

服务元网络体系结构下的 QoS 机制研究 *)

夏梦芹 叶娅兰 王 焱 曾家智

(电子科技大学计算机科学与工程学院 成都 610054)

摘 要 服务元网络体系结构采用改进树型网络拓扑结构,统一对数据类型进行划分,使用 QoS 区域路由方式转发数据,通过通信量整形彻底解决拥塞问题。本文着重分析介绍了服务元网络体系结构中上述 QoS 机制的基本原理和工作机制,并将其与现有 TCP/IP 网络对应机制进行了比较,得出了服务元网络体系结构的 QoS 机制更能满足当今分布式多媒体应用需求的结论。

关键词 服务元,服务质量,区域路由,资源预留,通信量整形

QoS Mechanism Research on Service-Unit Based Network Architecture

XIA Meng-Qin YE Ya-Lan WANG Yan ZENG Jia-Zhi

(School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054)

Abstract Being different from TCP/IP network, Service-unit based network architecture adopts several novel QoS mechanisms, which contain enhanced tree topology, unified type of data classification, QoS partial route and resource reservation in virtual circuit establishment, traffic shaping. This paper focus on introducing and analyzing the above QoS mechanisms in Service-unit based network architecture. After comparing these QoS mechanisms to the corresponding in TCP/IP network, the conclusion that QoS mechanism in Service-unit based network can serve current distributed multimedia applications better is drawn.

Keywords Service unit, Quality of service, Partial route, Resource reservation, Traffic shaping

1 引言

随着高速网络技术和多媒体技术的飞速发展,人们越来越多地提出了包括多媒体通信在内的综合服务要求。传统分组交换网络,如 Internet,是面向非实时的数据通信而设计的,采用的 TCP/IP 协议主要是为了优化整个网络的数据吞吐量并保证数据通信的可靠性。而当今分布式多媒体应用不仅包括文本数据信息,还包括语音、图形、图像、视频这些类型的多媒体信息。分布式多媒体应用不但对网络有很高的带宽要求,而且要求信息传输的低延迟和低抖动,同时大都能够容忍一定程度的信息丢失和错误,对网络提出了不同于数据应用的服务质量要求^[1],需要提供端到端的 QoS 控制和保证。

服务元网络体系结构^[2]是服务元及其相互之间交互使用的规则的集合。该结构中不再划分层次,其网络功能部件是服务元。服务元是能够提供服务而又隐藏内部细节的最小实体。各个服务元完成相对独立的网络功能,相互之间并不传递服务。不同服务元可以协调配合实现各种不同的网络功能,为应用提供各种不同的服务。

服务元网络体系结构的非层次型设计避免了层间交互和服务传递带来的开销^[3],能够为上层应用提供效率更高的服务;除此之外,它还在网络拓扑结构、数据类型划分、数据转发方式和拥塞控制算法等各方面采用了新型 QoS 机制,进一步提高了网络的服务质量,更能适应各类多媒体应用的新要求。

本文主要讨论了服务元网络体系结构下的新型 QoS 机制的基本原理和工作机制,并分别将其与现有主流 TCP/IP

网络对应机制进行了比较,最后得出了服务元网络体系结构的 QoS 机制更能满足当今分布式多媒体应用的 QoS 需求的结论。

2 改进树型网络拓扑结构

现有 TCP/IP 网络属于分组交换网络,采用不规则拓扑结构^[4],以分组“存储-转发”技术为基础,可动态复用物理信道,提高了信道利用率,适合传送突发数据,但可能出现拥塞,较难保证服务质量。

为解决互联网 QoS 问题,IETF 定义了综合服务模型^[5]和区分服务模型^[6]。但前者主干节点管理操作负载过重,难以应用于大型网络;后者缺乏建立连接过程和资源预留,不能完全确保满足实时 QoS 需求。目前 Internet 网络 QoS 研究的重点是寻找既能适用于大规模网络,又能保证满足实时 QoS 要求的解决方案。

服务元网络体系结构使用的改进树型网络拓扑结构参考了电路交换网络树型拓扑结构实时性强,能够保证服务质量的优点,将互联网通信子网目前的不规则型拓扑结构规范为分级树型结合同级不规则型的拓扑结构。改进树型拓扑结构如图 1 所示。

图 1 中同级节点(如 L1 中的 A1、B1、C1、D1 等节点)之间的虚线连接表示两个含义:一是同级节点没有全部画出,只用了 4 个节点来代表同级的数百或者数千个节点;二是同级各节点之间采用不规则拓扑结构,同级两个节点之间可能存在多条不同的通信线路(如 A1 可以通过直连和 C1 进行通

*)国家自然科学基金资助项目,编号:69871005。夏梦芹 博士研究生,主要研究方向:新型网络体系结构。叶娅兰 博士研究生,主要研究方向:新型网络体系结构。王 焱 博士研究生,主要研究方向:新型网络体系结构。曾家智 教授,博导,主要研究方向:计算机网络与通信。

信,也可以通过 B1 或者 D1 进行中转再和 C1 进行通信)。需要注意的是:由于图中每级只画出了 4 个节点,所以各节点之间是完整型的连接拓扑,但是实际上每级包含的节点数成百甚至上千,不可能完全采用完整型拓扑,应该根据实际情况选择相应的不规则拓扑结构。

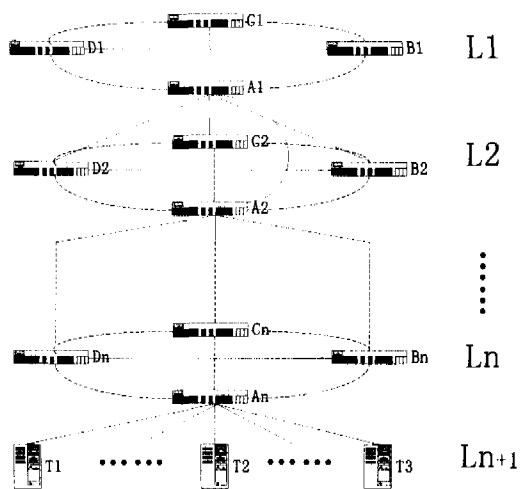


图 1 改进树型网络拓扑结构

第 1 级节点除了和同级节点相连外,还和它所属的全部第 2 级节点相连,如图 1 中的 A1 节点,除了和同级的 B1、C1、D1 等节点相连外,还和它所属的 A2、B2、C2、D2 等节点相连。当然, B1、C1、D1 等节点也分别和它所属的第 2 级节点相连,不过在图中没有画出。

中间节点(除第 1 级节点和终端节点以外的节点)除了和同级节点相连外,还与其直接上级节点和全部直接下级节点相连。如图 1 中的 A2 节点,除了和同级的 B2、C2、D2 等节点相连外,还和其直接上级节点 A1 节点相连,同时还和其全部直接下级节点 A3、B3、C3、D3 等节点相连。

终端节点的情况则根据实际情况的不同而有所不同:如果终端节点属于一个局域网,则部分终端节点之间存在直接互连关系,同时还要和其上级节点相连;如果终端节点采用拨号等点到点连接方式,则只和其上级节点相连,不与其他终端节点直接互连。

改进树型网络拓扑综合了电路交换网络树型拓扑和分组交换网络不规则拓扑的优点,结合后述的区域路由数据转发方式,在保证实时应用服务质量的同时还可以灵活利用网络资源,同时能够解决实时应用下主干节点负载过重问题和完全保证实时服务 QoS 问题,特别适用于大规模网络。

3 数据类型的讨论

TCP/IP 网络设计实现的初期很少考虑 QoS 方面的问题,因此基本没有提供数据类型和服务类型划分方面的实用机制。

随着网络应用对服务质量提出了越来越多不同类型的需, TCP/IP 网络不得不提供相关机制支持对不同类型应用需求提供不同的服务。但是限于层次型网络扩展性较差的特点, TCP/IP 网络不得不在各层分别打下一系列的“补丁”,如传输层的 RTP、RTCP 和 RSVP,网络层的 DS 字段的使用,数据链路层的 802.1d,这些 QoS 机制并不完全一致,极易造成混乱。更为重要的是 QoS 需要的是基于数据流的端到端的保证,这不是层次型的 TCP/IP 网络能够提供的,因此目前最流行的 QoS 解决方案是用 MPLS 来统一划分数据类型和服

务类型。

服务元网络体系结构显式划分了数据类型,在实时音视频和文本的基础上增加紧急数据。紧急数据指紧急并且重要的数据。例如反导弹警戒雷达发出的数据等实时控制数据。它们数据量小,但紧急并且重要。这三种数据类型的特点和处理方法如表 1 所示。

表 1 三种数据类型

数据类型	紧急数据	实时音视频	文本
容量特征	小	固定	可变
预留方法	按最大值	按容量预留	按平均值
丢弃策略	不	一般不	可以
套接字类	有连接	无连接	有连接

服务元网络体系结构下的服务类型采用 5 位标志,其中 3 位分别表示三种数据类型,另外两位可以作为紧急数据的优先级或预留资源的大小。

这种统一显式标记各个分组数据类型的方法真正做到了传递端到端 QoS 需求信息,使得下层网络能够对分组数据进行相应级别的不同处理,提供不同级别的服务质量。

4 QoS 区域路由数据转发方式

在采用改进树型拓扑结构的服务元网络体系结构中,数据分组从源节点传送到目的节点的路径只有部分需要路由,从源节点到传送最上级节点和从接收最上级节点到目的节点的路径都不需要路由,只有最上级节点之间的通路需要路由,区域路由因此得名^[7]。

用一个实时通信的具体例子来看看 QoS 区域路由数据转发方式工作原理。

在图 2 中,终端节点 A 要和终端节点 B 进行实时通信,从 A 到其发送最上级节点 D2 和从接收最上级节点 B2 到 B 的通信线路唯一固定,从 D2 到 B2 的通信线路则需要同级内通过路由选择来获得。

在实时通信开始以前,需要建立实时连接,建立实时连接过程完成的主要工作是确定通信线路和资源预留。在本例中,从 A 到 D2 和从 B2 到 B 的各中间节点在建立实时连接过程中完成的工作是根据通信量大小在和上下级节点之间的通道上预留带宽,同时在本节点内部预留相应缓冲;D2 到 B2 则需要根据路由协议选择合适的通路,然后在该通路上建立虚通道并预留资源。

实时连接建立以后就可以开始实时通信。从 A 到 D2 和从 B2 到 B 的各中间节点只需根据目的地址就可向下一节点进行转发,无需查找路由表项;D2 和 B2 之间的中间节点则查找已经建立的虚通路由表项,找到转发的下一节点,使用预留的资源完成满足 QoS 要求的转发操作。

实时通信结束以后,需要拆除连接释放预留的资源。可以由任一方发出拆除指令,沿传送通道传送一次后即可释放全部资源,如需保证可以确认一次。

分组从源节点到传送最上级节点和从接收最上级节点到目的节点的路径固定,转发过程不需进行查表操作,效率大大提高;最上级节点之间的通路在小范围内根据 QoS 路由建立虚通道,沿路所需资源全部进行预留,因而能够完全满足实时通信在延迟、抖动等方面的 QoS 要求。

在建立实时通信连接时,可以根据所在级别实际通信统计情况设置虚通道的大小。如在最低级别上每次连接建立的虚通道可供 1 个实时通信应用使用,而在高级别节点之间则

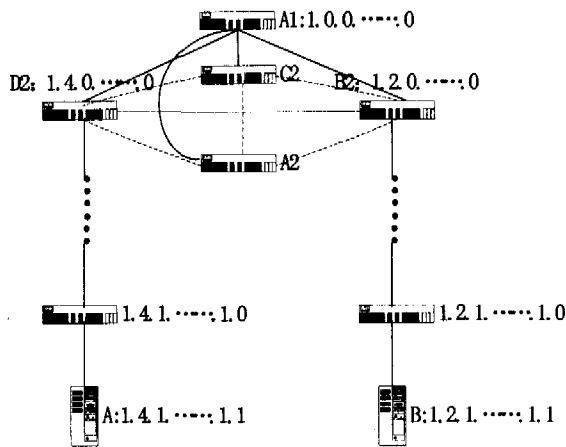


图2 区域路由数据转发工作原理实例图

可以考虑每次连接建立的虚通道可供多个(如 10000 个)实时通信应用使用,由于虚通道的大小设置是基于实际应用统计,其实际利用率应该非常高;而且两个建立连接高级别节点的下层节点间的实时通信都可以使用该通道,高级别节点只需要管理 1 个而不是多个虚通道,管理操作大为简化。

较之 TCP/IP 网络的查表路由转发方式, QoS 区域路由数据转发方式效率更高,能够提供实时 QoS 服务,对实时通信管理复杂度大大降低,适用于大规模网络,因而更易于满足当前分布式多媒体网络应用的需要。

5 通信量整形拥塞控制算法

通常认为在任何时间间隔内生成的包数服从泊松分布。在分析控制拥塞保证 QoS 的条件时,如果采用泊松分布和排队论将使问题复杂化。在服务元网络体系结构中,我们采用下述通信量整形算法来简化问题。

首先,我们来分析一条链路的上游节点如何才能不“淹没”下游节点。办法是对上游节点进行通信量整形,使其输出的包的间隙时间 ΔT 大于下游节点对包的服务时间 T_s :

$$\Delta T \geq T_s \tag{1}$$

这种通信量整形算法不同于令牌桶算法^[8]和漏桶算法^[9],虽然简单,但能够可靠地进行流控,保证下游节点不被“淹没”。并且条件(1)能够保证链路的输出流量等于输入流量,所以我们把满足公式(1)的链路称为线性链路。流量被定义为单位时间内传送的包数。

如果几条虚电路经过同一条线性链路,各数据流同时流经该链路产生的总输出流量等于各数据流单独流经该链路产生的输出流量之和。

由于链路中传送包时必须存在间隙时间 ΔT ,所以链路的容量 R 将小于传输率 v ,我们有:

$$R = v \frac{T_d}{T_p + \Delta T} \tag{2}$$

其中, T_d 为数据(净荷)的传送时间; T_p 为包的传送时间; ΔT 为传送包的间隙时间。

对于链路 k 按照其传输率 v 给定的容量 $R_k = f(v)$ 例如 v 的 90%, 来确定本链路的固定的 ΔT :

$$\Delta T = \Delta T_k = \frac{v T_d}{R_k} - T_p \tag{3}$$

链路的容量应该大于或等于各数据流预留容量之和,假设各数据流的虚电路在链路中预留的容量总和为:

$$r = \sum_{i=1}^n r_i \tag{4}$$

在链路上预留容量的总和 r 必须小于或等于 R :

$$r \leq R \tag{5}$$

由式(3)和(1)可知,当 $\Delta T = \Delta T_{\min} = T_s$ 时,在传输率为 v 的线性链路能够得到最大容量:

$$R = R_{\max} = v \frac{T_d}{T_p + T_s} \tag{6}$$

如果我们把链路上实际预留的容量作为链路的容量,我们得到:

$$\Delta T = \Delta T_{\max} = \frac{v T_d}{\sum_{i=1}^n r_i} - T_p \tag{7}$$

公式(7)和公式(1)分别限制了选择 ΔT 的上下限,通常 $\Delta T_{\max} \geq \Delta T_{\min} (T_s)$ 。如果选择的 $\Delta T < \Delta T_{\min} (T_s)$,则下游节点可能被“淹没”;如果选择的 $\Delta T > \Delta T_{\max}$,则不能保证公式(4)预留的容量;如果 $\Delta T_{\max} > \Delta T > \Delta T_{\min}$,则链路是线性链路,并且可以保证公式(4)预留的容量。由于各种实际因素可能造成的误差, ΔT 应选择为某一个中间值。为了简化操作,我们没有必要按预留资源的实际多少去动态地调节 ΔT 的大小。这条链路无论预留了多大容量,只要在 R_k 范围内,它都是线性链路。如果新的数据流所要求的容量使总的预留容量超过 R_k ,将被拒绝,新的虚电路将不能建立。线性链路 k 的容量为 R_k 。对于容量为 R_k 的线性链路 k ,超过 R_k 的数据流是不可能注入的。

实际上,实时音视频通常是要压缩的,而且压缩比是变化的。如果实时音频数据流预留的容量按非压缩值设置(为 64kbps);实时视频 MPEG1 数据流预留的容量也按非压缩值设置为 1.5 Mbps;文本数据流预留的容量按规定值(例如平均值的几分之一)预留资源。

这样,通过 QOSPF 协议得到约束路由,并建立虚电路,预留容量 r_i ,对每一段链路按上述规则确定的 ΔT_k 进行通信量整形,那么这条虚电路上的所有链路都是线性的。同样,在这样的网络中,所有虚电路流经的链路都是线性的,所有节点都不会被“淹没”。

对于源主机而言,按照 ΔT_k 进行通信量整形,当发送时刻到来,按规则“发送完紧急数据才发送实时音视频数据,二者都发送完后才发送文本数据”发送一个包;如果没有包发送,可以不发送。多产生的包放在缓冲区中。缓冲区放不下,可以暂停产生包的程序。由于虚电路上的包都是由源主机注入的,由于其需要的容量都在虚电路各节点上预留了的,所以只要有缓冲区,理论上是不需要丢包的。因为所有链路都是线性的,路由器在递交数据包时只需执行简单的递交规则:“传送完紧急数据才传送实时音视频数据;二者都传送完后才传送文本数据”。

不同于令牌桶算法和漏桶算法,通信量整形算法中的所有链路都是线性链路,因此所有节点都不会被“淹没”;可以从源头上避免网络拥塞的产生。

小结 服务元网络体系结构采用的改进树型网络拓扑引进了电路交换网络实时性强的优点,统一数据类型划分显式标志了端到端的 QoS 服务要求, QoS 区域路由数据转发方式提供了更高的转发效率和更简单的管理控制,通信量整形拥塞控制算法可以从源头上避免网络拥塞的产生。

较之 TCP/IP 网络,服务元网络体系结构的 QoS 机制更能满足当今大规模网络条件下分布式多媒体应用需求。

参考文献

1 林闯,单志广,任丰原. 计算机网络的服务质量(QoS). 北京:清华大学出版社, 2004

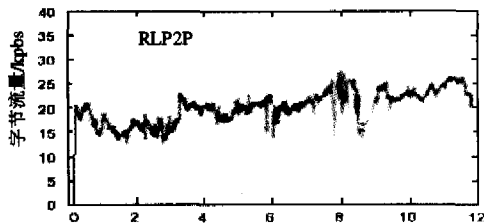


图7 RLP2P网络12小时的流量分布

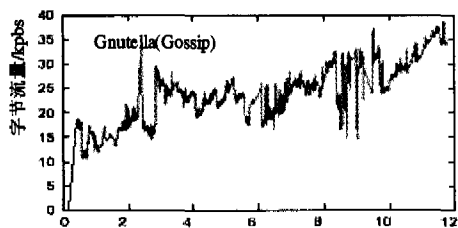


图8 Gnutella网络12小时的流量分布

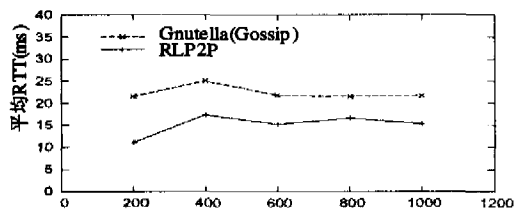


图9 查询返回平均时间

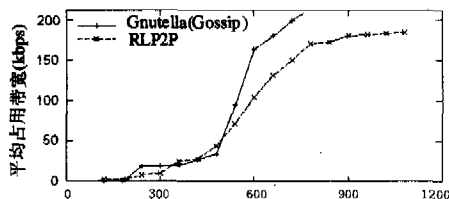


图10 平均带带占用量

仿真结果表明：当网络规模较小时，泛洪搜索算法 Gossip 在 Gnutella 网络中的综合性能非常优异，但随着网络规模的扩大，连接拓扑中“环”的数目增加和泛洪式消息扩散，使网络中的冗余消息呈指数级增加，消耗大量的处理时间和网络带宽，随着这部分节点失效，Gnutella 网络被分片，使查询访

问只能在网络很小的一部分进行，导致网络的可扩展性和搜索性能下降。本文提出的搜索算法对改善 Gnutella 网络的搜索性能非常有效。

结论 结合了传统集中式网络易于管理与分布式网络具有良好的区域自治、负载均衡以及健壮性的优点，以域为基本逻辑管理单位，将大规模节点在逻辑上划分成小区域，使域内任意两个节点都有很小的平均距离，节点间有很大的相识系数，在域内部找到满足服务请求的概率很大，节点出现跨域请求服务的概率很小，使定位某种服务的工作量、查询范围从网络中的所有节点数降低到域内的节点数，将大规模网络消息扩散划分成小规模消息扩散，在域规模相同的理想状况下，算法路由链上界为 $O(\log \sqrt{n/m})^{1+\epsilon}$ (m 为域内节点数)，其扩散效率明显由于 Gossip 算法；网络系统控制开销为常数 $O(K)$ ；提出的搜索包最小冗余扩散算法使扩散冗余包趋于一个很小的常数，算法时间复杂度为 $O(n^2)$ ，极大地减轻了网络的压力，有效地防止了请求洪，提高了网络的鲁棒性和扩展性，具有更高的综合性能指标。

参考文献

- 1 Parameswaran M, Susarla A, Whinston A B. P2PN Networking: An Information Sharing Alternative. Computing Practices, July 2003. 1~8
- 2 Parameswaran M, Susarla A, Whinston A B. P2P networking: An Information-Sharing Alternative [J]. Computer, 2003. 34 (7): 31~38
- 3 Traversat B, Arora A, et al. Project JXTA 2.0 Super-Peer Virtual Network. <http://www.jxta.org/project/www/docs/JXTA2.0protocols1.pdf>. 2004/08/10
- 4 Super-Peer Architectures for Distributed Computing. <http://www.fiorano.com/whitepapers/superpeer.pdf>. 2004/08/10
- 5 Kleinberg J. The Small-World phenomenon: An algorithmic perspective. ACM Symp on Theory of Computing, 2000. 820~828
- 6 Kempe D, Kleinberg J, Demers A. Spatial gossip and resource location protocols. In: Proc. of the 33rd ACM Symp on Theory of computing [C], Crete, Greece, 2002. 163~172
- 7 Gnutella: To the bandwidth barrier and beyond. <http://www.dss.clip2.com/gnutella.html>. 2004/08/10
- 8 Magoni D. nem: A Software for Network Topology Analysis and Modeling. In: Proc. 10th IEEE Intl. Symposium on 11-16 Oct. 2002. 364~371
- 9 Palmer C P, Steffan J G. Generating Network Topologies That Obey Power Laws. Global Telecommunications Conference, 2000. GLOBECOM '00. IEEE, 2000, 1: 434~438

(上接第40页)

- 2 Zeng Jiazhi, Xu Jie, Wu yue, et al. Service Unit Based Network Architecture. In: Proc. PDCAT'03, 2003
- 3 曾家智, 徐洁, 吴跃, 等. 服务元网络体系结构和微通信元系统构架. 电子学报, 2004, 32(5): 745~749
- 4 Tanenbaumk A S. 计算机网络(第3版). 北京: 清华大学出版社, 1998
- 5 RFC1633. Integrated Services in the Internet Architecture, an Overview. 1994

- 6 RFC2475. An Architecture for Differentiated Services. 1998
- 7 夏梦芹, 易发胜, 曾家智. 互联网区域路由 QoS 机制研究. 计算机学报, 2004, 31(11): 38~39
- 8 Perry Tang Puqi, Charles Tai Tsung-Yuan. Network traffic characterization using token bucket model. Proc IEEE INFOCOM, 1999
- 9 Au T M, Mehrpour H. Leaky bucket based scheduling algorithm for real time traffic. Australia National Conference Publication - Institution of Engineers, 1994