量监管

例：标记为green的报文通过， yellow报文通过，red的报文丢弃。
标记为red的报文丢弃。

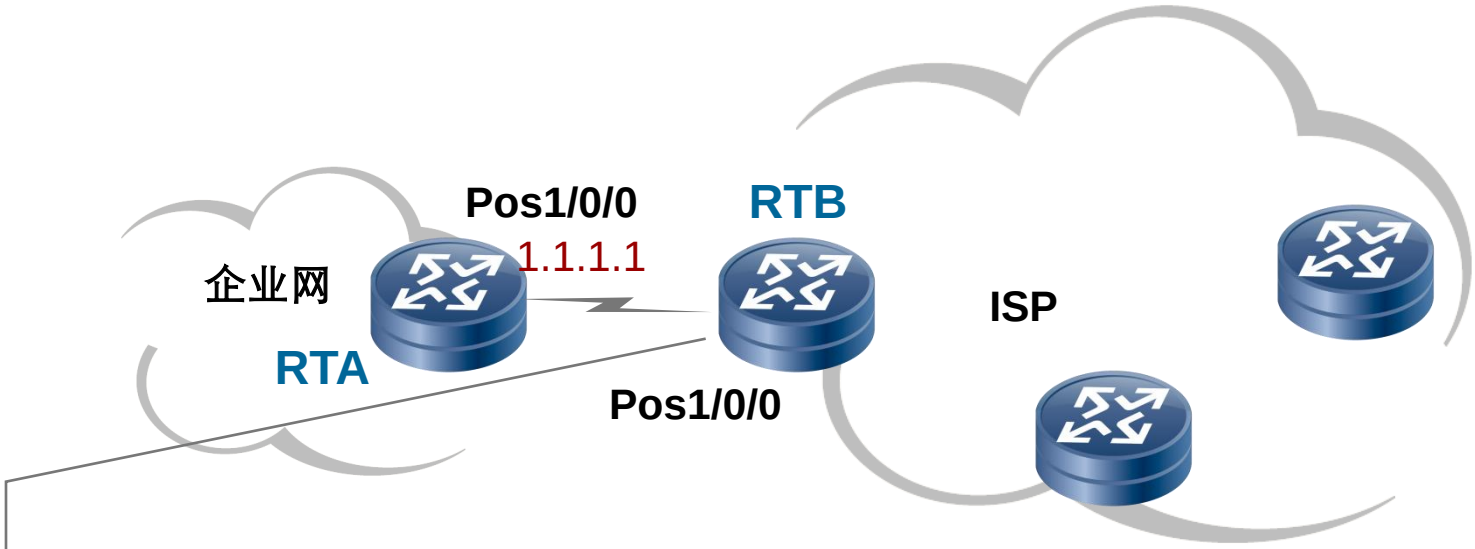


基于接口的流量监管配置举例



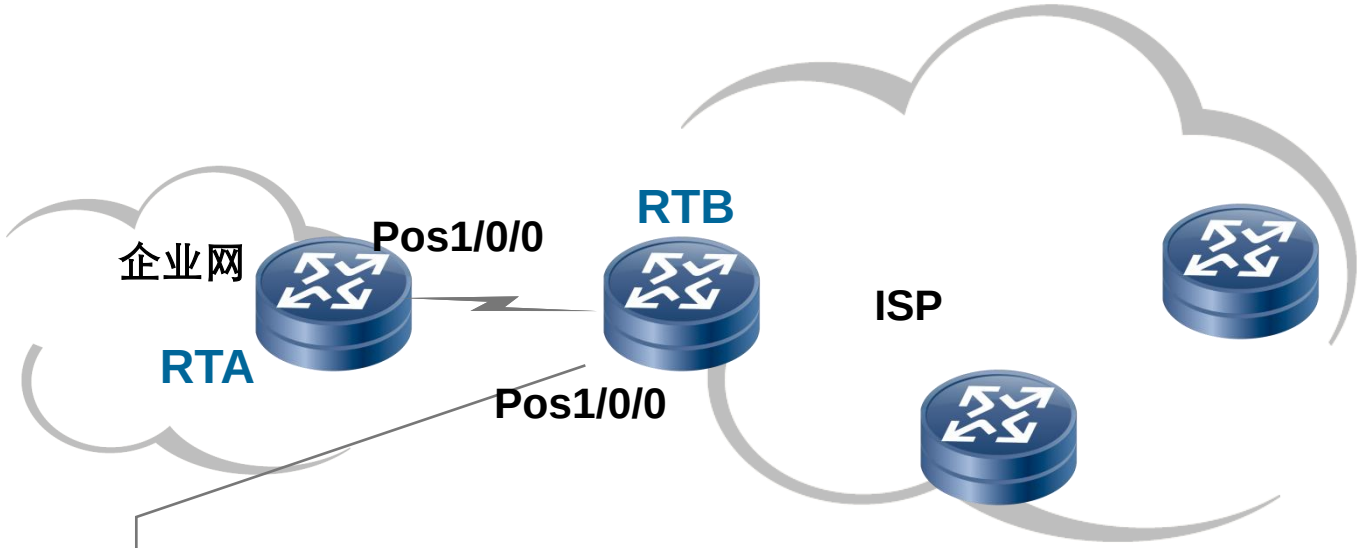

```
<RTB>system-view
[RTB]interface pos1/0/0
[RTB-pos1/0/0]qos car cir 100 pir 10000 green pass yellow
pass red discard inbound
```

基于复杂流分类的流量监管配置举例




```
[RTB] acl number 2001
[RTB-acl-basic-2001] rule permit source 1.1.1.1 0.0.0.0
[RTB] traffic classifier aa
[RTB-classifier-a] if-match acl 2001
[RTB] traffic behavior bb
[RTB-behavior-a] car cir 5000 pir 6000 green pass yellow pass red discard
[RTB-behavior-a] quit
```

基于复杂流分类的流量监管配置举例（续）



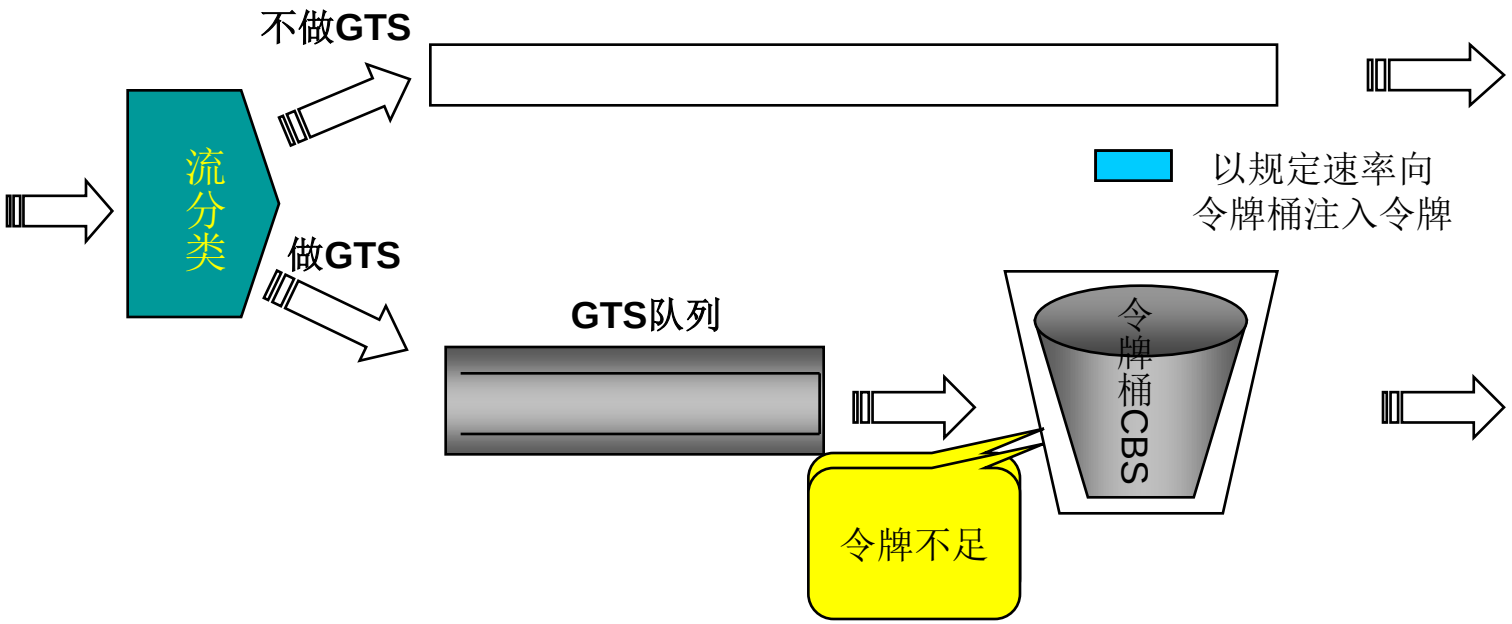

```
[RTB]traffic policy 1
[RTB-trafficpolicy-1]classifier aa behavior bb
[RTB]interface pos1/0/0
[RTB-pos1/0/0]traffic-policy 1 inbound
```

流量整形介绍

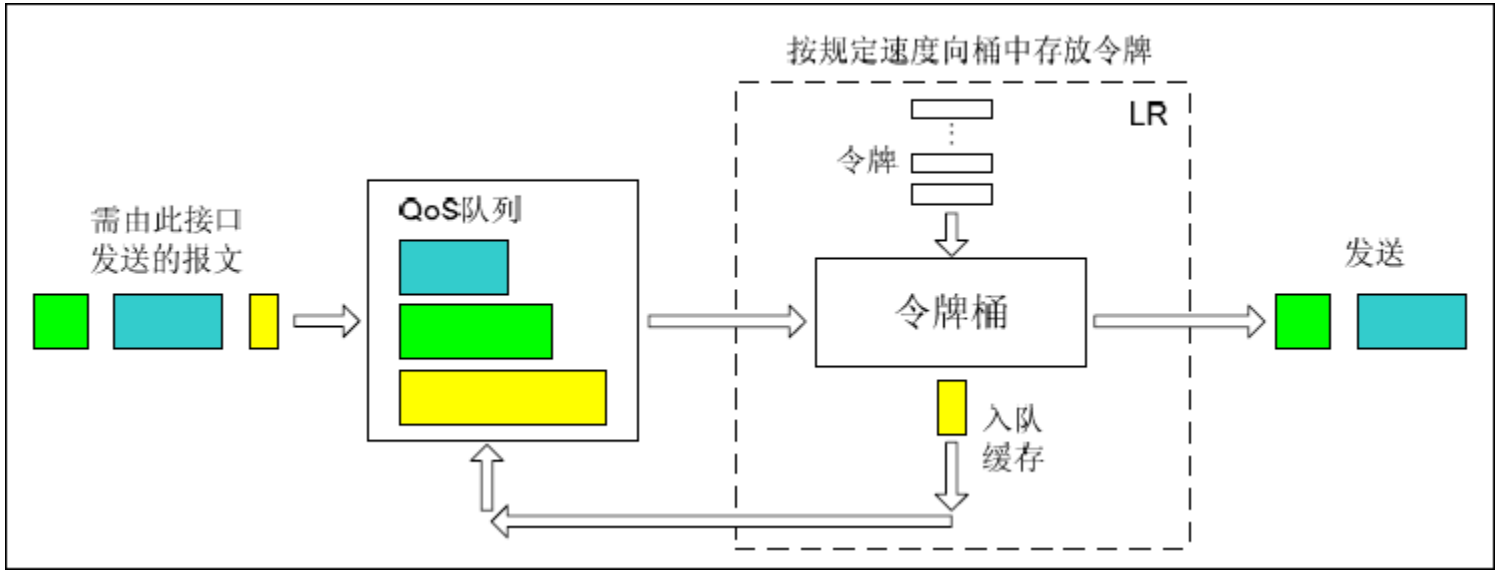
流量整形（traffic shaping）的典型作用是限制流出某一网络的某一连接的流量与突发，使这类报文以比较均匀的速度向外发送。流量整形通常使用缓冲区或队列和令牌桶来完成，当报文的发送速度过快时，首先在缓冲区或队列进行缓存，在令牌桶的控制下，再均匀地发送这些被缓冲的报文。

流量整形通常采用的技术有：Generic Traffic Shaping（通用流量整形，简称GTS），Line Rate（物理接口总速率限制，简称LR）。它们可以对不规则或不符合预定流量特性的流量进行整形，以利于网络上下游之间的带宽匹配。

GTS：Generic Traffic Shaping

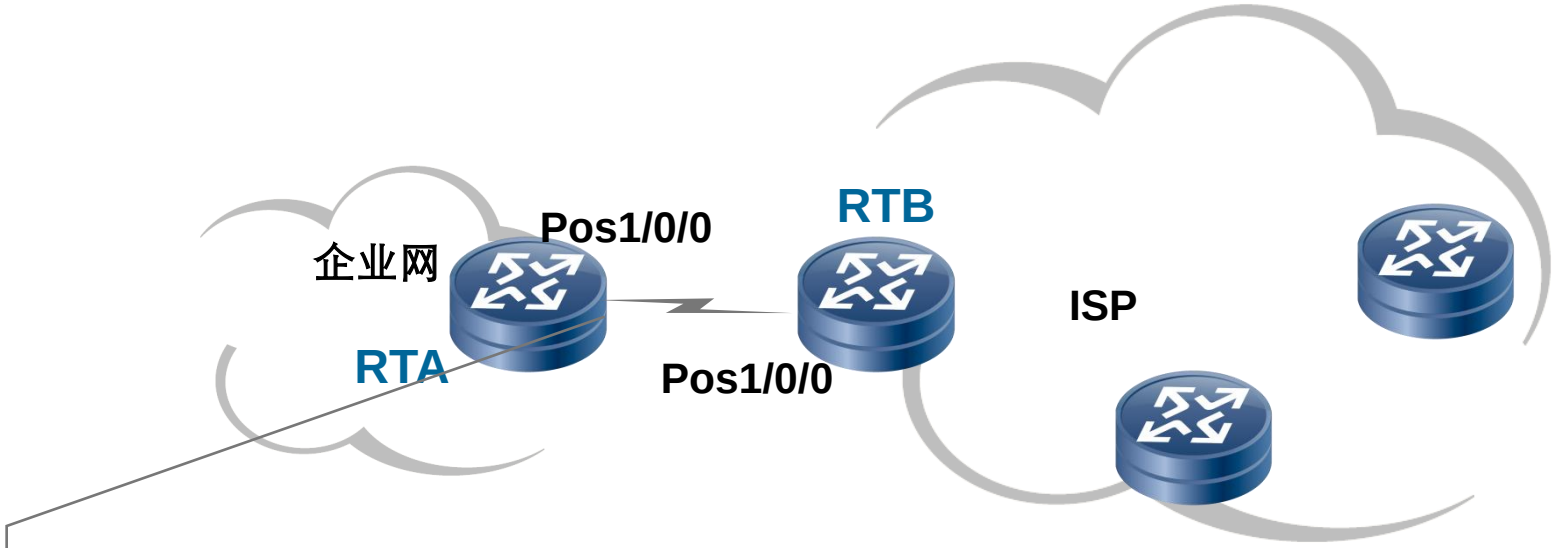


LR：Line Rate



LR限制的是接口发送报文的总速率。

流量整形配置举例

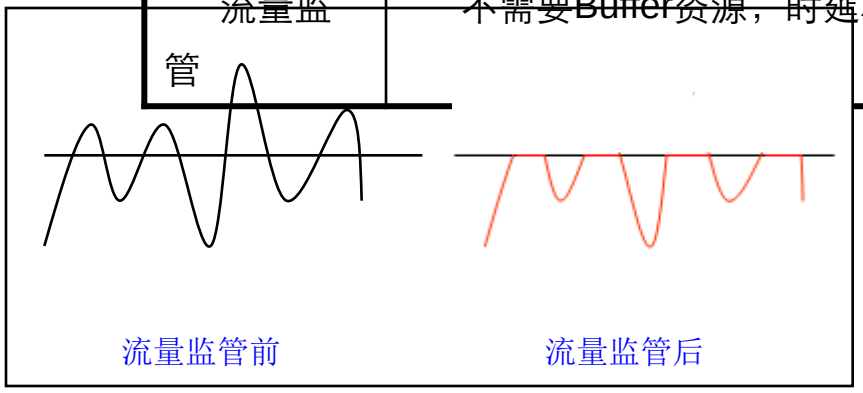


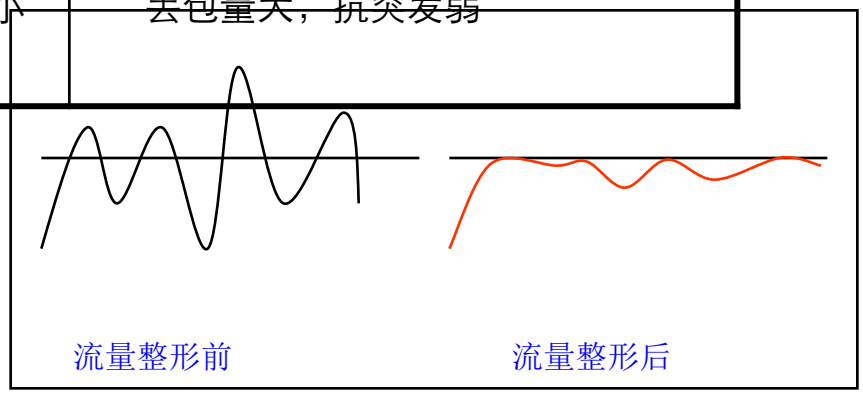

```
[RTA]interface pos1/0/0
[RTA-pos1/0/0]port-queue be wfq weight 10 shaping 20 outbound
[RTA-pos1/0/0]port-queue ef pq shaping 100 outbound
```

流量监管与流量整形的区别

GTS、LR与CAR三者均采用了令牌桶技术来控制流量。它们的主要区别在于：在进行报文流量控制时，CAR对超过流量限制的报文进行丢弃；而GTS则将报文缓存在GTS队列中。 相较于GTS，LR不但能够对超过流量限制的报文进行缓存，并且可以利用QoS丰富的队列来缓存报文。

类型	优点	缺点
流量整形	丢包量少，抗突发强	时延大，需要Buffer资源缓存报文
流量监管	不需要Buffer资源，时延小	丢包量大，抗突发弱





问 题

什么是流量监管?

流量监管的实现方法包括哪些?

流量整形的作用是什么?

流量监管与流量整形的区别是什么?

总 结

完成本节的学习后，您应该掌握以下几点：

流量监管和流量整形的基本概念。

流量监管与流量整形的实现原理。

流量监管与流量整形的区别。

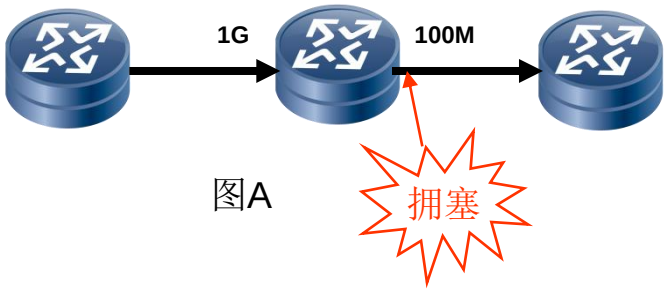
目 录

拥塞与拥塞管理

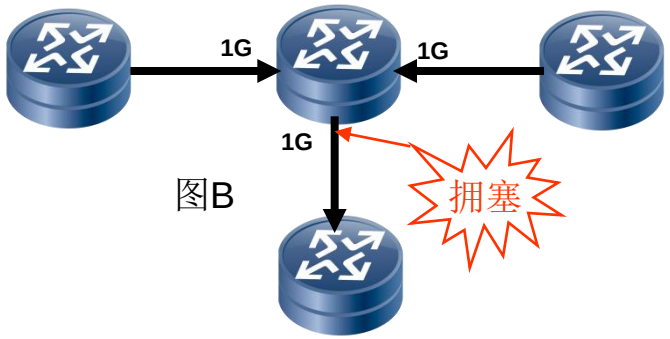
流量从高速端口流向低速端口会在低速端口上产生拥塞，如图A；

流量从多个端口流向同一个端口会在汇聚端口上产生拥塞，如图B

拥塞管理是指网络在发生拥塞时，如何进行管理和控制。处理的方法是使用队列调度技术。将所有要从一个接口发出的报文进入多个队列，按照各个队列的优先级进行处理。通过适当的队列调度机制，可以优先保证某种类型的报文的QoS 参数，例如带宽、时延、抖动等。



图A



图B

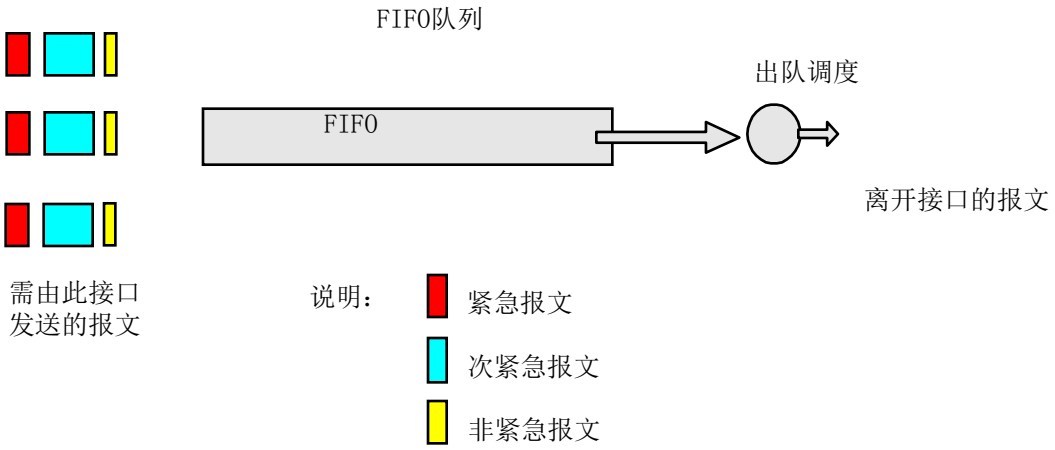
HUAWEI

队列技术

当拥塞发生时，多个报文会同时竞争使用资源， 导致得不到资源的某些业务报文被丢弃，尤其不能保证关键业务的带宽、时延、抖动等 QoS参数。此时如何制定一个资源的调度策略决定报文转发的处理次序，就是拥塞管理的中心内容。对于拥塞管理，一般采用队列调度技术，常见的队列调度技术有以下几种：

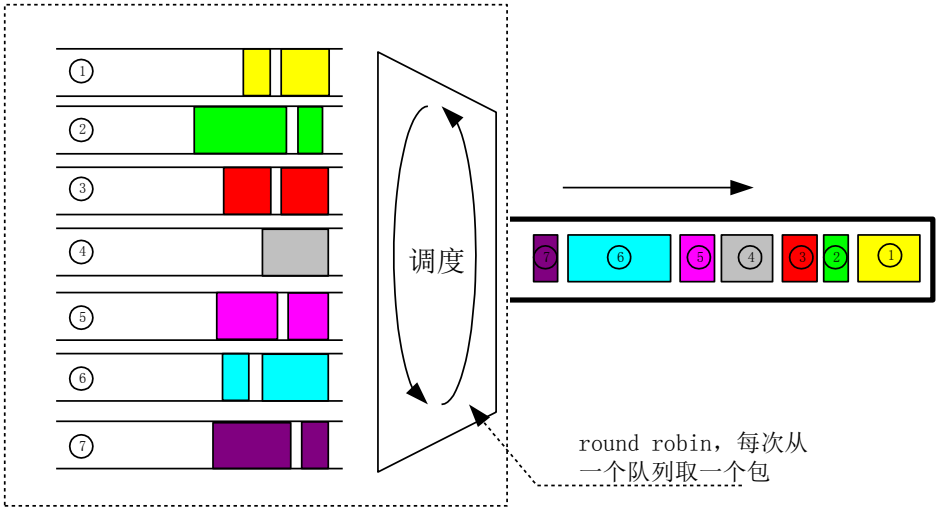
- FIFO：First In First Out，先进先出队列
- RR：Round Robin，轮询队列
- WRR：Weight Round Robin，按权重轮询队列
- PQ：Priority Queuing，优先级队列
- CQ：Custom Queuing，自定义队列
- WFQ：Weighted Fair Queuing，加权公平队列

FIFO：First In First Out



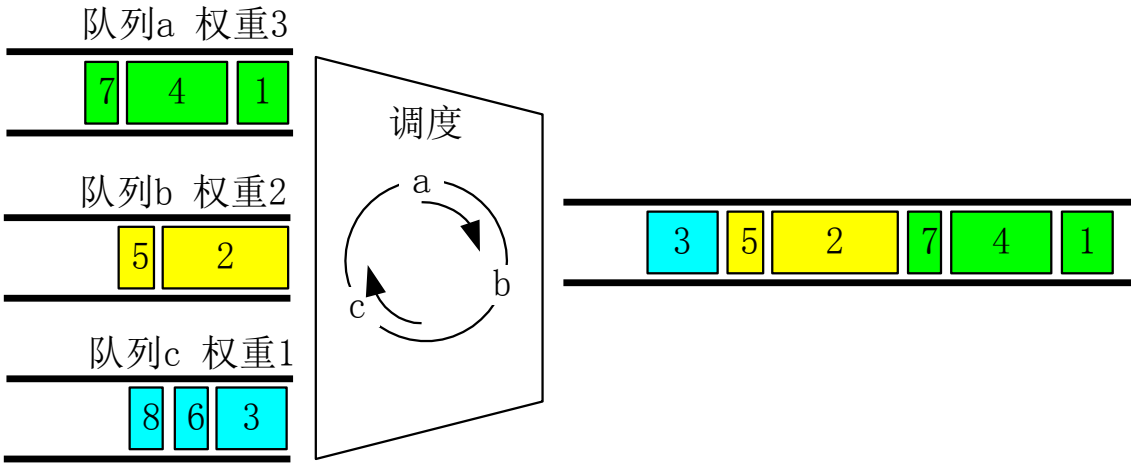
从接口输出的报文，按照到达的先后顺序进入接口的FIFO队列，调度器按照先进先出的原则，从队首开始，依次发送报文。所有的报文在发送过程中，没有任何区别，也不对报文传送的质量提供任何保证。

RR：Round Robin



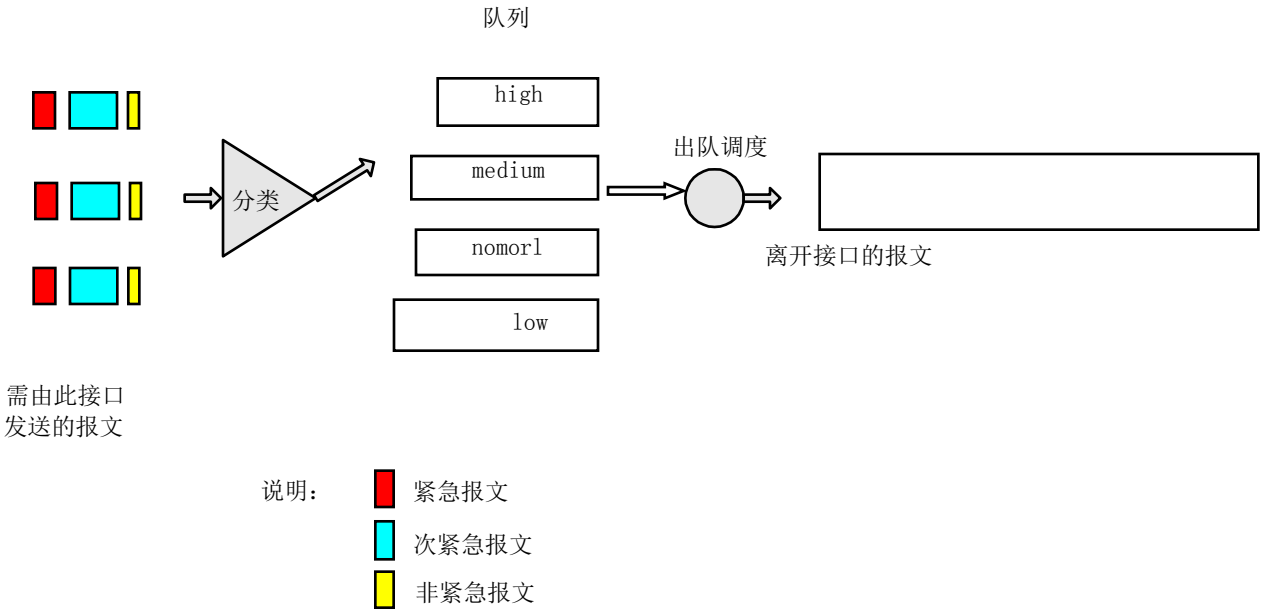
RR是Round Robin的缩写，是一种简单的调度方法，采用轮询的方式，对多个队列进行调度RR以环形的方式轮询多个队列。如果轮询的队列不为空，则从该队列取走一个报文；如果该队列为空，则直接跳过该队列，调度器并不等待。

WRR：Weight Round Robin



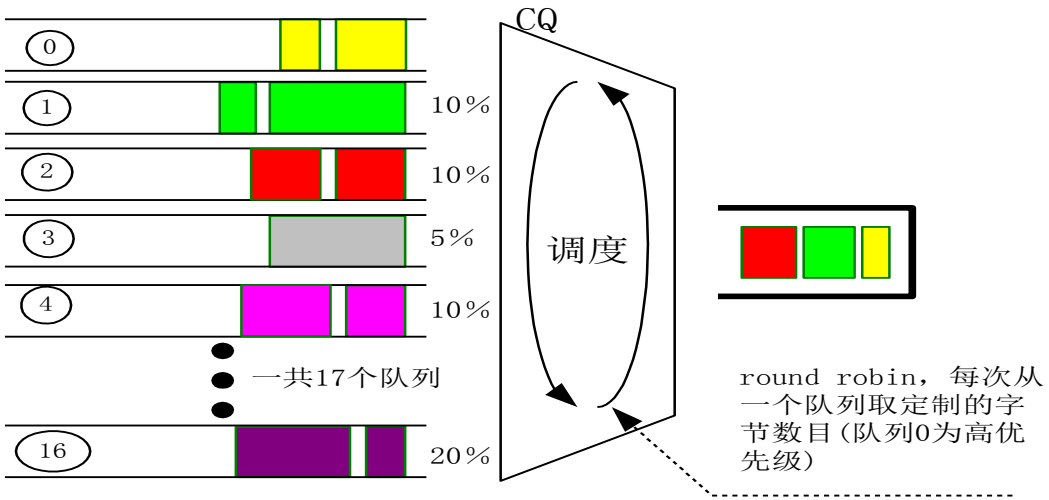
WRR（Weighted Round Robin）主要针对RR不能设置权重的不足，在轮询的时候，每个队列享受的机会和该队列的权重成比例。WRR最初是针对固定包长（ATM）设计的调度算法。WRR对于空的队列直接跳过，调度一周结束的时间变短，因此当某个队列的流量小的时候，剩余带宽能够被其他队列按照比例占用。

PQ：Priority Queuing



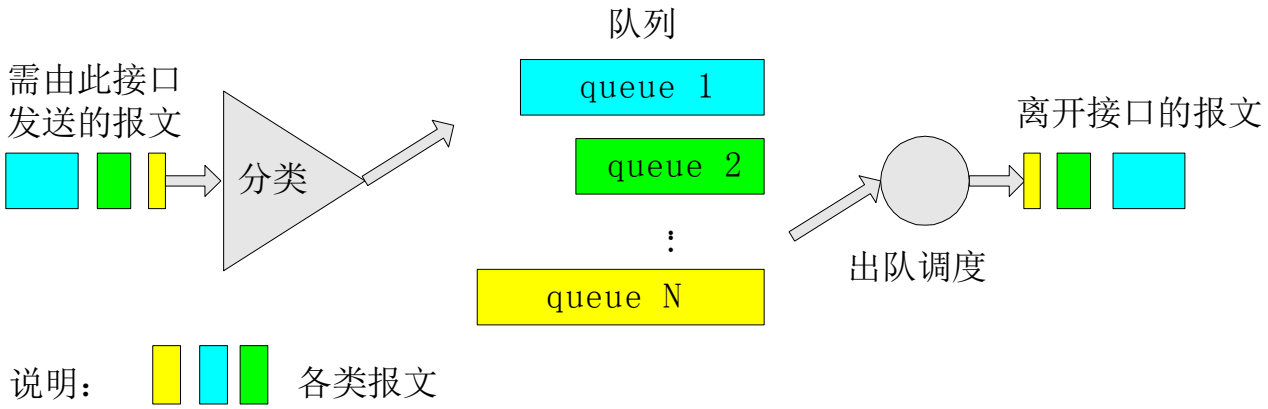
PQ（Priority Queuing）是一种按照严格优先级（SP，Strict Priority）进行调度的队列。PQ对队列划分等级，只有高优先级的队列排空以后才会从低一级的队列调度报文，这样重要的业务比其他业务提前获得服务

CQ：Custom Queuing



CQ（custom Queuing）可以支持17个队列，队列0用于系统队列，队列0和其他队列之间是SP的关系，只有队列0排空以后才能为其他队列提供服务，队列0一般用于协议报文。队列1至16没有优先级关系，采用轮询的方式，每次调度的时候从队列中调度固定字节数（预先配置），在轮询下一个队列之前，将数据包发出去。当某个队列已经调度了规定的字节数，或者该队列已经空，则轮询下一个队列。

WFQ：Weighted Fair Queuing



报文到达接口后，首先对报文进行分类，不同的流分入不同的队列。在出队的时候，WFQ按流的权重分配每个流应占的带宽。权重数值越小，所得的带宽越少。权重数值越大，所得的带宽越多。这样就保证了相同优先级业务之间的公平，体现了不同优先级业务之间的权值。

各种队列调度技术对比

队列技术	调度的时延/抖动（**在速率低的时候明显，速度绝对高的时候可忽略**）	公平性
FIFO	差	无
RR	差	依赖包长
WRR	差	依赖包长
PQ	高优先级队列的时延控制非常好	无
CQ	配置字节数小的时候，带宽分配不准确，当 配置字节数大的时候，时延抖动比较大差	一般
WFQ	时延控制较好，抖动小	好

队列技术在产品中的实现

目前产品中，主要使用FIFO、WFQ、PQ三种队列技术来实现拥塞管理。

对于队列配置，用户无须关心采用什么抽象的调度算法，只需关心队列所承载业务的外在流量参数特征，比如保证多少兆的带宽、峰值最多多少兆的带宽、要占剩余带宽的比例权重等。根据配置的流量参数选用不同的调度算法来严格保证用户的配置。

端口队列调度采用PQ+WFQ调度算法，采用这种调度优势在于，既能对时延敏感的实时业务得到保证，对优先业务的报文的带宽占用可以绝对优先，又可以为不同优先级的流根据配置的权重分配不同的带宽。

问 题

什么是拥塞管理?

常用的队列调度技术有哪些?

总 结

完成本节的学习后，您应该掌握以下几点：

拥塞管理的基本概念；

常用的队列技术；

各种队列调度技术的优缺点。

目 录

拥塞避免机制

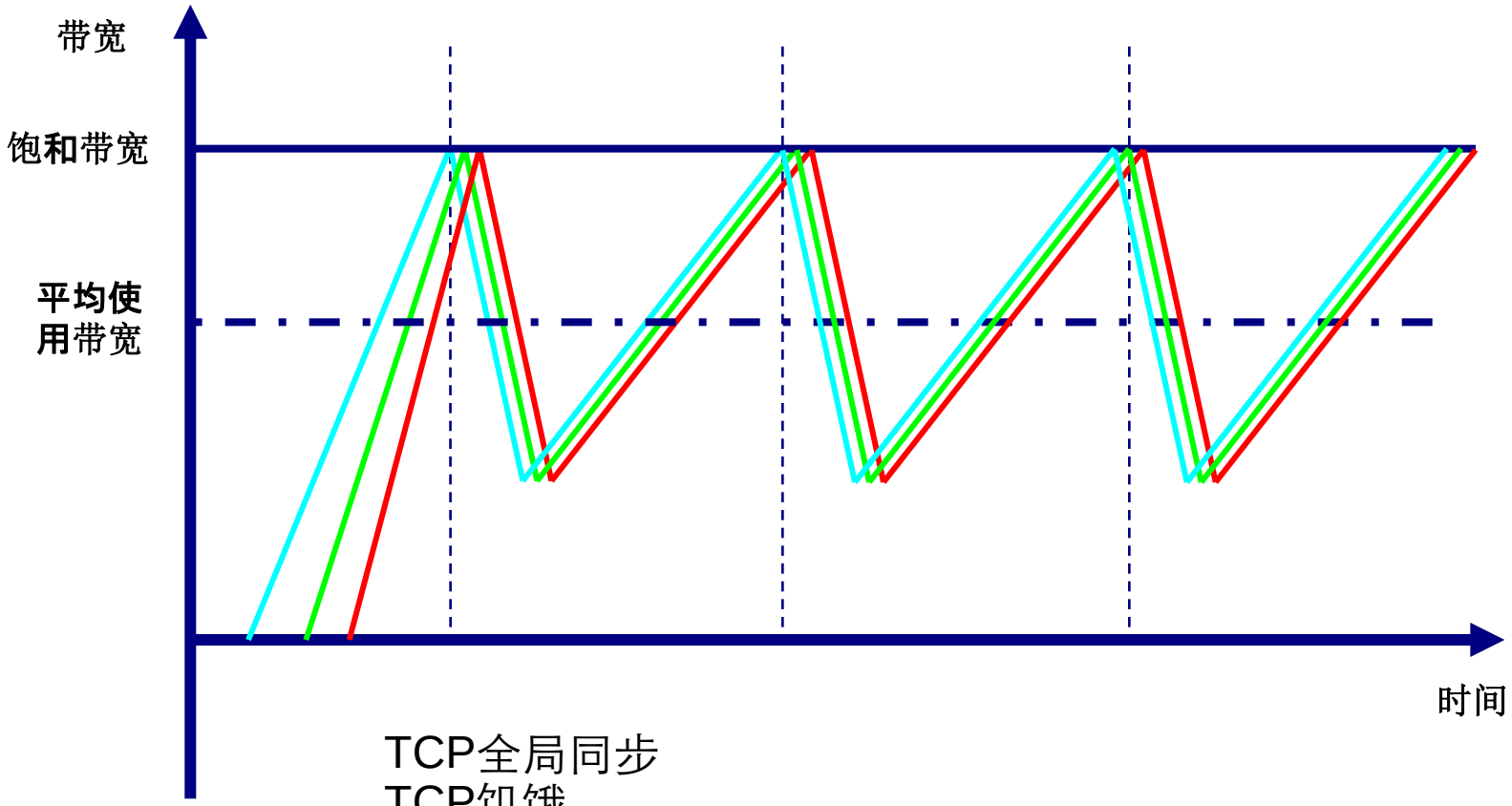
拥塞避免是一种流控机制，它可以通过监视网络资源（如队列或内存缓冲区）的使用情况，在拥塞有加剧的趋势时，主动丢弃报文，通过调整网络的流量来解除网络过载。

传统的丢包策略采用尾部丢弃（Tail-Drop）的方法。当队列的长度达到某一最大值后，所有新到来的报文都将被丢弃。这种丢弃策略会引发TCP全局同步现象。

为避免TCP全局同步现象，可使用RED或WRED。

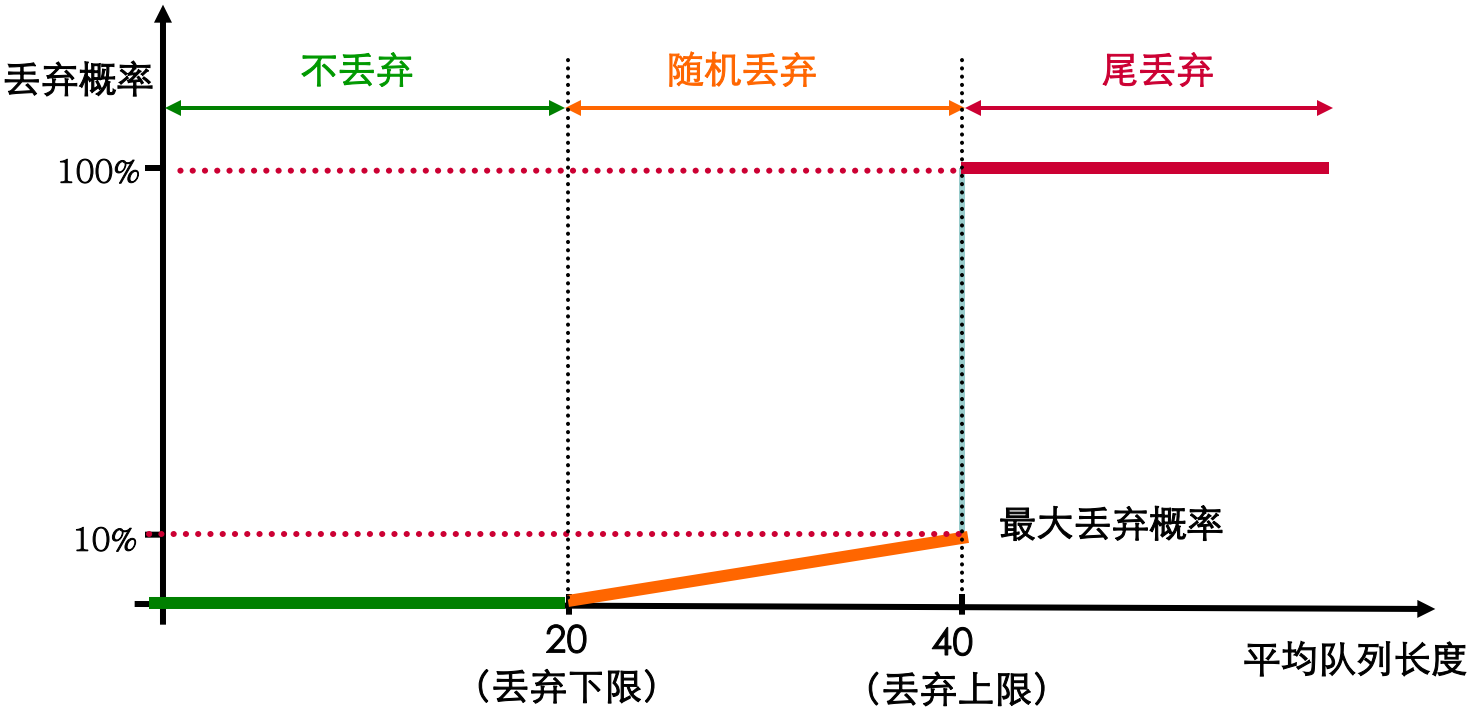
- RED：Random Early Detection，随机早期检测
- WRED：Weighted Random Early Detection，加权随机早期检测

尾丢弃（Tail-Drop）的缺点

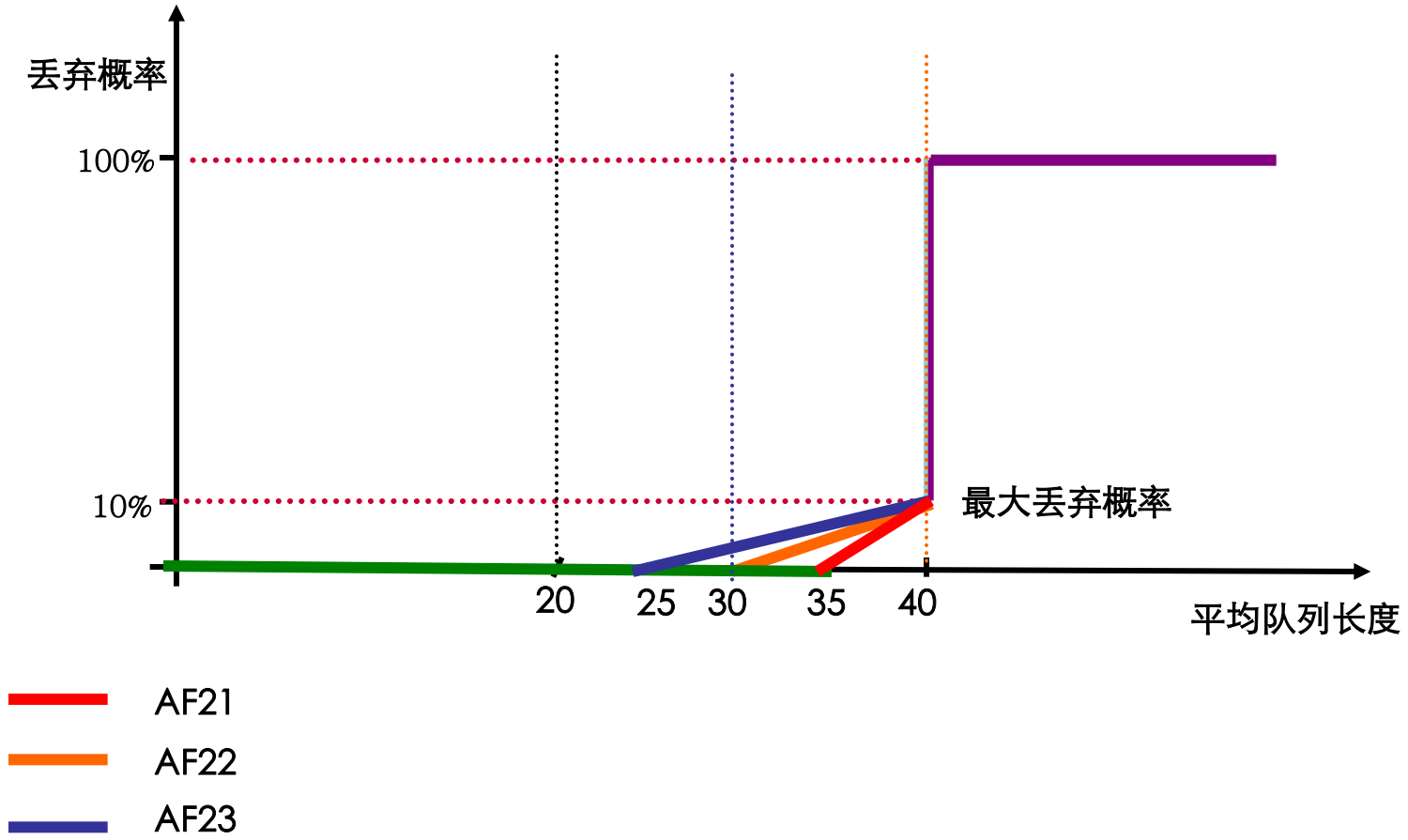


TCP全局同步
TCP饥饿
高延时和高抖动
无差别的丢弃

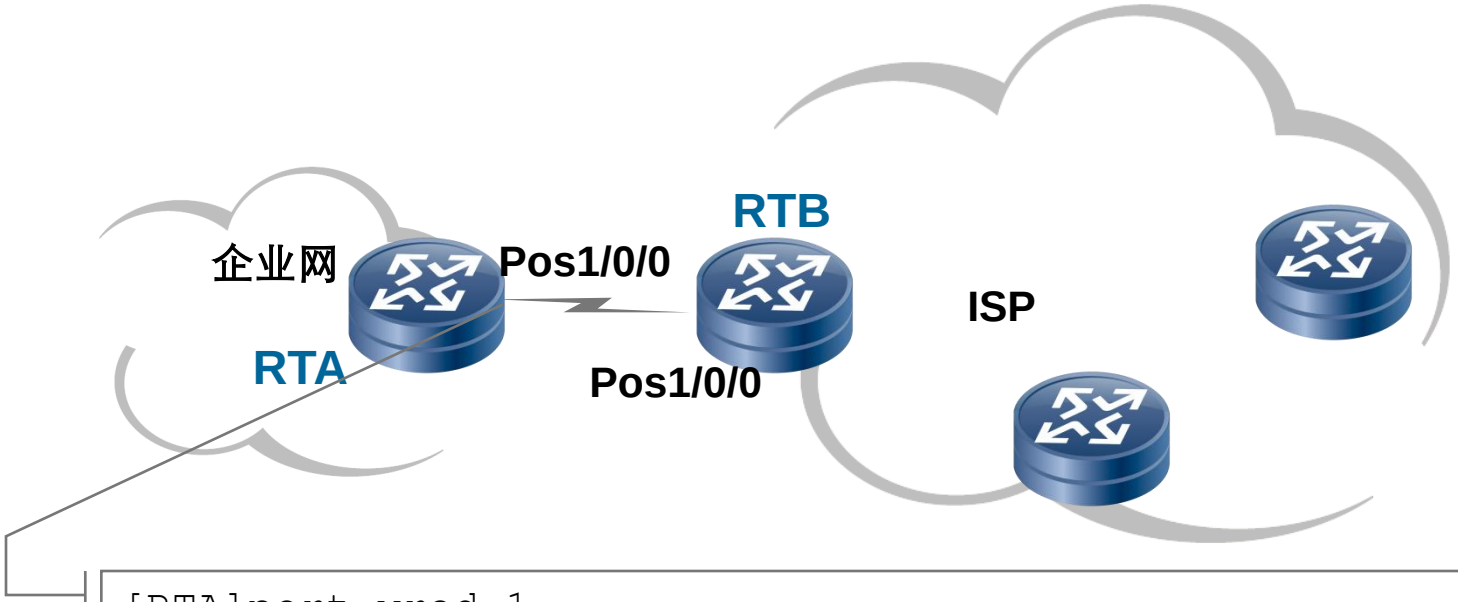
RED：Random Early Detection



WRED：Weighted Random Early Detection



拥塞避免机制配置举例




```
[RTA]port-wred 1
[RTA -port-wred-1] color green low-limited 30 high-
limited 80 discard-percentage 50
[RTA -port-wred-2] color green low-limited 20 high-
limited 80 discard-percentage 80
[RTA]inteface pos1/0/0
[RTA-pos1/0/0]port-queue ef port-wred 1 outbound
[RTA-pos1/0/0]port-queue be port-wred 2 outbound
```

问 题

什么是拥塞避免?

常用的拥塞避免有哪些?

总 结

完成本节的学习后，您应该掌握以下几点：

什么是TCP全局同步；

如何避免网络拥塞。

目 录

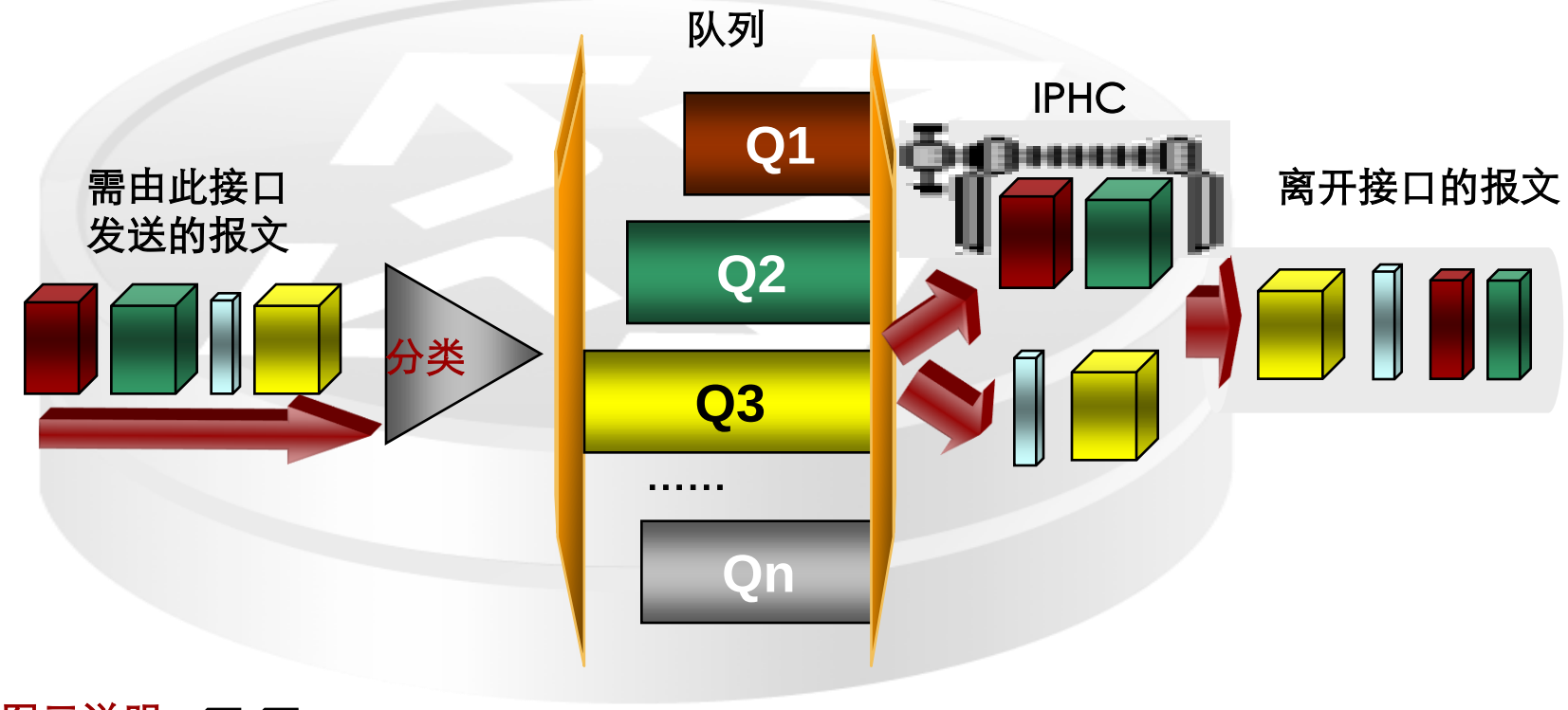
链路效率机制

IPHC：IP Header Compression，IP报文头压缩

- RTP报文头压缩
- TCP报文头压缩

LFI：Link Fragmentation and Interleaving，链路分片与交叉

IPHC：IP Header Compression



图示说明：

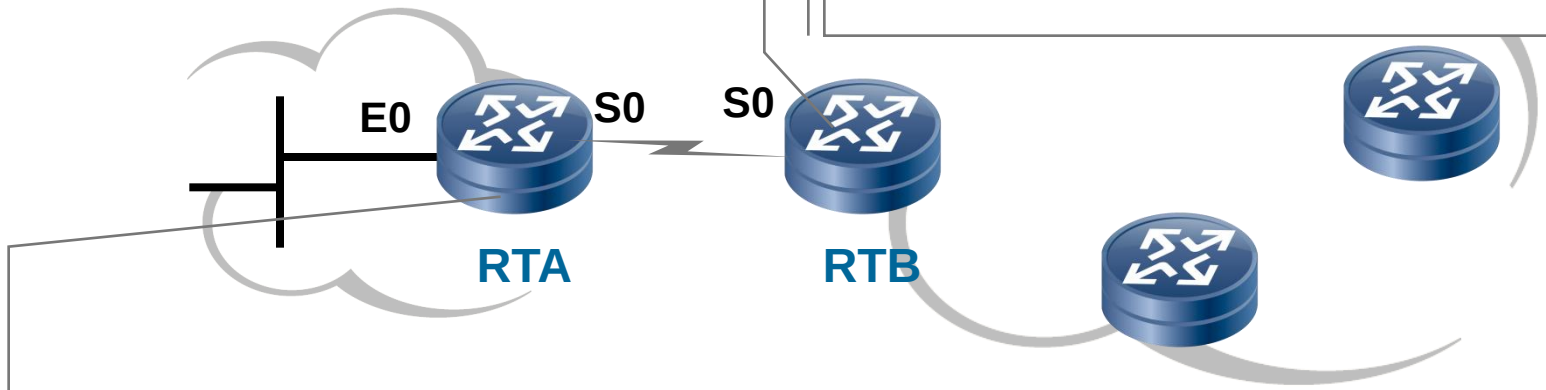
RTP和TCP报文

其它各类报文

HUAWEI

配置IPHC

```
interface Serial0
 link-protocol ppp
 ip address 12.12.12.2 255.255.255.252
 ppp compression iphc
 ppp compression iphc rtp-connections 20
 ppp compression iphc tcp-connections 20
```



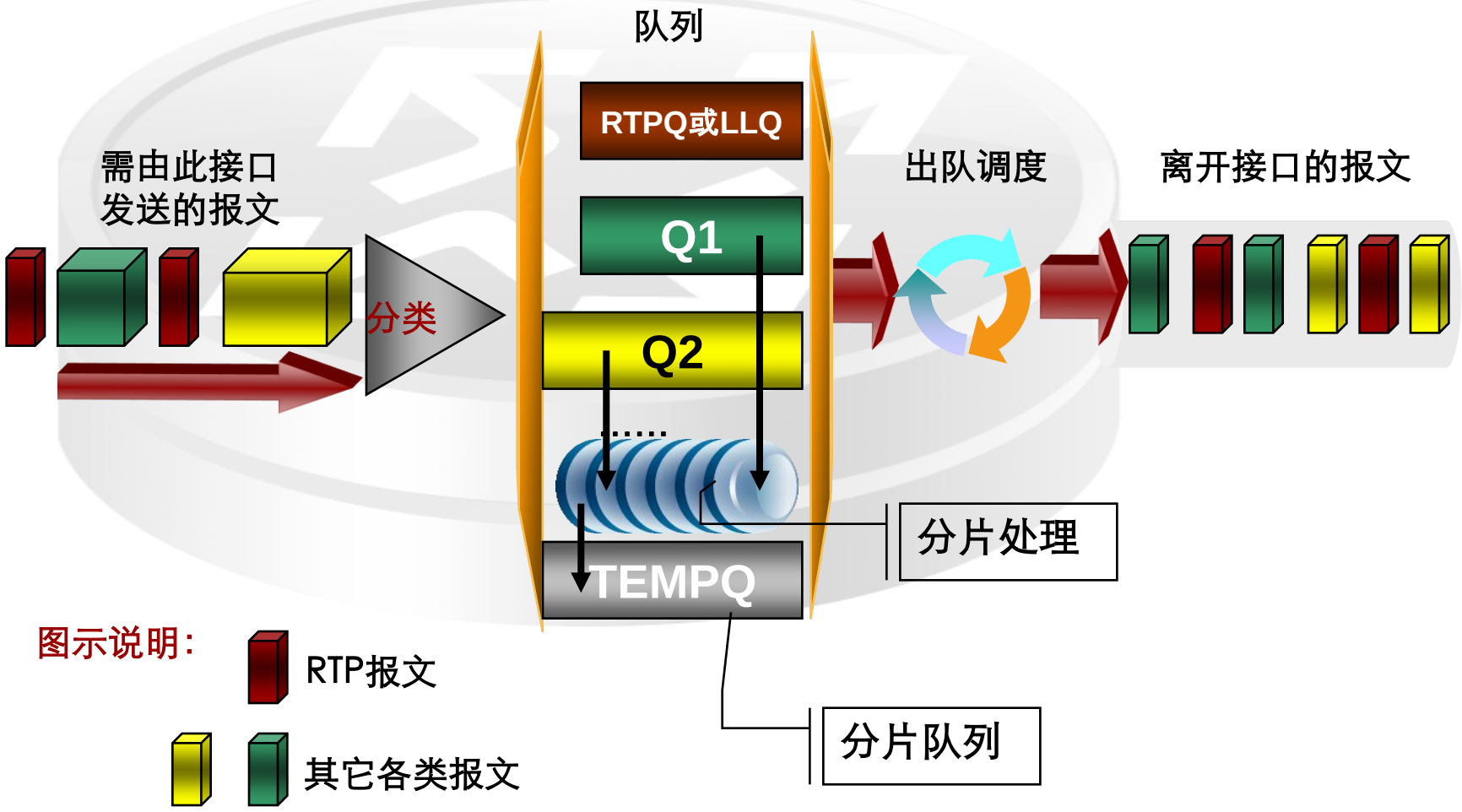

```
interface Serial0
 link-protocol ppp
 ip address 12.12.12.1 255.255.255.252
 ppp compression iphc
 ppp compression iphc rtp-connections 20
 ppp compression iphc tcp-connections 20
```

配置IPHC

```
[RTA]display ppp compression iphc tcp
IPHC: TCP/IP header compression
  Interface: Serial0
    Received:
      Compress/Error/Discard/Total: 0/0/0/0 (Packets)
    Sent:
      Compress/Total: 0/0 (Packets)
      Send/Save/Total: 0/0/0 (Bytes)
    Connect:
      Rx/Tx: 20/20
      Long-search/Miss: 0/0

[RTA]display ppp compression iphc rtp
IPHC: RTP/UDP/IP header compression
  Interface: Serial0
    Received:
      Compress/Error/Discard/Total: 0/0/0/0 (Packets)
    Sent:
      Compress/Total: 0/0 (Packets)
      Send/Save/Total: 0/0/0 (Bytes)
    Connect:
      Rx/Tx: 20/20
      Long-search/Miss: 0/0
```

LFI：Link Fragmentation and Interleaving



问 题

常用链路效率机制有哪些?

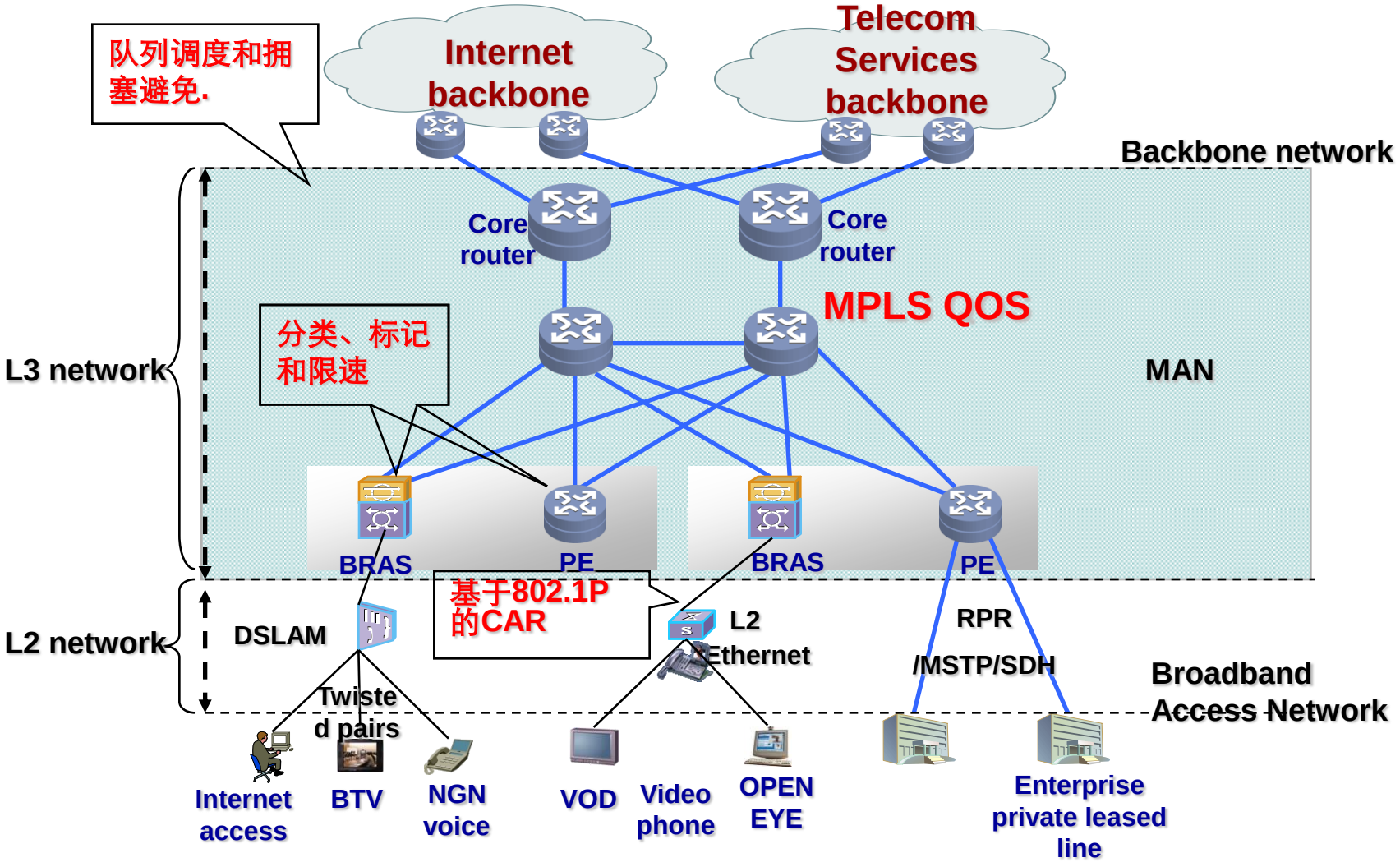
总 结

完成本节的学习后，您应该掌握以下几点：

常用链路效率机制。

链路效率机制应用。

参考：IP城域网QoS解决方案



HUAWEI

谢谢

www.huawei.com