

LVS集群系统模型及其任务分配算法的研究

郑勇明, 吴国才

(厦门大学 计算机科学系 福建 厦门 361005)

摘要】 对 Linux 虚拟服务器集群的结构和原理进行了分析和描述, 包括 LVS 的框架, 负载均衡技术, 并对 LVS 中的 8 种任务分配算法进行了分析, 最后提出了对任务分配算法的一种改进思想。

关键字】 集群; LVS; 负载均衡; 任务分配

0. 引言

Internet 的兴起和不断发展, 给网络带宽和服务器带来了巨大的挑战, 特别是对服务器的性能提出了更高的要求。由于高性能的服务器往往价格非常昂贵, 而且单台服务器即使性能再好, 也已经满足不了 Internet 的需求。为了在减少服务器成本的同时更好的满足 Internet 的需要, 通过高性能网络或局域网互联的服务器集群就应运而生了。

集群是由很多连接在一起的独立计算机组成, 对外界是透明的, 在外界看来, 它就像一台独立的服务器一样对外提供服务。集群已成为了实现高可伸缩的, 高可用网络服务的有效结构。Linux 虚拟服务器(Linux Virtual Server, 简称 LVS) 集群就是属于这种结构。LVS 把很多服务器连接在一起, 利用基于 IP 层的负载均衡技术把客户请求均匀的分发到各个服务器上, 从而高效的对外提供服务。

1. LVS 的结构框架

一般来说, LVS 集群采用三层结构, 使用 LVS 方式架构的集群系统结构模型如图 1 所示, 三个主要组成部分为:

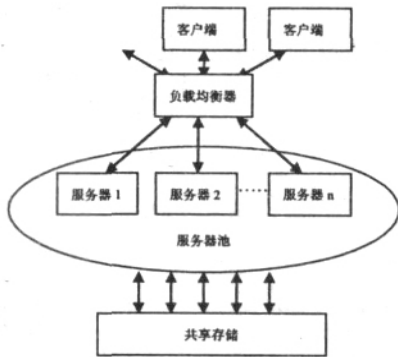


图 1 LVS 集群系统结构图

(1) 负载均衡器

它是整个集群对外的前端机, 也是 LVS 集群系统的唯一入口点, 所有的客户请求都会被发送到负载均衡器, 负载均衡器接收到客户请求后, 根据系统设定的任务分配算法从后台服务器中选择一台来执行客户的请求。不管选择哪台服务器提供服务, 客户机都认为服务是来自同一个 IP 地址上的, 这个 IP 地址称为虚拟 IP 地址(VIP)。因为所有的操作都是在 Linux 操作系统核心空间中将要完成的, 它的调度开销很小, 所以它具有很高的吞吐率。

(2) 服务器池

它由一组真正执行客户请求的服务器组成, 接收从负载均衡器发送来的请求, 并根据请求提供服务。服务器池的结点数目是可变的。当整个系统收到的负载超过目前所有结点的处理能力时, 可以在服务器池中增加服务器来满足不断增长的请求负载。对大多数网络服务来说, 请求间不存在很强的相关性, 请求

可以在不同的结点上并行执行, 所以整个系统的性能基本上可以随着服务器池的结点数目增加而线性增长。

(3) 共享存储

共享存储通常是数据库、网络文件系统或者分布式文件系统。它为服务器池提供一个共享的存储区, 当网络服务需要有相同的内容时, 共享存储是很好的选择, 当系统存储的内容越多, 共享存储的优势就越明显, 如果不采用共享存储, 则每台服务器需要一样大的存储空间, 任何的更新都需要涉及到每台服务器, 系统的维护代价会非常高, 而且资源的浪费也比较严重。

2. LVS 的负载均衡技术

前面已经提到过 LVS 采用基于 IP 层的负载均衡调度方法, 将 IP 层的 TCP/IP 请求均匀的转移到服务器池中不同的服务器上。LVS 主要有三种不同的负载均衡技术, 分别为 VS/NAT(网络地址转换), VSTUN(IP 隧道), VSDR(直接路由)。

(1) VS/NAT 技术

当请求报文到达负载均衡器时, 负载均衡器根据连接调度算法从一组真实服务器中选出一台服务器, 将报文的地址改写成选定服务器的地址, 报文的端口改写成选定服务器的相应端口, 最后将修改后的报文发送给选出的服务器。服务器响应完后将响应报文发送回给负载均衡器, 负载均衡器将报文的源地址和源端口改为系统的虚拟 IP 地址和相应的端口, 再把报文发给用户。由于请求和响应报文都要经过负载均衡器, 所以当客户请求增加到一定程度时, 负载均衡器就会成为整个 LVS 集群系统的瓶颈, 因此采用 NAT 方式实现 LVS 的系统服务器数量不宜过多。

(2) VSTUN 技术

由于 VS/NAT 具有上述的缺点, 所以如果能将请求和响应分开处理, 即负载均衡器只负责调度请求而响应直接返回给客户, 将极大地提高整个集群系统的性能。VSTUN 技术就实现了这种想法, 它利用了 IP 隧道技术, 当负载均衡器收到请求后, 根据各个服务器的负载情况, 动态地选择一台服务器, 根据 IP 隧道协议将请求报文封装在另一个 IP 报文中, 再将封装后的 IP 报文转发给选出的服务器; 服务器收到报文后, 先将报文解封获得原来目标地址为 VIP 的报文, 服务器发现 VIP 地址被配置在本地的 IP 隧道设备上, 所以就处理这个请求, 然后将响应报文直接返回给客户。这种技术要求所有的服务器都支持 IP 隧道协议。

(3) VSDR 技术

跟 VSTUN 方法一样, 在 VSDR 技术中, 负载均衡器只负责调度请求, 响应数据也是直接返回给客户。这可以极大地提高 LVS 集群系统的伸缩性。跟 VSTUN 相比, 这种方法没有 IP 隧道的开销, 但是要求负载均衡器与实际服务器都有一块网卡连在同一物理网段上。

3. LVS 的 8 种负载调度算法及其改进思想

3.1 八种调度算法

LVS支持八种任务分配算法,这些任务分配算法是以连接为粒度进行分配的,下面对这八种算法进行简单介绍和分析。

(1) 轮叫调度:以轮叫的方式依次将请求调度到不同的服务器,它假设所有服务器处理性能均相同,不管服务器的当前连接数和响应速度。因此不适用于服务器组中处理性能不一的情况,很容易导致服务器间的负载不平衡。

(2) 加权轮叫调度:为每个服务器指定一个权值,性能越高则权值越大。算法按权值的高低和轮叫方式分配请求到各服务器,权值高的服务器比权值低的服务器处理更多的连接。

(3) 最小连接调度:把新的连接请求分配到当前连接数最小的服务器。这是一种动态调度算法,它通过服务器当前所活跃的连接数来估计服务器的负载情况。

(4) 加权最小连接调度:同样为每台服务器指定一个权值,权值大小与性能成正比,当新请求到来时,选择一个权值较大且连接数较小的服务器提供服务。

(5) 基于局部性的最少连接调度:专门为高速缓存集群设计的,它尽量将对同一目标地址的连接请求分配到同一个高速缓存服务器中,这样可减少缓存节点向由目标地址确定的远端站点的访问次数。

(6) 带复制的基于局部性最少连接:与(5)类似,也是专门为高速缓存集群设计的,不同之处在于它要维护从一个目标地址到一组节点的映射,而(5)则只维护从一个目标地址到一个节点的映射。

(7) 目标地址散列调度:根据目标地址查找事先设定的静态散列表来分配连接。

(8) 源地址散列调度:根据请求服务客户端的地址查找设定的静态散列表来分配连接。

3.2 算法改进思想

通过分析可以得知这八种算法都是静态的,也就是调度策略事先已经确定了,而不考虑系统运行时的状态。这样很难保证

系统内服务器真正地实现负载均衡。为了解决这个问题,可以在负载均衡器中增加一个统计模块,用于在固定的时间间隔内动态地统计各个服务器节点的负载值,并把这些负载信息保存在负载均衡器中。当有新的连接请求到达时,负载均衡器就根据统计模块中保存的各个节点的负载信息选择一个负载最轻的服务器节点提供服务,这样可以更有效的实现真正的负载均衡,提高整个系统的性能。影响节点负载的因素很多,我们可以采用下面的公式计算节点负载:

LOAD CONN, CPU_U, CPU_S, MEM, NET

其中,CONN为服务器节点的当前连接数,CPU_U和CPU_S分别为服务器节点CPU的用户时间和系统时间,MEM为内存使用率,NET为网络带宽使用率,上述各个参数根据它们对负载影响的大小而被赋予不同的权值。

4. 结语

本文主要介绍了LVS的体系结构以及3种负载均衡技术和8种任务调度算法,并提出了对调度算法的一种改进思想。LVS具有高可靠性,高可用性,高可伸缩性,因此可以使用LVS来实现各种实际的应用,比如构建大型的视频点播系统以及电子商务应用等等。

参考文献:

1. Zhang Wensong. Linux Virtual Server Web Site. <http://www.linuxvirtu-alserver.org>
2. Rajkumar Buyya. 高性能集群计算:结构与系统(第1卷)[M],北京:电子工业出版社,2001
3. 姚耀文,茹建斌,缪炯焯. 基于Linux的服务器群集方案. 计算机工程, 2001, 27(4):139-141
4. Borzemi L, Gajewski D. A Load Balancing System for Unix-based Local Area Networks. Micro processing and Micro-programming, 1993, 39(2): 205-208

(上接第40页)

的分布式虚拟环境,清华大学国家光盘工程研究中心所做的布拉宫实现了大全景虚拟现实等。

现在,虚拟现实系统已突破了过去航空航天、军事、娱乐等几个特定应用领域,打破了只有政府才能用得起来的技术,渗入到了生活的各个方面,比如航空、航天、铁道、建筑、土木、科学计算可视化等各个领域。

三. 虚拟现实的趋势展望

虚拟现实技术是20世纪80年代才兴起的一门崭新的综合性信息技术,尚处于新生阶段,远未达到成熟阶段,同时,虚拟现实技术还需渗透到信息系统的各个领域,改变信息系统的人机交互方式,甚至改变现代信息系统的设计原则,使之更能适应以用户为中心的应用需求,只有这样,虚拟现实技术才能够得到快速健康的发展。从目前信息技术的发展来看,网络已经进入了人们的生活、工作和学习中,虚拟现实与网络的结合也日趋紧密和完善。目前,国内外已经将虚拟现实与网络结合,建成了多个商业宣传网站,甚至用于游戏。

同时,计算机图像压缩技术和显示技术的不断进步以及相应硬件设备的日趋完善,我们不难预测未来虚拟现实技术的发展动向,这种动向大致包括两个方向:一方面是朝着桌面级虚拟现实发展,特别是在商业领域。目前,已有数百家企业正在致力于桌面级虚拟现实的开发,其主要用途有商业展示教育培训及仿真游戏等;另一方面是朝着高性能沉浸式虚拟现实发展。很多高科技领域,如航空航天军事训练、模拟训练等,由于各种特殊要求,只能进行仿真试验,简单的桌面级虚拟现实根本无法满足,

因此只能建立高性能的沉浸式虚拟现实,目前国内外已建成多个高科技领域虚拟现实仿真系统。

四. 结束语

经过几十年的发展,我们可以看到,虚拟现实逐步从萌芽状态成长为今天日趋成熟的综合信息技术,并在各个领域得到了越来越多的应用发挥。然而,虚拟现实技术还是一门年轻的科学技术,还存在许多问题有待解决,比如,在计算机的虚拟环境中,操作者每次转动头部,计算机都必须更新三维图像,由于更新的数据量大,以致计算机还无法完成实时计算。这就造成了系统滞后。还有,美国空军的虚拟现实模拟器产生的视觉运动信号与人的感觉之间之间存在差异,容易引起头痛、眩晕等。

虽然如此,但它独特的优势特长为各个领域的发展提供了一个全新的突破口,同时,我们也相信,随着计算机技术,尤其是网络技术的飞速发展虚拟现实技术将会得到更为长足的发展,必将更为广泛地应用于各个领域为人类的生产生活带来全新的面貌。

参考文献:

1. 张茂军,虚拟现实系统. 科学出版社,2002.
2. 党保生,虚拟现实及其发展趋势,中国现代教育装备,2007.4
3. 张占龙等. 虚拟现实技术概述[J]. 计算机仿真,2005,3
4. 牟长军、董伟志,综述虚拟现实技术及其应用,装备制造技术,2007.1
5. 罗坚,虚拟现实及发展趋势,浙江工贸职业技术学院学报,2004.12
6. 陈雁飞、马建民、高丽,虚拟现实技术综述,综述与展望,2004.5