

基于代理的协同计算模型研究与实现

李 永

(盐城师范学院 信息科学与技术学院, 江苏 盐城 224002)

摘 要: 针对大规模多数据源实时协同计算结构模型的可扩展进行研究. 分析目前协同计算体系结构的缺点, 提出一个三层的协同计算结构模型. 然后给出该模型中最重要的覆盖层的构建方法, 即以自治域为单位征集或部署代理节点, 并在每个自治域选出一个或多个代理作为核心代理, 核心代理节点采用点对点的 Pastry 协议互联构成自治域间的覆盖网. 实验结果表明, 该模型具有良好的可扩展性.

关键词: 协同计算; 覆盖网; 代理部署

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2012)08-0091-03

Research and Implementation of Collaboration Computing Model Based on Proxies

LI Yong

(School of Information Science and Technology, Yancheng Teachers University, Yancheng 224002, China)

Abstract: Aiming at the scalability of real-time collaboration computing which is large scale and multiple data sources, a three-layer architecture model is presented after analyzing the shortcomings of existing models. Then the method of constructing overlay layer is given, namely, by collecting servers and deploying some proxies in every autonomy system, and kernel proxies are selected in every autonomy system. The kernel proxies run Pastry protocol and form overlay layer. Experimental results show that the model has good real-time performance and scalability.

Key words: collaboration computing; overlay network; proxy placement

1 引言

大规模的、世界范围内集成式的协作逐步变成现实, 协同计算的典型应用如视频会议、CSCW、虚拟环境、分布式交互仿真、多人在线游戏等^[1-2]. 其特点是具有高带宽、实时性、多数据源、交互性、一致性等, 这些特点对目前的网络结构提出严峻的挑战^[3].

大多数商用的协同计算采用服务器集群结构, 服务器之间通过高速网络连接. 在这种结构中, 将分布式协同计算的虚拟空间划分为多个区域, 每个服务器负责维护某一个区域的虚拟对象和端结点, 通过这种方式将整个系统的负载分配到各个服务器^[4]. 多服务器结构具有集中控制的优点, 且将整个系统的负载分配到多台服务器, 具有更好的可扩展性. 但也面临以下三个方面的问题. 第一: 端结点在

服务器间的迁移问题. 需要支持用户在区域间进行平滑迁移, 这种迁移不能对端结点的正常交互产生影响, 不能引入过多的服务器间的网络通信开销. 第二: 服务器间的动态负载平衡. 由于端结点在虚拟空间中的不均匀分布, 可能导致某些服务器的负载过重需要提供一种有效的动态负载平衡策略, 将过载服务器的部分负载调整到其他服务器, 以避免负载过重服务器成为瓶颈导致系统整体性能下降. 另外, 频繁的负载调度和平衡算法也极大地占用了服务器的计算资源. 第三: 多服务器结构中, 每个服务器与其负责区域内的端结点采用客户服务器方式连接, 由服务器保存共享状态信息和协调端结点之间的通信, 端结点只是一个接收命令的终端系统, 没有充分利用端结点的资源.

另外也有一些协同计算采用 P2P 结构, 利用

P2P 网络的无中心、自组织和高效路由等特点,构建基于 P2P 结构的分布式协同计算,所有端结点很自然地形成了一个覆盖网络,每个端结点通过贡献自己的计算资源和带宽资源在各端结点之间共享整个分布式协同计算的状态^[5]. 每个结点完全自治,不会因为一个节点崩溃而影响整个系统. P2P 结构不需要昂贵的中心服务器支持,从而可以有效地减少系统的经济费用. 但是基于 P2P 结构的分布式协同计算由于其端结点具有异构性、不稳定性、自私性,规模变大时会给整个系统带来一些难以克服的问题,如一致性、安全性、不易管理等问题.

一方面,协同计算缺少一个可扩展的支撑平台;另一方面,网络体系结构是网络基础理论研究的核心问题和最基本的研究课题,对网络协议的研究和相关算法的实现起着指导性的作用^[6]. 因此,本文从网络体系结构入手研究构造高性能的可扩展的协同计算的结构模型,通过自治域内代理节点的部署和基于代理节点的域间覆盖层的构建,实现一个可扩展的分布式协同计算平台 DCCON (Distributed Collaboration Computing Based on Overlay Network).

2 DCCON 模型

2.1 Overlay Network 技术

Overlay Network^[7] 技术的基本思想是在不改变或部分改变互联网底层基础设施的前提下,在部分网络节点上对已有的软硬件进行扩充,使这些特殊节点能够利用底层网络提供的基本服务协同配合,进而达到支持和实现各类新型网络应用和服务模式的目的. 不仅可增强已有基于 Internet 应用的 QoS,且可使现有网络基础设施对新应用提供支持.

2.2 DCCON 的三层结构模型

基于 Overlay Network 的分布式协同计算的网路体系结构模型 DCCON 由三个功能层构成:用户应用层、DCCON 覆盖层和基础物理网络层,如图 1 所示.

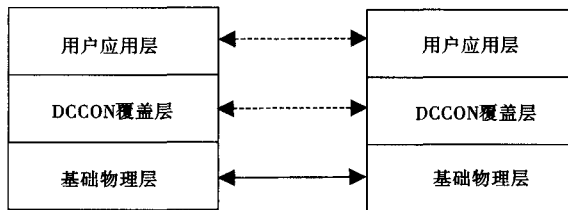


图 1 DCCON 模型

2.2.1 基础物理网络层

该层是 DCCON 覆盖层的构造基础,利用现有 TCP/IP 协议族在物理链路上提供基本的物理层传输、网络通信和网络层路由等功能. 它向上层(DCCON 覆盖层)提供必要的通信节点、通信协议和物理连接支持.

2.2.2 DCCON 覆盖层

DCCON 覆盖层是构建在底层基础物理网络层上的虚拟层. DCCON 覆盖层生成并维护覆盖网络拓扑结构,同时为其上层(用户应用层)提供实现分布式协同计算所需的功能. 由于 DCCON 覆盖层是构造在基础物理网络层上的虚拟层,因此不必对现有物理网络进行全网改造,而只需要在网络中部署一定的代理结点,易于部署和实现.

2.2.3 用户应用层

用户应用层负责向端用户提供实现分布式协同计算所需的各种服务,为端用户和应用程序提供应用接口. 用户应用层是端用户与 DCCON 网络之间的相互接口,它在 DCCON 覆盖层的支持下,实现端用户所需的分布式协同计算服务.

3 DCCON 模型覆盖层的构造

3.1 以自治域为单位来构造覆盖网络

在各自治域内部,通过征集并部署少量的代理服务器的方法构建域内覆盖网络,征集节点和代理节点共同形成了覆盖节点. 设无向图 $G_0 = (S, V, M, E)$, S 表示征用节点的集合, V 表示与征用节点直接相连的路由器的集合, M 表示代理候选位置集合, E 表示任意两个节点 $u, v \in S \cup V \cup M$ 之间边的集合.

定义 1 代理部署的标志函数

$$x(i) = \begin{cases} 1 & \text{在 } m_i \text{ 部署代理} \\ 0 & \text{在 } m_i \text{ 不部署代理} \end{cases} \quad (1)$$

定义 2 路由器与的代理之间的路径标志函数

$$y(i, j) = \begin{cases} 1 & \text{如果路由器 } v_i \text{ 和代理 } m_j \text{ 连通} \\ 0 & \text{其他} \end{cases} \quad (2)$$

定义 3 代理的部署代价函数

$$f: M \rightarrow [0, \infty) \quad (3)$$

$f(i)$ = 在 m_i 部署代理的花费, $m_i \in M$.

一个自治域中部署的代理越多,覆盖网络的性能就越高,花费就越大. 因此在覆盖网络性能达到基本要求以后,代理节点部署花费越小越好,代理部署花费表示为

$$cost = \sum_i f(i) \times x(i) \quad \forall m_i \in M, i \in [1, |M|] \quad (4)$$

与路由器 $v_i \in V$ 直接相连的征用节点数量用 $w(i)$ 表示,则 $\sum_i w(i) = |S|$. $e_D(i, j)$ 表示代理候选位置 m_i 和路由器 v_j 之间路径的延迟. 代理 m_i 和路由器 v_j 之间的路径延迟代价,表示为: $d(i, j) = x(i) \times y(i, j) \times e_D(i, j)$. 代理 m_i 和其连通的所有征用节点的延迟代价表示为

$$d(i) = \sum_{j=1}^{|M|} d(i, j) = \sum_{j=1}^{|V|} x(i) \times y(i, j) \times e_D(i, j) \times w(j) \quad (5)$$

覆盖网络中所有征用节点到与它距离最近的代理的延迟总和表示为

$$Delay = \sum_{i=1}^{|M|} d(i) = \sum_{i=1}^{|M|} \sum_{j=1}^{|V|} x(i) \times y(i, j) \times e_D(i, j) \times w(j) \quad (6)$$

代理部署问题的优化目标应综合考虑覆盖网络延时和代理花费,在两者之间取折中,因此代理部署代价目标函数为: $\omega \times cost + (1 - \omega) \times Delay$, $0 \leq \omega \leq 1$. 综上,代理的部署问题可以建模为公式所示的多约束优化问题:

$$\begin{cases} \min & \omega \times cost + (1 - \omega) \times Delay \\ \text{s. t.} & delay(v_i, v_j) \leq D, D \in R^+, v_i, v_j \in S \cup M, i \neq j \\ & \sum_j y(i, j) = 1 \quad \forall i \in [1, |V|] \\ & x(j) \leq \sum_{i=1}^{|V|} y(i, j) \quad \forall j \in [1, |M|] \\ & x(j), y(i, j) \in \{0, 1\} \quad \forall i \in [1, |V|], \\ & \quad \forall j \in [1, |M|] \end{cases} \quad (7)$$

其中优化模型的约束条件为覆盖层任意两个节点间的延时不超过 D , 每个路由器都要与一个部署的代理相连通; 每个部署的代理至少与一个路由器相连通. 代理部署问题可以抽象为设施定位问题, 是 NP-hard 类组合优化问题, 这里采用启发式算法进行求解.

3.2 核心代理节点层覆盖网结构

为了实现域间的分布式协同计算, 自治域内的覆盖网络构造完成以后, 进一步构造域间覆盖网络. 在各自域内选取一个或几个代理节点作为核心代理节点, 然后各自域的核心代理节点采用点对点的 Pastry^[8] 协议互联构成自治域间的覆盖网, 负责

维护域间覆盖网络拓扑信息, 并协同域间的分布式协同计算.

4 仿真实验

采用通用的拓扑产生器 Brite 生成网络拓扑数据, 模拟底层的物理网络. 实验中生成包含 5 000 个节点的层次结构网络拓扑图, 分成 10 个自治域, 每个自治域包含 500 个节点, 自治域内的节点之间延时代价 $[0, 10]$, 每个自治域能够征集节点数在 $[0, 3]$ 上均匀分布, 平均每个节点的度数为 4. 除了被征集作为代理的节点位置外, 都是代理部署的候选位置, 代理的部署代价是 $[100, 200]$ 之间的一个随机数.

图 2 所示为节点间的平均延时随网络规模的变化情况. DCCON 的域间覆盖层由核心代理节点采用 Pastry 协议互联构成自治域间的覆盖网, 比在域间覆盖层采用随机结构的平均延时性能优越.

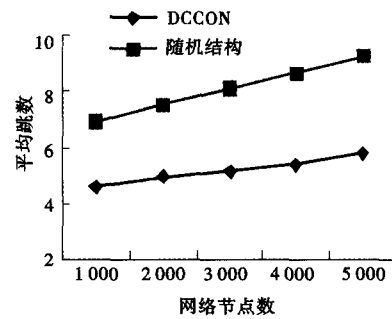


图2 平均跳数随网络规模的变化情况

5 结束语

针对协同计算的可扩展性问题, 结合 Overlay Network 技术的优点构造了一种协同计算模型 DCCON, 该模型由三个功能层构成: 用户应用层、DCCON 覆盖层和基础物理网络层. 其中最关键的是 DCCON 覆盖层的构建, 以自治域为单位, 通过征集并部署少量的代理节点的方法构建域内覆盖网络, 征集节点和代理节点共同形成了覆盖节点; 自治域间的覆盖层由核心代理节点采用点对点的 Pastry 协议互联构成. 最后, 通过实验验证了此模型的性能.

参考文献:

- [1] 刘淑芬. 协同 Internet 计算将成为协同计算主流[J]. 国际学术动态, 2008(5): 45-46.
- [2] 卢瞰, 顾宁. 多种协同方式集成的一致性维护[J]. 中国计算机学会通讯, 2010, 6(12): 15-21.

(下转第 97 页)

- clouds[J]. J. Photogrammetry & Remote Sensing, 2007 (61): 355-370.
- [4] Chen Y, Medioni G. Object modeling by registration of multiple range images[J]. Image and Vision Computing, 1992, 10(3): 145-155.
- [5] Fan K C, Tsai T H. Optimal shape error analysis of the matching image for a free-form surface[J]. International Journal of Robotic and Computer-Integrated Manufacturing, 2001, 17(3): 215-222.
- [6] Ristic M, Brujic D. Efficient registration of NURBS geometry[J]. Image and Vision Computing, 1997, 15 (12): 925-935.
- [7] David W Eggert, Andrew W Fitzgibbon, Robert B Fisher. Simultaneous registration of multiple range views for use in reverse engineering of CAD models [J]. Computer Vision and Image Understanding, 1998, 69(3): 253-272.

作者简介:

姚晓山 男, (1975-), 讲师. 研究方向为计算机可视化、机械设备状态监测与故障诊断.

刘健鑫 男, (1983-), 博士研究生. 研究方向为动力装置自动化与仿真.

柯维 男, (1984-), 博士研究生. 研究方向为工程信息可视化、振动与噪声控制.

(上接第93页)

- [3] 陈祥涛, 商建东. 协同设计网格平台研究[J]. 微电子学与计算机, 2009, 26(11): 127-130.
- [4] Michael Kwok, Johnny W Wong. Scalability analysis of the hierarchical architecture for distributed virtual environments[J]. IEEE Transactions on Parallel and Distributed Systems, 2008(19): 408-417.
- [5] Sheng Pang, Changjia Chen, Chao Xu. Structure analysis of social network on multi-activities in MMOG[C] // ISECS International Colloquium on Computing, Communication, Control, and Management, Sanya, China, 2009.
- [6] Clark D, Tennenhouse D. Architectural considerations for a new generation of protocols[C] // ACM SIGCOMM. USA: Philadelphia, 1990.
- [7] 杜丽娟, 余镇危. 覆盖网体系结构及应用研究[J]. 计算机工程与应用, 2009, 45(28): 102-104.
- [8] 薛涛, 石美红, 王会燃, 等. 应用 P2P 网络实现基于内容的发布订阅系统[J]. 微电子学与计算机, 2007, 24 (9): 89-91.

作者简介:

李永 男, (1978-), 讲师. 研究方向为网络体系结构、覆盖网、分布式应用.