

利用网格技术实现流媒体服务迁移^{*})

周 俊 李文中 陆桑璐 陈道蓄

(南京大学计算机软件新技术国家重点实验室 计算机科学与技术系 南京 210093)

摘 要 流媒体服务正在成为当今网络服务的代表性应用。为了提高流媒体服务的可用性,保证服务质量,快速高效的流迁移是研究中的重点和难点。本文探讨了流媒体服务迁移面临的关键问题,利用网格技术在实现流媒体服务迁移方面的优势。结合 RTSP 会话和流媒体传输监控技术,文中在自主研发的流媒体服务网格 NDSSG 中实现了一个流媒体服务迁移系统,解决了因服务器故障造成的服务不可用问题以及可能的负载不平衡问题。在 NDSSG 中的实践证明:利用先进的网格技术可以实现高性能的流媒体服务迁移,可以有效提高流媒体服务的性能和质量。

关键词 网格技术,流媒体服务迁移,RTSP

Implementing Hand-off of the Streaming Media Service with Grid Technology

ZHOU Jun LI Wen-Zhong LU Sang-Lu CEHN Dao-Xu

(Department of Computer Science and Technology, Nanjing University, Nanjing 210093)

Abstract Streaming media service is getting to be a typical application among today's network services. In order to improve the availability of the streaming media service and guarantee the quality of service, it is important but difficult to make the stream hand-off quick and efficient. This paper simply introduces the key problems which the hand-off of streaming media service is facing and the advantage of the grid technology to implement the hand-off of streaming media service. Combining RTSP sessions and the monitoring technology on streaming media transmission, it implements a hand-off system of the streaming media service in the Globus-based streaming media service grid NDSSG, which can support the hand-off that occurs when the server breaks down or load-balancing happens. It proves by the practice in NDSSG that applying the advanced grid technology can implement highly efficient hand-off of the streaming media service, can effectively improve the performance and quality of the streaming media service.

Keywords Grid technology, Hand-off of the streaming media service, RTSP

1 引言

进入 20 世纪 90 年代以后,Internet 高速发展,已经从文字时代进入了多媒体时代。大量的影音数据在网络上传播,流媒体服务应运而生。流媒体服务具有数据量大,传输速率要求高,低延时,高稳定等特点,因此对网络的 QoS 要求非常高。流媒体迁移是流媒体服务迁移的简称,是流媒体技术中的一个重要部分,是移动终端服务,负载均衡,系统容错和服务透明性等领域的关键技术。流媒体服务迁移的研究包括判断流媒体服务何时需要迁移,如何寻找迁移的目的地,如何与目的地协作启动迁移任务,迁移哪些内容等问题,这项技术的最高要求是实现无缝迁移,即在客户不会发觉的情况下实现迁移。传统上,由于流媒体迁移的网络开销很大,对网络的 QoS 要求太高,难以做到无缝迁移。

网格是当今分布式计算领域的一个前沿,被称为是新一代的 Internet。对于网格,目前学术界还没有一个明确的定义,一般认为网格就是一个集成的计算与资源环境,或者说是一个计算资源池。网格将能够充分吸纳各种计算资源,并将它们转化成一种随处可得的、可靠的、标准的同时还是经济的计算能力。这里的计算资源除传统意义上的各型计算机外,还包括网络通信能力、数据资料、仪器设备,甚至是人等各种相关的资源^[5]。所谓网格计算就是将地理上分布的、异构的各种高性能计算资源、存储资源、数据资源和其它特殊资源通过高速网络连接起来,实现高性能联合计算,共同完成重大应

用问题,即基于网格的问题求解。网格研究致力于用高速网络将强力的计算资源连接起来为用户提供透明的超级计算环境,由于网格可以将计算资源尽可能地利用起来,因此它的结构及采用和开发的技术如果用于现有资源的互联和现有网络的改造上就可望大幅提高现有网络的 QoS,这一点对于传统上对 QoS 要求较高的网络服务非常有意义。

我们在长期对分布式计算和应用进行研究的基础上,选择流媒体服务作为应用背景,设计并开发了 NDSSG(南大流媒体服务网格),以求在不升级服务器的情况下尽可能提高整个服务网络的服务能力,提高整个流媒体服务的服务质量。本文将重点介绍 NDSSG 中流媒体服务迁移的设计和实现。本文第 2 节介绍研究背景和相关工作研究,第 3、4、5 节介绍我们在流媒体网格中实现流媒体服务迁移的详细工作,最后是总结。

2 网格的发展与现状

2.1 背景与相关工作研究

现代科学与工程问题的计算量越来越大,基因工程、量子力学、天文学等等许多领域都渴望获得高性能的计算能力。同时,新兴交叉学科的出现使得参与计算过程的学科和系统也日趋复杂,以往依靠单台计算机提供计算能力的方法已经难以满足需要,多台高性能计算机协同、多个系统合作、多种科学实验仪器设备相连的网络虚拟超级计算机被认为是解决问题的希望;而另一方面计算机尤其是微机的硬件发展很快,

^{*}) 本文得到国家八六三技术研究发展计划(编号 2001AA113050)资助。周 俊 硕士研究生,主要研究方向为分布式计算。李文中 硕士研究生,主要研究方向为分布式计算。陆桑璐 教授,主要研究领域为分布与并行计算。陈道蓄 教授,博士生导师,主要研究领域为分布计算与并行处理。

网络上存在大量闲置的计算能力。20 世纪 90 年代初,根据 Internet 上主机大量增加但利用率并不高的情况,NFS 将其四个超级计算中心构筑成一个元计算机,逐渐发展到利用它研究解决具有重大挑战性的并行问题,这被认为是网格计算的雏形。与传统的分布式计算不同的是网格计算关注的是涉及大规模资源共享的、新兴的应用,以及某些情况下的高性能方向^[3],如何在动态的、多机构的虚拟组织(Virtual Organization)中实现灵活、安全、协同的资源共享和问题求解是它的研究内容。这里提到的资源共享并不是简单的文件交换,而是指直接访问计算机、应用软件、数据库及其它资源,在我们的流媒体服务网格中资源主要是指存放在服务器上的影片以及服务器的存储空间和网络带宽。

因为利用网格技术来提高现有网络的服务能力和质量必须做到在尽可能不触动各个资源自主性和网络协议的情况下实现异构屏蔽性、可扩展性和动态适应性,所以网格系统的研究重点集中在了中间件技术上,甚至有一种看法认为网格就是中间件。

网格研究有很多具体的实现或实验平台,其中最具有代表性的是 Globus 系统^[2]。Globus 2.x 版由三大单元组成:GRAM(Grid Resource Allocation and Management),MDS(Meta Directory Service),GridFTP(Grid Data Management)。此外这三大单元还共同基于一个分布式安全设施 GSI(Grid Security Infrastructure)^[1]。其中 MDS 是信息的收集、管理和查询机构,它好比是一个 Grid 系统的神经系统,运行在各台服务器上的信息提供程序(或脚本)将信息汇总到其所属的 GRIS(Grid Resource Information Service)数据库中,GRIS 数据库再进一步将信息汇总到 GIIS(Grid Index Information Service)数据库中,这样系统或用户就可以方便地查询到所需的信息,如图 1 所示。MDS 是工作在 LDAP(Lightweight Directory Access Protocol)上的,采用了动态层次结构。

Globus 系统目前已经发展到了 3.x 版本,新的 Globus 系统引入了 Grid Services 的概念将各功能模块都看作是“服务”,它采用 Web Service 的协议(WSDL, SOAP, UDDI 等)规范 Globus 中的定义、描述、注册、查询和消息交互,称之为 OGSA(Open Grid Service Architecture)^[4]。

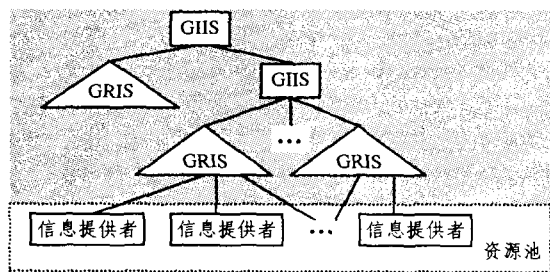


图 1 Globus 2.x MDS 的组织结构

2.2 流媒体服务迁移面临的问题与研究现状

近年来随着 Internet 的发展,网络已经逐渐代替电视成为人们获取信息的第一渠道。在网络上传播的信息和服务已经不再是简单的文本而是内容丰富、更容易被人们所接受的图片、视频、音乐等媒体信息,多媒体技术的迅速发展正在为人们展现一个全新的虚拟世界。信息高速公路的普及使得以 VoD 和视频会议为代表的流媒体服务在网络应用中得以迅速发展。流媒体服务是指媒体数据采用字节流的方式在网络上传输的服务,数据来源既可以是文件也可以是数字设备,它具有数据量大、带宽占用高、传输要求稳定、服务时间长等特点。

点。由于要长时间保持“大、快、稳”的服务质量,因此对网络服务能力的要求极高。网络不同于电视,它具有开放性、多功能性和复杂的异构性,单纯依靠升级硬件来提高网络的服务质量代价高昂而且很多问题也无法解决,因此研究关于流媒体服务的软件技术非常必要,流媒体服务迁移便是其中一项关键技术。

流媒体迁移技术直接影响流媒体服务系统的服务质量,它可以有效提高系统的容错能力,负载均衡性,服务终端的移动性和服务透明性。流媒体服务迁移根据不同的标准可以有多种分类,如由于服务器失效和网络分割迫使系统进行的被动迁移和由于系统为了均衡负载而发起的主动迁移,根据迁移发起者的不同可分为服务器发起、客户端发起或第三方发起三种,根据迁移发起的层次可以分为传输层、应用层、用户层迁移,此外根据系统是否可以预知迁移的目的地还可以分为前摄式迁移和反应式迁移。

无论是根据什么标准,无论是哪一种流媒体服务迁移,都必须解决一些关键问题:

- a. 如何判断须要引发流媒体服务迁移;
- b. 流媒体服务迁移由谁发起;
- c. 如何寻找服务迁移的目的地(或服务的替换者);
- d. 如何迁移;d. 1. 迁移什么内容;d. 2. 从什么地方开始迁移;d. 3. 参与迁移的客户端、服务器或中间件如何协同合作;d. 4. 如何优化服务迁移,以较低的开销尽可能实现无缝迁移。

一般研究流媒体服务迁移时都可以认为客户只关心所获服务的质量而并不关心是谁在为他服务。由于 TCP 传输附带很多保障可靠性的开销且不易做流量控制,因此流媒体服务多采用 UDP 传输媒体数据。为了便于测试和评估迁移工作的效果,人们主要研究的是应用层的迁移(传输层迁移涉及网络底层,而用户层迁移研究的是服务在不同的用户终端间的迁移);同时为了尽可能避免对现有的网络设施和服务器、客户端提出特殊的要求,因此许多研究者选择了研究和开发中间件来支持流媒体服务迁移。

在当今的流媒体服务系统中,客户端一般都配备一个适应性缓冲区,这个缓冲区至少具有三个作用:平滑数据输入,重组收到的数据包,当发现收到的数据包丢失率过高时向服务系统申请发起迁移。此外由于对服务质量的要求是由用户提出的,因此一般都是由客户端向服务器发送服务控制信息,如流量控制信息。在网络条件好的情况下,流量控制加上一定的缓存就可以达到资源预留的效果^[10]。进一步,客户端还可以获取用户正在享受的服务的具体情况,生成描述具体服务的状态记录^[11],对服务的迁移提供支持。

流媒体服务系统中一般都有许多台高性能的服务器,因为单台服务器很难满足大量用户的并发请求。由于服务器自身需要尽可能地将服务能力应用在为客户提供数据上,因此系统应该尽量减轻服务器的负担。有一种服务器维护的流媒体服务迁移很具有代表性^[10],系统内的所有服务器会定期广播自己的运行状态(包括有哪些客户,这些客户的信息和为其服务的状况),当一台服务器发生故障没有按时广播自己的状态时,其他服务器会自动根据上个周期对这台服务器的了解来为它的客户提供替代服务。显然这一设计造成的系统开销比较大而且各服务器间的协作较难实现,系统的可扩展性也受到了制约。服务器有时也会发起迁移,如重组组播树,但一般而言服务迁移前后对于服务器来说可以看作是两次不同的服务请求,不必额外考虑。

由中间件或服务的第三方来发起服务迁移是当前流媒体

服务系统中最为普遍的情况,因为这是最佳的选择,这样的迁移可以提高服务的透明性,而且可操作性更好^[8]。网络监视设施发现拥塞时发起的迁移,服务监控设施发现某台服务器死机、过载或新加入时发起的迁移,用户地点检测设施发现用户移动后发起的迁移和数据转发设施发现用户更改服务质量要求时发起的迁移等都属于这类迁移。

已经实现的流媒体服务迁移有很多,但性能都无法令人满意。从以往的研究工作中我们可以总结出制约流媒体服务迁移的几个环节:迁移信息的维护需要兼顾系统开销和迁移性能;系统结构中要存在便于监测服务的位置;系统资源能够被合理地量化或抽象为服务能力;服务器可以自由地加入和退出系统。在下文中我们可以看到网格技术是如何在这些环节上提高流媒体服务迁移的性能的。

2.3 流媒体服务相关协议

顺应流媒体服务的兴起,一系列与流媒体服务相适应的传输层和应用层协议相继出现,其中对于本文工作最重要的协议是 RTSP(Real Time Streaming Protocol, RFC2326)。

RTSP^[7]是一个用以控制具有实时属性的数据传递的应用层协议。RTSP 提供了一个可扩展的框架,这个框架能够实现受控的、按需传递的实时数据,如音频和视频等。这个协议为的是控制多数据传送会话,提供一种途径来选择传送渠道,如 UDP、组播 UDP 和 TCP,以及提供一个途径来选择基于 RTP^[6]的传送机制。RTSP 负责建立和控制单个或若干时间同步的连续流,但是其本身并不直接传送这些流,而是通过生成一个控制流来工作,如同实现了对多媒体服务器的“网络远程控制”。它支持的操作包括:从媒体服务器上检索媒体,要求一个媒体服务器加入一个会议,在已经存在的流中加入一个新的媒体。

RTSP 控制的是由另一个协议发送的、独立于控制通道的媒体数据流,譬如 RTSP 控制可以使用 TCP 连接而媒体流则使用 UDP。因此,即使媒体服务器收不到 RTSP 请求仍然可能继续发出数据,而一个正在服务的媒体数据流也可能顺序地被来自不同 TCP 连接的 RTSP 请求控制。所以,服务器须要维护“会话状态”来将 RTSP 请求关联到一个媒体数据流。

RTSP 的控制能力非常强大,因为它定义了一批与流媒体服务控制相关的方法,这些方法包括: DESCRIBE、ANNOUNCE、GET_PARAMETER、OPTIONS、PAUSE、PLAY、RECORD、REDIRECT、SETUP、SET_PARAMETER、TEARDOWN。这些方法有些是必须的(如 PLAY),有些则是可选的(如 GET_PARAMETER),但是多数可选的方法对优化流媒体服务和流媒体服务迁移来说却是必要的,譬如 DESCRIBE 方法可以得到某个流媒体服务的详细静态信息(包括名称、格式、长度,由几个流合成等),GET_PARAMETER 方法可以获得客户端、服务器和数据流的动态信息(如所占带宽、偏移等),SET_PARAMETER 的作用则是动态设置服务中的各项参数,而 REDIRECT 方法则使得服务器可以直接要求客户端更换服务器。不仅如此,应用开发者还可以根据具体须要进行扩展,如增加某个方法的参数或增加新的方法等,这非常重要,在流媒体服务网格 NDSSG 中,我们选择了 Real Server 实现的 RTSP 作为研究流媒体服务迁移的环境。

3 流媒体服务网格 NDSSG

我们设计并开发了一个流媒体服务网格 NDSSG,目标是依靠现有的网格技术以及我们自身研究网格和流媒体服务的

成果开发一个高性能、可扩展性强、多服务器协同的流媒体服务系统,从而实现“利用网格技术提高网络性能”。NDSSG 的基本结构如图 2 所示。

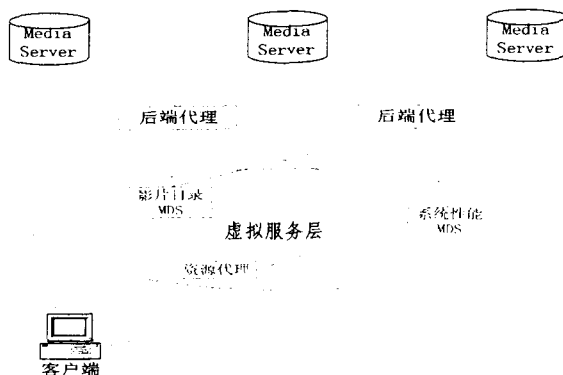


图2 NDSSG 全局结构图

整个系统分为两层:资源池层和虚拟服务层。资源池层主要包括媒体服务器和后端代理的后台模块,是系统中真正提供流媒体数据,向客户提供流媒体服务的部分,由于后端代理的屏蔽作用这部分对用户是不可见的,其中每个后端代理屏蔽的服务器群在系统内就表现为一个统一的、整体的服务节点。虚拟服务层主要包括资源代理、影片目录 MDS、系统性能 MDS 和后端代理的前台模块,是系统中调度用户请求、与用户进行直接交互的部分,对用户是可见的。显然后端代理在整个系统中是一个关键,它横跨两层,屏蔽后台服务器差异,监控后台服务器群的运行情况,是流媒体迁移中不可避免的一个环节。后端代理有两个最基本的功能:转发流媒体数据和介入流媒体服务会话,前者是为了屏蔽不同媒体服务器间的差异,后者则是为了实现对流媒体服务的控制。凡是符合系统规范(如服务端口等)的媒体服务器和客户端都可以接入和访问 NDSSG,不过为了更好地提高流媒体迁移的效果,我们选择 RTSP 协议作为客户端和媒体服务器间的交互协议(通过后端代理)。

在 NDSSG 中,一个流从请求到建立的过程如下:客户端首先向资源代理提交服务请求,资源代理从请求中获取影片标号和对服务质量的要求,然后查询影片目录 MDS 和系统性能 MDS,选择一个可以为用户提供服务同时接入成功率又是最高的节点将它的网络地址返还给客户端,客户端将通过后端代理的中转来和媒体服务器进行 RTSP 交互,当 RTSP 交互经过 SETUP 和 PLAY 阶段后媒体数据流被建立并开始传输数据包,一个媒体流就建立起来了(一个媒体流中包含了 RTSP 交互和媒体数据流两个会话)。

4 NDSSG 对流媒体服务迁移的功能支持

4.1 流媒体服务迁移的分类

在 NDSSG 中我们总结出了四种流媒体服务迁移,按照迁移发生的原因分为:

(1)失效迁移:当某个服务器无法继续运行时,系统将把该机服务的流迁移到节点内其它服务器上;

(2)过载迁移:当某个服务器发生过载时,系统将把某些流从该机迁移到节点内的其他服务器上;

(3)增容迁移:当一个新的服务器加入某个服务器群时,系统将把某些流从节点内其他服务器迁移过来;

(4)跨群迁移:指服务流从一个机群迁移到另一个机群,当一个节点内某个流无法正常运行且该节点内又没有合适的

服务器可以迁移时,把这个流迁移到另一个节点。

在这几种迁移中,失效迁移、过载迁移和增容迁移都是节点内部迁移,跨群迁移为节点间的迁移,考虑到 Internet 的规模节点间的迁移代价相对要大得多,因此真正对提高系统性能有帮助的应是改善节点内部迁移。为了在我们的 NDSSG 中实现这四种流迁移,系统必须具备某些功能,根据本文第 3 节的总结我们可以按照流迁移必须解决的关键问题来引出这些功能。

4.2 流媒体服务迁移要求的系统功能

首先,系统必须要在客户觉察前发现流媒体服务须要迁移的原因。失效迁移要求服务器发出的数据经过第三方转发或监视后才到达客户端,这样第三方将可以在客户端之前发现服务器无法到达;过载迁移更是要求监视流传输的第三方能够判断、发现甚至是预测服务器发生过载;增容迁移要求系统具备成员管理机制,同时又可以对同一节点内所有的流进行监控;跨群迁移要求后端代理具备和资源代理交互的能力,而且可以通知或控制客户端改变流服务的请求目标。

其次是由谁发起流媒体服务迁移。在这个问题上为了缩短流迁移的完成时间,我们确定了谁发现谁发起的原则。

再其次,系统须要为流媒体服务发现、创造和查询、选择迁移目的地或服务替代者。这就要求系统具备资源搜索和发现功能、资源注册功能、副本管理功能、信息收集、维护和查询功能,以及时间短、成功率高的调度功能。

最后是迁移的内容和起始点。从迁移的代价来看,在迁移数据流、控制流和服务的状态记录三者中迁移数据流的代价无疑是最大的,而迁移控制流则很难保证客户端和目标服务器之间的同步性,因此我们选择了流媒体服务的状态记录作为迁移内容。这就要求系统能够很好地定义和规范流媒体服务的状态记录的格式,监视和记录流媒体服务的状态信息,维护这些信息并保证它们的实时性和一致性,提供方便快捷的查询。此外,系统还必须能够获取流媒体服务中数据传输的实时偏移。

4.3 支持流媒体服务迁移的 NDSSG 各模块

对流媒体服务迁移要求的系统功能的实现质量直接决定了一个流媒体服务系统中流迁移的性能。在研究网络结构和技术的我们发现一般网络尤其是 Globus 平台的两级结构,资源池层的组织和对应用服务器异构性的屏蔽,网络中的资源发现,信息服务,副本管理,作业调度与管理等技术可以给实现上述功能提供强大的支持。NDSSG 是基于 Globus 平台构建的流媒体服务网格系统,因此 NDSSG 也相应地分为资源池层和虚拟服务层两个层次。从功能模块上看,NDSSG 分为后端代理、信息服务、调度管理、副本管理等几个功能模块,它们共同提供了最基本的流媒体服务支持,包括对流媒体服务迁移的支持,下面会详细说明。再结合我们对流媒体服务协议和流媒体数据传输方面的研究,NDSSG 很好地实现了这四种流媒体迁移。

4.3.1 后端代理模块 网格的最大特点就是要向用户提供统一的、标准的服务,因此网格的虚拟服务层必须能够屏蔽资源池层内的异构性,这就要求后端代理必须成为访问节点的统一门户,这样的结构也使得后端代理成为监控流媒体服务的最佳位置。

NDSSG 的后端代理模块由数据包转发模块、日志记录模块和控制台模块三个子模块组成。数据包转发模块是一个可以加载到 Red Hat Linux 操作系统中的内核模块,它可以根据需要在运行主机上开辟两个端口,分别负责与一个流媒体服务两端的媒体服务器和客户端交互并在两个端口之间转发数

据包。日志记录模块负责记录若干时间窗口内经由数据转发模块的各个流的传输信息,包括协议、端口、速率等等,并且将这些信息实时地发送给控制台模块。控制台模块从日志记录模块读取各个流的传输信息后将完成一系列的统计工作,包括统计当前监控的流的总数,将这些流进行分类并计算接入某个分类的一个流的系统开销,判断接入某个流后是否会使得系统过载(即影响到正在运行的流的传输),并根据历史记录估算一个流的接入概率。经过这些量化工作,节点及节点内各服务器的服务能力便被抽象出来,从而为调度模块提供了调度依据。

由于后端代理模块监视着所有接入流的状态,因此它可以发现系统中发生的失效迁移和过载迁移,一旦这两种迁移发生,便可以由它发起流媒体迁移。

4.3.2 信息服务模块 是 NDSSG 中一个最为重要的功能模块,是整个流媒体网格的神经。NDSSG 实现了一个基于 Globus 的流媒体网格信息服务系统,在 Globus2.x 的 MDS 基础上开发了一个两级信息服务模块,结构如图 1 所示,其中 GRIS 数据库分布在后端代理上,主要存放各个节点内服务器发来的资源注册信息,而 GIIS 数据库则分布在虚拟服务层中,主要存放影片目录信息和系统性能信息。在 GRIS 数据库中信息分别按照影片和服务器分类存放,按影片存放的信息包括了影片在节点内的标号、影片质量和影片的 URL 等,按服务器存放的信息则包括了服务器在节点内的标号、服务器的 IP 地址和服务器的最大服务能力等。在 GIIS 数据库中影片目录信息是从各个 GRIS 的影片信息汇总而来的,主要是将重复的项目合并为带副本信息的同一项目并对影片的标号进行了重新编排,这些信息是触发式更新的,即当某个节点的影片信息发生变更时产生信号进行刷新;而系统性能信息则是向各个后端代理中的控制台模块获取的,主要是各个节点当前的总负载和某个流的接入概率,这些信息采用时钟过时更新,即每个信息项都有一个不断减小的生命时钟,如果时钟不为零则该项信息有效,如果时钟为零则刷新该项并重置时钟。与 Globus 系统相同,GRIS 和 GIIS 之间的信息传递和对两个数据库的查询是通过 LDAP 来完成的,而且两级数据库的结构是动态可扩展的,新的 GRIS 数据库可以方便地随新节点的加入而扩展到信息服务系统中。在信息查询方面,NDSSG 引入了缓存技术和协同查找技术,与 Globus2.x 的 MDS 相比查询速度取得了极大的提高。所有这些特点使得我们的系统可以为提高迁移性能而维护更多、更实时的信息,同时又不会在系统中造成太大的开销。

信息服务模块的主要功能是网格资源信息的注册、发现和维护,所以信息服务模块可以发现系统中的增容迁移和跨群迁移。同时,信息服务模块也是流媒体迁移所需的服务状态记录最佳的搜集者、维护者和提供者,方便高效的信息查询接口使它可以为流媒体服务迁移提供各种信息作为决策的依据。

4.3.3 调度模块 快速而且成功率高的流调度是提高流媒体迁移性能的一个重要方面。在 NDSSG 中调度分为两级,第一级是发生在虚拟服务层的调度,调度目标是选择一个合适的节点接受客户提出的流媒体服务请求,由资源代理完成;第二级是发生在节点内部,调度目标是选择一个合适的媒体服务器提供所要求的流媒体服务,由后端代理完成。第一级调度的依据是虚拟服务层中的影片目录信息和系统性能信息,第二级调度的依据则是节点内部的影片信息和后端代理中控制台的统计信息,可见第一级调度主要依靠静态信息调度同时兼顾了一定的实时性以求在松耦合的 Internet 环境中

取得较快的调度速度,而第二级调度则主要依靠动态信息调度从而在调度开销可以接受的情况下保证了调度的成功率。

调度模块负责为流媒体服务选择迁移的目的地,在迁移流媒体服务时,总是要根据调度模块的调度结果寻找合适的替代服务器。

4.3.4 副本管理模块 流媒体网格中,同一个媒体文件往往有多个副本,需要一个副本管理模块来统一管理。以 Globus2.x 中的副本管理模块为基础,以影片的热门程度和地理分布合理度为依据,开发出了 NDSSG 的副本管理系统,它根据影片的热门程度来判断是否要生成副本以及生成多少副本,然后根据流媒体服务请求的地域特征来分布这些副本,最后借助 Globus Toolkits 中的 GridFTP 在系统负荷较低的时段中(如凌晨)将这些副本分散存放到须要拥有这些副本的媒体服务器上。

副本管理模块对副本的数量和副本的分布进行统一的管理,它是流媒体服务能否成功迁移的前提,因为用户对某个媒体文件的服务请求总是被迁移到该文件的一个副本上。

综上所述,在 NDSSG 中,信息服务模块为流媒体服务迁移提供了依据,而后端代理、信息服务、调度管理、副本管理等几个功能模块则为流媒体服务迁移的发现和决策提供了相应的支持。

5 NDSSG 中流媒体服务迁移的实现

当某台 Real Server 因为退出节点或死机而导致无法继续服务时,后端代理中的数据包转发模块将会发现某些流的媒体服务器中断了数据包的发送而发起失效迁移。后端代理会模仿服务器端发送一个含 GET_PARAMETER 方法的 RTSP 消息给客户端播放器查询媒体流的播放偏移(格式请参考 RFC2326,下同),同时向信息服务模块查询这个流服务的状态信息(如媒体文件标号、质量要求等),然后调用二级调度模块寻找合适的媒体服务器作为迁移目标,最后向目标服务器发起 RTSP 会话请求提供相应偏移之后的流媒体数据,在新的流建立起来后将新服务器的地址交给数据包转发模块并更新信息服务模块中的记录,完成迁移。

当后端代理的控制台模块发现某台服务器发生过载时(一般由传输波动造成)它就会发起过载迁移。它首先查询控制台模块该迁移哪些流,以降低该服务器的负载,然后向信息服务模块查询这些流的状态记录,调用二级调度寻找合适的迁移目标,模仿客户端播放器发送 TEARDOWN 消息给原服务器终止服务,最后向目标服务器发起 RTSP 会话完成迁移(后续步骤和失效迁移相同)。

当一台新的 Real Server 注册加入某个服务节点时,信息服务模块会发起扩容迁移并通知控制台模块,控制台会查找当前节点内哪些流可以迁移到新加入的服务器上,并判断哪些流值得迁移到新加入的服务器上,然后通知后端代理,其后的迁移过程将和过载迁移相似。

跨群迁移往往在节点内部迁移找不到合适的媒体服务器作为迁移目标时发生。这时后端代理会向资源代理发送相应的流媒体服务调度请求,资源代理会查询信息服务模块中的影片目录和系统性能并通过一定的算法为该服务流调度一个新的后端代理,当原后端代理收到调度结果后会发送一个 REDIRECT 消息给客户端播放器,这个消息中包含了目标后端代理的地址和当前的影片播放迁移。

上述几种迁移其实有着统一的流程。如果我们把后端代理、信息服务模块和调度模块中与流媒体服务迁移相关的功能在逻辑上重新组织,那么就得到如图 3 所示的迁移流程。

其中,迁移监控模块负责捕捉系统中引发流媒体服务迁移的各种事件并发起迁移,在具体实现中该模块的功能被分散在后端代理和信息服务模块中;迁移信息服务模块负责为流媒体迁移提供所需的相关信息(如流媒体服务状态记录),实现上它是信息服务模块和控制台模块的一部分。迁移策略库包含多种迁移策略,这些策略包含了选择什么样的流进行迁移或选择什么样的服务器为迁移目标等设定,实现时只制定了一条迁移策略;迁移调度模块负责为流媒体迁移查找和选择一个媒体服务器作为迁移目标,在实际系统中它由节点内部调度模块替代。

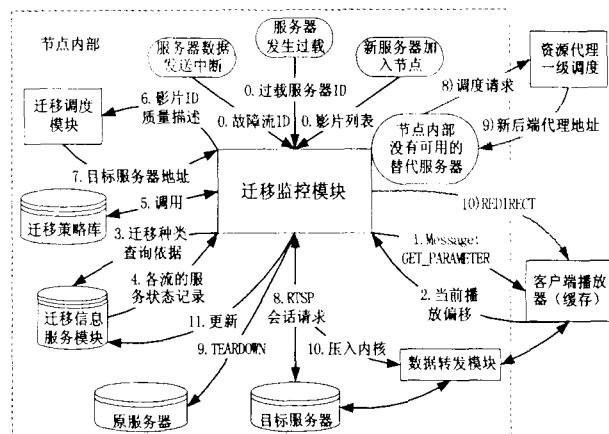


图3 流媒体迁移流程图

NDSSG 系统流迁移的流程可以描述如下:

(1) 迁移监控模块监控媒体流的服务状态,当失效迁移、过载迁移、扩容迁移和跨群迁移发生并抛出相应的系统事件时,它会捕捉到这些事件并发起流媒体迁移;

(2) 迁移监控模块从客户端播放器获取媒体流的当前播放偏移,然后向迁移信息服务模块查询迁移所需的有关流的服务状态记录(媒体文件 ID 和用户请求的服务质量)并根据查询结果从迁移策略库中选择一种迁移策略;

(3) 迁移监控模块根据迁移策略请求迁移调度模块提供替代服务器作为迁移的目标;

(4) 如果在本节点内能够找到合适的替代服务器则向它发起 RTSP 会话请求,当媒体流建立起来以后就终止原服务器的服务并将把流压入内核中的数据转发模块,最后更新信息服务模块;如果节点内部没有可用的替代服务器,那么就向资源代理发起流媒体服务调度请求,然后通知客户端重定向到新节点上。

跨群迁移其实还有一种特殊情况:某个后端代理发生故障无法继续服务,这时就不能由后端代理来发现和完成迁移了。解决的方法是由客户端缓存来发现服务器端中断数据包的发送而发起迁移,由它模仿服务器端发送一个含 GET_PARAMETER 方法的 RTSP 消息给客户端播放器查询影片播放的当前偏移,然后向资源代理重新发起流媒体服务请求。从资源代理的角度上看,它将为客户端建立一个新的流媒体服务。

为了完成上述几种迁移的实现步骤,有几点具体工作是值得注意的。客户端的缓存既要具备在客户端播放器和虚拟服务层间转发数据的功能,也要具备在迁移时合并收到的重复数据包的功能^[9]。实验证明 Real Server 和 Real Player 之间的媒体数据流没有经过加密是可以介入管理的。Real

(下转第 137 页)

改进提供一定的指导。在分析的基础上,给出了“群核”的概念,并提出了群核进化粒子群优化方法(SCEPSO),这种方法算法相对稳定,即在 50 次运行中所得最好结果和最差结果相差不多,同时,与文[14]提供的数据比较,在计算条件相同的前提下,计算结果有很大提高,这表明:SCEPSO 方法能够较好地平衡局部“搜索”和“开拓”新领域之间的关系,在解决复杂非线性优化问题上,有较强的应用价值。

参考文献

- 1 Eberhart R C, Kennedy J. A new optimizer using particle swarm theory. In: Proc. of the sixth Intl. Symposium on Micro Machine and Human Science. Nagoya Japan, 1995, 39~43
- 2 Kennedy J, Eberhart R C. Particle Swarm Optimization. In: Proc. IEEE Intl. Conf. on Neural Networks. IEEE Service Center, Piscataway, NJ, IV, 1942~1948
- 3 Hirotaka, Yoshida, Kenichi. A particle Swarm Optimization for Reactive Power and Voltage Control Considering Voltage Stability. In: IEEE Intl. Conf. on Intelligent System Applications to Power Systems (ISAP'99) Rio de Janeiro, April, 1999
- 4 Voss M S, Feng X. Arma Model Selection Using Particle Swarm Optimization and Aic Criteria. In: 2002 15th Triennial World Congress, Barcelona, Spain
- 5 Parsopoulos K E, Vrahatis M N. Particle Swarm Optimization Method in Multiobjective Problems. In: Proc. of the 2002 ACM Symposium on Applied Computing (SAC 2002), 2002, 603~607
- 6 van den Bergh F, Engelbrecht A P. Cooperative Learning in Neural Networks using Particle Swarm Optimizers. South African Computer Journal, 2000, 26: 84~90
- 7 Clerc M. TRIBES-Aparameter Free Particle Swarm Optimizer. <http://clerc.maurice.free.fr/PSO> 2002
- 8 Hu Xiaohui, Eberhart R C. Adaptive Particle Swarm Optimiza-

tion; Detection and Response to Dynamic Systems. Computational Intelligence, Hawaii, IEEE Press, May, 2002

- 9 Salman A. Discrete Particle Swarm Optimization for Heterogeneous Task Assignment Problem. In: Proc. of World Multiconference on Systemics, Cybernetics and Informatics Cybernetics And Informatics (SCI 2001)
- 10 Clerc M. Discrete Particle Swarm Optimization; A Fuzzy Combinatorial Black Box. <http://clerc.maurice.free.fr/PSO/Fuzzy-Discrete-PSO/Fuzzy-DPSO.htm> 2000
- 11 Eberhart R C, Shi Y. Comparison between Genetic Algorithms and Particle Swarm Optimization. In: Proc. of the Seventh Annual Conf. on Evolutionary Programming. Springer Verlag, 1998, VII: 611~618
- 12 Angeline P J. Evolutionary Optimization Versus Particle Swarm Optimization; Philosophy and Performance Differences. In: Proc. of the Seventh Annual Conf. on Evolutionary Programming, Springer Verlag, 1998, VII: 601~610
- 13 Parsopoulos K E, Vrahatis M N. Recent approaches to global optimization problems through Particle Swarm Optimization. Natural Computing, 2002(1): 235~306
- 14 Clerc M, Kennedy J. The Particle Swarm: Explosion, Stability, and Convergence in a Multi-Dimensional Complex Space. IEEE Transactions on Evolutionary Computation, 2002, 58~73
- 15 Clerc M. Particle Swarm Optimization and Information. <http://clerc.maurice.free.fr/psa/> 2002-08-10/2003-10-08 2002
- 16 Clerc M. Some math about Particle Swarm Optimization. <http://clerc.maurice.free.fr/psa/> 2002-08-10/2003-10-08
- 17 van den Bergh F, Engelbrecht A P. Effects of Swarm Size on Cooperative Particle Swarm Optimisers. In: Proc. of the Genetic and Evolutionary Computation Conf. (GECCO)-2001
- 18 van den Bergh F. An analysis of Particle Swarm Optimizers. [Ph D dissertation]. Pretoria: University of Pretoria, 2001

(上接第 113 页)

Server 对 RTSP 的支持并不完整,有些方法没有实现,而且和 Real Player 间的 RTSP 会话受到了挑战数的保护(RN5 加密)^[12],所以在介入这两者间的 RTSP 会话前必须要遵循挑战数加密机制,或者通过特殊的方法避免挑战数的影响。

在 NDSSG 的实验系统上做的测试证明本文考虑的几种流媒体迁移都可以实现,尤其是三种节点内部迁移性能相当好,在客户端缓存足够大(不小于 2MB)的情况下可以实现无缝迁移,即在客户没有察觉的情况下完成流媒体服务迁移。此外在引入分类服务的思想后,尽量选择优先级低和质量差的流媒体服务进行迁移,会让系统获得更好的整体性能。

由于信息服务模块的支持和依靠尽可能介入 RTSP 会话来实现对流媒体传输的控制,在 NDSSG 中迁移一个流的系统开销被控制得很小,不容易影响到其他正常运行的流媒体服务,保证了整个系统的平稳运行。

总结 对网格的研究为我们提供了新的组织结构和许多新的技术,合理地运用这些技术可以帮助我们改造现有网络系统,在不升级硬件设备的情况下获得更高的性能,也可以帮助我们在设计开发新的分布式系统时优化系统的服务质量。流媒体服务是目前流行的网络服务中对 QoS 要求最高的,而流媒体服务迁移所需的系统开销更是让人望而生畏。在 NDSSG 中,我们成功地实现了流程统一的失效迁移、过载迁移、扩容迁移和跨群迁移等流媒体服务迁移。这不仅体现了网格技术对流媒体服务的有力支持,而且体现了网格技术在

提高网络服务的性能和质量方面的重要作用,显示了网格技术对于优化和推广各项网络应用的积极意义。

参考文献

- 1 The International Technical Support Organization, Austin Center. Introduction to Grid Computing with Globus. IBM References, Dec. 2002
- 2 The Globus Project. <http://www.globus.org>
- 3 Foster I, Kesselman C, Tuecke S. The Anatomy of the Grid: Enabling Scalable Virtual Organizations. Intl. Journal of Supercomputing Applications, 2001
- 4 Foster I, et al. The physiology of the grid: An open grid services architecture for distributed systems integration. [Technical report]. Open Grid Service Infrastructure WG, Global Grid Forum, June 2002
- 5 都志辉,陈渝,刘鹏. 网格计算. 清华大学出版社, 2002
- 6 Schulzrinne H, et al. RTP: A Transport Protocol for Real-Time Applications. IETF RFC1889, Jan. 1996
- 7 Schulzrinne H, et al. RTSP: Real Time Streaming Protocol. IETF RFC2326, June 2003
- 8 Roy S, Shen B, Sundaram V. Application Level Hand-off Support for Mobile Media Transcoding Sessions. NOSSDAV'02, May
- 9 Karrer R, Gross T. Dynamic Handoff of Multimedia Streams. 11th NOSSDAV, June 2001
- 10 Anker T, Dolev D, Keidar I. Fault Tolerant Video on Demand Services. 19th IEEE ICDCS, 1999
- 11 Cui Y, Nahrstedt K, Xu D. Seamless User-level Handoff in Ubiquitous Multimedia Service Delivery. Multimedia Tools and Applications, 2004, 22: 137~170
- 12 The Real System SDK. <http://www.reálnetworks.com/resources/sdk/index.html>