

两比特区分服务实现策略研究

邓亚平 谭 强

(重庆邮电学院计算机科学与技术学院 重庆400065)

Two-Bit Diffserv Implementing Policy Study

DENG Ya-Ping TAN Qiang

(College of Computer Science & Technology, Chongqing University of Posts & Telecommunications, Chongqing 400065)

Abstract In this paper we discuss the Two-Bit Diffserv Architecture, study the RIO packet dropping policy and FMPA packet marking algorithm, develop the FMPA, compare the two FMPA By simulation on NS.

Keywords Diffserv, Dropping policy, Marking policy, QoS

1. 引言

区分服务是近两年的一个研究热点,其目的是在传统的IP分组交换网络上为用户提供较大粒度的服务质量保证。采用的实现手段是简化网络内部节点的服务机制和简化网络内部节点的服务对象。区分服务的基本思想是在网络边缘节点保存流状态信息,并根据用户的流规定和资源预留信息对流入分组进行分类、整形、聚合为不同的流聚集,并将处理结果信息保存在IP包头的DS字段中,称为DS标记(DSCP)。内部节点根据包头的DSCP来选择调度策略来提供相应的服务,其外特性称为逐跳行为PHB。内部节点与状态无关。

目前IETF已经定义为Diffserv定义了三种类型的PHB:尽量做好型BE、确保型AF、加速型EF。另外还有大量的PHB类型处于讨论中,并且这些大多是处于概念层次的讨论之中,主要是因为Diffserv是一个处于开放状态的课题。但是也提出了一些建议,这些建议利用Diffserv概念,说明网络如何利用一些具体的策略实现一些具体的服务。两比特(Two-Bit)区分服务体系结构就是其中一种。下面我们就两比特区分服务及其实现策略,主要对分组丢弃和分组标记策略进行研究,对文[1]中提出的改进分组标记算法作进一步改进,并以文[1]中的仿真环境进行仿真实验。

2. 两比特区分服务体系结构

两比特区分服务实际上是在Diffserv的优质服务PS和确保服务AS的基础上,将两者统一在一个框架中而提出来的。对PS和AS的服务导致了PF和EF的产生。

优质服务是目前定义的最高服务级别的区分服务种类,为用户提供低延迟、低抖动、低丢失率、保证带宽的端到端或网络边界到网络边界的服务。由于延迟、抖动、丢失主要是由于分组在传输途中排队导致。因此“三低一保证”实际意味着业务流在传输过程中几乎不排队。因此PS服务的实现上要保证在每一个节点处PS分组的输入速率要小于输出速率。在网络边缘,必须对PS流进行调节以保证其符合约定的流规定,不超过额定最高速率,传输路径上也要以最高速率进行资源预留。

确保服务是这样的:在网络拥塞的情况下,确保服务数据

流用户能够拥有一定量的预约带宽;在网络负荷轻,有空闲资源时,可以为用户分配额外的带宽。用户实际使用的带宽是预约带宽和与其它流竞争网络空闲带宽所得之和。AS的实现思路是:分组进入网络时,在边界节点给数据包做标记,预约带宽内的流标记为“in”,超出的标记为“out”;拥塞时,包头的标记决定分组的丢失率,out的丢失概率高于in的丢失概率,从而在一定程度上保护in流。中间节点在调度转发时,要保证源头相同的数据流不乱序,不管分组是in或out。

两比特区分服务技术只支持三个级别的服务:BE服务、AS服务、PS服务。在分组头用两比特来指示业务类型,两比特是P比特和A比特。用“00”标记BE类型流,“10”标记PS类型流,“01”标记AS类型流。优质服务级别是为传输适时流设计,确保流是为用户提供统计服务质量,BE服务是为了兼容传统IP流。

在边缘节点,分组被分类处理的方式如图1所示。优质服务流通过一个漏斗桶进行严格的整形,使数据流完全符合流规定,并把数据包标记上“10”。确保服务流通过一个令牌桶进行调节,把符合流规定的数据包标记为“01”,不符合流规定的数据包标记为“00”,当然也可以用基于时间滑动窗口机制来估计分组到达的速率。文[2]指出,采用令牌桶实现标记的方法,对TCP流来说,有比基于时间滑动窗口机制来估计流速率有更好的平滑突发分组的性能。

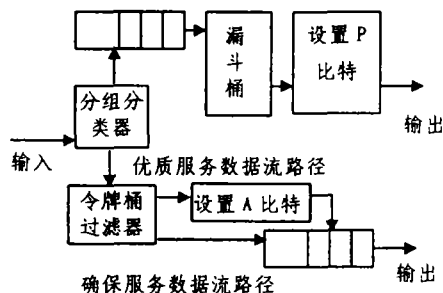


图1 Two-Bit 区分服务边缘路由器结构

在核心节点处,分组以图2所示的方式进行处理。根据分组的业务类型标记将分组分为两类:一类是优质服务流,一类是确保服务流和BE服务流。PS服务流有一个缓冲空间,

邓亚平 教授,主要研究方向为计算机网络与通信。谭 强 硕士生,研究方向为计算机网络。

主要是因为当一个 PS 数据包到达时,可能有其它分组正在占用输出链路,或多个 PS 流同时到达某一个核心节点而导致的分组丢失。对于 AS 流和 BE 流,首先采用 RIO 丢弃算法,避免网络拥塞发生,保证网络性能。

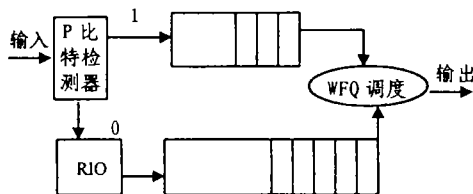


图2 Two-Bit 区分服务核心路由器

3. 两比特区分服务的丢弃策略

两比特区分服务推荐采用RIO分组丢弃算法,是RED丢弃算法的扩展。RED算法叫随机提前检测算法,最初是为了避免网络拥塞改善TCP性能而提出。其基本思想是为路由器设置两个缓存门限参数low和high,把拥塞控制分为三个阶段:正常传输、拥塞避免和拥塞控制。RED计算平均缓存长度avg,如果avg小于low,为正常运行阶段;如果avg大于low而小于high,进入拥塞避免阶段,路由器就以概率 $F(avg)$ ($F(avg)$ 随着avg增大而增大)丢弃接收到的每一个分组;如果avg继续增长并超过了high,则进入拥塞控制阶段,路由器丢弃每一个新接收到的分组。

在RED的实现中,利用指数加权移动平均值来计算平均队列长度,保证突发业务量或瞬时拥塞引起的队列长度不会导致平均队列长度的显著增长,平滑突发流量。选择恰当队列权值,如果权值太大,则平滑突发流量的能力太差,如果太小,将会导致路由器不能及时发现拥塞的出现,通常取值为0.002。low和high的取值决定于网络的最佳平均队列长度。对于典型的突发型业务网络,low值不能太小,否则导致链路的利用率太低,high一般应至少大于low两倍。

RIO算法对处于同一队列中的两个级别的数据流实施两种不同的服务,其中一个数据流中的分组比另一个流中的分组有更高的丢弃率。RIO可以看成是两个有不同丢弃曲线的RED的组合。其参数表示为(low-out,high-out,maxp-out)、(low-in,high-in,maxp-in),分别表示两个不同的丢弃曲线。

对于out分组,如果队列长度小于low-out,则不丢弃分组;但如果平均队列长度大于low-out,则对到达分组以某一概率进行丢弃,该概率根据平均队列长度大小从0到maxp-out线性增长;如果队列长度大于high-out,则丢弃所有到达的out分组。平均队列大小是基于队列中所有的分组数目,而不管其标记如何。

对于in分组,则参考(low-in,high-in,maxp-in)参数执行RED操作。平均队列长度是基于队列中的in分组的数目,并且参数设置也不同,保证首先丢弃out分组,然后丢弃in分组。

如果只有out分组,则RIO退化为RED。out分组的RIO参数设置原则同RED一致。在文[2]仿真表明,in参数和out参数不必差异太大。就基于平均队列长度的计算方法的差异已经足够对in分组提供一定的服务质量。

4. 两比特区分服务的分组标记策略

标记工作可在两个地方完成,一个是集成在用户端,另一

个集成在边界路由器中,分别称为源集成和源透明集成的标记策略。源集成的方式一般用于数据源端系统中,这样用户可以实现一些管理策略,比如将某些重要应用的数据都标记为in;透明的方式一般用于用户域边界和ISP网络边界处。源透明标记是研究的重点。并且PS服务流采用的是漏斗整形算法,资源是按峰值行进行预留,只是需要确定漏斗桶算法的参数,几乎没有更多的策略研究必要。由于AS数据流可能分为带内数据和带外数据两部分,因此对AS流分组标记的策略有进行研究的必要。两比特区分服务中对分组标记策略的研究主要是针对AS流中的分组。

分组标记通常有两种方法:一种是令牌桶标记方法,边界节点以固定的速率产生令牌,对于每一个到达的分组,如果令牌桶有足够的令牌,则可以把该分组标记为in,否则标记该分组为out;另一种方法是采用基于时间滑动窗口机制来估计分组到达的速率。当该速率超过规定的阈值时,边界节点按照某一概率把到达的分组标记为out,否则标记为in。

现有的分组标记算法的基本思想可以用下面的方程(1)表示:

$$P_{out}(t) = (V_{AG}(t) - V_{RESV}) / V_{AG}(t) \quad (1)$$

其中 $P_{out}(t)$ 表示时刻t分组被标记为out的概率, $V_{AG}(t)$ 为t时刻聚集流分组到达边界节点的平均速率, V_{RESV} 为预留带宽。这种标记算法是等比例地标记分组,即不考虑分组的数据源,只要分组到达速率高于规定的阈值,就按照某一相同的概率 $P_{out}(t)$ 把到达的分组标记为out。这样将使某些具有带宽嗜好的数据源得到更好的服务,而遵守规矩的数据源则无法得到应有的服务。为了维护Diffserv中各数据流的公平性,保证AS流的服务质量,我们需要一种新的标记算法。

文[1]提出了一个新的分组算法并对算法进行了仿真。如图3所示,用户域向ISP域申请预留接入带宽为B,M为标记器,R2处的标记任务是对来自用户域的分组进行监控,以维持流入的in分组符合约定。用户域有n个数据源,R1分配给各数据源的预留带宽为 B_i 。各数据源在t时刻的实际发送速率为 $V_i(t)$, $V_{AG}(t)$ 为经R1的聚集速率,有式(2)成立:

$$V_{AG}(t) = \sum_{i=1}^n V_i(t) \quad (2)$$

文[1]简称提出的分组标记算法为FMPA。FMPA算法如下:

```

For 每一个到达的分组,计算聚集流平均速率  $V_{AG}(t)$ 
If ( $V_{AG}(t) < B$ )
    标记该分组为 in
else
    计算该分组所属数据源的平均速率  $V_i(t)$ 
    if ( $V_i(t) > B_i$ )
        计算该分组被标记为 out 的概率  $P_{out-i}(t)$ 
        以  $P_{out-i}(t)$  为概率标记该分组为 out
    else
        标记该分组为 in
    Endif
Endif
End

```

算法采用滑动时间窗口来估计各数据源的发送速率 $V_i(t)$, $P_{out-i}(t)$ 由式(3)计算:

$$P_{out-i}(t) = (V_i(t) - B_i) / V_i(t) \quad (3)$$

当数据源i的发送速率超过预留带宽并且聚集流速率超过阈值时,算法并没有立即以概率1把该数据源随后的分组标记为out,而是以概率 $P_{out-i}(t)$ 标记随后的分组,可以防止出现分组连续被标记为out的情况。

文[1]在NS上用等比例标记算法和FMPA算法进行仿真比较,图4为仿真拓扑。 S_i 和 D_i 分别表示源和目的端;路由器R1和R2之间带宽为33Mbps,预留8Mbps用于标记in分

组;R2和 R3之间带宽为10Mbps,R1分配给各数据源 Si 的预留为2Mbps;各路由器均采用 RIO 分组丢弃策略,in 分组的RIO 参数为(60,100,0.02),out 分组的RIO 参数为(30,60,0.2);S1~S4分别以固定的速率向 D1~D4发送数据,选用了3组数据,如表1。

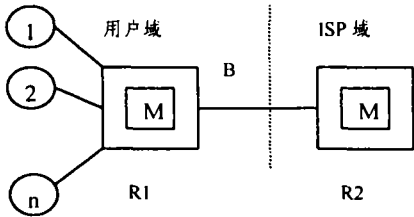


图3 边界节点的分组标记

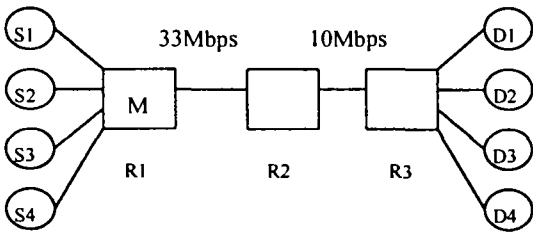


图4 仿真拓扑

表1 仿真数据(单位:Mbps)

	S1	S2	S3	S4
第一组	1	1	2	2
第二组	1	2	3	4
第三组	1	3	5	7

结果表明,因为 R2、R3之间的链路带宽,使用第1、2组数据进行仿真,无论采用哪种标记算法,均没有分组丢弃,包括out 分组,用第3组数据进行仿真时,有部分分组被丢弃。从第2、3组数据仿真结果表明,采用 FMPA 后,可以保证未超过预留带宽的数据没有被标记为 out 的,但标记为 in 的分组的总的吞吐量反而比等比例算法的要小。这是因为,对于超过预留带宽的数据流,其被标记为 in 的分组的吞吐量被限制为预留带宽;而没有超过预留带宽的数据源没有用完的带宽没有被其它数据源利用。

文[1]对 FMPA 的改进是:任何时刻 t,把所有数据源分为两类,一类为数据发送速率超过预留带宽 B_i 的数据源,一类为数据发送速率没有超过预留带宽 B_i 的数据源;V_i(t),U_i(t)分别表示前后两类数据的发送速率,k1和 k2分别表示两类数据源的个数;后一类数据源还可以继续利用的带宽为 B_{un}(t)。

$$B_{un}(t)=\sum_{i=1}^{k_2}(B_i-U_i(t)) \tag{4}$$

把 B_{un}(t)分配给前一类数据源,使前一类数据的 in 分组可以适当超过阈值。结合(3)式,得(5)式,在 FMPA 中,用式(5)来标记分组得到改进后的 FMPA。

$$P_{out-i}(t)=Max((V_i(t)-B_i-B_{un}(t)/k_1)/V_i(t),0) \tag{5}$$

仿真表明,对于前一类数据源,有的数据源分配到的额外带宽超过了它的需求,而有的数据源的带宽仍然不足。针对这种情况,我们提出了对改进了的 FMPA 算法的改进,简称再改进的 FMPA 算法。主要的策略是:将带宽 B_{un}(t)不再按数据源的个数平均分配,而是按各数据源超过预留带宽的数据

量的按比例分配 B_{un}(t)。前一类数据流分别超过的量为 V_i(t)-B_i,超过的总量为 B_{sn}(t)。

$$B_{sn}(t)=\sum_{i=1}^{k_1}(V_i(t)-B_i) \tag{6}$$

按比例分配给后一类数据源的带宽是:B_{un}(t) * (V_i(t)-B_i)/B_{sn}(t),带入(3)式,得(7)式:

$$P_{out-i}(t)=Max((V_i(t)-B_i-B_{un}(t) * (V_i(t)-B_i)/B_{sn}(t))/V_i(t),0) \tag{7}$$

用文[1]的仿真环境,用表2中数据进行仿真。仿真结果为表3和表4。

表2 仿真数据(单位:Mbps)

	S1	S2	S3	S4
第一组	1	1	2	2
第二组	1	2	2.1	4
第三组	1	3	5	7

表3 改进的 FMPA(文[1])算法仿真结果(单位:Mbps)

	S1	S2	S3	S4
第一组	1 0	1 0	2 0	2 0
第二组	1 0	2 0	2.1 0	2.5 1.5
第三组	1 0	2.33 0.17	2.33 0.67	2.33 1.17

表4 再改进的 FMPA 算法仿真结果(单位:Mbps)

	S1	S2	S3	S4
第一组	1 0	1 0	2 0	2 0
第二组	1 0	2 0	2.05 0.05	2.95 1.05
第三组	1 0	2.11 0.22	2.33 0.67	2.55 1.11

第2组数据表明再改进 FMPA 算法可以充分利用预留带宽,从2、3组数据结果表明,该算法还有一个特点:在保证预留带宽的情况下,数据量超过预留带宽越多,分配到的额外带宽越多,而且在带宽不足时,每一个数据量超过预留带宽的数据源都会受到“惩罚”。

总结 本文以两比特区分服务体系结构为基础,讨论了支持 PS、AS、BE 的区分服务网络的实现策略,重点讨论了分组丢弃和标记策略,具有实现简单,兼容传统 IP 服务好的特点。下一步工作是讨论 RIO 参数和标记策略的相互影响,并进一步提出整体的实现策略。

参考文献

1 马小俊,严峻,顾冠群.区分服务的分组标记策略研究.软件学报,2001,12(10):1472~1478
2 Preliminary Simulation Evaluation of an Assured Service. www. Google. com
3 林闯,单志广,盛立杰,吴建平. Internet 区分服务及其几个热点问题的研究. 计算机学报,2000,23(4)
4 Nichols K, Jacobson V, Zhang L. Atwo-bit differentiated services architecture for the Internet. [Technical Report]. RFC2638.1999
5 Clark D D, Fang W. Explicit allocation of best-effort packet delivery service. IEEE/ACM Transactions on Networking, 1998, 6(4): 372~373
6 Network simulator (NS). University of California at Berkeley, CA, 1997. http://www. nrg. ee. lbl. gov/ns.
7 Feng W, et al. Adaptive packet marking for providing diffrentiated services in the Internet. In: proc. of the ICMP'98. Austin, 1998. 108~117