|  |  |
| --- | --- |
|  |  |

硕士研究生学位论文阶段报告

学 号: 2012111499

姓 名: 赵晓森

学 院: 网络技术研究院

专业(领域): 计算机科学与技术

研究方向: 网络技术与应用

导师姓名: 王志谦

北京邮电大学

2014年10月13日

|  |  |  |  |
| --- | --- | --- | --- |
| 论文题目 | 一种适用于广电网的云应用服务管理方案的研究与实现 | | |
| 论文类型 | 应用研究 | 选题来源 | 企、事业单位委托项目 |
| 开题日期 | 2013-12-02 | 是否开题题目 | 是 |
| 论文开始日期 | 2013-12-02 | 报告日期 | 2014-10-13 |
| 报告地点 | 北京邮电大学信息网络中心 | 报告时间 | 上午 8:30-9:00 |
| **研究内容简介**  **选题背景：**  伴随着国内三网融合的不断推进，越来越多的用户家庭拥有具有双向交互能力的数字电视。同时由于电脑和手机、平板电脑等智能终端的发展，用户可以接触到的应用和内容越来越丰富，用户也越来越多的向互联网方向靠拢，国内电视机的开机率也不断降低。在这样一个内容与体验为王的时代，如何利用现有平台基础，通过技术的改进创新，为用户提供更多的应用、内容，重新获得用户的关注，成为广电运营商亟待解决的问题。  我国国内广电运营现状比较复杂，地区与地区之间差距大，呈现“军阀割据”的现状。国内用户机顶盒平台多种多样，并且普遍配置低、运算能力差。如果运营商对终端进行整体升级，这样耗时太长、投入太大，是无法接受的。  当下提出了一种新的解决方案，即采用云技术，将终端的运算能力剥离到云端，终端只需要具备基本的网络接入、视频解码和交互处理能力。把应用部署到云端，应用程序的处理、显示图像的渲染等在云端完成，运行结果以音视频流的方式通过网络传送到终端，终端完成音视频的解码并将结果呈现。用户的操作指令通过网络传送到云端运行的应用，应用处理后将响应结果再次以音视频的形式下发。该解决方案对终端能力要求低，现存的多数终端设备都能满足要求，避免了对终端的升级维护。而且在云端部署应用，可以实现应用的快速部署、内容的可管可控、应用软件的版权保护，实现对资源统一有效利用，是解决目前广电运营商面临的难题的最佳选择。  **研究内容**：  （1）NGOD架构扩展  NGOD(Next Generation On Demand，即下一代交互电视架构)是由美国规模最大的有线电视服务商 Comcast公司提出的一种新的开放的交互式服务网络框架结构。该架构合理细致的划分出了各个逻辑功能组件，而且将各组件间的协议交互和通信接口规定的更加全面和详尽。目前，业内以及开始逐步的转向使用 NGOD 架构进行开发和部署广电网内的视频点播系统。  然而NGOD架构主要是为视频点播业务服务的，并没有考虑到在广电网上的云应用服务。为此，本课题需要对NGOD架构进行扩展，在不影响现有功能的基础上，能够在广电网上部署云应用为用户服务。  （2）应用服务器管理  云端应用部署在应用服务器上，应用服务器的性能、分配、资源管理直接关系到用户的体验效果。  用户选择应用的类型不同，对服务器硬件资源的要求就不同。例如，3D游戏对CPU、内存和显卡性能要求远高于一般的2D游戏。又或者，有些应用对内存要求高，但对CPU要求低，而另外一些应用恰恰相反。目前有两种简单的分配方法。一种是将新的请求应用部署到负载最小的服务器上，这样可以保证服务器间负载均衡，但服务器资源碎片化严重，造成资源的浪费；一种是将新的请求应用部署到负载最大的服务器上，这样可以减少一部分碎片，但是服务器间负载均衡性很差，如果一些服务器宕机，会造成很多会话的丢失，服务质量差。在对服务器资源进行分配时，要考虑如何将不同类型、不同需求的应用部署到多台服务器上，既能保证对服务器资源的充分利用、减少碎片，又能保证服务器间负载均衡以及应用服务器群的可扩展性，是进行服务器资源管理的核心问题。  （3）应用部署  应用部署到云端，需要对云端资源虚拟化，本课题选择了应用虚拟化技术。  应用虚拟化，将应用程序与操作系统解耦合，为应用程序提供了一个虚拟的运行环境。在这个环境中，不仅包括应用程序的可执行文件，还包括它所需要的运行时环境。从本质上说，应用虚拟化是把应用对低层的系统和硬件的依赖抽象出来。  **关键技术**  一、对NGOD架构的扩展。在不影响原架构功能的基础上对NGOD架构合理扩展。  二、提出合理的算法，综合考虑用户选择应用的类型、用户的位置、服务器的负载情况等因素选定应用服务器。  三、建立应用与推流服务器、边缘设备和客户端的连接。  四、在云端实现虚拟化，并部署运行应用。  **论文计划、进度与目标:**  (1)2014年1月－2014年3月：通过对系统进行设计并对所用技术进行选择和分析，设计出可行的方案。在这个过程中要充分考虑计划的合理性，并及时进行调整，完成详细设计。  (2)2014年4月－2014年7月：在前期设计完成的基础上进行相关模块的编码实现，在此期间进行大量的软件工程相关资料的学习。  (3)2014年8月－2014年10月：对系统进行测试，找出和分析系统出现的各种问题，完成各种业务的测试以及压力测试，完善系统的功能，实现最终目标。  (4)2014 年11月－2014年12月：整理相关的文档和数据资料，完成论文的写作。 | | | |

|  |
| --- |
| **论文进展情况**  **报告工作计划:**  (1)2014年1月－2014年3月：通过对系统进行设计并对所用技术进行选择和分析，设计出可行的方案。在这个过程中要充分考虑计划的合理性，并及时进行调整，完成详细设计。  (2)2014年4月－2014年7月：在前期设计完成的基础上进行相关模块的编码实现，在此期间进行大量的软件工程相关资料的学习。  (3)2014年8月－2014年10月：对系统进行测试，找出和分析系统出现的各种问题，完成各种业务的测试以及压力测试，完善系统的功能，实现最终目标。  (4)2014 年11月－2014年12月：整理相关的文档和数据资料，完成论文的写作。  **实际工作进展情况**：  (1)2014年1月－2014年3月：阅读及整理相关文档，包括NGOD架构标准规范、已实现的NGOD 模块的相关文档、研究论文。完成了系统设计并对所用技术进行了选择和分析，完成了概要和详细设计，并完成小论文。  (2)2014年4月－2014年7月：在前期设计完成的基础上进行相关模块的编码实现，实现了主要的模块通信接口、模块功能和最小资源分配策略。  (3)2014年8月－2014年10月：对系统进行测试，发现并分析了系统出现的一些问题，完成了各种业务的测试以及压力测试，完善系统的功能。  (4)2014 年11月－2014年12月：与其它同学完成项目的整体测试，整理相关的文档和数据资料，完成论文的写作。  从目前的进展和工作情况看，可以定期完成论文 |
| **工作成果**  **已完成学位论文工作的内容**   1. 学习NGOD官方文档，掌握NGOD对各个模块的功能划分、各个接口的定义实现，以及NGOD官方对RTSP协议的扩展和VREP协议的定义。   VREP协议是NGOD在TRIP（RFC 3219）的基础上扩展而来的协议。   1. 完成对NGOD架构的扩展，新增加了ASM和AS两个功能模块，S7、R8和D8三个接口。   如图1所示，是对NGOD架构进行扩展后的部分结构图：  图1 扩展后的NGOD架构图  ASM（Application Server Manager，应用服务器管理器）的功能包括资源管理功能和设备发现功能。  AS（Application Server，应用服务器）是用户请求应用运行的服务器。在AS上实现了应用虚拟化，提高了应用程序的兼容性和应用部署的速度。  ASM与SM之间定义实现了S7接口，负责在应用服务器上对会话请求所需的服务器资源进行商议。S7接口实现了RTSP协议，主要包括SETUP、TEARDOWN、ANNOUNCE等消息信令。图2为S7接口的SETUP消息交互流程：  图2 S7接口的消息交互流程  ODC（On Demand Client，点播客户端）通过S1接口向SM发出应用服务请求后，SM首先通过S6接口向ERM（Edge Resource Manager，边缘资源管理器）发送SETUP会话建立请求，选择ED（Edge Device，边缘设备）。ERM向SM返回SETUP RESPONSE消息，包括ED的IP、端口等信息。然后SM通过S3接口向ODRM（On Demand Resource Manager，点播资源管理器）发送包括ED的IP、端口等信息的SETUP会话建立请求，选择SS（Streaming Server，推流服务器）[5]。之后，SM作为RTSP客户端通过S7接口向ASM发出服务器资源请求SETUP。在SETUP报文中，除了点播客户端的信息、用户请求的应用信息外，还包括ODRM返回的SS的信息，以便于建立会话的下行通路。ASM在选定AS（Application Server，应用服务器）部署应用之后，返回SETUP RESPONSE消息，包括选定的AS的IP和部署的应用的进程号，SM将其转发给点播客户端，用来建立点播客户端到AS的上行通路。  TEARDOWN消息用于释放会话占用的资源。ANNOUNCE用于向SM汇报会话工程中发生的异常。  ASM与AS之间定义实现了R8接口。ASM通过使用该接口，依据选定的资源调度和负载均衡策略来管理应用服务器的资源。R8接口实现了RTSP协议，包括SETUP、TEARDOWN、ANNOUNCE、GET\_PARAMETER、PING等信令。  当SM 向ASM发出SETUP会话建立请求后，ASM结合点播客户端的请求与各应用服务器的状态，依据负载均衡策略，选定应用服务器来部署应用，并通过R8接口向选定的AS发送SETUP报文，包括客户端请求的应用的信息。AS加载部署应用后，向ASM发送SETUP RESPONSE报文，包括IP、所部署应用的进程号。  ASM收到SM发送的TEARDOWN报文后，进行一些内部处理，例如更新AS信息表，然后向AS发送TEARDOWN报文告知结束会话。AS结束会话后返回TEARDOWN RESPONSE报文。  在会话过程中发生异常时，AS可以通过发送ANNOUNCE报文向ASM反映异常。ASM可以通过GET\_PARAMETER Request和Response来获取和AS会话的参数消息。ASM通过周期性地向AS发送PING报文来维护会话的生命周期。在NGOD RTSP Usage Specification中，默认的会话超时时间被设为3个小时。  ASM和AS之间定义了服务发现与注册接口D8。AS通过该接口向ASM注册并提供可用资源的详细目录。D8接口实现了VREP协议，包括OPEN、UPDATE、NOTIFICATION、KEEPALIVE四种消息类型。OPEN消息主要用于VREP传输会话的建立；UPDATE消息主要用于资源设备向其资源管理器上报自身配置参数；NOTIFICATION消息主要用于错误的报告；KEEPALIVE消息用来恢复和维持心跳。  AS接入系统后主动向相应的ASM发起连接请求，传输层连接建立后，两者首先通过OPEN消息协商参数信息，包括VREP协议版本、超时时间、VREP会话标识、支持的路由类型以及组件的收发能力等信息。OPEN消息被接受之后，会向对方发送KEEPALIVE消息确认。  在通过OPEN消息对组件的收发能力协商后，AS只具备发送UPDATE消息的能力，ASM的D8接口只具备接受UPDATE消息的能力。AS通过向ASM发送UPDATE消息报告自己的服务器资源详细信息，包括CPU信息和内存信息。ASM发送KEEPALIVE消息对它进行确认。  当AS和ASM在收发消息的过程中发现消息有误时，任何一方都会利用NOTIFACATION消息报告错误并断开连接。   1. 完成应用虚拟化，并实现了将应用部署到应用服务器上。通过应用虚拟化技术，将应用程序与操作系统解耦合，每个应用程序生成一个文件包。该文件包中包括应用程序的可执行文件和所需要的运行时环境。应用服务器收到ASM发送的用户应用请求后，会向存储服务器请求该应用的文件包，实现在本地的快速部署。   4．实现了云端应用服务器群的调度管理策略，能够针对用户不同请求，选择负载最小的应用服务器部署应用。每台应用服务器会定时向服务器管理模块报告自己的负载情况，服务器管理模块根据各服务器的负载情况，综合考虑用户位置等信息，选择负载最小的服务器为新的用户请求提供服务。  **阶段性成果**  已完成论文《一种基于NGOD架构的云应用服务管理方案》投送《电视技术》杂志，已发表。  **主要创新点**  1．对NGOD架构进行扩展。NGOD架构主要是针对视频点播服务的，并没有考虑在广电网上部署云应用。广电网上云应用服务与视频点播服务的区别在于，云应用服务对时延要求更高、云应用服务器管理更复杂，会话的上下行通路以及客户端操作命令的传输解析都与视频点播服务不同。本次课题在不影响NGOD现有功能的基础上，通过增加新的模块和接口，实现了对NGOD架构的扩展。  2．提出了一种新的算法，在云端实现对应用服务器的调度和资源管理。对服务器资源的调度管理需要综合考虑用户选择应用程序的类型、用户的位置、服务器的负载情况等因素。目前，有人已实现NGOD架构中ERM、ODRM模块，在选择推流服务器和边缘资源管理器时采用靠近用户的原则选择设备。但是，广电网中的云应用对传播时延要求很高，在选择应用服务器的位置时需要综合考虑上行和下行通路的时延。不同的应用类型对服务器资源的需求不同，有些应用可能对CPU要求高，对内存、显卡等要求低，有些应用可能对显卡要求较高，例如3D游戏，显卡必须具备较高的性能才能对游戏画面完成渲染。服务器间负载平衡，有利于保证服务质量。假如部分服务器负载远高于其它服务器，如果这部分服务器宕机，会造成大部分会话丢失，降低了系统的QoS。为此，实现了一种算法，每台应用服务器会定时向服务器管理模块报告自己的负载情况，服务器管理模块根据各服务器的负载情况，综合考虑用户位置等信息，选择负载最小的服务器为新的用户请求提供服务。 |
|  |

|  |
| --- |
| **计划及进度安排**  10月底 完成对应用服务器管理模块的改进。  11月1日-11月10日 进行单元测试和整体测试，然后根据测试结果进行修改和优化。  11月10日-11月30日 整理文献资料、代码和数据等，完成论文初稿。  12月 完成论文终稿。  从目前的完成情况看，剩余工作主要是测试、改进和论文撰写工作，可以按照开题的进度安排按时完成论文。 |
| **问题及整改方案**  云端应用服务器群的调度管理方法完善：现在实现的策略是最小分配策略，即根据每台服务器的负载情况和客户端服务器位置，选择负载最小的服务器分配用户请求。该算法有一些不足，尚需进一步完善，需要综合考虑到用户请求应用所需负载不确定以及需要同时为多个用户请求服务。 |

|  |
| --- |
| **参考文献**  [1] Joeng Kim,Ricardo A.Baratto and Jason Nieh. An Application Streaming Service for Mobile Handheld Devices. IEEE International Conference on Services Computing(SCC'06), 2006.  [2] 李华宇. 运营商拓展增值业务利器——视频云计算系统. 广播电视信息, 2013年7月总第255期.  [3] 付平武. 应用虚拟化技术应用与研究. 电脑知识与技术, Vol.9, No.20, July 2013.  [4] 李永涛. 基于 NGOD 的 VOD 系统的会话管理和边缘资源管理的设计与实现. 北京邮电大学. 2012.  [5] 宫伟俊. 双向广电网络中统一边缘资源管理器的研究与实现. 北京邮电大学. 2013.  [6] 杨娴,陈麟. 云计算环境下的应用虚拟化的研究. 软件, 2012年第33卷第4期.  [7] 张京立.浅析NGOD标准架构的设计理念和优点. 有线电视技术, 2011(9):98-99.  [8] Comcast. Next Generation On Demand (NGOD 2.0) RTSP Usage Specification. 2006.  [9] 李永涛,胡朋,王志谦.基于NGOD架构的VOD系统边缘资源会话管理方案.电视技术, 2011, 35(19):1-3.  [10] 胡朋.基于NGOD架构的VOD系统资源管理和服务发现研究与应用.北京邮电大学,2012.  [11] Valeria Cardellini, Michele Colajanni, Philip S.Yu. Dynamic Load Balancing on Web-Server Systems. IEEE Internet Computing(1999 vol.3):28-39. |
| |  |  |  |  | | --- | --- | --- | --- | | 姓 名 | 职 称 | 职务 | 工 作 单 位 | | 黄小红 | 副教授 | 组长 | 北京邮电大学 | | 吴志刚 | 高级工程师 | 成员 | 北京邮电大学 | | 王志谦 | 高级工程师 | 成员 | 北京邮电大学 | | 徐明昆 | 高级工程师 | 成员 | 北京邮电大学 |   **评审小组** |

|  |
| --- |
| **导师评语**  该生已按开题报告所制定的研究方向，依据研究计划开展了相关工作。目前工作进展正常，同意通过阶段检查。 |
| 导师：  日期： 年 月 日 |
| **阶段报告小组意见：** |
| 负责人：  日期： 年 月 日 |
| **学院意见：** |
| 负责人：  日期： 年 月 日 （签章） |