

分布式均衡算法在防火墙集群中的实现

Distributed Load - Balanced Algorithm Apply to Firewalls' Cluster's

吴璇 隋红建 (中国科学与技术大学管理学院 100049)

摘要:集中式的负载均衡算法虽然能提高防火墙集群系统的工作效率,但是负载平衡服务器本身承担了信息收集和权值计算等任务,会成为网络流量新的瓶颈。分布式负载均衡算法,将各结点信息的收集和权值的计算等任务交由结点机自己完成,结点机主动将计算出的指标值定期发送给负载均衡服务器。负载均衡服务器只运行负载均衡算法来均衡网络访问,减少了中心负载均衡服务器的工作量,更有效地提高了防火墙集群系统的工作效率。

关键词:防火墙 集群 负载均衡算法

防火墙系统是当今最有效的网络安全解决方案之一。由于它在单位时间内允许通过的信息流量是有限的,所以常常成为整个网络系统的瓶颈,而且单台防火墙的可靠性也比较差,为此人们采用了防火墙集群技术来解决这个问题。集群就是把两个或多个提供相同服务的系统通过网络结合起来一起工作,在外部看来整个集群表现为一个单一的系统。它建立在现有的设备和网络结构上,提供了一种廉价的方法,来增强服务器的处理能力。集群需要一个负载均衡器把多个负载请求均衡地分配到不同的结点上,以充分利用集群中各结点的效能,从而提高系统的总带宽和容错性能。负载均衡服务器是一个工作在网络层上的平衡系统,整个服务器集群体现为一个单一的IP地址被用户访问。负载均衡服务器可静态或动态配置,用一种或几种算法决定哪个结点将获得网络服务请求。

1 现有的负载均衡算法

负载均衡算法可分为静态算法和动态算法两大类。为了能够充分保证负载的均衡,还可以通过加权值来决定服务器集群中各结点的优先级,按照优先级的先后顺序进行负载的合理分配。权值的获取分为基于流量的权值计算和基于负载的权值计算。基于

流量的权值计算是通过分析当前网络的平均流量或者是服务器结点的平均流量来获得的;基于负载的权值计算是通过负载信息采集工具定时获得结点的CPU、RAM、I/O的运行状况,来判断结点是否处于过载状态。

1.1 静态负载均衡算法

负载均衡算法设计的好坏直接决定了集群的工作效率,具体在负载均衡器上的表现是,设计不好的算法会导致集群的负载失衡。有些简单平衡方法可以独立使用,有些必须和其它方法组合使用。一个好的负载均衡算法也并不是万能的,它一般只在某些特殊的应用环境下才能发挥最大效用。因此在考察负载均衡算法的同时,应注意算法本身的适用面,在采取集群部署的时候根据集群自身的特点进行综合考虑,把不同的算法和技术结合起来使用。常见的静态负载均衡算法有:轮转法(Round - Robin);散列法(HASH);最少连接数法(Least - Connection);最低缺失法;最快响应法。

1.2 动态负载均衡算法

客户访问集群系统时,提交的访问请求所需的时间和所要消耗的计算资源是千差万别的。一些负载比较重的请求需要进行密集计算的查询、进行数据库的访问、有很长需要响应的数据流;而负载比较轻的请求

只需要读一个小文件或者进行很简单的计算。请求处理时间的不同可能会导致处理结点利用率的倾斜 (Skew), 即集群中各服务器的负载不均衡。

因此, 产生了根据负载不同随时调整整个结点分配的动态负载均衡算法。在运行动态负载均衡算法的集群内部, 负载均衡服务器上运行服务器端监控程序, 它负责监视和收集集群内各个结点的负载信息; 而防火墙客户端结点上运行客户端程序, 负责定时向负载均衡器报告自身的负载状况。监控进程根据收到的全部结点的负载信息来进行同步操作, 对将要分配的任务按照权值的比例重新分配。权值的计算主要根据各个结点的 CPU 利用率、可用内存以及磁盘 I/O 状态计算出来。

2 分布式负载均衡算法

现有的动态负载均衡调度算法通常由负载均衡器收集各结点的负载信息, 按照服务器中预先设置好的算法计算出权值, 比较计算所得值找出最优结点, 将网络访问发送到选定结点上。这是一种集中式调度的工作方式。在这种工作方式中, 负载均衡服务器承担了对负载的均衡、对各结点的信息收集和权值计算等任务, 易成为网络流量新的瓶颈。为了解决这个问题, 我们设计了分布式负载均衡算法, 将结点信息的收集和权值计算等任务交由运行客户端的结点机来分析计算, 再主动将计算出的值反馈给服务器。而运行服务器端模块的均衡服务器只运行负载均衡算法来分配网

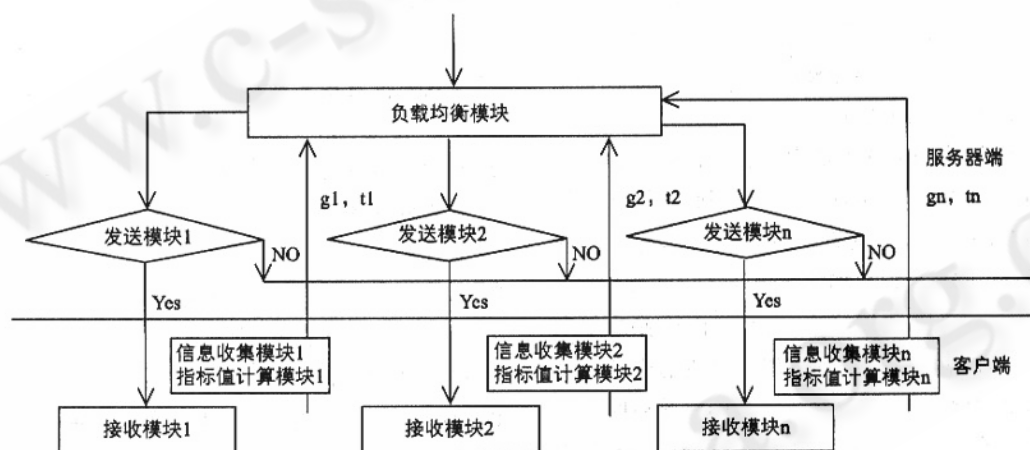


图 1 分布式负载均衡算法流程图

1.3 权值计算

动态权值是由结点运行时各方面的参数计算出来的。我们可以选取最重要的几项: CPU 资源 (cpu), 内存资源 (mem), 连接数 (cnt), 响应时间 (rtme) 等信息, 都可以作为计算权值的因子。动态权值的目的是要能更准确地反映出结点负载的当前状况, 来预测将来可能的负载变化。不同系统中, 不同参数的重要性也是不同的, 可以为每个参数因子设置一个常量系数 λ , 表示各个负载因子的重要程度。在实际应用中, 可以对参数不断修正, 来更好地反映系统的负载情况。

络访问, 大大地减少了中心结点的工作量, 更有效地提高了集群系统的工作效率。

在分布式动态负载均衡算法中, 我们在服务器端模块上只运行: 监听、接收、负载均衡和发送四个子模块; 客户端模块上运行接收、发送、信息收集和指标值计算四个子模块, 如图 1 所示。服务器端负载均衡模块只需要根据简单的均衡算法分配网络访问即可, 不必再去主动收集各结点机的信息, 降低了对信息收集和计算所增加的额外开销, 减轻了负载均衡服务器的负担。各结点权值因子的收集和指标值的计算都由客户端的信息收集和指标值计算两个子模块来完成。根

据当前已有的各结点发来的指标值 g_i 结合权值 t_i 来进行调度决策。

把信息的收集和权值的计算等任务分配到结点机上,极大地减轻了负载均衡服务器的工作压力,有效地提高了整个负载均衡系统的工作效率。

3 结论

集群中的客户端结点初次投入系统中使用时,可以为每个结点都设置一个初始权值 t_{i0} ,随着结点负载的变化,负载均衡器再根据当前指标值对权值进行调整。结点的动态权值计算公式可描述为:

$$t_i = [g_1, g_2, g_3, g_4] \begin{bmatrix} \text{cpu}_i \\ \text{mem}_i \\ \text{cnt}_i \\ \text{rtime}_i \end{bmatrix}$$

其中 $i=1, 2, \dots, n \sum g=1$

通过实验,当防火墙集群中只有一台防火墙服务器时,平均网络响应时间为 18ms;采用轮转算法平均网络响应时间几乎没有变动;采用了集中式动态均衡算法,由于算法复杂性,平均网络响应时间反而有所下降;采用分布式动态均衡算法时,同样由于算法复杂性,平均网络响应时间也有所下降。当集群中有两台防火墙服务器时,采用轮转算法平均网络响应时间提高了 25%;采用集中式算法的平均网络响应时间提高了 41%;采用分布式算法时的平均网络响应时间提高了 41%。当集群系统中有三台防火墙服务器时,采用轮转算法平均网络响应时间提高了 49%;采用集中式算法平均网络响应时间提高了 62%;采用分布式算法时平均网络响应时间提高了 70%。可见,分布式动态负载均衡算法确实能够有效地提高集群系统的工作效率,而且集群中防火墙服务器数量越多就越能体现分布式动态负载均衡算法的优势。

参考文献

- 1 Kai Hwang & Zhiwei Xu, Scalable Parallel Computing Technology, Architectue, Programming, Published by McGraw-Hill, May. 2000.
- 2 Ananth grama, Anshul Gupta, George Karypis&Ipin kumar, Introduction to parallel Computing (Second Edition), Published by Pearson Education Ltd., 2003.
- 3 傅清祥、王晓东, 算法与数据结构(第二版), 电子工业出版社, 2001年8月。
- 4 王霜、修保新、肖卫东, Web 服务器集群的负载均衡算法研究, 计算机工程与应用, 2004年第25卷, 78-80。
- 5 刘安丰、陈志刚、曾志文、曾碧卿, 基于数据挖掘的 Web 集群负载均衡算法, 计算机工程与应用, 2003年第25卷, 59-61。
- 6 刘红、白栋、丁炜, 应用于 MPLS 网络负载均衡的启发式自适应遗传算法研究, 通信学报, 2003年第24卷, 39-45。
- 7 周佳祥、郑纬民、杨广文, 一种基于进程迁移的自适应双阈值动态负载均衡系统, 清华大学学报, 2000年第40卷, 121-125。
- 8 林国辉、马正新、王勇前、曹志刚, 基于蚂蚁算法的拥塞规避路由算法, 清华大学学报, 2003年第43卷, 1-4。
- 9 高鹏、张德远、刘刚、李庆海, 并行电子邮件监听代理的负载均衡算法, 西安交通大学学报, 2003年第37卷, 372-379。
- 10 刘健、徐磊、张维明, 基于动态反馈的负载均衡算法, 计算机工程与科学, 2003年第25卷, 65-67。
- 11 李煜、李汉菊, 基于 CompactPCI 的防火墙并行技术研究, 计算机工程, 2003年5月, 第29卷, 第8期, 47-61。
- 12 骆宗阳、董玮文、杨宇航、王澄, 一种具有高可用性的通用负载均衡技术, 计算机工程, 2003年2月, 第29卷, 第2期, 134-136。
- 13 纪翀、胡静、陈春铃, 分布式主动型防火墙, 微机发展, 2004年2月, 第14卷, 第2期, 97-99。
- 14 吴璇、隋红建, 负载均衡技术在防火墙系统中的应用, 计算机应用, 2003年12月, 第23卷增刊, 22-24。
- 15 中国科学技术大学硕士学位论文, 负载均衡技术在防火墙的应用, 吴璇 2004年7月。