

服务质量

分类学

Internet网

计算机科学2000Vol. 27No. 11

多媒体

服务质量(QoS)的分类学^{*}

Taxonomy for QoS Parameters

郭乐深 刘锦德 张乃靖 郭乐群¹
(电子科技大学计算机学院微机所 成都610054)

Abstract In order to design a comprehensive QoS-driven distributed system where heterogeneous applications can compete for heterogeneous resources, we have developed a QoS taxonomy that captures, in a generic fashion, the various facets of QoS. The primary categories within this taxonomy are metrics and policies. Metrics specifies quantifiable QoS parameters. Metrics can be further grouped into the following classifications: performance specifications, security levels, and relative importance. Policies are divided into service guarantees, and management policies.

Keywords QoS Taxonomy, QoS-driven, Integrated Service Architecture, QoS Guarantee, Metric, Policy

当前分布式多媒体系统技术在飞速发展,其核心问题是多媒体流的服务质量(QoS)保证。对于由 QoS 控制来实现 QoS 保证,国际上不同组织和团体提出了不同的控制机制和策略,比较著名的有:1)ISO/OSI 提出的基于 ODP 分布式环境的 QoS 控制^[1,2],但至今仍停留在只给出了用户层的 QoS 参数说明和编程接口阶段,具体实现 QoS 控制策略并未提出;2)ATM 论坛也提出了 QoS 控制的策略和实现^[3],ATM 控制是“连接预定”型(connection and reservation),其核心内容是在服务建立之前,通过接纳控制和资源预留来提供服务的 QoS 保证,而在服务交互的过程中,用户进程和网络之间要严格按照约定的 QoS 实现服务 QoS 保证。3)Internet 的 IETF 相应的工作小组提出的 RFC2115, RFC2117 以及 1998, 1999 年提出的 RFC26xx 系列中的 IS(Integrated Service)和 DS(Differentiated Service)模型均用于解决 Internet 网络的 QoS 控制和管理。

如果建立复杂异质的分布式系统,由于上述组织中的 QoS 参数不同,使得在系统中的高层应用程序和下层不同层次中单元的 QoS 参数含义、表示和度量不同,例如,在资源层次中 QoS 参数取决于使用和设计资源类型以及其上的控制参数上(如在 ATM 中 QoS 参数以 cell 为单位,表示 cell 的速率、cell 的延迟、cell 的抖动和 cell 的丢失率等,而在 IETF 的 IS 中以 pack-

et 速率、packet 延迟、packet 抖动和 packet 丢失率等表示)。具体的分布式多媒体应用程序需要在分布式系统上运行,它则要求下层(如操作系统层、网络层等)的资源按照在指定的应用层 QoS 参数范围之内进行服务操作;分布式系统的下层资源需要知道上层应用 QoS 参数,并能在自己的资源上把上层的 QoS 参数映射为自己可以使用的 QoS 参数类别。例如,Internet 网络的通信资源使用是数据包延迟和数据包抖动等 QoS 参数,而具体的分布式视频程序使用帧延迟和帧抖动等 QoS 参数。如果就存在 n 个标准和一个应用层标准,那么 QoS 参数的转换存在 $(n+1)!$ 种可能性,而这样建立的分布式系统很复杂而且效率低。

要建立一个支持 QoS 保证的分布式系统必须建立统一的 QoS 框架,该框架包括 QoS 体系结构和统一的 QoS 说明分类,其中 QoS 体系结构完成把复杂的资源对象结合起来,以此来建立复杂高效的分布式系统;而 QoS 说明分类是在详细研究 QoS 参数和其间的相互关系的基础上,建立统一的 QoS 参数,详细定义 QoS 参数的含义、表示、度量单位和相互关系,最终使得在分布式的各个系统资源组件之间,在 QoS 层面上进行统一的转换——即 QoS 映射。这样建立的基于 QoS(QoS-based)框架将是下一代的分布式多媒体和实时系统的基础,于是人们逐渐关心 QoS 分类学技术,对各类别的 QoS 参数进行统一分类^[4,5],来简化

^{*} 本文得到电子科学预研项目资助。郭乐深 博士生,主要研究领域为开放系统技术和多媒体技术。刘锦德 教授,博士生导师,主要研究领域为开放系统技术、多媒体技术和虚拟现实应用。张乃靖 硕士生,主要研究领域为开放系统技术和安全技术。郭乐群 工程师,主要研究领域为系统安全技术。

QoS 说明分类,使得 QoS 的映射得到简化和易于实现,从而为建立统一的异质分布式多媒体和实时系统打下基础。

1 QoS 分类学

在 QoS 分类学中主要的分类标准是尺度(metric)和规则(policies)两个方面(详见图1)。尺度指定可度量的 QoS 参数分类集合,它能进一步分为:性能说明、安全性级别、其他有关的重要性参数,性能尺度用来说明高层应用中,与其性能有关的参数集合,例如端到端延迟、总计的传输量、误码率(BER)等等性能度量, QoS 中性能尺度的参数具体可以通过时间性(timeliness)、精确性、准确度等三个方面来表示;安全级别定义高层应用所需要提供的数据安全级别;其他有关的重要 QoS 参数有:价格(或费用),不同 QoS 域之间的统一 QoS 衡量(主要应用在不同 QoS 域(如 Intserv 和 Diffserv)之间)。

QoS 中的规则参数具体又划分成服务保证和管理规则等两类。服务保证对于具体的应用程序,分布式系统定义它的 QoS 承诺类型(如保证服务类型(GS, guaranteed service)、受控负载服务类型(CLS, controlled load service)和最大努力服务类型(BES, best effort service))。QoS 管理规则是指应用在不同的状况下,可以采取的管理措施,例如,在资源的缺乏的情况下,一些应用程序可以通过再协商方式,接受层次较低的服务质量参数;而另一些应用采取拒绝服务方式等;另外它同时也定义了应用程序和系统资源管理之间描述交互的语言。

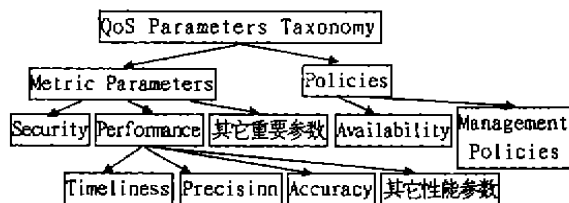


图1 QoS 参数分类图

2 QoS 分类学中尺度(METRICS)分类

2.1 QoS 性能尺度分类

现在,当人们在谈论服务质量(QoS)时,通常提及的是 QoS 分类学中性能尺度,因为它们与上层的应用程序或服务的行为直接有关,性能尺度包括时间性、精确和准确度三个方面。

2.1.1 时间性 QoS 的时间性参数定义了分布式系统中各个实体的时间尺度,时间性表示为时间的

单元(unit)。具体的时间性参数可以表示为执行某个服务时其时间上的属性,它包括六个方面:

1)完成某个服务的总时间完成(开始到结束)。这作为操作延迟时间、传输延迟时间或预定完成时间等参数来度量。

2)服务起始时间。

3)服务截止时间。

4)完成任务的时间变化性——完成任务的延迟抖动(Delay jitter),通过它可以度量时间性参数的内部的一致性。

5)服务的截止时间和和其他任务的起始时间的相互关系(即同步关系),通过它可以度量时间性参数的相互之间的一致性。

6)上述各个参数的统计特性,即均值、方差和协方差等特性。

其中前三个时间性参数是绝对数值,由分布式系统中各个资源调度程序使用;而参数4和5是时间性的一致性详细说明,是用于维护分布式客户/服务器进程之间的相互关系;而最后参数6是对上面五个参数的统计,描述参数在服务过程中的变化。

2.1.2 精确性和准确度 QoS 精确性和准确性参数说明交互数据的精确程度和准确程度,由于系统的不同实体对同样的数据存在不同的理解方法,所以在 QoS 分类中,对于数据的精确性和准确性参数要求更多的详细的定义,通过精确性和准确性参数可以随数据的属性具体化,其中数据的属性包括:数据内容、数据表示和数据数量。我们定义数据内容是数据的意义,数据内容代表数据是实时数据、音频数据或视频数据等等,数据在应用程序之间流动,数据的内容区分是重要的。数据表示是数据在计算机内的表示,同样的数据内容可以存在一种或多种不同的数据表示,数据数量表示数据的大小,数据是通过数据内容、数据表示和数据数量共同来完整地表达出来。所以应用程序和下层的资源管理进程在支持 QoS 时,QoS 分类学要求必须提供媒体数据的精确性和准确性描述。

另外由于多媒体数据存在大量的压缩算法,可以使用户的原始信息数据在允许范围中丢弃,但是分布式系统中 QoS 保证机制必须满足用户在应用层次上对于 QoS 精度和准确度的要求,这使得在分布式系统中不同层次上的实体,在具体的 QoS 分类时必须充分考虑这一问题,同时精确性和准确度在不同资源服务功能变换时,定义了不同的转换条件,但是现在大多数的 QoS 讨论没有涉及到这个方面及其影响。

QoS 分类学中涉及到的精确度参数包括:

1)输入、输出时的数据内容的精度;

2)输入、输出时的数据表示的精度;

3) 流(flow)内部一致性的精度(即精度的抖动数值);

4) 流之间的外部一致性的精度(即精度的同步特性);

5) 上面参数的统计分布。

参数1和2是绝对数值,由分布式系统中各个资源调度程序使用;而参数4和5是时间性的一致性详细说明,是用于维护分布式进程之间的相互关系;而最后参数6是对上面五个参数的统计,描述参数在服务过程中的变化。

QoS 分类学中涉及到的准确度参数包括:

1) 输入、输出时的数据内容的准确度;

2) 输入、输出时的数据表示的准确度;

3) 上面参数的统计分布。

参数1和2是绝对数值,由分布式系统中各个资源调度程序使用;而参数4是对上面两个参数的统计,统计数值描述参数在服务过程中的变化。

2.2 QoS 分类学的安全性尺度分类

具体地说,安全性参数尺度是为应用服务提供与其交互数据的安全性相互联系的安全机制和安全服务等方面的分类。其中安全机制包括加密机制、数字签名机制、访问控制机制、数据完整性机制等等;安全服务包括对象认证安全服务、访问控制安全服务、数据保密性安全服务、数据完整性安全服务等等。用户在执行具体的应用时,对于安全服务质量提出相应的要求,安全服务质量的参数包括存取控制功能、安全审计、认证功能、加密功能、不可否认、密钥管理功能等等安全服务的级别的分类。存取控制功能阻止未授权的服务对象进行访问;安全审计功能监视和收集与活动有关的安全信息,并允许分析这些信息;认证功能确保服务对象是已声明的该对象;加密功能有防止对服务信息进行未授权的发布;不可否认功能防止交互中的服务对象否认交互中牵连到的事情;密钥管理功能为密钥的管理提供方便。

总之,用户应用的安全 QoS 要求是符合 C1, C2, C3, B1, B2, A1, A2 的安全级别,或这么分类:秘密、机密、绝密等不同分类的安全服务质量的参数。

3 QoS 分类学的规则分类

3.1 有效性

为了有效使用资源,为应用提供其 QoS 所需要承诺的级别要求,资源管理器必须按照一定的规则来提供服务。不同的规则用于不同的媒体数据,例如资源的调度算法, RM、EDF 用于实时服务,而 WFQ (Weighted-Fair Queuing)、FQS (Fair Queuing based on Start-Time)、SFQ (Start-time Fair Queuing)^[4~7] 用于多媒

体数据。为了进一步提高资源利用率,要通过媒体流进行划分,不同的媒体流作不同对待。目前 IETF 提出的 Internet 综合业务结构 (ISA, Integrated Service Architecture^[1]) 定义了进行媒体流的划分来实现应用 QoS 保证的结构,通过 ISA 划分使得按照不同的规则为上层服务提供其要求的端到端的 QoS 保证,使得服务根据其服务数据包的 QoS 要求选择服务类型。ISA 的分类有: QoS 保证服务类型 (GS, guaranteed service)、受控负载服务类型 (CLS, controlled load service) 和最大努力服务类型 (BES, best effort service)。其中 QoS 保证服务类型是要求分布式系统保证服务的传输速率和延迟,并保证不丢失数据包,该业务不能控制固定延迟(即传输延迟,它决定于服务所选的路由),但它能保证排队延迟。网络的调度算法是加权公平排队算法 (Weighted Fair Queuing), 在资源预约建立的 Tspec 中给出对数据流的描述,在 Rspec 中对传输速率和延迟提出要求;受控负载服务类型提供最小数据包的延迟保证,不要求网络的排队算法,应用可以假设网络传输差错率近乎等于链路层的差错率。最大努力服务类型就是现在通常使用的无 QoS 保证的服务类型,无传输速率和延迟 QoS 保证。

由此可见,不同特定的服务规则对应不同的服务保证规则(从尽力型规则到服务保证型规则)。我们使用有效性这一服务质量尺度对规则进行分类,其中有效性服务质量尺度的含义是分布系统对应用服务及其数据的 QoS 保证要求的可能性,可以用概率函数表示,通过有效性服务质量尺度我们可以把服务分类为不同的服务级别,不同级别对应不同的服务规则。

3.2 管理规则

QoS 管理规则是由应用程序专用(application-specific)的规则,由下层的资源管理器具体执行。具体例子是在未预见的资源缺乏而导致的多媒体交互失败,上层的应用程序通过定义是否愿意通过重新协商来接受另一个的服务质量级别还是拒绝服务。由于没有具体对管理规则的可量化尺度,很难对管理规则划分分类等级,但是管理规则可以属于元级别分类。例如,元级别可能是“renegotiation 允许”和“renegotiation 不允许”两级。

结论 通过对 QoS 参数进行分类,使得 QoS 参数可以在用户界面和资源对象之间传递,复杂的分布式系统的各个对象之间可以保持对 QoS 含义和其参数的数值统一的理解,从而成为一个完整的 QoS 驱动分布式系统。

建立统一 QoS 驱动的分布式系统的关键之一是,

(下转第 45 页)

流量的高效一体化传输,从而降低成本,提高性能。目前电信终端的智能化、数字化、多媒体化已成为发展方向,国外的一些高智能电话机已可同时处理声音、数据、图像,如英国技术伙伴公司与另一家公司合作研制出具有浏览 Internet 功能的数字式移动电话,装上特定的软件后,能方便地浏览 Internet 上的文本信息;双向有线电视必然要求其电视终端设备走向智能化,使电视机具备交互功能和处理文本和数据的功能,WebTV 和机顶盒将成为时尚;在计算机上安装一定的设备后,接收电视信号、进行传真、语音通信、图形图像通信也已成为现实;具有类似功能的个人数字助理(PDA)装置也已研制成功。由此可知,具备显示多种媒体的网络通信终端也将随着技术的发展、用户的需求而逐步走向融合。

当然管理维护“合一”的网络,有待于进一步的研究。

结束语 社会对信息的需求使得信息服务已成为 21 世纪的重点课题,信息化已将计算机、电信、电视融合为一个不可分割的整体,三网融合是必然的发展趋势。若采用宽带 IP 骨干网技术(包括全国、省内宽带 IP 广域网和市内宽带 IP 城域网),可以以非常低廉的

价格提供宽带多媒体业务服务,这将大大促进我国信息化的进程,同时宽带接入网的出现不仅解决了现有接入网的“瓶颈”限制,而且使得通信网络向宽带数字化的过渡成为可能。

信息化推动技术发展,大目标就是实现电信、计算机和电视三种技术、业务、市场、行业、网络和终端的融合。而三网本身也将通过不同途径向融合的可持续发展的全业务网方向演进。

参考文献

- 1 马丁·德·普瑞克(程时端译). 异步传递方式. 邮电出版社, 1995
- 2 谢希仁. 计算机网络(第二版). 大连理工大学出版社, 1994
- 3 Tanenbaum A S. Computer Networks. 清华大学出版社(影印版), 1996
- 4 赵慧玲,等. ATM: Internet 和企业网. 电子工业出版社, 1998
- 5 林瑞,等译. 用 TCP/IP 进行网际互连,第 1 卷:原理、协议和体系结构(第 3 版). 电子工业出版社, 1998
- 6 刘心松,等. 通向三网融合之路. 电脑信息技术, 1999(11): 1~4

(上接第 31 页)

必须使分散的分布式系统中各个组件对 QoS 概念的定义有统一的认识,使得建立后的分布式系统各组件不会存在牛头不对马嘴的现象。所以通过 QoS 分类学标识,我们对不同的 QoS 参数进行分类,分布式系统各个元件建立统一的尺度,而具体到各个组件实现 QoS 保证时,先要进行 QoS 映射(QoS mapping),在统一的 QoS 尺度与本身的 QoS 尺度进行转换,具体实施 QoS 保证,所以建立统一的 QoS 分类十分重要。

参考文献

- 1 Braden R, Clark D, Shenker S. Integrated Services in the Internet Architecture; IETF RFC 1633. June 1994
- 2 ISO/IEC JTC 1/SC21. Quality of Service in ODP—Attachment 1 January 1997
- 3 ATM Forum. ATM User—Network Interface Specification Version 4.0. 1995
- 4 Kaneko H, Stankovic J A. Integrated scheduling of multimedia and hard time task. In: Proc of the 17th Real-time System Symposium, 1996. 206~217
- 5 Adelberg B, et al. Emulating Soft Real-Time Scheduling Using Traditional Os Scheduler: [Technical Report]. Stanford University, 1994
- 6 Demer A, et al. Analysis and Simulation of a Fair Queuing Algorithm. In: Proc. of ACM SIGCOMM, 1989. 1~12
- 7 Stocia I, et al. A Proportional Share Resource Allocation Algorithm for Real-Time, Time-shared System. In: IEEE Real-Time System Symposium, Dec 1996
- 8 Sabata B, et al. Taxonomy for QoS Specifications. In: Proc. of IEEE Computer Society 3rd Intl. Workshop on Object-oriented Real-time Dependable Systems (WORDS '97), Newport Beach, California, 1997
- 9 Sabata B, et al. Hierarchical Modeling of Systems for QoS based Distributed Resource Management: [Technical Report]. SRI International, Menlo Park, California 1.. 54