一种分布式并行服务器模型的性能分析与改进①

陈 宁 (西安工程大学 计算机科学学院 陕西 西安 710048)

摘 要: 设计一个分布式并行服务器，必需全面考虑网络层和应用层流量以及响应特性对性能的影响。分布式并行服务器模型(DPS)具有良好的响应特性和较强的服务能力，对它进行了全面的分析，重点研究了网络层流量对模型性能的影响，并且在理论分析和实验测试的基础上，将网络层流量和应用层流量按协议栈分布式处理,提出了改进的 DPS 模型。改进的 DPS 模型能够更好的处理网络层流量，有效地改善模型的性能。

关键词: 分布式并行服务器；流量控制；性能评价

An Analysis and Improvement on Network Performance of Distributed Parallel Server

CHEN Ning

(College of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: To design an effective distributed parallel server, the two principal metrics of networking should necessarily be considered: throughput and delay of both network layer and application layer to improve the original power of the network. Distributed Parallel Server (DPS) is characterized by the response characteristic and service capability of distributed parallel server. This paper makes an in-depth study on DPS, and based on the analysis and experiments it introduces an intermedium in DPS, namely distribution device, to partition flux in network layer according to working ability of application server. Consequently performance is improved.

Keywords: distributed parallel server; flux control; performance evaluation

1 引言

当前，激烈的竞争要求企业能够迅速调整自己的业务，快速实现业务变更和重组，提供高效的服务系统，这给企业 IT 系统的开发带来了巨大的挑战。新的 IT 系统不仅需要集成已有的信息系统和数据管理软件，而且要能够随着业务的变更被快速实现。性能、成本、可扩展性和可靠性都是企业信息系统的开发的目标。

分布式并行服务器被认为有望缓解这一难题,它为新一代的 B2B 业务和企业应用集成提供一整套标准和基础设施。近年来，分布式并行服务器得到了广泛的研究，许多实用的分布式并行服务器结构和调度算法被提出[1]。文献[2]针对当前多服务器系统透明性和任务调度研究中存在的问题，提出一种分布式并行服务器的网络服务透明性实现机制和相应的任务调度算法，并取得了良好的效果。

设计一个分布式并行服务器模型，必需全面考虑网络层和应用层流量对模型性能的影响，本文在文献[2]提出的 DPS 模型的基础上，通过理论分析和试验，提出了改进的 DPS 模型，将网络层流量和应用层流量按协议栈分布式处理，有效的改善了模型系统的性能。

2 分布式并行服务器模型

2.1 DPS 模型[2]

DPS 模型针对当前多服务器系统透明性和任务调

① 基金项目:西安工程大学基础研究项目(XGJ07008);陕西省教育厅专项科研计划(07JK264);国家科技支撑计划资助项目(2006BAF01A44)
收稿时间:2009-06-05

度研究中存在的问题，提出一种分布式并行服务器的网络服务透明性实现机制和相应的任务调度算法[2]。该透明性机制修改服务器结点的 ARP 地址解析协议以及客户端到服务器端的连接和数据请求处理，使得整个服务器系统对外界表现为惟一的 VIP 地址和 VMAC 地址。相应的任务调度算法则根据负载和阈值设置将服务器结点分成两个链：有效服务器结点链和过载服务器结点链，然后由量值循环法对有效服务器结点链进行任务调度。 在修改 Linux 内核网卡驱动程序和部分底层网络协议的基础上进行了实现，测试结果表明其具有良好的响应特性和较强的服务能力。

经过透明性处理和任务调度之后，从客户端发送到 VIP 服务器的数据都能够被一个优化调度得到的服务器结点进行处理，并且对外界完全透明。

利用多服务点损失制排队理论，对 DPS 模型的任务损失概率进行分析。传统的多服务点损失制排队模型 M/ M/ n/ n 假定系统内有 n 个服务结点，任务按泊松流到达系统，其强度为 λ。各服务结点的服务时间服从负指数分布，强度为常数 μ。 如果任务到达系统时发现 n 个服务结点均忙着，就离开系统放弃服务。分布式并行服务器的任务处理与上述模型的差别在于其每个服务结点可同时对多个任务服务，并且其服务强度 μ 不是常数，因此对模型进行一定的转换。得到了很好的分析结果。

2.2 模型分析

对单个性能目标的量度和评价并不能全面地反映分布式并行服务器的整体性能，因此需要有一个综合的性能评价[3,4]。按照文献[2]中的分析和测试，DPS 模型具有良好的响应特性和较强的服务能力。我们在大业务量条件下测试，发现存在着服务性能饱和区。在性能分析中发现，这是由于网络层流量处理的局限性造成的，因此不仅要对系统的应用层业务进行了分析，还要对网络层的流量进行分析。

应用层测量可以使我们对整个应用的性能有一个清楚的认识，而这是很难从低层测量数据综合得到的。同时应用层测量也能提供客户机和服务器之间、网络链路之间的性能参考。比如，Web 下载是一种网络业务。而且测量应用层性能也能间接反映网络层的性能。

比如想比较某一特定业务连续几天或几星期的性能，假定由服务器负载引起的变化比由网络拥塞引起的变化小，则测量该业务的总体性能就足够了。这种方法常用于对不同提供商提供的业务进行性能比较。

采用应用层测量的另一个原因是一些 ISP 在其网络内使用通信过滤(Traffic Filtering)技术，比如阻止 ICMP 响应包或限制其传送速率。虽然用 ping 作一些网络测量还是有用的，但通信过滤技术的使用日益广泛，在一定程度上减少了此类测量的使用。但是，DPS 模型是一个自洽的系统，仅仅采取应用层的测量，是远远不够的。还要采用网络层测量，以评估其提供的网络链路或路由器、服务器等网络节点的性能[5、6]。

2.3 网络层流量分析

在 DPS 模型中，采用了虚拟 IP 地址的方式。DPS 模型中所有主机都会接受目的地址为这个 IP 的数据包。因此，在网络层，DPS 模型中每个主机的流量是一样的。假定仅仅考虑 WEB 业务的情况下，每个主机的流量可以表示为：

B = B1 + B2

B 表示整体流量

B1 表示背景流量

B2 表示目标流量，在这里是 WEB 业务流量。

背景流量 B1 是在平均值上下波动的流量，在一段时期内基本特征不会发生变化。所以可以采用平均值来表示。

目标流量 B2，是和业务有关的流量。在网络层所有主机都会接受所有的流量。所以各个主机的流量是一样的。采用多个泊松分布模型的叠加来模拟 WEB 业务，对于单个主机可以采用 M/ M/ 1 模型模拟。其中：λ 表示平均到达率，即单位时间到达顾客数，P_k 表示(服务员)随机观察队长为 k 的概率，$\bar{\tau}$ 表示平均服务时间，而定义 $\mu = \frac{1}{\bar{\tau}}$。

则根据排队论可以得到如下方程组：

$$\lambda p_{k-1} + \mu p_{k+1} - (\lambda + \mu) p_k = 0$$

$$\mu p_1 - \lambda p_0 = 0$$

令 $\rho = \frac{\lambda}{\mu}$， 方程组转化为：

$$p_1 = \rho p_0$$

$$p_{k+1} = \frac{1}{\mu}[(\lambda + \mu) p_k - \lambda p_{k-1}] = (\rho + 1) p_k - \rho p_{k-1}$$

令 k=1，可以得到：

$$p_2 = (\rho + 1) p_1 - \rho p_0 = (\rho + 1) \rho p_0 - \rho p_0 = \rho^2 p_0$$

令 k=2，可以得到 $p_3 = \rho^3 p_0$。同理可得 $p_k = \rho^k p_0$。

使用归一化条件后，得到平均队长：

$$\bar{k}=\sum_{k=0}^{\infty}kp_k=(1-\rho)\sum_{k=0}^{\infty}k\rho^k=(1-\rho)\rho\frac{1}{(1-\rho)^2}=\frac{\rho}{(1-\rho)}$$

则 $\bar{k}$ 只于 ρ 有关。

因此，在网络层业务流量不能像在应用层那样采用 M/ M/ n/ n 模型，必须采用 M/ M/ 1 模型模拟。由此可以得到以下结论：

应用层：M/ M/ n/ n 模型　　流量：各个主机根据调度模式处理不同的流量

网络层：M/ M/ 1 模型　　流量：各个主机流量一样。

从以上分析可以知道，在网络系统受限的情况下，DPS 模型性能存在上限。即使采用最好的网卡，DPS 模型的性能也存在着饱和区。

按照 DPS 模型，假设每个主机均采用百兆网卡，HTTP 服务包长 10k。在网卡满负荷发挥最大功能的情况下，而且不考虑其他业务的影响，则系统的最大节点数为(每个节点提供一个 WEB 会话)：100M/10K = 10k.

3 实验方案

3.1 实验原理

在文献[2]一文中已经对系统的应用层进行了详尽的理论分析和试验测试。因此，这里仅对网络层进行测试[7]。由于 DPS 模型中所有主机采用同一个虚拟的 IP 地址，和网络的广播接受方式类似，我们可以将主机设置为广播接受方式来模拟 DPS 模型。实际上，每一个服务器主机接受的流量为广播包流量和目的为虚拟 IP 地址的 IP 包流量之和。

我们将网卡设置为广播方式，各个服务器之间采用 SWITCH 连接。由于 SWITCH 中有一个地址表记录哪一个 MAC 地址在哪一个端口，因此对于非广播包，SWITCH 不必像 HUB 那样将该包发送到所有端口，只须将该包发送到对应的端口，从而使不相关的端口可以并行通信，不会相互干扰。这样服务器在广播方式的流量和设置虚拟 IP 地址的方式是一样的。

* 测试主机采用　PIV 1.5G 265M　Windows XP　Sniffer Pro 4.5

* 服务器采用　PIV 1.5G 265M　Redhat LINUX 7.2　Tcpdump　Apache

3.2 实验工具

为了降低了测试成本，采用 Sniffer Pro 4.5 对网络和服务器的流量进行测试，在测试的过程中 Sniffer Pro 4.5 会自动将服务器设置成广播模式，模拟原系统的流量。业务流量的生成也采用 Sniffer Pro 4.5，通过网络抓包，对捕捉到的 HTTP 业务包编辑后按照一定的速度重复发送，生成测试所要求的业务流量。在 Sniffer Pro 4.5 中设置捕获主机到某一 WEB 服务器的流量，然后启动 IE 浏览器，浏览某一网站，然后保存 Sniffer Pro 4.5 捕获的网络包，即是主机和 WEB 服务器之间的 HTTP 协议的交互过程。图 1 是捕获的主机和百度 WEB 网站的一次 HTTP 协议的交互。

3.3 实验方案及结果分析

如图 2 所示，测试主机利用 Sniffer Pro 4.5 发送捕获的 HTTP 协议包，DPS 模型的每个主机利用 tcpdump 设置主机为广播模式，同时设置过滤模式，收到包后丢弃掉，仅仅模拟网络层的处理过程。

(1) 实验方案：

测试主机产生不同速率的包流量，然后比较发出的包的数量和收到的包的数量，测试丢包率。

测试主机产生不同速率的包流量，然后利用 IE 浏览器访问服务器，利用 Sniffer Pro 4.5 测试响应时间[8]。

(2) 实验结果：

在应用层业务处理能力受限的情况下，每个服务器主机能够处理的有效网络层流量。当测试主机满负荷发送时，分布式服务器对于另一台测试主机的 IE 浏览器没有反应，出现丢包现象。并且发送的数据包和收到的数据包大小不一致。WEB 服务的响应时间增加。

(3) 实验结果讨论：

在网络层，由于调度策略使业务分配到每个服务器上，因此每个服务器的流量中有效流量仅仅是 1/n (n 表示服务器的数量)。随着 DPS 模型的业务量增长，每个主机的网络层性能急剧下降。

表 1　响应时间和流量的关系

测试流量(M)	最大响应时间(毫秒)
0	110
10	140
20	210
30	260
50	300
70	350
90	3000

4　改进的DPS模型

DPS 模型的网络层和应用层在业务和性能方面存在冲突，造成了性能饱和区，因此模型改进主要集中在网络层和应用层流量的分布式处理。我们采用网络流量按协议栈分布式处理的方式解决这个矛盾。如图 3 所示。应用服务器模块采用 DPS 模型，主要考虑应用层的业务处理；网络分流器是新增加的模块，主要处理网络层的分流，限制网络层的流量，使 DPS 模型处于正常的工作状态。网络分流器在网络层工作于负荷分担模式。负荷分担模式已经有很多商业应用，可见参考文献[9,10]。

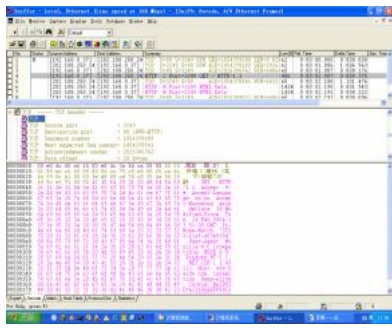

图 1　捕获的用于测试的 HTTP 协议包的流量

F5 Networks 公司采用 BIG-IP 负荷分担技术平衡 IBM 的应用服务器WebSphere Servers 的流量[10]，这个功能恰好适合我们的改进模型，因此网络分流器采用 BIG-IP 负荷分担技术实现。

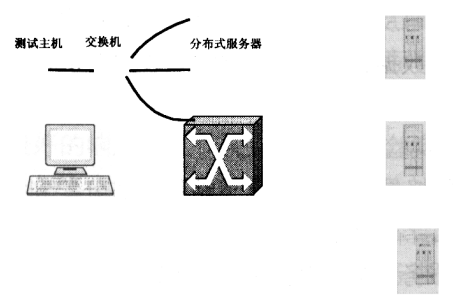

图 2　测试实验模型

5　结论

分布式并行服务器的设计是一项复杂的工作。随着网络技术的日益发展、网络业务的日益更新，基于分布式并行服务器的性能测量与分析，更成为今后网络性能研究的重要内容。对于不同的应用，有必要建立不同的性能评价模型，以保证实现不同的业务质量；而对于多种不同应用所基于的网络平台，更需要确立一种综合的性能体系架构，为进一步实现基于性能策略的网络结构提供基础。如分布式并行服务器系统体系上 WEB 业务的性能问题，由于WEB 业务的广泛性，对各种体系上 WEB 业务质量的测量和评估需要结合多方面因素来考虑，如何客观而又真实地评价各种体系上WEB 业务的质量，又如何在数据、语音、视频业务融合的 IP 网络上保证各种业务的质量，对于正处于激烈竞争状态下的运营商而言，这些都是十分迫切且重要的问题，因而，也成为下一步网络性能研究的重点之列。

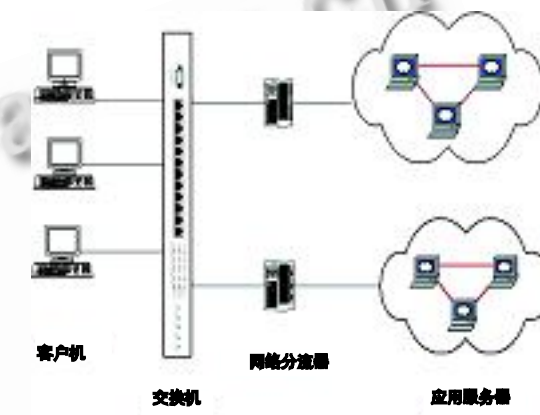

图 3　改进的分布式并行服务器模型

DPS 模型在应用层具有良好的响应特性和较强的服务能力的模型，但是在大流量条件下，存在服务能力的饱和区。通过全面的理论分析和测试，提出了改进的 DPS 模型，分布式处理网络层和应用层的流量，改善了模型的性能。

参考文献

1　杨洪勇,宗广灯,武玉强.并行服务器信息流切换的模糊控制.控制与决策, 2002,17(Supp1):758－760.
2　杨峰,刘心松,左朝树,唐续.分布式并行服务器透明性及任务调度研究.计算机研究与发展, 2003,40(9):1319－1325.
3　江勇,林闯,吴建平.网络传输控制的综合性能评价标准.计算机学报, 2002, 25(8):869－877.
4　Jain CR. The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling. New York: Wiley-Interscience, April 1991.
5　曹阳,陶舒,吴逵.基于模拟技术与测量技术的网络性

(下转第 191 页)

(上接第 122 页)

能评测方法.计算机学报, 1999,22(5):508－512.

6 刘曲明,顾桔.网络性能分析评价方法及其计算机仿真方法讨论.计算机仿真, 2000,17(1):53－59.

7 林闯,周文江,田立勤. IP 网络传输控制的性能评价标准研究.电子学报, 2002,12 (12A):1973－1977.

8 张奇智,张彬,张卫东.基于网络演算计算交换式工业以太网中的最大时延.控制与决策, 2005,20(1):117－120.

9 Rumsewicz M, Dwyer M. Preferential load balancing for distributed Internet servers. Proc. Of The 1st IEEE/ ACM Int'l Symp on Cluster Computing and the Grid, Brisbane, Australia, 2001.

10 F5 Networks, Inc. Load Balancing IBM WebSphere Servers with F5 Networks BIG-IP System. [2006-1-13]http://www.f5.com/solutions/deployment/pdfs/websphere_bigip42_dg.pdf.