

服务等级(CoS)的应用

Juniper 网络公司，爱立信公司，2001 年 3 月

内容提要	2
角度	2
服务等级机制	2
分组延迟	3
实现 COS 特性	4
限速	5
策略管理程序	5
输出队列选择	5
加权循环	6
优先级字段重写	6
随机早期检测	7
在 IP 语音中应用 COS	8
VoIP 配置实例	8
队列管理	11
结论	12
缩略语	12

内容提要

服务等级 (CoS) 可以定义为把流量分配到不同的等级,其中每个等级可以提供不同的时延、抖动和丢包特点。

本文介绍了爱立信公司的 AXI 系列路由器中实现的、通过JUNOS™ Internet软件配置的CoS特性。它还介绍了CoS特性解决的部分问题。另外还包括了多个实例,说明了怎样使用这些特性提供新服务。

角度

典型的IP网络独立路由分组,没有任何端到端吞吐量保障;这些网络提供的服务称为尽力而为的服务,其服务质量取决于网络设备和网络链路,其结果对大多数IP应用还是满意的。但是,某些IP应用,如实时视频和音频应用,要求在网络拥塞时提供优于尽力而为的服务。CoS针对的就是这些类型的应用,以保证它们一直正常运转。

服务等级机制

为了满足Internet对CoS的需求,业内提供了许多解决方案,包括利用服务类型(ToS)字段、资源预留协议(RSVP)及IETF DiffServ工作组最近完成的工作。

CoS特性同时利用ToS字段设置和DiffServ (DS)字节,因为DS字节是由ToS字段内部的六位构成的。RFC 791中定义的IP包头说明了ToS字段的位置(图1)。

$600 \text{英里} * 2 / (0.6 * 186,000 \text{英里/秒}) = 0.01075 \text{秒} = 10.75 \text{毫秒}$

实现 CoS 特性

CoS特性包括策略管理、队列选择、加权循环(WRR)、优先级字段重写和随机早期检测(RED)。

您可以实现一种或多种CoS特性,具体取决于网络要求。在构成网络的路由器上启动时,AXI系列路由器的CoS特性可以彼此及与其它厂商的平台很好地一起运行。

在配置 CoS 前,很重要的一点是认真考虑整体 CoS 设计和每个网络设备上的配置,因为它们必须一起运行,如图 2 所示。

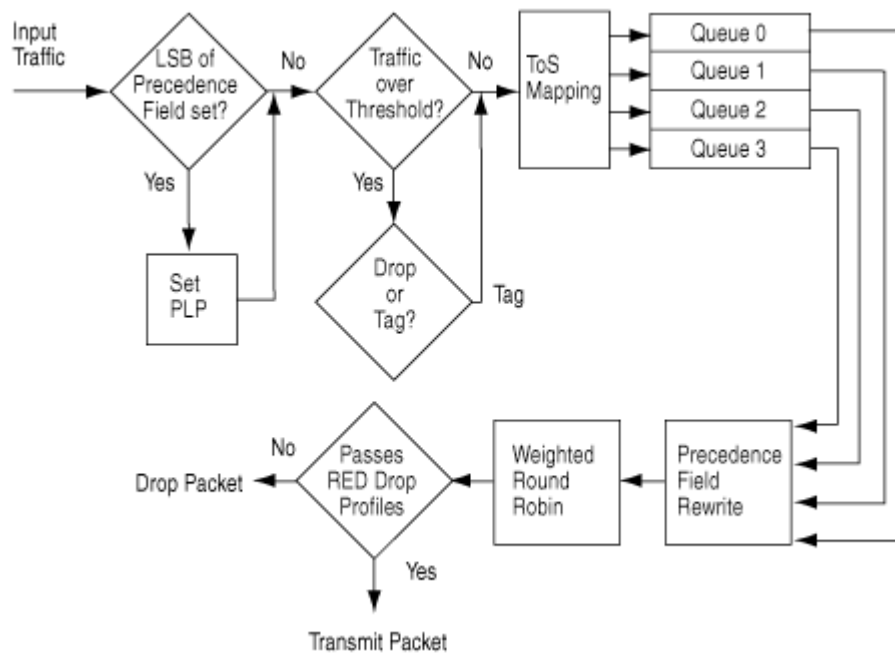


图 2 : CoS 特性的逻辑图

Input Traffic : 输入流量

LSB of Precedence Field set? : 是否设置优先字段的LSB ?

Traffic over Threshold? : 超过门限的流量

ToS Mapping: ToS映射

Queue 0: 队列0

Set PLP: 设置PLP

Drop or Tag? : 丢弃不是标记 ?

Tag: 标记

Drop Packet: 丢弃分组

Passes RED Drop Profiles: 通过RED丢弃概况

Transmit Packet: 传送分组

Weighted Round Robin: 加权循环

Precedence Field Rewrite: 重写优先级字段

限速

所有采用Internet Processor II ASIC处理器的AXI系列路由器都可以在任何接口上，在硬件中提供策略管理。过去，所有SONET/SDH和T3接口在硬件中提供了限速功能，现在它们仍然提供这种功能。但是，Internet Processor II ASIC策略管理功能进一步得到了改善，其变得更加灵活，原因如下：

- 其采用的令牌桶算法在处理突发流量的同时，在调节长期平均传输速率时更加有效。SONET/SDH和DS-3接口上I/O管理器ASIC执行的泄漏桶算法只能在接收桶上提供有限的突发功能。而通过使用Internet Processor II ASIC，传输桶和接收桶上都可以保持更大的突发灵活性。
- 在使用Internet Processor II ASIC执行的令牌桶算法时，接收桶不会丢弃流量，而在泄漏桶算法中，在传输分组的信用不足时，则会丢弃分组。
- Internet Processor II ASIC可以在所有接口上执行入局和出局限速功能。而I/O管理器ASIC执行的限速功能只能用于SONET/SDH和DS-3。
- 策略管理功能从防火墙过滤策略内部调用，这意味着限速的流量可以基于为防火墙过滤器提供的任何匹配条件。这种改善的细密度非常宝贵，因为在决定对哪些流量进行限速时，现在可以把关键事务型流量与其它流量区别开来。

如需匹配条件列表，请参阅接口和机箱软件配置指南。

策略管理程序

为定义策略管理程序，应编写防火墙过滤器，其中包括一条策略管理程序语句。这条策略管理程序语句定义了最大带宽和最大突发参数；带宽以每秒比特数为单位，突发以每秒字节数为单位。

策略管理程序中还定义了一个超过操作。超过操作可以规定一项操作，如丢弃、输出队列选择、或把丢包优先权(PLP)位设为1。PLP位是通知包头中的一种内部位，在拥塞过程中可供输出队列中的RED使用，以提高分组被丢弃的概率。通过重新编写优先级字段，确保设置了有效性最低的位(LSB)，还可以下行检查PLP信息。另外也可以通过泄漏桶配置和防火墙过滤器操作，来设置PLP。

输出队列选择

所有AXI系列路由器都为每条物理链路提供了4条出局传输队列，每个灵活PIC集中器(FPC)支持最多64个出局传输队列。您可以把流量分配到不同的队列，以实现不同的服务水平。您可以在CoS配置下，根据进入接口、ToS字段设置或信宿IP地址，把IPV4流量分配给特定的出局传输队列。您还可以通过防火墙过滤器配置把流量分配到出局队列上。在这种情况下，您可以根据匹配分组的任何标准选择出局队列。

您可以忽略IPV4分组的初始进入分类，而是根据信宿IP地址选择输出队列。多协议标记交换(MPLS)分组可以静态匹配到与MPLS包头中的CoS字段设置相对应的队列上。也就是说，如果IPV4分组映射到入口路由器的队列2，然后沿着标记交换路径(LSP)转发，MPLS CoS字段

的前两位将被10占用，分组将沿着LSP发送到每个出局接口上的出局队列2。作为备选方案，可以在LSP的入口编写CoS字段，MPLS CoS字段中两个最有效位的值将与分组在沿着LSP的每个出局接口上使用的出局队列相对应。

下面是队列默认值。

- 所有IPv4流量都放在队列0中，除非ToS位被设置为110或111，在这种情况下，流量被放在队列3中。路由协议控制流量使用这些设置。
- 所有MPLS流量都放在队列0中。
- 在缓冲分配中，队列0获得95%的缓冲总量，队列3获得5%的缓冲总量。
- 在带宽分配中，队列0获得95%的带宽总量，队列3获得5%的带宽总量。您可以通过WRR计算这种设置。

加权循环

AXI系列路由器的WRR方案可以控制四个传输队列中每个队列的服务方式。您可以为每个队列分配一定比例的总可用带宽。为了进行这种分配，可以为每个队列分配一个加权值。加权值可以是比率或百分比，所有加权值的总和必须等于100。在默认状态下，队列0获得95%的加权值，队列3获得5%的加权值。我们建议把队列3保留为默认设置，这意味着您可以在队列0、队列1和队列2之间分配其余95%的带宽。

每个插槽都具有128 MB的内存，这是系统内存池的一部分。大约1/5的内存用于分组通知排队，其余部分是用于系统缓冲池的分组内存。记住，并不是正在排队的分组，而是通知包头。通知包头包括分组存储的内存位置指针。因此，AXI系列路由器在每个分组上只执行一次读操作和一次写操作，而不管是单路广播分组还是多路广播分组。

所有正队列都以配置的比率进行检查，将一直为正队列提供服务，直到它变为负。也就是说，如果一个队列是负的，而其它队列是正的，那么只服务正队列，直到所有队列都变为负。当所有队列都为负时，先检查最低的队列。这种能够传输负信用的能力意味着如果一个队列拥塞，而其它队列拥有剩余的缓冲空间，那么拥塞队列可以从利用率低的队列中借用带宽。这还意味着，AXI系列路由器一直尝试传输，即使没有要传输的分组时。

优先级字段重写

您可以在任何出局队列上重写IPv4分组的优先位。您可以只在LSP的入口上，重写MPLS包头中的实验CoS字段。在选择重写到优先级字段中的值时，有许多关键的考虑因素：

- 是否设置PLP位？
- 这个分组是否应该优先对待？

对MPLS和IP流量，如果设置了优先级字段的LSB，那么下行路由器上的进入接口将设置PLP位。这种配置的目的是提高下行路由器上RED的丢弃概率。因此，明确设置了PLP的分组还应在优先级字段中，重新写入已经设置了优先级字段的LSB的值。

- 001
- 011
- 101
- 111

相反，将重写应优先对待的分组；对没有把PLP设置成在优先级字段中不包含1的值的任何流量，也将重写其分组。

- 000
- 010
- 100
- 110

对MPLS分组，除与下行处理有关的优先级字段的LSB外，前两位用来识别在所有下行出局接口上使用的出局传输队列。您可以通过配置MPLS CoS来管理MPLS CoS字段。

随机早期检测

RED是一种避免拥塞机制，作为AXI系列路由器的CoS特性集的一部分提供。RED允许路由器在队列已满前丢弃分组。如果没有采用这种避免拥塞机制，一旦传输队列已满，它们就会开始丢弃分组，这将带来某些负面影响。最大的问题之一是，交付的分组通常是时间最长的分组，这些分组可能已经请求进行重新传输。结果是导致更加严重的拥塞。另一个结果是全局同步化，也就是一条链路开始变慢时，共享这条链路的所有TCP会话几乎都会同时启动放慢。启动放慢是指控制可以传输的TCP分组数量，而不接收ACK的滑动窗口的小型初始尺寸。如果链路采用先进先出法，从尾部开始丢弃，那么一旦队列拥塞，采用该链路的所有会话都会同时发现拥塞。

RED在拥塞发生前预测拥塞的链路，并从队列前面以随机的方式丢弃分组。RED的随机性保证了不会一直影响具体的会话，从而在会话丢弃分组过程中实现公平性。

Internet Processor II ASIC采用的RED实现技术允许配置哪些分组的丢弃概率更高，从而提高了灵活性。通过网络边缘配置策略管理，可以丢弃或标记超过门限的分组。如果标记，则在发生拥塞时，它们被丢弃的概率要更高一些。您可以为设置了PLP的分组及清除了PLP的分组简单地配置不同的丢弃概况。

可以配置三种丢弃概况：一种用于两条数据流(物理接口的集合)，两种用于各个队列。其中一种队列丢弃概况用于设置了PLP位的分组，另一种用于清除了PLP的分组。丢弃概况是带有丢弃概率的索引匹配缓冲器深度(或满的程度)。在每个丢弃概况中，您可以配置最多64套满时丢弃概率对。两个丢弃概况在队列级对每个分组检查一次，这取决于是否设置了PLP位，在数据流级则检查一次。对于要丢弃的分组，两种概况必须协商一致。这种协议可以防止分组被丢弃，除非链路发生拥塞，而且分组所在的队列也发生拥塞。

每个队列的缓冲器深度是可以配置的，这直接影响着RED。例如，如果您为队列0分配40%的缓冲器空间，为队列1到队列3各分配20%的缓冲器空间，那么队列1到队列3每个队

列都将先于队列 0 达到装满水平。这种配置适用于实时视频、音频流量和语音。为服务于低时延应用的队列分配比例较小的缓冲器空间，可以强制取消交付陈旧的分组。

在 IP 语音中应用 CoS

由于采用Internet承载语音流量较采用传统电信网络承载语音流量的成本要低，因此IP语音 (VoIP)正越来越普及。与语音应用相比，数据应用对延迟和丢失的容忍程度较高。数据应用可以处理某些重传和延迟，但仍提供可以接受的性能。而语音则不能很好地处理延迟、抖动和丢失，这是因为如果语音流不实时和不连续，则会导致通话不完整。

当在公共网络上同时承载语音和传统数据时，必须保证在拥塞过程中优先处理语音流量。通过AXI系列路由器的CoS特性，可以实现这种优先权分配，其中涉及的基本配置任务如下：

- 使用输出队列选择和优先位重写，把语音流量与尽力而为的流量分开；
- 把WRR配置成为包含语音的队列分配更高的优先权，保证在拥塞过程中先传输该队列中的分组。
- 把语音流量的缓冲器深度配置成较小的百分比，以强制消除交付陈旧分组的概率。

您可以通过两种方式把语音流量和尽力而为的流量分开。您可以把语音流量连接到与所有其它流量不同的一个单独接口上，也可以把语音流量的优先位设置成与所有其它流量不同的值。在实践中，同时采用这种方法可以达到最佳效果。

生成流量的应用负责设置优先位。当这种应用处于您的控制之中时，您可以管理优先级设置。在这种情况下，不要把优先位设置成 111 或 110，因为这些设置用于协议控制流量，在默认状态下放在队列 3 中。此外，应选择带有不等于 1 的 LSB 的优先级字段值，因为您希望把这些设置留给配置的限速功能设置了 PLP 的流量。因此，可用值是 000、010 和 100。如果按优先级字段分隔流量，那么必需使用一种机制，保证尽力而为的流量采用的优先级设置不同于语音流量。如果尽力而为的流量位于单独的进入接口上，您可以使用 AXI 系列路由器的 CoS 特性的优先级重写功能。如果使用的是同一个物理接口，那么您必须采用外部机制。基于这种原因，最好把语音网关放在与尽力而为的流量不同的进入接口上。

VoIP 配置实例

本配置实例说明了可以怎样使用AXI系列路由器的CoS特性提供VoIP。第一步是把语音流量与尽力而为的流量分开。

在本例中，我们使用下述参数：

- 语音流量的优先级字段设置为000。
- 所有尽力而为的流量使用的进入接口为t3-1/0/0。

注意：本例中的取值不一定对每个网络都是最好的，建议进行独立测试，以优化配置。

1. 把语音放在队列0中，把尽力而为的流量放在队列1中。

- A. 针对每种可能的优先级设置，明确配置流量应放入的队列。
- B. 把网络协议管理流量明确配置成放入队列3中，这是默认状态。
- C. 把尽力而为的流量明确配置成放入队列1中。

```
[edit class-of-service input]
  precedence-map separate-traffic {
    bits 000 queue 0;
    bits 001 queue 1;
    bits 010 queue 1;
    bits 011 queue 1;
    bits 100 queue 1;
    bits 101 queue 1;
    bits 110 queue 3;
    bits 111 queue 3;
  }
```

- 2. 把进入t3-1/0/0接口的所有流量都分配给尽力而为的队列1。

```
[edit class-of-service input]
  interfaces {
    t3-1/0/0 {
      unit 0 {
        output-queue 1;
      }
    }
  }
```

- 3. 通过在进入网络核心的每个接口上，把尽力而为的流量的优先位重写为010，把语音流量与尽力而为的流量分开。例如，如果这是一台通过一对千兆位以太网接口连接到核心路由器的接入路由器，那么您应对两个千兆位以太网接口同时应用这种配置。这种重写是必要的，这样，下行路由器就不会把尽力而为的流量放在语音队列中。

```
[edit]
class-of-service {
  output {
    interfaces {
      ge-1/0/0 {
        unit 0 {
          precedence-rewrite {
            output-queue 1 {
              plp-clear rewrite-bits 010;
              plp-set rewrite-bits 010;
            }
          }
        }
      }
    }
  }
}
```

```

    }
  }
}
ge-1/1/0 {
  unit 0 {
    precedence-rewrite {
      output-queue 1 {
        plp-clear rewrite-bits 010;
        plp-set rewrite-bits 010;
      }
    }
  }
}
}
}
}

```

4. 把出局队列上的WRR配置成为语音分配的可用带宽百分比高于尽力而为的队列。在所有网络接口上应用这种配置，保证一致的行为。注意，队列3是网络协议控制流量队列，默认状态是分配5%的可用带宽。建议保持这一配置。

```

[edit class-of-service output]
interfaces {
  ge-1/0/0 {
    weighted-round-robin {
      output-queue 0 weight 70;
      output-queue 1 weight 25;
      output-queue 2 weight 0;
      output-queue 3 weight 5;
    }
  }
}

```

5. 把RED配置成管理网络内部的拥塞。在可能发生拥塞的任何链路上配置RED。这样，您通常可以配置把POP连接起来的所有远距离链路，因为拥塞通常不会发生在一个枢纽站点内。

在配置RED时，记住以下要点：

- 必需为队列3分配5%的总缓冲器空间，队列3是网络协议控制流量队列。
- 缓冲器深度直接关系到总体延迟。对实时应用，如 VoIP，应较尽力而为的队列分配更低的缓冲器百分比。
- 不管在入口是否使用了策略管理来标记超过门限的分组，都要为 PLP 设置和 PLP 清除这两种情况配置一个成对的满时丢弃概率表。

```

[edit class-of-service output]
drop-profile red-conf {

```

```
stream-profile {
    fill-level 60 drop-probability 50;
    fill-level 80 drop-probability 75;
    fill-level 95 drop-probability 100;
}
plp-set-queue-profile {
    fill-level 60 drop-probability 70;
    fill-level 80 drop-probability 90;
    fill-level 95 drop-probability 100;
}
plp-clear-queue-profile {
    fill-level 60 drop-probability 40;
    fill-level 80 drop-probability 60;
    fill-level 95 drop-probability 100;
}
}
fpc 1 {
    drop-profile red-conf;
}
interfaces {
    so-2/0/0 {
        unit 0;
        transmit-queues {
            output-queue 0 buffer-percentage 35;
            output-queue 1 buffer-percentage 60;
            output-queue 2 buffer-percentage 0;
            output-queue 3 buffer-percentage 5;
        }
    }
}
}
```

队列管理

通过使用防火墙过滤器，可以监视队列利用率。如果您根据优先级设置把流量专门分配给一个队列，那么可以计算与这一优先级设置相符的分组数或字节数。出于许多原因，可能都需要这种计数，包括计费 and 容量规划目的。

例如，把优先级设置为000的所有分组放在出局队列0中。

```
[edit class-of-service input]
  precedence-map separate-traffic {
    bits 000 queue 0;
    bits 001 queue 1;
    bits 010 queue 1;
```

```
bits 011 queue 1;
bits 100 queue 1;
    bits 101 queue 1;
bits 110 queue 3;
bits 111 queue 3;
}
```

对接口的出局一边应用防火墙过滤器，以计算队列0服务的分组数量。

```
[edit firewall filter count-packets-in-q1]
term a {
  from {
    precedence 001 000;
  }
  then {
    count q0-counter;
    accept;
  }
}
```

结论

AXI 系列路由器的 CoS 特性允许您进一步配置更低的等级，同时在客户需要时，尽最大可能保证客户获得约定的带宽。通过区分对待流量，您还可以在拥塞时优先处理特级服务。这种分层的等级方法保证了您能够提供特级服务，满足服务水平协议，并向客户收取相应的费用。

缩略语

ASIC	专用集成电路
CoS	服务等级
DS	DiffServ
FIFO	先进先出
IETF	Internet工程任务小组
IP	网际协议
LSB	有效性最低的位
LSP	标记交换路径
MPLS	多协议标记交换
PLP	分组丢弃优先权
RED	随机早期检测
RFC	请求评论
RSVP	资源预留协议
TCP	传输控制协议
ToS	服务类型

VoIP IP语音
WRR 加权循环