一种新的 IP 分组丢弃策略 PRDP [1]

金明晔 黄永明 李乐民

(电子科技大学传输与宽带通信国家重点实验室 成都 610054)

摘 要 该文分析了 IP 业务在 MPLS/DiffServ 上运行的特点，从考虑业务量工程性能的角度出发提出了一种新的 IP 分组优先级的划分方法，并在此基础上给出了一种新的 IP 分组丢弃算法——基于优先级的丢弃策略 (PBDP)．考虑到随机早检测策略 (RED) 的优点，将 PBDP 与 RED 相结合，得到一种改进的算法——基于优先级的 RED 策略 (PRDP)，仿真结果表明这种算法在提高全网性能上有优势，其性能优于传统的“尾丢弃”策略．

关键词 MPLS, DiffServ, 优先级，丢弃策略

中图号 TN919.3

1 关于 IP 服务质量

业务量工程是一个可以带来高性能服务质量的重要方法．现在的 IP 网络中，仍然使用传统的标准的 IP 路由协议来传输业务，如 OSPF (Open Shortest Path First) 和 BGP(Border Gateway Protocol)．这些路由算法会将业务集中在某些连接和接口上，这样将导致某些连接或接口处的拥塞以及网络业务的不平衡分布，从而降低了网络资源利用率和网络的总体性能．由于对业务缺乏有效的控制，因此，对业务可能的处理以及网络的可能性能缺乏前瞻性，这是问题的症结所在．而现在的无连接网络不能够及时有效地反映网络的状态和性能，使得网络管理者很难采取有效的策略和措施来优化网络路由从而提高网络资源利用率和网络性能．因此，缺乏业务量工程的网络将没有能力或者很难为网络的运行和资源的利用提供有效的管理和控制，因而导致了较低的网络总体性能．为了更好地实现 IP 服务质量，必须研究一个基于业务量工程的运行机制．

目前，IETF(Internet Engineering Task Force) 提出了两种服务模型和机制来提供 IP 的服务质量，即“综合服务 / 资源预留协议” (InterServ/RSVP: Integrated Services/Resource Reservation Protocol) 和“区分服务 / 多协议标签交换” DiffServ/MPLS(Differentiated Services/Multi Protocol Label Switching) 模型．其中 DiffServ/MPLS 以其灵活的可扩展性、简单控制管理和高效的传输转发性能逐渐成为当前学术界研究的热点，并被认为是实现 IP 服务质量的较合适的结构 [1,2]．

在 DiffSrv/MPLS 结构下的 IP 网络是一个面向连接的网络．当有数据流到达时，建立一条“标签交换路由” (LSP: Label Switched Path)，该数据流将从这条 LSP 上被传输至目的节点．这样一个类似与电路交换网络中“直接路由” (direct path) 和“备用路由” (alternate path) 的概念就出现了．在传统的电路交换网络中，直接路由一般是指从初始节点到目的节点之间跳数最少的路由，而备用路由则是指其他有较多跳数的路由．一般来讲，在一个全连接网络模型中，直接路由指的是只有一跳的那些路由，而备用路由则是指有两跳的路由，三跳或三跳以上的路由由于性能较差不予考虑 [3]．一般而言一条直接路由都有一条或一条以上备用路由，这与网络的拓扑结构有关，如一个 5 节点全连接的网络，每个初始节点和目的节点之间都有一条直接路由和 3 条备用路由．当直接路由上业务负荷太重或是直接路由不能够再承载其他的业务时，网络将会从备用路由中选择一条来进行业务的传输 [4]．在 IP 网络中，情况则有所不同，而在

[1] 2001-01-02 收到，2001-08-24 定稿
国家自然科学基金资助 (69990540, 60002004)

DiffServ/MPLS 网络结构下，与传统的电路交换网络有相似之处．由于网络业务类型不同及网络的性能要求不同，研究的方法也将不同．

本文将在 DiffServ/MPLS 的网络结构下，研究支持多种类型业务的 IP 网络的分组丢弃策略，该策略将考虑分组传输时的路由特性 (即考虑其与业务量工程有关的性能)，并在实时和非实时业务情况下，对该策略进行仿真实验，并与传统的分组丢弃策略进行比较．本文的结构如下：第 2 节将简单回顾 IP 网络中分组丢弃算法，并且将提出 DiffServ/MPLS 结构下的 IP 业务模型；第 3 节将提出一种基于业务量工程的 IP 分组优先级划分方法，并在此基础上给出了基于优先级的 IP 分组丢弃策略—— PBDP(Priority-Based Discarding Policy)；第 4 节将介绍随机早检测 (RED, Random Early Detection) 与 PBDP 相结合后的改进了的 RED 策略，提出了适合 Premium 服务的分组丢弃算法——基于优先级的 RED 策略 (PRDP， Priority-based RED Discarding Policy)；仿真实验和性能分析将在第 5 节给出．最后是结论．

2 IP 网络的常规丢弃策略及基于 DiffServ/MPLS 的 IP 业务服务模型

在 DiffServ/MPLS 结构上目前已经提出了两种主要的分组丢弃策略，即尾丢弃策略 TDDP (Tail-Drop Discarding Policy) 和随机早检测策略 RED[1]．

TDDP 是指当先进先出队列 (FIFO: First In First Out) 满的时候，在队列尾部的分组首先被丢弃，称之为“尾丢弃”．由于在 TCP/IP 协议中， TCP 源是利用分组的丢失作为网络拥塞的暗示信号，并相应地降低相关的信息传输率．所以，尾丢弃策略能够避免队列发生溢出，队列也不会干涸，但它也同时导致网络的资源利用率降低以及网络业务通过量 (throughput) 的降低，此后 TCP 源检测到可用的网络容量并逐渐调高数据的发出，于是又会重新出现队列满，尾丢弃，降低数据发出率 ……，如此周而复始．这种低效率的循环称为“全局同步” (global synchronization)．解决的方法是采用一种称为 RED 的分组丢弃机制．其思想为：分组将以一定的概率随机被丢弃，丢弃的概率随着队列长度的增加而增加．这种基于概率的分组丢弃策略能够很好地引导 TCP 源调整其数据发送率，使得路由器里的队列既不溢出也不干涸，从而避免了尾丢弃所产生的现象．这使得路由器能够支持新的 TCP 连接，能处理周期性突发的数据，并且将网络利用率维持到一个比较高的水平．

尾丢弃算法是一种最基本分组丢弃算法，也是其他丢弃算法的基础，而 RED 策略只是从任何克服全局同步的角度进行考虑，这两者都没有从业务量工程的角度进行考虑，基于这种状况，我们先提出一种基于业务量工程考虑的丢弃算法—— PBDP 算法之后考虑到 RED 策略和 PBDP 的设计出发点不同，再综合两者的优点，改进 RED 算法，又提出了一种新的改进了的 RED 算法—— PRDP．

DiffServ/MPLS 结构提供三种基本服务类型： Best-effort 服务， Premium 服务及 Assured 服务 [5,6]．

Assured 服务为那些需要可靠服务的用户提供所需要的服务质量，即使在网络拥塞和繁忙时；Assured 服务的丢弃策略要靠 RIO(RED with In and Out) 来完成 (RIO 是一种改进的 RED 策略，通常它在运行时有两个 RED 算法同时进行，一个为“ IN ”分组服务，一个为“ OUT ”分组服务 [7]； Premium 服务为那些传送固定峰值速率业务的用户提供低时延以及低时延抖动的服务．这些业务包括 Internet 电话，视频会议，以及为虚拟专用网 (VPNs: Virtual Private Networks) 创建虚拟租用线路等等．但是 Premium 服务并没有在分组时延以及时延抖动上有特别的保证．在网络拥塞或是繁忙时，其相应的性能也会降低，但是它相对于低等级业务的性能总是受到保障．

由于 Premium 服务是今后实时业务的主要服务类型，为了找到一种适合 Premium 服务的分组丢弃算法以及考虑到非实时业务的影响，本文将重点研究同时支持 Premium 服务和 Best-effort 服务的网络协议.

3 优先级 IP 分组丢弃策略— PBDP

本节提出一种新的分组丢弃策略——一种基于业务量过程的优先级丢弃策略 PBDP . 众所周知，在 DiffServ/MPLS 结构下的 IP 网络是一个面向连接的网络，每个到达业务流都将被分配到不同的路由上 (有的是直接路由，有的是备用路由) . 由于直接路由的路由跳数最少，因而占用的网络资源较少 (如链路资源，路由器的缓存资源等) . 而备用路由则相对占用较多的资源. 当发生网络拥塞，需要进行分组丢弃的时候，被分配到备用路由上的分组将首先被丢弃，因为这样可以释放相对更多的网络资源，加快缓解网络拥塞的现象. 这样在进行分组丢弃时，不仅仅依赖分组本身的服务类型，而且还将根据其相应的路由特性对不同的分组进行区分.

当分组进入网络时，入口路由器 (ingress router) 将给每个分组插入一个 MPLS 头，在每个分组的 MPLS 头部 (即标签) 定义一个 2 bit 的 “分组优先级” DP(Discarding priority) 域. 2 bit 的 DP 域的高位比特为业务类型标识，“ 1 ” 表示 Premium 服务，“ 0 ” 表示 Best-effort 服务；低位比特为路由特性标识，“ 1 ” 表示直接路由，“ 0 ” 表示备用路由. 这样分组优先级标识 DPI(DP Identifier) 就有 4 种类型码： 11 ， 10 ， 01 ， 00 ，它们代表了不同的分组丢弃优先等级， DPI 的数值越大，其等级越高，越晚被丢弃. 在这种定义下，一种新的丢弃分组优先等级就被确定下来了，从最高等级到最低等级依次为：在直接路由上的 Premium 业务 (“ 11 ”)；在备用路由上的 Premium 业务 (“ 10 ”)；在直接路由上的 Best-effort 业务 (“ 01 ”) 及在备用路由上的 Best-effort 业务 (“ 00 ”) .

这样， PBDP 就应如下所述：当队列满时，新到达分组将在队列中找到比自己丢弃优先级低的分组或 DPI 低的分组，如果找到则将该分组丢弃，并将新的分组插入队列；如没有找到则丢弃该分组. 如果有多个这样的分组，将从中选择最近期进入队列的分组被丢弃. 其具体算法如下：

当队列不能容纳新到达的分组 i
　读取分组 i 的 DPI 值，即 DPI(i)；
　若 DPI(i)=0 ，则丢弃该分组；
　否则
　　找到一个分组集 X ，对所有 $x \in X$ ，有 DPI(x <DPI(i)；
　　若 $X \neq$ 空集
　　　找到 X 中最近到达队列的分组 k；
　　　丢弃分组 k ，插入分组 i；
　　否则，丢弃分组 i .

4 一种改进的基于优先级的 RED 丢弃策略 PRDP

RED 机制的提出是为了提高网络拥塞时的性能以避免 TCP 协议对流量控制的恶性周期重复. RED 机制能够有效控制队列的长度以及突发性业务的丢失率. 为了降低 TCP 协议对流量控制的干涉程度，并同时提高网络资源利用率，从而提高网络的总体性能，我们将 PBDP 与 RED 的优点结合起来，并进行改进形成了一个新的分组丢弃策略——基于优先级的 RED 分组丢弃策略 PRDP ，该策略是 RED 策略的改进.

4.1 RED 机制

RED 最早是由 Floyd 和 Jacobson 在文献 [8] 提出的。它的基本思想是：随着队列长度的增加，分组随机丢弃的概率也随之增加。其具体算法如下：

首先为队列定义两个门限：一个最小长度门限 $\min_{\text{th}}$ 和一个最大长度门限 $\max_{\text{th}}$。当队列长度 q 小于 $\min_{\text{th}}$ 时，不丢弃任何分组；当队列长度 q 大于 $\max_{\text{th}}$ 时，所有到达分组都将被丢弃；当队列长度 q 介于 $\min_{\text{th}}$ 和 $\max_{\text{th}}$ 之间时，分组以概率 p_b 随机被丢弃。p_b 被定义为

$$p_b = \frac{q - \min_{\text{th}}}{\max_{\text{th}} - \min_{\text{th}}}, \quad q > \min_{\text{th}} \tag{1}$$

RED 算法简述如下：

如分组 i 到达，
  得到当前队列长度 q；
  若 $\min_{\text{th}} < q < \max_{\text{th}}$
    计算出丢弃概率 p_b；
    以概率 p_b 丢弃分组 i；
  若 $\max_{\text{th}} \le q$
    丢弃分组 i；
  否则，将分级 i 插入队列。

4.2 PRDP 机制

我们借用文献 [9] 中采用的方法来定义最小长度门限 $\min_{\text{th}}$ 和最大长度门限 $\max_{\text{th}}$，该方法也是 RED 算法中计算丢弃概率最常用的一种方法。

假设路由器缓存里的队列最多能够容纳 B 个分组，则最大长度门限就是 B，最小长度门限为 B_{th}。则丢弃概率是队列长度的增函数，$\alpha\{0,\cdots,B\} \to [0,1]$，同时 $\alpha(0)=0$ 且 $\alpha(B)=1$。因此当队列长度 q 大于 B_{th} 时，

$$p_b = \alpha(q) = (q - B_{\text{th}})/(B - B_{\text{th}}) \tag{2}$$

这里 B_{th} 通常被设置为 $B/2$。

当队列长度 $q > B_{\text{th}}$，分组将以不同的概率被丢弃。不论丢弃概率如何，一旦丢弃发生，分组将会根据其 DPI 的值或丢弃优先级按从低到高的次序被丢弃，若有多个相同优先级的分组存在，则选择最近期进入队列的那个进行丢弃。这样 RED 算法就和 PBDP 策略吻合起来了。RED 算法将决定丢弃是否及何时发生，而 PBDP 将决定那个分组被丢弃以及如何被丢弃。这种新的丢弃算法被命名为 PRDP。其具体算法描述如下：

分组 i 到达
  得到当前队列长度 q；
  若 $\min_{\text{th}} < q < \max_{\text{th}}$
    计算丢弃概率 p_b；
    以概率 p_b 开始进行分组丢弃 {
      得到分组 i 的 DPI 值，即 DPI(i)；
      若 DPI(i)=00，丢弃分组 i；
      否则
        找到一个分组集 X，对所有 $x \in X$，有 DPI(x) < DPI(i)；
        若 $X \neq$ 空集
          找到 X 中最近到达队列的分组 k；

丢弃分组 k，插入分组 i；

否则，丢弃分组 i； }

若 $\max_{\text{th}} \leq q$

得到分组 i 的 DPI 值，即 DPI(i)；

若 DPI(i)=00，丢弃分组 i；

否则

找到一个分组集 X，对所有 $x \in X$，有 DPI(x) < DPI(i)；

若 $X \neq$ 空集

找到 X 中最近到达队列的分组 k；

丢弃分组 k，插入分组 i；

否则，丢弃分组 i；

否则，插入分组 i。

5 仿真实验及性能分析

我们在一个 5 个节点全连接的 IP 主干网模型上对实时与非实时两种业务进行仿真实验，实时业务被作为 Premium 服务处理，而非实时业务被作为 Best-effort 业务处理。实时业务为一个 On-Off 到达过程，On 和 Off 的持续时间分别服从均值为 m_{on} 和 m_{off} 的指数分布。在 Off 期间没有分组到达，在 ON 期间的分组以速率 v_c 到达。非实时业务以到达率 λ 服从 Poisson 分布。每个出端口对应一个缓存区，缓存区大小为 B，缓存区里的队列服从“固定优先级排队策略” (SPQM, Strict Priority Queuing Method)[9]。当 $B = 200$, $m_{\text{on}} = 1000$, $m_{\text{off}} = 20000$, $v_c = 100$ 时，比较 TDDP，PBDP，RED 和 PRDP。各项指标如图 1 所示。

图中的曲线显示 PBDP 相对于传统的丢弃算法能够明显提高网络的总体性能，特别是实时业务的性能。在 PBDP 下，两种业务的分组丢失率以及分组延时好于 TDDP，与 RED 接近，但略优于 RED。从而证明了：考虑分组的路由特性，并将备用路由上的分组先于直接路由的分组被丢弃，确实能够提高网络资源的利用率，从而达到降低分组丢失率以及分组延时的作用。而 PRDP 是所有算法中性能最好的一种，无论是对实时业务还是非实时业务。这说明将 RED 和 PBDP 两者结合起来形成的 PRDP 算法确实吸取了两者的优点，达到了更好的性能指标。同时发现，当网络不很拥塞时，非实时业务不是当前主要的业务类型，在 PRDP 和 PBDP 下，网络性能都维持在一个较好的程度。但是在网络负荷较大，网络明显拥塞时，PRDP 的性能明显比 PBDP 好。由于本实验所采用的 On-Off 模型来构造 Premium 服务类型的实时业务，其业务具有很强的突发性，而采用 Poisson 模型的非实时业务，其业务相对平滑，随着网络拥塞程度的增加，非实时业务在当前网络中所占的比例逐渐增大，这个结果再次证明了：RED 策略在突发性业务下能够很好地避免分组地丢失，同时这个特性随着平滑业务比例的增加而更为明显。Floyd 和 Thomas 已经在文献 [8, 10] 里证明了这个结论。

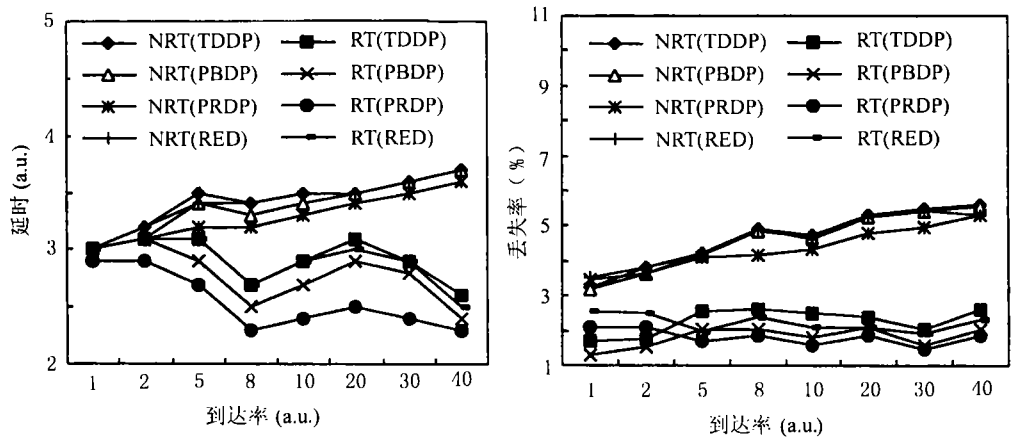

图 1 TDDP, PBDP, RED 和 PRDP 的延迟和丢失率比较

6 结 论

本文从分析实现 IP 服务质量的所用方法出发，论述了在 IP 网络上实现服务质量的两个主要方面：网络结构和网络运行机制．其中网络结构经 IETF 以及相关研究证明，DiffServ/MPLS 的结构是目前较为理想的一种方案．而为了更好地实现不同类型业务有不同的服务质量，业务量工程是网络运行时必须考虑的机制．有许多技术都从业务量工程的角度进行研究和设计．本文正是从此角度研究了 DiffServ/MPLS 结构下的 IP 分组丢弃策略．在考虑丢弃分组时，将分组按其业务类型以及路由特性进行优先级划分，优先级越高的分组越晚被丢弃，在此基础上提出了基于优先级的丢弃策略 (PBDP) ．实验证明 PBDP 较原来的尾丢弃策略能更好地提高网络总体性能，其性能略优于 RED ．考虑到 RED 和优先级策略的不同特点，为了把两者的优点更好地结合，最终提出了一种改进的 RED 算法——基于优先级的 RED 算法 (PRDP)，即 RED 策略用来决定是否及何时进行丢弃，而 PBDP 决定如何进行分组丢弃．经实验证明该算法相对于以上 3 种方法能够更好地提高网络资源利用率，并达到相对最好的网络性能．同时实验也再次证明了： RED 策略在突发性业务下能够很好地避免分组的丢失，同时这个特性将随着平滑业务比例的增加而更为明显．这也说明， PRDP 可以成为一种最适合 Premium 服务的分组丢弃策略，就像 RIO 确保 Assured 型业务的服务质量一样．

参 考 文 献

[1] D. Ferrari, L. Delgrossi, Charging for QoS, IEEE/IFIP IWQOS'98, Keynote Paper, Napa, CA, May 1998, 469–473.

[2] Francois Le Faucheur, *et al.*, MPLS Support of Differentiated Services, Internet draft, draft-ietf-mpls-diff-ext-05.txt, Jun 2000.

[3] G. Ash, Dynamic Routing in Telecommunications Networks, New York, McGraw-Hill, 1998, 439–440.

[4] K. W. Ross, Multiservices Loss Models for Broadband Telecommunication Networks, New York, Springer, 1996, 111–112.

[5] K. Nichols, V. Jacobson, L. Zhang, A Two-bit Differentiated Service Architecture for the Internet, Internet draft, draft-nichols-diff-svc-arch-00.txt, Nov, 1997.

[6] Y. Bernet, *et al.*, A Framework for Use of RSVP with Diff-serv Networks, Internet draft, draft-ietf-diffserv-rsvp-00.txt, June, 1998.

[7] D. Clark, J. Wroclawski, An Approach to Service Allocation in the Internet, Internet draft, draft-clark-different-svc-alloc-00.txt, July 1997.

[8] Sally Floyd, Van Jacobson, Random early detection gateways for congestion avoidance, IEEE/ACM, Trans. on Networking, 1993, 1(4), 397–413.

[9] L. Kleinrock, Queuing Systems, New York, International Spencification Publishing, 1976, chapter 8.

[10] T. Bonald, M. May, Drop Behavior of RED for Bursty and Smooth Traffic, IEEE IWQoS, 1999.

A NEW PRIORITY-BASED IP PACKETS DISCARDING POLICY

Jin Mingye Huang Yongming Li Lemin

(*Nat. Key Lab of Fiber Transm. and Broadband Comm. of UESTC, Chengdu* 610054, *China*)

Abstract Concerning about the feature of IP traffic, this paper propose a new Priority-Based Discarding Policy (PBDP) from the view of traffic engineering, and further more gives a advanced RED algorithm PRDP (Priority-based RED Discarding Policy), which combining the virtue of RED and PBDP. Considering real-time and non-real-time traffic, simulation shows the PBDP is better than traditional discarding policy, and the PRDP is the best of all.

Key words MPLS, DiffServ, Priority, Discarding policy, RED

金明晔： 女，1974 年生，博士生，目前的主要研究方向为：IP QoS 技术，MPLS 技术，WDM 光网结合技术，网络路由算法，网络体系结构等.

黄永明： 男，1965 年生，1994 年获美国麻萨诸塞大学电子与计算机工程系博士学位，香港城市大学电子工程系助教，IEEE 高级会员. 研究方向主要为：高速网络，视频点播、卫星通信以及动态路由技术.

李乐民： 男，1932 年生，博士生导师，中国工程院院士，研究方向为信息传输与通信网.