

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷
H04L 12/56



[12] 发 明 专 利 说 明 书

[21] ZL 专利号 01125587.0

[45] 授权公告日 2004 年 10 月 20 日

[11] 授权公告号 CN 1172491C

[22] 申请日 2001.8.16 [21] 申请号 01125587.0

[30] 优先权

[32] 2000.8.18 [33] EP [31] 00402306.5

[71] 专利权人 阿尔卡塔尔公司

地址 法国巴黎

[72] 发明人 奥玛·艾罗米 斯特凡·D·科诺德

审查员 朱少华

[74] 专利代理机构 北京市中咨律师事务所

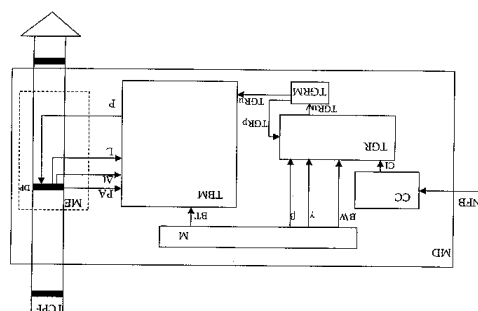
代理人 杨晓光 于 静

权利要求书 3 页 说明书 7 页 附图 1 页

[54] 发明名称 标记数据传输流的分组的方法和执行这个方法的标记设备

[57] 摘要

在一个标记涉及分组网络内部某个端到端连接的数据传输流的数据分组的方法中，根据至少一个属于数据传输流的服务质量参数和一个指示分组网络内部阻塞状态的网络反馈参数计算优先权数值。还描述了一个适于执行上述方法的标记设备。



I S S N 1 0 0 8 - 4 2 7 4

1. 一种标记设备(MD), 用于产生一个优先权(P)并将其插入一个属于一分组网络内部某个端到端连接的数据传输流(TCPF)的一个数据分组(DP), 其中所述优先权(P)是基于至少一个属于上述数据传输流(TCPF)的服务质量参数(BW, BT), 以及基于指示上述分组网络内部的阻塞状态的网络反馈参数(NFB)产生的, 其特征在于:

所述标记设备(MD)包括拥挤检查设备(CC), 用于接收所述网络反馈参数(NFB)并从中确定到令牌产生速率装置的触发信号(CI), 所述令牌产生速率装置包括在标记设备(MD)中并用于确定速率(TGR_u), 利用该速率, 在包括在标记设备(MD)中的令牌桶标记设备(TMB)中产生所述优先权(P), 所述令牌桶标记设备(TMB)被进一步送到包括在标记设备(MD)中的设备(ME), 用于将所述优先权(P)插入到所述数据分组(DP)。

2. 根据权利要求1的标记设备(MD), 其中, 所述用于将所述优先权(P)插入到所述数据分组的设备(ME)还包括用于接收所述数据传输流(TCPF)的装置, 用于从所述数据传输流(TCPF)的所述数据分组(DP)中提取出数据参数(PA, Δt , L), 并将所述数据参数发送给所述令牌桶标记设备(TMB), 所述令牌桶标记设备(TMB)进一步从所示数据分组参数(PA, Δt , L)中产生所述优先权。

3. 根据权利要求1或2的标记设备(MD), 其中所述令牌产生速率装置(TGR)还根据属于所述端到端连接的减少系数(γ)以及所述至少一个服务参数(BW)产生所述优先权(P)。

4. 根据权利要求1的标记设备(MD), 其中所述令牌产生速率装置(TGR)还根据属于所述端到端连接的增加系数(β)以及所述至少一个服务参数(BW)产生所述优先权(P)。

5. 根据权利要求1的标记设备(MD), 其中所述令牌桶标记设备(TMB)产生所述优先权(P)以便具有一个预定优先权数值集合(HI, LO)的一个数值。

6. 根据权利要求5的标记设备(MD), 其中所述令牌产生速率装置(TGR)根据所述减少系数(γ)减少具有所述集合的第一预定数值(HI)的第一类优

先权的产生速率。

7. 根据权利要求6的标记设备(MD), 其中所述令牌产生速率装置(TGR)根据上述增加系数(β)增加上述第一类优先权的产生速率。

8. 根据权利要求1的标记设备(MD), 其中所述至少一个服务质量参数包括一个一致带宽数值(BW)。

9. 根据权利要求1的标记设备(MD), 其中所述令牌桶标记设备(TMB)还基于另一个服务质量参数(BT)产生所述优先权。

10. 连接到一个分组网络并且适于将涉及所述分组网络内部的一个端到端连接的数据传输流(TCPF)的数据分组发送到至少一个与所述分组网络相连的其它终端的终端, 所述终端包含用于产生一个优先权(P)并且将其插入所述数据传输流(TCPF)的一个数据分组(DP)内的标记设备(MD), 其中所述标记设备(MD)还包括根据权利要求1, 2, 4-9中的任何一个的特征。

11. 连接到一个分组网络并且适于将涉及所述分组网络内部的一个端到端连接的数据传输流(TCPF)的数据分组发送到至少一个与所述分组网络相连的其它终端的终端, 所述终端包含用于产生一个优先权(P)并且将其插入所述数据传输流(TCPF)的一个数据分组(DP)内的标记设备(MD), 其中所述标记设备(MD)还包括根据权利要求3中的特征。

12. 一种分组网络接入路由器装置, 所述接入路由器装置适于从一个与上述接入路由器相连的终端接收数据分组, 上述数据分组涉及上述分组网络内部一个介于上述终端和另一个与上述分组网络相连的终端之间的端到端连接, 上述接入路由器装置包含用于产生一个优先权(P)并且将其插入属于所述端到端连接的数据传输流(TCPF)的一个数据分组(DP)内的标记设备(MD), 其中所述标记设备(MD)还包括根据权利要求1, 2, 4-9中的任何一个的特征。

13. 一种分组网络接入路由器装置, 所述接入路由器装置适于从一个与上述接入路由器相连的终端接收数据分组, 上述数据分组涉及上述分组网络内部一个介于上述终端和另一个与上述分组网络相连的终端之间的端到端连接, 上述接入路由器装置包含用于产生一个优先权(P)并且将其插入属于所述端到端连接的数据传输流(TCPF)的一个数据分组(DP)内的标记设备

(MD)，其中所述标记设备 (MD) 还包括根据权利要求3的特征。

标记数据传输流的分组的方法和执行这个方法的标记设备

技术领域

本发明涉及标记设备，用于产生一个优先权并将其插入属于一分组网络内部某个端到端连接的数据传输流的一个数据分组。

背景技术

在作为分组网络的一个例子的因特网中，标记设备结合路由器中基于数据分组优先权的丢弃机制和 TCP/IP 协议的内在性质允许根据其服务质量和/或传输参数区别出若干传输流。在进行实际的数据传输之前，这些服务质量参数通常在用户和因特网服务提供商之间得到同意，此后因特网服务提供商简称为 ISP。

在题为“Method and apparatus for adjustment of TCP sliding window with information about network conditions”的国际申请 WO00/02395 中，提到了用于产生优先权的一个特殊机构。该现有技术文献的第 8 页描述了每个数据分组都有一个有优先级，它是根据一个服务质量参数确定的，例如：分配给源的标称比特速率。现有技术网络还提供优先级级反馈，其中一个典型的接受的优先级级作为源节点的输入参数，它据此借助于该现有技术文献的其余部分所描述的 TCP 传输窗口尺寸方法调整其发送速率。作为调整源传输速率的结果，数据分组被以不同的速率发送过来又控制优先权的计算，因为这是根据源的瞬时实际比特速率与源的标称比特速率的比率确定的，如在该现有技术文献第 8 页第 4 行所述的那样。

该优先权确定或标记方法的缺点是它涉及一个间接方法，即首先控制源的发送速率，该更新的数据然后确定新的优先权。这种方法复杂而且慢，因为在根据网络反馈参数实际调整优先权值之前需要一定的时间。

发明概述

因此，本发明的目的是提供一种上述类型的标记设备，但它能允许根据网络反馈参数以更快和更简单的方式调整优先权。

根据本发明，该目的是通过这样的标记设备实现的，该标记设备的特征在于包括拥挤检查设备，用于接收所述网络反馈参数(NFB)并从中确定到令牌产生速率装置的触发信号，所述令牌产生速率装置包括在标记设备中并用于确定速率，利用该速率，在包括在标记设备中的令牌桶标记中产生所述优先权，所述令牌桶标记被进一步送到包括在标记设备中的设备，用于将所述优先权插入到所述数据分组。

这样，网络反馈参数被实际输入到一个拥挤检查设备，它据此确定作为到令牌产生速率装置的触发信号的输出信号，它进一步控制令牌桶标记。后一设备实际产生优先权并将其送到另一设备，该设备将其插入相应的数据分组。这样就可以预见到一种直接和简单的方法，该方法不在依赖源的速率调整用于响应网络反馈参数调整优先权。

通过考虑属于端到端连接自身的其它参数，例如：减小系数和增加系数，获得一个更好的关于端到端连接自身的级别划分，其中这些参数与已经提到并使用的服务质量参数和指示阻塞的网络反馈参数不同。当在试图发送关于单个TCP流的数据的用户和ISP之间就服务级别达成一致时，可以协商这些其它参数。

根据本发明，这些其它参数中属于端到端连接的参数被用于影响产生优先权的速率。例如减小系数被用于减少产生第一类优先权，例如高优先权的速率。通过所述的阻塞信息和这个属于端到端连接的减少参数来减少产生高优先权的速率，获得了更快速并且得到区分的流减少。类似地，增加参数被用于在非阻塞期间再次增加速率，其中按照上述这率产生这个第一类优先权，例如高优先权。再次区分其它端到端连接。

由于因特网中的端到端连接通常与不同终端相关，因而获得的标记机制也允许在考虑阻塞信息的情况下区分终端。

根据本发明，将已达成一致的一个带宽数值和一个脉冲串容差数值用作服务质量参数允许与诸如令牌桶机制的现有标记机制兼容。

本发明还涉及引入这种标记设备的终端和边缘路由器。

附图说明

参照下列结合附图对一个实施例的描述可以更好地理解本发明的上述和其它目标与特性以及本发明自身，其中图例表示基于本发明的标记设备 MD 的一个实施例。

具体实施方式

本发明的标记设备 MD 被用于诸如因特网的分组网络，目的是区别服务。区别服务意味着支持若干服务或服务质量类，其中在各个数据分组的 IP 报头内部可以将一个命名为服务类型并且简称为 ToS 字段的特定字段设置成特定数值。这允许彼此区分不同类型的数据分组或属于一个特定连接的数据分组，使得可以以不同方式传递这些数据分组。对于标准化 DiffServ 模型，这个 ToS 字段被重命名为 DiffServ 代码点并且简称为 DSCP 以便指示数据分组的服务质量要求。各个数据分组包含其自身的服务参数，上述参数部分确定了其各个转发段或路由器中应当采取的处理方式。另外还根据这个 DSCP 字段中的另一个位组确定这种路由器或转发段中处理具有不同服务参数的不同数据分组的方式，其中这个位组表示一个优先权值。在标记设备内部完成数据分组中优先权的归属和插入，其中标记设备位于数据分组发送终端自身或数据分组网络的路由器内。根据其在因特网内的位置可以在这些路由器中间区分接入或边缘路由器与核心路由器。因而接入路由器被直接连接到用户终端并且位于用户终端和核心路由器之间。

本发明涉及数据分组的初始标记过程，这个过程发生在数据发送终端自身或与这个数据发送终端相连的接入路由器内。被放入核心路由器并且适于调整入局分组的标记值的标记设备不是本发明的主题。

这些优先权值与其服务质量类相结合在路由器内部被用于确定所谓的逐转发段行为。通常通过也被称为活跃队列管理机制的缓冲区接受机制和调度程序实现这些逐转发段行为，其中通过上述机制和调度程序根据其优

先权，服务质量和涉及队列填充的其它参数对涉及若干服务类的分组进行丢弃或排队。这些缓冲区接受机制中的一个简称为 RED 算法的所谓随机早期检测机制。在文献中可以找到有关这种机制的大量描述，例如 S. Floyd 和 V. Jacobson 的文章“用于阻塞回避的随机早期检测网关”，IEEE/ACM 网络学报，V.1 N.4，1993 年 8 月，p397-413。由于这个算法可以可本发明的标记设备配合用于获得服务水平级别，所以在下面段落中仍然简要解释了这个算法。

RED 是一个活跃队列管理方案，该方案尝试在保持较小的平均队列长度的同时保持较高的总体吞吐量。RED 具有 4 个可配置参数。存在两个缓冲区阈值，之间会发生概率性分组丢弃。当平均队列占有率低于最小阈值时，不丢弃分组。当平均队列占有率超过最小阈值时，根据一个增加的概率丢弃分组。当平均队列尺寸超过最大阈值时，丢弃所有的到达分组。为了区分服务级别，也使用直接控制平均队列长度的参数。思路是具有 n 组阈值，一个最大丢弃概率和一个权重，其中 n 最小为 2 并且涉及在标记分组时使用的优先权数量。在指定一个平均队列长度的情况下参数数值的差异会导致不同的丢弃概率。

本发明的标记机制与一个丢弃机制结合使用，这个丢弃机制具有至少两个但通常与优先权同样多的不同丢弃概率曲线，这些丢弃概率曲线用于丢弃涉及服务类的队列中具有不同优先权的分组。现在讨论本发明的标记机制以及标记设备。

图中示出了一个标记设备 MD。它包含一个存储器 M，该存储器存储诸如一致带宽 BW，一致脉冲串容差数值 BT，和两个也属于所考虑的端到端连接的参数 β 与 γ 的服务质量参数数值。这个因特网中的端到端连接包括一个 TCP 连接，对应的 TCP 流被记为 TCPF 并且由进入和离开标记设备的粗箭头表示。所有 4 个参数 BW，BT， β 与 γ 均得到用户和 ISP 的同意，并且可以被认为是用于某个具体端到端连接的标记设备的操作的静态参数。然而可以定期对其进行更新，在进行更新时新的数值将被存储在 M 内。

其中若干分组被表示成黑矩形的 TCP 流进入一个设备 ME 内的标记

设备。设备 ME 适于在新数据分组 DP 到达时为另一个设备 TBM 指定一个触发器 PA 以便提取这个数据分组的字节长度 L 并且计算以前分组到达和当前到达之间的时间 Δt 。这个数值和分组长度 L 也被交付给设备 TBM，下面会描述设备 TBM 的操作。

标记设备 MD 还从分组网络接收具有网络反馈参数 NFB 的形式的信息，其中包含涉及网络内部阻塞状态的信息。这个参数可以涉及端到端连接上的阻塞状态，并且可以包括 TCP 双重或正常确认信号。网络内阻塞信息的另一个例子在网络具有 ECN 能力的情况下可以包括 ECN 位。ECN 是显式阻塞通知的缩写，在当前的文献中也描述了这个机制，例如在 K. K. Ramakrishnan 和 S. Floyd 的文幸“向 IP 增加显式阻塞通知(ECN)的建议”，RFC 2481。然而这个网络反馈参数 NFB 也可以包含涉及一个核心路由器内部的阻塞状态，由这个核心路由器自身产生并发送的信息。在后一种情况下由若干被阻塞的核心路由器产生的信号可以进入标记设备。

下面会描述标记设备的一个适于接收各个发送分组的 TCP 连接的正常和双重确认的实施例。当标记设备接收到这些信号中的一个时，在一个阻塞检查设备 CC 内对其进行检查以发现阻塞。根据这个检查的结果，这个阻塞检查设备的输出信号 CI 会指示是否有阻塞。这个信号是令牌产生速率设备 TGR 的一个输入信号，这个设备根据上述信号，和一致带宽 BW，涉及所考虑的 TCP 连接的参数 β 与 γ 的值，以及一个令牌产生速率的以前或起始值导出这个令牌产生速率的一个更新数值 TGR_u。这个更新的数值被临时存储在一个令牌产生速率存储器 TGRM 内。当一个新的网络反馈信号 NFB 到达时，更新 CI 并且触发 IGR 计算 TGR_u 的一个新值，因而 NFB 信号的到达触发或激活了 IGR。

为了区分在 TGR 内部使用和产生的令牌产生速率变量的以前数值 TGR_p 与更新数值 TGR_u，旧的数值 TGR_p 被表示成一个与 TGR_u 不同的信号。

当一个新数据分组 DP 从 TCP 流 TCPF 到达时，也从 TGRM 请求更新数值 TGR_u 并且由一个令牌桶标记设备 TBM 来使用。如本文以前的段

落所述，由信号 PA 触发这个到达。这个令牌桶标记设备根据 TGR_u ，分组长度 L 和上述 Δt 产生一个优先权 P 。这个优先权或标记被回送到设备 ME，设备 ME 将其插入最近检测到达的数据分组 DP 中。应当注意，当到达数据的传输速度最小时进行这个优先权产生，因而分组缓冲被减少到最小。

下面描述一个计算令牌产生速率和优先权的可能算法。

令牌产生速率的初始值被设置成 BW 。

当一个诸如 TCP 确认或双重确认的网络反馈信号到达时，在 CC 内部执行一个检查以确定是否存在阻塞。因而产生变量 CI 的一个更新值。

在无阻塞的情况下，通过将其以前的数值 TGR_p 与 $(P+1)$ 相乘来计算令牌产生速率的更新数值 TGR_u ，其中 β 是一个大于零的正数。所以 β 也被称作一个增加系数。

在有阻塞的情况下，通过将以前的数值 TGR_p 与 γ 相乘来计算令牌产生速率的更新数值 TGR_u ，其中 γ 大于零并且小于一。所以 γ 也被称作减少系数。

当一个数据分组 DP 到达时，PA 信号触发 TBM。接着根据下列等式更新一个内部变量，即令牌计数 TC：

$$TC \leftarrow \min(TC + TGR_u \cdot \Delta t, BT)$$

其中 TGR_u 表示 TBM 从存储器 TGRM 请求的数值。

接着将 TC 与 L 相比较。在 TC 低于 L 的情况下，优先权 P 被设置成低优先权。在 TC 超过或等于 L 的情况下，TC 的值被减 L ，并且优先权被设置成高优先权。

注意 TC 的初始值被设置成 BT。

上述实施例和算法使用了两个优先权数值，即高或低。为了在一个服务区分方法内使用这个标记算法，路由器内的逐转发段行为需要符合两个优先权，因而需要提供两个不同的丢弃行为，每个优先权一个。因而为此可以使用上述 2-RED 方法和代表 RED 入和出算法并且也在本文中描述的 RIO 方法。

然而应当注意，基于本发明的标记方法和标记设备可以有超过两个的优先权。对于 n 个优先权，路由器内部的逐转发段行为可以包括 n -RED。还应当注意，还存在其它算法和其它标记设备实施例，其中单纯根据一个网络反馈阻塞参数并且结合 BW 和 BT 来产生优先权。

虽然已经结合特定装置描述了本发明的原理，但显然应当理解这种描述仅仅是为了举例并且不对如所附权利要求书定义的本发明范围产生限制。

