

保证 QoS 的路径级网络资源分配算法研究

史 琰, 刘增基, 邱智亮, 刘亚社

(西安电子科技大学 ISN 国家重点实验室, 陕西西安 710071)

摘 要: 随着越来越多的实时业务接入因特网, 如何在满足业务端到端 QoS 要求的前提下, 使网络接纳更多用户成为一个严峻的挑战. 本文首先对路径级的网络资源最优分配问题建立数学模型. 理论推导证明: 该问题可以通过一个简单的最优规划问题求解. 随后, 根据求得的最优解的特点, 本文提出了一种结合网络资源状态和用户业务 QoS 要求的资源分配算法) ERA 算法. ERA 算法不仅运算简单, 而且仿真结果也表明, 在相同网络资源总量的前提下, 其接纳的用户数目可以达到理论的上界.

关键词: 网络资源分配; QoS 划分; 最优规划; 等比例资源分配

中图分类号: TN915103 **文献标识码:** A **文章编号:** 03722112 (2005) 071158205

Study of a Path Level Network Resource Allocation Algorithm with QoS Guarantees

SHI Yan, LIU Zengji, QIU Zhiliang, LIU Yashu

(ISN National Key Laboratory, Xidian Univ., Xi'an 710071, China)

Abstract: As more and more services with definite QoS requirements go into the Internet, ISPs face a severe challenge that how maximum number of users can be admitted into networks with QoS guarantees. In this paper, a mathematical model is built first for describing the problem of optimum network resource allocation. Secondly, it is proved that the problem above can be simplified into an easy optimization programming, the unique optimal solution to which is just that of the original problem, and whose optimal value is the upper bound of the number of admitted users. At last, a novel path level network resource allocation algorithm, ERA algorithm, is developed based on the characteristics of that optimal solution. The ERA algorithm possesses the good simplicity and efficiency, and the numerical results also show that it outperforms other algorithms dealing with the same problem.

Key words: network resource allocation; QoS partition; optimization programming; equal ratio allocation (ERA)

1 引言

随着因特网的发展, 越来越多具有 QoS 要求的业务接入因特网. 对于用户而言, 其主要关注业务的服务质量是否被满足, 而对因特网业务提供商 (ISP) 来说, 它不仅要考虑如何满足用户业务的 QoS 指标, 还要使网络尽可能地接纳更多的用户^[1]. 保证业务的 QoS 依赖于网络资源的分配, 而且用户业务的 QoS 指标越高, 其需要的网络资源也就越多^[2]. 网络提供 QoS 保证的过程, 就是网络资源的合理分配、适当使用的过程. 网络资源, 无论是链路带宽、存储器容量, 还是处理器能力都与某一网络结点或链路相关, 因此我们描述网络资源的单位是链路级⁰的. 但是, QoS 指标描述的是用户业务从发送端经过网络到达接收端的性能⁰下限, 即 QoS 指标是路径级⁰的, 那么很自然的想到一个问题: 如何分配链路级的网络资源去满足的路径级的用户业务 QoS 指标, 使网络可以接纳最多的用户^[1], 这也是本文的研究目标.

本文首先建立了以接纳用户数目为优化目标的网络资源最优分配问题的数学模型, 然后根据详细的理论分析证明: 此问题可以简化为一个简单的最优规划问题, 而且此规划的唯一最优解就是网络资源最优分配方案之一, 其最优值是网络可以接纳用户数目的上限. 而且, 根据此唯一最优解的特点, 本文提出了一种结合网络资源状态, 满足带宽和时延要求的路径级网络资源分配算法- ERA 算法 (Equal Ratio Allocation). ERA 算法不仅运算简单, 而且计算机仿真结果也表明, 它可以达到接纳用户业务数目的理论上限.

2 网络资源分配及其优化目标

所谓网络资源分配, 就是在用户业务端到端路径的各段链路上分配适当的资源, 不仅可以确保用户业务的端到端的服务质量, 同时满足网络资源最优分配的要求. 网络资源分配算法的复杂度与用户业务 QoS 指标的类型和数量密切相关^[2,3]. 根据文献[4], 在这些指标中, 带宽和时延是对于服务

质量的影响最大. 因此, 本文以带宽和端到端时延作为用户业务的 QoS 指标.

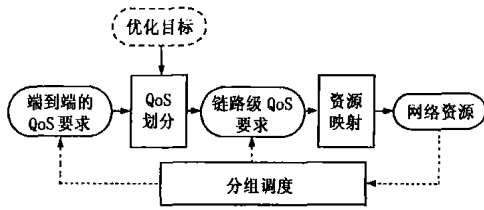


图 1 网络资源分配关系图

图 1 为网络资源分配关系图. 在网络资源分配过程中, QoS 划分依据优化目标, 将端到端 QoS 指标分解为链路级 QoS 指标. 然后, 再根据链路级 QoS 指标到网络资源的映射算法求得所需网络资源的数量. 分组调度算法使用网络资源为用户业务服务, 最终影响用户业务的端到端性能, 在满足相同的 QoS 指标前提下, 不同的分组调度算法(如 EDF 或 WFQ)需要不同数量的网络资源^[5]. 可见, 虽然 QoS 划分是网络资源划分过程的关键步骤, 但是不同的分组调度算法和优化目标, 也会导致 QoS 划分结果不同, 并最终影响网络资源分配结果.

对于 ISP 而言, 它非常关注如何使有限的网络资源可以为更多的用户服务, 从而获得更高的经济效益. 因此, 在本文中, 假设所有的网络结点都采用 WFQ like(包括 VC, PGPS 和 WF²Q) 分组调度算法. 在这种分组调度算法下, 一个受漏桶参数(R, Q)约束的用户业务的端到端时延上界由式(1)表示^[5, 6], R 为用户业务的最大突发长度, Q 为用户业务的平均速率, L_{max} 为用户业务的最大分组长度, C_j 为链路 j 的发送速率, 而 Q 为链路 j 上分配给用户业务的带宽.

$$D = \frac{R}{Q} + \sum_{j=1}^m \left(\frac{L_{max}}{Q} + \frac{L_{max}}{C_j} \right) \quad (1)$$

从式(1)可以看出, 对于那些端到端时延要求不高的用户业务, 只需要在各段链路上分配等于其平均速率的带宽即可; 反之, 就必须在路径链路上分配更多的带宽, 这也是 ERA 算法的主要解决目标. 在 WFQ like 的分组调度算法下, 用户业务的端到端时延只与分配给该业务的链路带宽有关, 因此后文中的网络资源分配就是带宽分配.

3 网络资源最优分配问题分析

3.1 假设条件

(1) 网络路径模型: 如图 2 所示, 由 m+1 个网络结点 N₀, N₁, ..., N_m 和 m 条链路 L₀, L₁, ..., L_m 构成的网络, 其中每条链路的带宽分别为 C₀, C₁, ..., C_m, 而且相对于单个用户业务请求, 链路带宽很大. N₀ 和 N_m 分别为用户业务的源结点和目的结点.

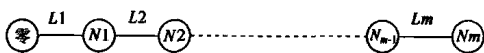


图 2 网络路径模型

(2) 用户业务模型: 从结点 N₀ 进入网络的用户业务满足漏桶参数(R, Q_{avg})的约束, 分组最大长度为 L_{max}.

(3) QoS 指标: 用户业务的端到端 QoS 指标为(R_B, R_D), 其中 R_{B} \setminus Q_{vg} 为最小的链路带宽, 而 R_D 为允许的最大端到端时延.}

(4) 变量定义: 算法为每个用户业务 i 在链路 L_j 上分配的带宽为 Q_j, N 网络接纳的用户业务数.

(5) 优化目标: 使网络接纳的用户数目 N 最大.

3.1.2 分析与推导

根据假设条件和优化目标, 上述网络资源最优分配问题可以使用规划(I)表示.

$$\begin{cases} \max N \\ \text{s. t. } \sum_{j=1}^m \frac{L_{max}}{Q_j} [R_D - \frac{R}{Q_{avg}} - \sum_{j=1}^m \frac{L_{max}}{C_j}] \leq 0, i = 1, \dots, N \quad (1) \\ \sum_{i=1}^N Q_{j,i} \leq C_j, \quad j = 1, \dots, m \quad (2) \end{cases} \quad (I)$$

在规划(I)中, 没有考虑约束条件 Q_{j} \setminus R_B, i = 1, \dots, N, j = 1, \dots, m. 在本文后面将证明, 即使满足该约束条件, 仍可使用类似的方法进行最优的网络资源分配. 在已知条件下, 所有用户业务的 QoS 指标相同, 这就产生一种直观的猜测: 如果所有的业务采用同样的资源分配方案, 是否存在规划(I)的最优解. 基于这种猜测, 将规划(I)转换为规划(II).}

$$\begin{cases} \max N \\ \text{s. t. } \sum_{j=1}^m \frac{L_{max}}{Q} [R_{CD}] \leq 0 \quad (1) \\ NQ_j \leq C_j, \quad j = 1, \dots, m \quad (2) \end{cases} \quad (II)$$

在规划(II)的约束条件中, R_{CD} = R_D - R/Q_{vg} - \sum_{j=1}^m L_{max}/C_j. 由于规划(II)的目标函数值要求是整数, 为了后文分析的需要, 这里将规划(II)转化为对应的目标函数值为实数的规划(IIa).}}

$$\begin{cases} \max n \\ \text{s. t. } \sum_{j=1}^m \frac{L_{max}}{Q} [R_{CD}] \leq 0 \quad (1) \\ nQ_j \leq C_j, \quad j = 1, \dots, m \quad (2) \end{cases} \quad (IIa)$$

规划(II)和规划(IIa)可行集相同, 目标函数也非常类似, 那么它们的最优解之间和最优值之间是否存在确定性关系? 如果存在, 这种关系是否利于计算规划(II)的最优解和最优值. 根据式(IIa-2), 可以得到式(2). 可见, 当 C_{1}/Q_{1} = C_{2}/Q_{2} = \dots = C_{m}/Q_{m} 成立时, 规划(IIa)最优值 n 具有最大的上界. 因此, 再次将规划(IIa)转化为规划(III).}}}}}}

$$\begin{cases} n^m [\sum_{j=1}^m \frac{C_j}{Q}] Z \leq n [\sum_{j=1}^m \frac{C_j}{Q}]^{\frac{1}{m}} [\frac{1}{m} \sum_{j=1}^m \frac{C_j}{Q}] \quad (1) \\ \max n = \frac{C_1}{Q_1} \\ \text{s. t. } \sum_{j=1}^m \frac{L_{max}}{Q} [R_{CD}] \leq 0 \quad (1) \\ \frac{C_1}{Q_1} = \frac{C_2}{Q_2} = \dots = \frac{C_m}{Q_m} \quad (2) \end{cases} \quad (III)$$

很自然我们有以下猜测: 规划(III)的最优解就是规划

(IIa) 的最优解. 为了证明上述猜测, 首先需要证明命题 1.

命题 1 如果规划(IIa)有可行解, 规划(III)必有可行解, 且存在满足式(III- 1)中等式的唯一最优解.

证明 假设 $\mathbf{x} = (Q_1, Q_2, \dots, Q_m)$ 为规划(IIa)的可行解, 如果其满足(III- 2), 则其为规划(III)的可行解; 否则, 必然存在 $Q_k, k \in \{1, 2, \dots, m\}$ 满足 $A = C_k / Q_k [C_j / Q_j, j = 1, \dots, m$. 根据 $Q_j = C_j / A, j = 1, \dots, m$ 来构造 $\mathbf{sd} = (Q_1, Q_2, \dots, Q_m)$. 可见, \mathbf{sd} 满足式(III- 1)和(III- 2), 因此它是规划(III)的可行解. 下面证明, 规划(III)存在满足式(III- 1)中等式的唯一最优解. 将式(III- 2)带入式(III- 1)中, 得到式(3), 它表明了规划(III)的可行集的范围.

$$Q_1 \setminus \frac{L_{\max}}{R_{CD}} \sum_{j=1}^m \frac{C_j}{C_j} \quad (3)$$

又根据规划(III)的目标函数, 目标函数值 n 与可行解中的 Q_1 成反比例关系, 所以当可行解满足式(3)中的等式时, 目标函数取得最优值. 显然, 最优解是唯一的.

命题 2 规划(IIa)和规划(III)的具有相同的最优解和最优值.

证明 使用反证法证明. 假设规划(III)和(IIa)具有不同的最优值. 由于规划(III)的可行集是规划(IIa)可行集的子集, 它们的目标函数又相同, 因此规划(III)的最优值一定不大于规划(IIa)的最优值, 则规划(IIa)的最优解(设为 $\mathbf{x} = (Q_1, Q_2, \dots, Q_m)$)一定不在规划(III)的可行集中. 可是, 如果规划(IIa)的最优值为 \mathbf{rc} , 那么 $\mathbf{s} = (Q_1, Q_2, \dots, Q_n) = (C_1 / \mathbf{rc}, C_2 / \mathbf{rc}, \dots, C_m / \mathbf{rc})$ 一定是规划(III)的一个可行解, 而且其目标函数值也为 \mathbf{rc} , 那么它就是规划(III)的最优值. 因此规划(IIa)和规划(III)具有相同的最优值.

结合式(IIa- 2)和式(III- 2)可知, $Q_j [Q, 1 [j [m$. 如果 $\mathbf{s} \in X$, 则必然存在 $k \in \{1, 2, \dots, m\}$ 满足 $Q_k < Q$, 这表明规划(III)的最优解 \mathbf{s} 不能满足式(III- 1)中的等式, 这是与命题 1 矛盾. 所以规划(IIa)和(III)必然具有相同的最优解.

命题 3 规划(IIa)的最优解就是规划(II)的最优解.

证明 由于规划(II)和规划(IIa)具有相同的可行集, 因此规划(II)的可行解 \mathbf{s} 也是规划(IIa)的可行解, 反之亦然. 如果 N 和 n 分别是 \mathbf{s} 对于规划(II)和(IIa)的目标函数值, $N [n$ 一定成立. (如果 $N > n$, n 一定不是 \mathbf{s} 对应于规划(IIa)的目标函数值.) 又由于 $n < N + 1$ (因为如果 $n \setminus N + 1$, 则 N 就不是 \mathbf{s} 对应于规划(II)的目标函数值), 所以 $N = 7n\delta$ 成立.

如果 \mathbf{s} 是规划(IIa)的最优解, 那么 n 为规划(IIa)的最优值, 而 $7n\delta$ 也就是 \mathbf{s} 对应于规划(II)的目标函数值. 根据上面的分析, 规划(II)可行集中的任一可行解的目标函数值都不可能超过 $7n\delta$, 因此 $7n\delta$ 就是规划(II)的最优值, 而 \mathbf{s} 也就是规划(II)的最优解.

命题 3 表明, 规划(IIa)的最优解一定是规划(II)的最优解. 如果规划(IIa)的最优值为整数, 规划(II)存在唯一的最优解. 如果规划(IIa)的最优值不是整数, 规划(II)的最优解存在什么情况. 这就是命题 4.

命题 4 如果规划(IIa)的最优值不是整数, 规划(IIa)的最优解集合是规划(II)最优解集合的真子集.

证明 命题 4 等价于: 如果 n 不是整数, 存在规划(IIa)的可行解(非最优解)是规划(II)的最优解. 假设规划(IIa)的最优解为 $\mathbf{s} = (Q, Q_2, \dots, Q_n)$. 根据命题 3, 规划(II)的最优值 $N = 7n\delta$. 根据式(4)构造 $\mathbf{x} = (Q_1, Q_2, \dots, Q_m)$, 显然 $\mathbf{x} \in X$.

$$Q_j = \begin{cases} C_1 / 7n\delta, & j = 1 \\ Q_j, & j = 2, \dots, m \end{cases} \quad (4)$$

由于 $7n\delta < n$ (即 $Q_1 > Q$), 因此 \mathbf{x} 满足式(II- 1). 而且, 其对应于规划(II)的目标函数值就是 $7n\delta$, 所以 \mathbf{x} 为规划(II)的最优解. 因此命题 4 成立.

命题 5 规划(II)的最优解集合为凸集.

证明 设 $\mathbf{x} = (x_1, x_2, \dots, x_m)$ 和 $\mathbf{y} = (y_1, y_2, \dots, y_m)$ 为规划(II)任意的两个最优解, 规划(II)的最优值为 N , 根据规划(II)的约束条件, 当 $j \in \{1, 2, \dots, m\}$ 时, $(C_j / x_j) \setminus N$ 和 $(C_j / y_j) \setminus N$ 成立. 对任意 $K, 0 [K [1$, 构造 $K\mathbf{x} + (1 - K)\mathbf{y}$.

$$\sum_{j=1}^m \frac{L_{\max}}{Kx_j + (1 - K)y_j} [\sum_{j=1}^m K \frac{L_{\max}}{x_j} + (1 - K) \frac{L_{\max}}{y_j} = K \sum_{j=1}^m \frac{L_{\max}}{x_j} + (1 - K) \sum_{j=1}^m \frac{L_{\max}}{y_j} [R_{CD} \quad (5)$$

$$\frac{C_j}{Kx_j + (1 - K)y_j} \setminus \frac{C_j}{K \frac{C_j}{N} + (1 - K) \frac{C_j}{N}} = N \quad (6)$$

式(5)和(6)表明 $K\mathbf{x} + (1 - K)\mathbf{y}$ 是规划(II)的最优解. 因此, 规划(II)的最优解的集合为凸集.

如果规划(IIa)的最优值不是整数, 则求解规划(IIa)无法得到规划(II)的所有最优解. 但是, 从命题 4 的证明过程中也看出, 可以使用规划(IIa)的最优解求得某些规划(II)的最优解. 又因为规划(II)的最优解集合为凸集, 又可以使用这些最优解构造更多的规划(II)的最优解. 下面考虑规划(II)和规划(I)的关系.

命题 6 规划(II)的最优值等于规划(I)的最优值.

证明 假设 $\mathbf{s} = (Q_i, j, 1 [i [N, 1 [j [m$ 为规划(I)的最优解, 规划(I)最优值为 N . 根据式(I- 1), 式(7)成立.

$$\sum_{i=1}^N \sum_{j=1}^m \frac{L_{\max}}{Q_j} [NR_{CD} Z \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^m \frac{L_{\max}}{Q_j} = \sum_{j=1}^m \left[\frac{1}{N} \sum_{i=1}^N \frac{L_{\max}}{Q_j} \right] [R_{CD} \quad (7)$$

$$\frac{1}{Q_j} = \frac{1}{N} \sum_{i=1}^N \frac{1}{Q_j}, 1 [j [m \quad (8)$$

根据式(8)构造 $\mathbf{x} = (Q_1, Q_2, \dots, Q_m)$, 可知 \mathbf{x} 为规划(II)的一个可行解, 且对任意的 $j = 1, \dots, m$ 式(9)成立.

$$\sum_{i=1}^N Q_i [C_j Z \frac{N}{C_j} \left[\left[\frac{1}{N} \sum_{i=1}^N Q_i \right]^{-1} \left[\left(\sum_{i=1}^N Q_i \right)^N \right]^{-1} \left[\frac{1}{N} \sum_{i=1}^N \frac{1}{Q_j} = \frac{1}{Q_j} \quad (9)$$

如果 \mathbf{x} 对应于规划(II)的目标函数值为 Nc , 由于规划(II)的可行集是规划(I)可行集的子集, 显然 $Nc [N$ 成立. 又根据式(9)可知, $Nc \setminus N$. 因此, $Nc = N$. 所以, 规划(I)的最优值等于规划(II)的最优值, 而且根据式(8)构造的 \mathbf{x} 为规划(II)得最优解.

可以看出, 规划(III)的最优值和规划(I)的最优值具有明确的数值关系, 规划(III)的最优解就是规划(I)的最优解, 而

求解规划(III)的难度远小于直接求解规划(I). 本节分析证明就是使用规划(III)来求解规划(I)的方法.

4 ERA 算法

4.1 算法特点

根据第三节的分析, 本节提出一种结合网络资源状态和用户业务的端到端 QoS 指标的资源分配算法, 实现的网络资源的优化分配. 根据这种算法的特点, 将其命名为等比例资源分配算法(ERA). 设 (Q, Q_1, \dots, Q_m) 为规划(III)的最优解, 根据命题 1, 规划(III)的最优解可以从式(10)求得, 其中集合 $P = \{1, 2, \dots, m\}$.

$$Q_j = C_j \prod_{i \in P} (L_{\max} / C_i) / R_{CD} \quad j \in P \quad (10)$$

使用式(10)可以非常简单地计算规划(I)的一个最优解, 求得的解具有以下特点: 在各段链路路上的分配带宽与链路资源是等比例的. 如果(10)中的 C_j 为了链路 j 上的可用带宽, 则这时的 ERA 算法是基于可用带宽等比的.

4.2 算法流程

在第三节中提到, 规划(I)中省去了约束条件 $Q_j \setminus R_B, i = 1, 2, \dots, N, j = 1, 2, \dots, m$, 这表明在分析过程中省略了用户业务对带宽的要求, 因此最优解中可能出现分配带宽小于用户业务带宽的情况. 因此, 在 ERA 算法在应用中必须考虑这种情形. 图 3 给出 ERA 算法的流程.

(1) 令 $Q = R_B, j = 1, 2, \dots, m$, 判断 $\frac{R}{Q_{vg}} + \prod_{j=1}^m (L_{\max} / Q) + \prod_{j=1}^m (L_{\max} / C_j) [R_D$ 是否成立.

如果上式成立, 表示每条链路为用户分配用户业务要求的最小带宽后, 就可以满足用户业务的时延要求, 这时不必再进行等比例地网络资源划分, 进入步骤(6);

(2) 令 $P = P^{(0)} = \{1, 2, \dots, m\}, i = 0$ 和 $R_{CD} = R_D^{(0)} = R_D - \frac{R}{Q_{vg}} \prod_{i \in P} (L_{\max} / C_j)$, 进入步骤(3).

(3) 根据式(10), 计算规划(III)的最优解 $(Q^{(i)}, j \in P^{(i)})$, 进入步骤(4).

(4) 是否存在集合 $A \subset P^{(i)}$, 对任意 $j \in A$ 满足 $Q_j^{(i)} < R_B$, 和任意的 $j \in P^{(i)} - A$ 满足 $Q_j^{(i)} \setminus R_B$. 如果 A 是空集, 表明某些链路上分配得带宽小于用户业务要求得最小带宽, 表示这次资源分配不成功. 令 $Q = R_B, j \in A$, 进入步骤(5); 否则, 则 $Q = Q_j^{(i)}, j \in P^{(i)}$ 进入步骤(6).

(5) 调整 $P = P^{(i+1)} = P^{(i)} - A$ 和 $R_{CD} = R_D^{(i)} - \prod_{i \in A} (L_{\max} / R_B)$, 并令 $i = i + 1$, 返回步骤(3).

(6) 得到网络资源的分配方案 $Q, j = 1, 2, \dots, m$, 结束.

图 3 ERA 网络资源分配算法流程

从 ERA 算法的流程可以看出, ERA 算法最多执行 $m - 1$ 次循环后就可以结束, 其中 m 为了路径上链路的数目, 而且上一步的运算结果可以直接被下一次循环使用, 所以此算法简单有效.

5 ERA 算法性能仿真与结果分析

5.1 仿真参数

在本节中, 通过计算机仿真对 ERA 算法和其他网络资源分配算法的性能进行比较. 与 ERA 算法进行比较的算法包括: EA (Equal Allocation) 算法^[1]、ESS (Equal Slack Sharing) 算法^[7]以及 GM (Greedy Move) 算法^[8]. 这些算法都实现如何给具有明确 QoS 指标的业务分配网络资源, 但是他们实现的方法差别很大. 其中, EA 算法是在路径的各条链路上分配相同的资源来满足业务的端到端的 QoS 指标. ESS 算法首先根据所有的可用资源算出网络所能提供的 QoS 指标, 然后将超出的部分平均分给路径上的各个链路, 这样得到一个网络资源分配方案. GM 算法的核心是一个代价函数, 在为一个业务分配资源时, 使这个代价函数值最小. 在本次仿真中, 使用的代价函数为所有链路利用率的平方和. 仿真网络与图 2 所示网络类似, 由四条链路组成. 其中, 链路 L1、L2、L3 的带宽为 40Mbps, 而链路 L4 的带宽为 20Mbps. 用户业务的发送结点 0, 接收结点 4, 业务的最大分组长度为 250Byte.

5.1.1 仿真结果及分析

首先, 在相同 QoS 指标下, 比较不同的资源分配算法最多可以接纳的用户数目. 假定用户业务的带宽需求为 0.5Mbps, 而时延要求从 17ms 到 33ms. 图 4 显示了不同的资源分配算法最多可以接纳的用户数目随端到端的时延变化的曲线. 图 5 表示, 在端到端时延要求为 24ms 时, 不同算法下各个链路的带宽利用率随接纳用户的变化曲线.

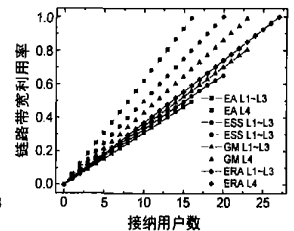
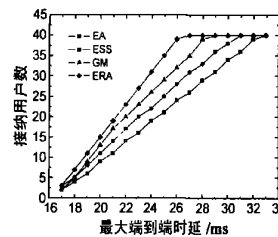


图 4 不同算法接纳的最多用户数 图 5 不同算法的链路利用率

从图 4 可以看出, 当用户业务的端到端时延要求非常小时, 各种算法的结果差别不大, 随着端到端时延增大, 虽然所有的算法都可以接纳更多的用户业务, 但是 ERA 算法接纳的用户数明显高于其他的算法. 当用户的端到端时延要求为 27ms 时, ERA 算法最多可以接纳 40 个用户, 但是随着端到端时延的继续增大, 其接纳用户数不再增加, 这是由于根据端到端时延计算出来的链路带宽小于用户业务最小带宽要求的缘故. 图 5 的结果可以解释 ERA 算法为什么比其他算法具有更高的性能. 在 ERA 算法下, 各条链路的带宽利用率相同, 避免了链路 L4 过早地成为网络瓶颈, 从而增加了网络接纳用户数. 从图 4 可见, ERA 算法对端到端时延要求较高的业务进行资源分配时, 更能体现其优势. 而且: 在不同的端到端时延下, 使用 ERA 算法接纳的用户数与根据规划(III)计算的最优值相同, 这表明 ERA 算法的确达到了理论上限.

同时, 为了说明 ERA 算法在线计算(on line)的性能优点, 假设业务到达时间间隔服从均值为 1 秒的指数分布, 而业务

的持续时间服从均值为 20 秒的指数分布. 这时, 图 6 显示不同的端到端时延要求下的业务接纳率, 而图 7 则显示在业务接纳率不小于 0.8 时, 各种资源分配算法在不同端到端时延下, 可以接受的最高业务到达率. 图 6 和图 7 表明了: 在相同的 QoS 要求下, ERA 算法可以实现最高的业务接纳率, 或者允许进入网络的最大业务到达率. 在相同的网络资源条件下, ERA 算法比其他资源分配算法可以接纳更多的用户业务, 的确可以更加充分的利用网络资源.

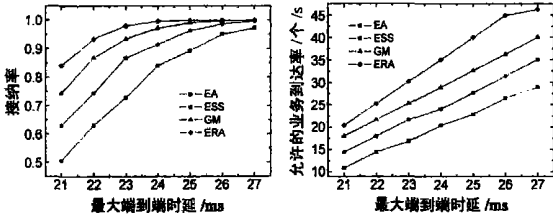


图 6 不同算法的业务接纳率

图 7 不同算法的允许地最大业务到达率

6 总结

本文通过对具有 QoS 保证的网络资源优化分配问题进行分析, 建立其数学模型. 理论推导表明: 在以接纳用户数目为优化目标的情况下, 可以通过求解一个较简单的最优规划问题得到原来问题的最优解, 而其最优值就是接纳用户数的上界. 根据最优解的特点, 本文提出了一种网络资源分配算法 ERA 算法, 它可以在满足用户业务端到端 QoS 要求的情况下, 有效的分配网络资源, 达到接纳最多用户的目的. 仿真结果证明, 相对于其他的网络资源分配算法, ERA 算法具有最高的性能, 并可以达到理论的上界.

参考文献:

- [1] K Gopalan. Efficient network resource allocation with QoS guarantees [D]. Ph. D. Dissertation Proposal, Department of Computer Science

- State University of New York at Stony Brook Stony Brook, NY, 11794-4400.
- [2] X Xiao, L M Ni. Internet QoS: a big picture [J]. IEEE Network, 1999, 13(2): 8- 18.
- [3] F Hao, E W Zegura. Scalability techniques in QoS routing [R]. Tech. Rep. GIT- CC- 99- 04, College of Computing, Georgia Institute of Technology, 1999.
- [4] C Pomavalai, G Chakraborty, N Shiratori. QoS Based Routing Algorithm in Integrated Services Packet Networks [A]. in Proc of IEEE ICNP. 97 [C]. Atlanta, GA, USA, IEEE press, 1997. 167- 174.
- [5] H Zhang. Service disciplines for guaranteed performance service in packet-switching networks [J]. Proceedings of the IEEE, 1995, 83 (10): 1374- 1396.
- [6] Q M Ma, P Steenkiste. Quality of Service Routing for Traffic with Performance Guarantees [A]. In Proc of 5th International IFIP Workshop on QoS [C]. New York, NY, USA, North Holland, 1997. 115- 126.
- [7] R Nagarajan, J Kurose, D Towsley. Local Allocation of end-to-end Quality of Service in High Speed Networks [A]. In Proc of IFIP Workshop on Perf. Analysis of ATM Systems [C]. Martinique, North Holland, 1993. 99- 118.
- [8] D H Lorenz, A Orda. Optimal partition of QoS requirements on unicast paths and multicast trees [J]. IEEE/ ACM Transactions on Networking, 2002, 10(1): 102- 114.

作者简介:

史 琰 男, 1975 年生于河南洛阳. 现于西安电子科技大学 ISN 国家重点实验室攻读博士学位, 研究方向包括现代通信网络中的路由技术, QoS 保障机制以及流量管理. E-mail: yshi@xidian. edu. cn.

刘增基 男, 1937 年生于浙江丽水, 现为西安电子科技大学教授, 博导, 中国通信学会会员, 当前主要从事宽带网络技术的研究. E-mail: zjliu@xidian. edu. cn.

邱智亮 男, 1965 年 5 月生于吉林长春, 博士, 西安电子科技大学综合业务网国家重点实验室教授, 主要研究领域为宽带综合业务网, ATM 接入与交换技术, 高性能路由器交换技术.