无建立无状态 QoS 路由

张宝贤, 樊秀梅, 陈常嘉
(北方交通大学通信与信息工程系, 北京 100044)

摘 要: 本文提出了无建立无状态 QoS 路由的概念, 并认证了实现无状态 QoS 路由的关键在于: (1) 如何将应用的 QoS 要求嵌入 IP 分组头中; (2) 如何有效地设计路由器的分组转发策略. 本文结合特定的路由算法, 提出了多种新颖的无建立无状态 QoS 路由具体实现方案. 本文提出的无状态 QoS 路由方案最大限度地利用了目前 IP 网络中采用的"尽力传送"路由表风格, 具有良好的可扩展性、鲁棒性、简单性和有效性.

关键词: QoS 路由; 无状态; 无建立
中图分类号: TN 919.1 文献标识码: A 文章编号: 0372-2112 (2001) 07-0881-04

Setupless and Stateless QoS Routing

ZHANG Bao-xian, FAN Xiu-mei, CHEN Chang-jia
(*Dept. of Communication and Information Engineering, Northern Jiaotong University, Beijing* 100044, *China*)

Abstract: The concept of setupless and stateless QoS routing was proposed in this paper. The key idea in implementation of stateless QoS routing lies in: (1) how to insert application's QoS requirements into IP packet header; (2) how to efficiently design forwarding strategy of a router. Combined with certain QoS-based routing heuristics, several novel solutions were proposed. The proposed stateless QoS routing schemes maximize the utilization of best-effort routing entries and are scalable, robust, simple and efficient.

Key words: QoS routing; stateless; setupless

1 引言

目前已知的 QoS 路由机制[1~3], 都采用面向连接的方式. 尽管这种方式能提供功能强大的服务, 路由器必须处理与每个流相关的预约请求, 并为该 QoS 流建立并保持一个表项, 以描述路由器对该 QoS 流的相应操作. 当流数量的增长难以控制时, 将给路由器带来难以承受的负担. 尽管有许多努力试图通过聚合、分层策略增强上述方案的可扩展性[4], 然而 IntServ 框架中的可扩展性依然是一个难题.

无状态是原本 IP 体系的一个重要特征, 正是这个特征的可扩展性和鲁棒性, 使得 Internet 得以成功. 最近克服 QoS 提供方面的可扩展性问题的努力主要集中在提供有差别服务 (DiffServ) 的研究[5,6], 在差别服务框架中, 核心网络采用无状态转发机制, 相应分组转发根据每个分组头中携带的 QoS 描述, 确定分组的送出时刻 (调度问题) 和分组的送出方向 (路由问题). 已经提出的无状态转发机制中, 仅考虑了分组的调度问题, 而未能解决分组的 QoS 路由问题. 若 QoS 路由问题不能有效地解决, 无状态转发机制提供 QoS 服务的能力将受到很大的限制. 这正是本文研究和解决的主要问题. 本文成功地提供了几种无状态 QoS 路由方案, 这些方案将尽力保存有状态转发机制的强大功能, 同时保持无状态网络的可扩展性和鲁棒性. 值得注意的是, 基于无状态网络条件提供 QoS 服务是一个全新的课题, 其研究包括调度和路由两个层面. 之前关于无状态转发机制的调度研究[5,6]过程中没有考虑 QoS 路由问题; 与此类似, 在我们的研究中, 只考虑了路由一个侧面而不涉及调度问题, 如何有效地将调度和路由结合起来尚未深入拓展, 希望本文的工作能为将来的研究提供有用的启迪.

定义 (无建立无状态 QoS 路由):

无状态: 路由器不需要对任何特定附带 QoS 要求的流保持表项, 转发机制也不需要区分每个分组所在的流, 路由器: (1) 仅保存针对关心的多个 QoS 测度的尽力传送路由表; (2) 依据分组头中附带的 QoS 要求及其类型, 采用统一的适用于每个分组的行为表转发分组, 从而取得 QoS 路由的效果.

无建立: 所谓无建立是指这种路由机制并不需要网络上的路由建立阶段. 实际上 QoS 路由往往需要在实际通信前的一个预计算阶段, 它包括实际路由的选择和/或 QoS 的测量 (针对于各种基于测量的 CAC 和资源分配) 过程, 对于用户来说这个过程可以看成是一个路由建立过程, 我们这里所说的无建立是要求这个过程单独由用户 (或边缘路由器) 完成, 而不需要网络管理层面的参与, 一旦路径选定, 也无须在该路径涉及的任何路由器上进行登记和通报过程.

无状态 QoS 路由的实现对于简化网络提供 QoS 服务具有重要意义. 实现无状态 QoS 路由的关键在于: (1) 如何将

收稿日期: 2000-04-24; 修回日期: 2000-09-11

QoS 信息嵌在数据分组头上; (2) 如何设计节点的转发机制. 实现无状态 QoS 路由的意义在于网络的可扩展性.

除了上述要求外, 本文提出的路由方案还具有如下特色: (1) 与现有网络技术兼容: 结合 TOS 转发机制, 附加适当的控制前端就可实现简单有效的 QoS 路由. 随着路由器网络状态的增加和 TOS 路由能力的增强, 通过逐步增加控制前端的功能, 丰富所支持 QoS 路由的内容; (2) 最大程度利用网络中现有的尽力传送路由表实现 QoS 路由; (3) 采用的路由算法简单而有效.

2 有限方向搜索和无状态 QoS 路由

本文所提出的无状态 QoS 路由方案中, 均是由所涉及的多个 QoS 分量中某个 QoS 分量的最小方向路段拼接构成一条 QoS 路径的, 例如在讨论时延和代价双约束时, 最终选择的路由总是先沿一段最小代价(或最短时延) 方向到达某个网络节点, 然后再从该节点沿最小代价(或最短时延) 方向到达信宿. 在 3.4 节中我们称这种方法为双方向搜索算法, 当扩展到多个 QoS 约束时, 就成为有限方向搜索. 虽然有限方向搜索开始成为一个新的研究方向, 但目前大多数的 QoS 路由研究工作仍是基于数学规划的, 因此有必要单劈一节来讨论有限方向搜索对无状态 QoS 路由的重要意义, 以期引起研究工作者的注意.

一个实施无状态 QoS 路由的路由器显然会保持一个路由表, 从目前的技术来看, 这个路由表会是基于距离向量的, 即一个路由器的路由表中为每一个 QoS 分量保持一个到其它路由器最优方向出口的列表, 实际上这正是尽力传送路由机制可以直接向我们提供的. 基于数学规划的 QoS 路由算法往往能够为一个特定的流找到几乎最优的路由, 但所找到的路由一般在每个分段的每个 QoS 分量上均不是最优的, 所以很难利用路由器的尽力传送路由表来简化对所选路由的实施, 因此对于这样的一个路由, 只有在每一个分组头上都记录该分组选定路由要经过的每一个路由器, 才能实现无状态 QoS 路由, 显然分组头开销为 $O(|V|)$. 与基于数学规划的 QoS 路由算法不同, 有限方向搜索与上述路由表有着相同的风格, 它所选择的路径正是由一段段在某个 QoS 分量上最优的路段组成的, 对于这些路段只要通知路由器是到哪个节点的哪个 QoS 分量就可以了, 其他工作路由器会根据本地路由表自动地完成分组转发, 如果组成路径的路段不太多, 则为了通知路由器如何行为的信息也不会太多, 完全有可能集成到每一个分组的分组头上, 这样就构成了无状态 QoS 路由的基础. 已有研究显示[2,7], 对于时延和代价双约束 QoS 路由问题, 甚至路径只由两段路段组成也能得到非常良好的平均性能, 正是这些研究工作构成了本文的基础.

3 无状态 QoS 路由方案

QoS 路由包括路由选择和分组转发机制两方面问题, 本文的方案中将同时对这两个问题提出解决方案, 因此是较为实用的 QoS 路由技术. 为了便于叙述和理解, 可通过典型的测度(metric) 和典型的 QoS 路由问题进行陈述和解释, 但不失一般性. 本文陈述的方案可以根据实际的 QoS 测度和 QoS 要求方便地扩展到其他应用. 本文采用的 QoS 测度是研究最多应用最广的时延和代价, 路由问题是时延约束下的最小代价问题, 这个问题中将时延约束下的最小代价作为用户所追求的 QoS, 这是一个 NPC 问题[1].

符号定义: (1) 节点: s 信源, d 信宿; (2) 路径: $P_{ld}(u, v)$、$P_{lc}(u, v)$, 分别表示从 u 到 v 的最短时延路径和最小代价路径; (3) 时延函数: 函数 $D(\ )$ 是将链路或路径映射到对应的时延值, 如 $D(P_{lc}(u, v))$ 指 $P_{lc}(u, v)$ 上的时延; (4) 路径方向: LD, LC 分别代表最短时延和最小代价路径方向属性; (5) 端口函数: 函数 $I(\ )$ 将属性{LC, LD} 映射到相应的端口, 如 I(LC) 指最小代价路径方向上的出行端口; (6) TOS 映射函数: 函数 $f(\ )$ 是将属性{LC, LD} 映射到 IP 分组头中的 TOS 内容.

3.1 采用 Sun Landgendorfer[2] 路由策略(方案 1)

Sun Landgendorfer 算法是一个分布时延受限最小的代价启发式路由算法. 在作者原始的方案中, 它是一个路径发现和构造过程. 下面对其寻径思想进行适当地改造, 使它成为一个可实施的无状态 QoS 路由机制.

时延边界标记 方案 1 中要求在分组头部分有标记时延边界的附加字段, 对于给定分组 P, 记这个字段的内容为 $T(P)$, 在用户发送一个分组 P 时, 将所要求的 QoS 相对时延边界约束值添入该分组的时延边界字段, 所谓相对时延边界值是指该分组到达信宿的最晚时间与当前时间之差. 分组经过的每一个路由器将相继修改 $T(P)$ 的内容, 使它总是反映当前时刻的相对时延边界值.

路由器的无状态转发机制 任何一个收到分组 P 的路由器, 如 v, 其 QoS 路由行为如下:

(1) 如果 $D(P_{ld}(v, d)) > T(P)$ 则丢弃该分组并用 ICMP 分组回送"目的节点不可达"(可适当增加 ICMP 内容用于指示给定的 QoS 要求下信宿不可达), 否则:

(2) 如果 $D(P_{lc}(v, d)) \le T(P)$ 则执行: {$T(P) = T(P) - D(v, lc\_nhop)$, 其中 lc_nhop 是当前节点到信宿 LC 路径上的下一跳节点; 采用 I(LC) 端口发送分组 P; *}

(3) 否则 $T(P) = T(P) - D(v, ld\_nhop)$, 其中 ld_nhop 是当前节点到信宿 LD 路径上的下一跳; 采用 I(LD) 端口发送分组 P;

其中 $T(P)$, $D(v, ld\_nhop)$, $D(v, lc\_nhop)$ 指平均意义上的时延.

实际上, 这里是把 Sun Landgendorfer 算法路径构造过程中采用的策略运用到了节点的数据分组转发机制中. 上述方案的优点是, 通过设计简单的节点转发策略, 结合针对不同 QoS 测度的尽力传送路由表实现了 QoS 路由; 用户侧协议简单, 只须对入网的分组设定一个时延边界和 QoS 要求类型(时延受限最小代价路由); 仿真显示 Sun Landgendorfer 算法平均代价性能较好[2]. 缺点: 分组转发需要路由器做一定的在线计算, 且与路由器状态有关, 增加了高速实现的复杂性.

---

* 此时可以在分组头中设定一个标记比特, 表征此分组在后续转发中持续采用 LC 方向.

3.2 基于单点中继的隧道方案(方案 2)

下面讨论的两种无建立无状态 QoS 路由方案, 是对文[7]中提出的一种信源路由算法的改造得到的, 文[7]中算法得到的启发式解满足如下形式: $P_x(s, v) \cup P_y(v, d)$, 其中 xy 表示路径方向, 其值为{LC, LD}之一, 而节点 v 是网络中的一个连接上述两段路径的节点, 记作中继节点. 例如: $P_{ld}(s, u) \cup P_{ld}(u, d)$ 表示以节点 u 为中继节点, $P_{ld}(s, u)$ 和 $P_{ld}(u, d)$ 进行路径组合作为最终的受限路径. 通过恰当选择 v, x, y, 并在所有可能组合情况中选择代价最小的一个, 可以得到无环时延受限代价优化路径. 当 v 的取值空间为 $\{P_{lc}(s, d) \cup P_{ld}(s, d)\}$ 时, 算法的执行时间为 $O(|V|)$, 相应算法记: SDCR-1. 仿真显示 SDCR-1 平均代价性能优于 Sun-Landgendorfer 算法.

网络假定　路由器具有 TOS 路由能力. 分组 P 采用 IP 隧道封装技术, 且采用 SDCR-1 算法确定中继节点 D_r 及 $s \rightarrow D_r$(首段), $D_r \rightarrow d$(末段)的路由方向(LC 和 LD 之一); 外层分组头目的地址 D_o 为中继路由器 D_r 地址, TOS 为首段标识 TOS_o(指示从 s 到 D_r 的寻径方向); 内层分组头目的地址 D_i 为原分组的实际信宿地址 D_p(信宿 d 的 IP 地址), TOS 为末段标识 TOS_i(指示从 D_r 到 d 的寻径方向), 由于路由算法的单节点中继特性, 因此最多只需要两层封装*.

无状态 QoS 路由机制　包括路径计算和分组转发两部分.

信源行为: (1) 首先信源采用 SDCR-1 算法求解(D_r, up, down), 其中 up 和 down 分别是 $s \rightarrow D_r$ 和 $D_r \rightarrow d$ 的路由方向, 取值空间{LD, LC}; (2) 封装分组: 采用内外两层隧道技术包封分组, 外层地址 $D_o = D_r$; $TOS_o = f(up)$; 内层地址 $D_i = D_p$; $TOS_i = f(down)$;

路由器的前处理　可以将每个分组 P 直接交付路由器, 让路由器按处理 IP-IP 分组的一般方法进行处理, 但一般路由器有可能将 IP-IP 分组放在较低优先级的位置, 所以也可以通过某些前处理来加快处理速度, 对于收到分组 P 的路由器的 QoS 路由前处理包括:

(1) 如果分组 P 的目的地址不是本路由器地址, 则将分组直接交付转发模块处理, 否则

(2) 如果本路由器是分组头上标记的目的地址, 且分组头表明分组 P 是 IP-IP 分组, 那么脱去外层分组头再将分组交付转发模块处理; 否则分组实际信宿已经到达.

路由器转发行为: 按(外层)分组头的目的地址和 TOS 指示查找路由表, 按相应路由项指示的端口送出分组.

上述方案的优点: 与现有网络技术适配; 明显解藕了路由器前处理和路由器本身的功能, 能够高速处理和简化路由器的功能; 仿真证明该方案性能很好. 缺点: 外层包封不含分组的实际到达地址, 较难对该分组实施许多与实际信宿地址有关的管理和资源分配(如在 VPN 环境对 CIR 的统计, 对 In/Out 分组的标记等)**; 隧道技术包封开销较大.

3.3 基于单点中继的单分组头方案(方案 3)

为了便于实现与信宿地址有关的管理和资源分配, 最自然的方式是分组头上的目的地址位置一直保存分组的实际信宿地址. 对于文[7]中路由算法, 可对分组采用 IP 选项技术, 即, 指示分组首先沿 $P_{up}(s, d)$ 到达中继节点 D_r, 然后从 D_r 沿 $P_{down}(D_r, d)$ 到达信宿. 仔细分析发现, SDCR-1 算法的解不满足上述特征, 原因是 SDCR-1 的组合路径解中首段 $P_{up}(s, D_r)$ 与从 s 到 d 在 $P_{up}(s, d)$ 上部分不一致.

最直接的解决办法是在分组头设置两个地址: 分组实际信宿地址和分组中继地址, 流量管制工作按分组实际信宿地址进行, 而节点路由按分组中继地址进行, 当分组到达中继节点, 将分组实际信宿地址拷入分组中继地址位置, 并将中继后 TOS 标识 TOS_r 拷入分组 TOS 字段. 由于当前的网络技术中, 流量管制工作和节点路由均针对分组头的特定地址位置实施的, 而它们实际上又要针对两个完全不同的实际地址实施, 因此无法完全基于当前的网络条件实现.

可行的办法是适当修改文[7]中算法, 使得路由算法的解满足采用选项技术对分组头的要求, 我们称这个方案为方案 3. 这样, 分组头的目的地址一直存放实际的信宿地址, 而在分组头选项中增加中继节点地址 D_r 中继后 TOS 标识 TOS_r, 整个方案可描述为:

信源行为　(1) 根据修改的文[7]中算法, 计算(Dr, up, down); (2) 封装分组如下: 分组头目的地址仍为 D_p, $TOS = f(up)$, 设定中继选项中包含中继节点地址 D_r 和从 D_r 到信宿的 $TOS_r = f(down)$;

路由器的前处理　任何一个收到数据分组的路由器的 QoS 路由行为:

(1) 如果本路由器地址与收到分组的中继选项中指示的中继地址不同则将分组直接交付转发模块处理, 否则

(2) 如果本路由器地址与收到分组的中继选项中指示的中继地址相同, 则按 TOS_r 内容修改分组头中 TOS 字段, 然后将分组交付转发模块, 否则分组接收节点已经到达.

路由器的转发行为　按(外层)分组头的目的地址和 TOS 指示查找路由表, 按相应路由项指示的端口送出分组.

方案 3 的优点: 通过适当的前处理, 可与当前网络技术适配达到 QoS 路由的目的; 明显解藕了路由器前处理和路由器本身的功能, 能够高速处理和简化路由器的功能; 可对分组实施许多与终端地址有关的管理和资源分配. 缺点是要稍稍修改目前的 IP 头选项, 且修改后的寻径算法的复杂性将上升到 $O(|V|2)$.

3.4 三种方案的评论和比较

分析发现, 方案 1, 2, 3 所根据的路由算法[2, 7]的一个共同特征就是最终受限路径是由 LC/LD 两段路径组合而成, 这种只在两个可行方向上搜索路径的方法我们称为双方向搜索. 存在许多基于双方向搜索 QoS 路由的研究工作[2, 3, 7], 大

---

* 如果算法选择的路径为 $P_{lc}(s, d)$ 或 $P_{ld}(s, d)$, 则此时没有必要对分组进行隧道封装.

** 实现基于方案 2 的流量管制需要承担管制任务的路由器识别数据分组的最终信宿, 而不是仅仅查看隧道式封装后分组的外层地址, 这需要引入附加模块实现这部分功能.

量的仿真工作也表明双方向搜索 QoS 路由能够得到相当好的平均路由特性, 但也存在对双方向搜索 QoS 路由的疑问, 这些怀疑出自如下一些考虑, 很显然, 如果只是限制在双方向上寻径, 最优搜索的计算复杂性仍然是指数地随网络节点数增长, 如果不采用某些简化了的启发性双方向搜索, 则双方向搜索 QoS 路由并没有在本质上减少计算的复杂性, 另一方面, 目前所有已知的双方向搜索 QoS 路由算法的性能评估都是在随机图上计算出的平均值, 几乎没有比较双方向搜索 QoS 路由的解与最优解的性能差别的理论界, 因此无法评估一个特定双方向搜索 QoS 路由结果的性能到底与最佳路由有多大差别. 但是本文所提出和讨论的无建立无状态 QoS 路由方案, 确实为双方向搜索 QoS 路由奠定了坚实的基础, 足以扫除所有对它的各种疑虑. 理由简单地来自无状态转发机制的路由表. 本文提出的方案指出: 正是依托双方向搜索方式寻径使得尽力传送路由表在 QoS 路由实现中得以最大程度的利用, 而不需要针对 QoS 路由增加任何其他路由表开销.

本文提出的 3 种方法虽然都满足无建立无状态的要求, 但仍存在如下的异同点. (1) 路由机制方面: 方案 1 不需要特定的寻径过程, 路由器的分组转发机制自然保证了分组的 QoS 要求; 方案 2、3 则需要特定的信源路由过程. 但在信源路由计算过程中, 信源可以同时计算出同一信源、信宿之间不同时延要求下的低代价路径; (2) 网络动态特性: 方案 1 不受网络动态影响, 方案 2、3 则可根据应用要求对受到影响的会话重新计算路由, 按照新结果对分组封装, 这些任务都在信源方完成, 对核心路由器无影响; (3) 多限制路由: 方案 1 不适合, 而根据文[7]中提出的多限制条件路由算法, 方案 2、3 可以扩展到对多限制路由的支持; (4) 兼容性: 方案 1 采用的分组封装形式与当前 IP 分组不兼容; 而方案 2、3 采用的隧道、选项技术是现有技术, 易于实现.

4 结论

本文中提出了无建立无状态 QoS 路由的概念, 并结合特定路由算法给出了 3 种包含寻径和分组转发机制在内的无状态 QoS 路由方案. 各方案实行过程中所采用的路由算法保证了本文提出的路由机制在实现 QoS 路由过程中: (1) 最大程度利用了现有尽力传送路由表; (2) 尽量采用现有网络技术, 因而对网络做了最小的假定. 鉴于实现基于每个流的有状态 QoS 路由转发的困难性, 本文提出的机制简化了路由器对 QoS 路由的支持能力, 是一种简单、实用、有效的路由机制. 提供的方案易于扩展和调测.

其它问题: (1) 本文提出的机制, 由于采用无建立, 因而不包含接入控制、资源预约功能. 针对这一问题, 一种方法是信源能够根据当前网络拥塞状况自动调节信源速率. 另外, 当网络路由器采用 WFQ 调度算法时, 结合文[6]中的接入控制和资源预约机制, 适当修改本文中的方案 1 可以在核心网络中实现无状态保证 QoS 机制. (2) QoS 业务与尽力传送业务共存条件下的资源分配问题.

参考文献:

[1] Z Wang, et al. Quality-of-service routing for supporting multimedia applications [J]. IEEE JSAC, 1996, 14(7): 1228- 1234.

[2] Q Sun, H Langendorfer. A new distributed routing algorithm for supporting delay-sensitive applications [J]. Computer Communications, May 1998, 21(6).

[3] Reeves, et al. A distributed algorithm for delay-constrained unicast routing [J]. IEEE/ACM Transaction on Networking, Apr. 2000, 8(2): 239- 250.

[4] F Hao, et al. On scalable QoS routing: performance evaluation of topology aggregation [A]. In Proceedings of IEEE INFOCOM [C], 2000.

[5] Ion Stoica, et al. Core-stateless fair queueing: achieving approximately fair bandwidth allocations in high speed networks [A]. in Proceeding of Sigcomm' 98 [C].

[6] Ion Stoica, et al. Providing guaranteed services without per flow management [A]. in Proceeding of Sigcomm' 99 [C].

[7] 张宝贤, 刘越, 陈常嘉. 基于信源路由的时延受限点到点路由算法 [J]. 投送电子学报.

作者简介:

张宝贤 男. 生于 1972 年, 2000 年在北方交大通信与信息工程系取得博士学位, 现在加拿大金斯顿大学做博士后研究. 主要研究方向是: QoS 路由, 组播通信.

樊秀梅 女. 生于 1967 年, 现于北方交大通信与信息工程系攻读博士学位, 主要研究方向是: 组播通信, 计算机网络.