

内容交换技术在服务器负载平衡中的应用

曾庆

(西昌铁路高级技校, 四川 西昌 615000)

【摘要】通过分析内容交换技术(Content Switching)在服务器负载平衡中的典型应用,展示了如何利用内容交换技术优化服务器负载平衡,说明了为提高服务器的性能,更有效的管理和控制网络,应该建立以内容交换技术为核心的网络架构。

【关键词】内容交换技术;服务器负载平衡

【中图分类号】TP393 **【文献标识码】**A **【文章编号】**1673-1891(2008)01-0076-04

引言

随着 Internet 的迅速发展,网络带宽的增加,新应用的涌现,网络中信息的流量更大、信息的内容也更丰富,给网络和服务器带来了极大的压力,新的瓶颈是应用程序和服务器。因此,需要在网络中部署内容交换技术,识别信息的真实内容,从而对信息进行更有效的管理,并实现更明确的负载平衡,完成网络向智能化的转变,提高网络和服务器的性能。

1 内容交换技术

内容交换是指第4层到第7层的交换,是针对 Web 流量和复杂性迅速增长的现象而提出的一种技术。内容交换技术可以智能识别各种不同高层的应用,并进行有针对性的转发。在多数情况下,内容交换技术根据各应用程序使用的 TCP/UDP 端口(如 FTP、HTTP、SMTP 等等)进行流量的引导,实现基本的负载平衡。除了考虑第4层的信息外,更重要的是,内容交换允许深入到第7层,能够根据统一资源定位符(URL)、统一资源标识符(URI)、文件类型、Cookies 等,将应用程序的请求转发到与之相匹配的服务器,与利用 TCP/IP 端口相比,提供更明确的负载平衡。

2 内容交换技术在服务器负载平衡中的应用

目前为止内容交换技术最常用的领域是服务器负载平衡,内容交换的一个重要任务是从一个单独的服务器模型转换为一种可以将负载均匀地分布到很多不同对象服务器的模型。

在第四层,可以根据源/目的 IP 地址以及应用程序所使用的 TCP/UDP 端口对流量进行引导,从而实现基本的服务器负载平衡。然而,在 Internet 迅速发展壮大的今天,仅仅部署第四层的负载平衡是不够的。对于某些应用,使用有效负载(实际的用户数据)中所含的信息来进行决策可以带来更多的好处,我们需要基于第七层(应用层)的内容识别服务

器负载平衡。在第七层,可以使用:HTTP URL 解析、HTTP 报头检查、HTTP Cookie、FTP 解析、DNS 解析、RTSP 流解析等,来实现更为明确的、基于内容的服务器负载平衡。下面,将探讨第七层内容服务器负载平衡是如何实现的。

仔细地查看下面的客户端 HTTP GET 请求,其中的内容会在后面的论述中引用。

```
Hypertext Transfer Protocol
GET /brochures/aboutus.html#whereweare HTTP/
1.0\r\n
Accept:image/gif, image/x-xbitmap, image/jpeg,
image/pjpeg\r\n
Accept-Language:en-gb\r\n
User-Agent:Mozilla/4.0(compatible;MSIE 5.01;
Windows NT 5.01)\r\n
Host:www.foocorp.com\r\n
Connection:Keep-Alive\r\n
\r\n
```

2.1 HTTP URL 解析和负载平衡

利用 HTTP URL 解析,可以区分静态内容和动态内容,通过解析 URL 中的文件扩展名(如*.gif、*.html、*.bmp 等)来识别静态内容。动态内容则可以通过文件扩展名(如*.asp 和*.cgi)以及字符串(如/cgi-bin/)来识别。网络管理员可以使用这种方法来维护两个不同的服务器配置文件,以便于处理请求类型上的差异。静态内容只需要访问静态内容服务器,其传输特点是对内存和 CPU 的占用较小,而动态内容则需要访问应用程序服务器和后端数据库,其传输需要消耗服务器较大的内存和 CPU 资源。这样,可以根据内容的不同,配置不同性能的服务器,从而更合理化的利用资源,实现更优化的负载平衡。图1给出了一个区分静态和动态内容类型的示例。

收稿日期:2007-12-04

作者简介:曾庆(1969-),男,汉族,四川中江人,讲师,本科,主要从事计算机专业教学与研究。

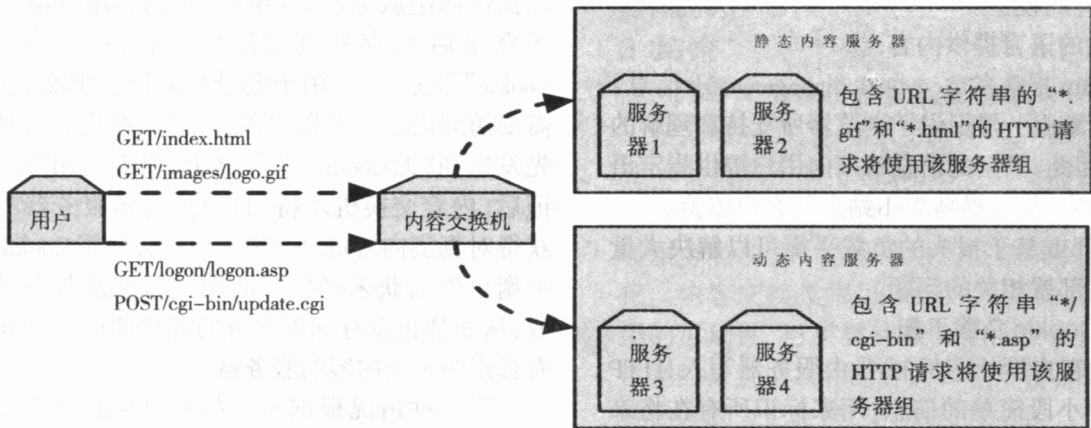


图 1 使用 URL 解析静态和动态内容

此外,另一个使用 URL 解析的用处是向现有站点添加新的部分。例如:某个站点需要增设促销部分,该部分可能只在有限的时间内应用,而且可能由单独的管理员来管理。那么,通过添加 URL 字符串 “/promotion/” 来识别这一新增内容,并使用一组新的对象服务器将该部分从站点分离出来,一旦促销结束,只须将这部分服务器撤走即可,不会影响到原来的服务器。

2.2 HTTP 报头负载均衡

2.2.1 使用 HOST:Header 实现虚拟主机

使用 Host:header(Host:www.foocorp.com), 可以更好的实现虚拟主机。虚拟主机一般指使用单个 IP 地

址来表现多个主机或 Web 站点。在具备 HTTP Host:header 这个功能之前,要实现虚拟主机,需要为客户的每台单独的主机分配一个 IP 地址,或必须在每个对象服务器上包含所有基于主机的 Web 站点内容。由于 Internet 地址的注册费用比较昂贵,因此这种方式的扩展性代价不菲。而使用 HTTP Host:header 是一个更为有效和可升级的部署模式。从内容交换的角度来看,可以使用单个 VIP 来表现一系列的站点,而站点的主机名则使用 Host:header 来区分。需要注意的是,此方法只有在 HTTP/1.1 规范中才可使用。图 2 给出了通过对虚拟主机使用 HTTP Host:headers 部署的共享资源模型。

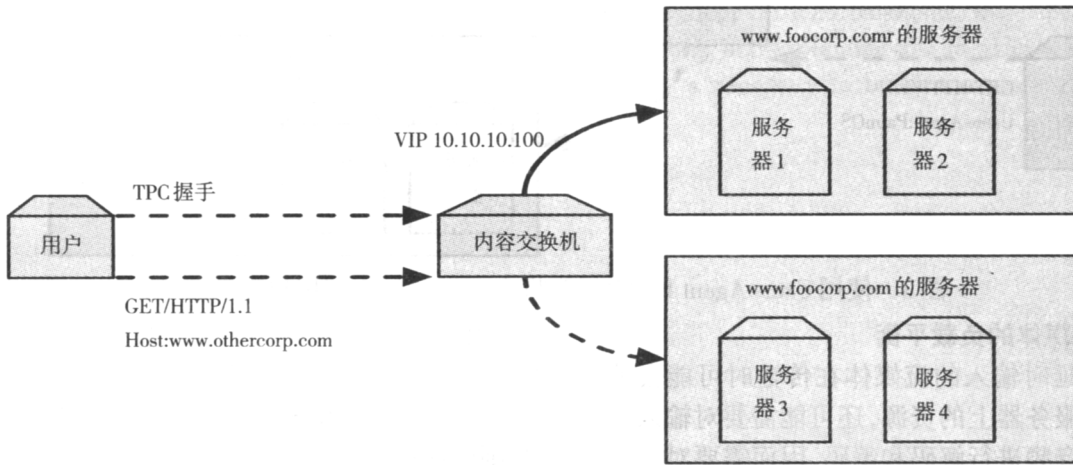


图 2 通过对虚拟主机使用 HTTP Host:headers 部署的共享资源模型

2.2.2 根据浏览器的类型优化内容

并非所有的浏览器都相同,同时,用户可能使用 WAP 电话或是 PDA 来查看站点。无疑的是,将站点内容以最适合用户的浏览器的形式发送到用户是一个需要解决的难题。

在 HTTP 规范中,对 User-Agent(用户代理)有一个标准的报头定义,User-Agent 提供了关于发出请

求的浏览器类型的信息以及有关兼容性和所使用的操作系统的信息。

如:User-Agent:Mozilla/4.0(compatible;MSIE 5.01; Windows NT 5.01)\r\n

这样,凭借读取该报头信息的能力,可以部署一个主机方案,其中同一内容对不同的浏览器和接收设备进行优化,可以极大地增强用户的体验。图

3展示了这样的模型。

2.2.3 以正确的语言提供内容

查看 Http 报头的 Accept-Language:en-gb 信息,该报头向对象服务器标识了浏览器所支持和理解的语言编码,因此,根据此信息可以向用户提供指定语言的 Web 内容,从而提高 Web 站点的客户体验。

总之,部署基于报头的负载均衡可以解决大量与 Web 站点部署相关的问题。

2.3 HTTP Cookie 负载均衡

Cookie(客户端状态机制)是由服务器写入 HTTP 内容中的一小段简单的信息,用来标识所有在将来要访问该站点的客户端信息。通常有两种类型的 Cookie—永久的和临时的。Cookie 可以很好的解决 Internet 上的持续性问题,Cookie 值用于决定客户端的 TCP 连接应当连接到哪一个对象服务器以便于保证用户的持续性访问,这通常用于电子商务中。但这里主要讨论 Cookie 在负载均衡的应用。

第一种情况是使用 Cookie 负载均衡对 Web 站点实施一种严格的登录和验证模式。例如:站点可

以提供两组服务器,一组称为“内容服务器”(仅用于登录后),存放真实的内容;另一组称为“无 Cookie”服务器组,用于登录和验证。那么,在用户需要访问内容服务器之前,内容交换机会将请求首先发送到“无 Cookie”服务器上,只有当用户经过验证后,内容交换机才将用户与“内容服务器”相连,获得对真实内容的访问权。因此,使用 Cookie 负载均衡可以捕获未经授权的用户会话及其登录和授权,从而禁止没有标识自身的客户使用其 IP 访问含有真正的站点内容的服务器。

第二种情况根据写入的 Cookie 值确定服务的质量或授权给用户的内容,提供优化服务。由于 Cookie 值实际上可以是任何字符串值,因此使用这一模式可以实现许多不同的服务级别。比如,可以使用“金卡”、“银卡”和“标准”服务级别。当用户成功登录并通过验证时,此 Cookie 将允许内容交换机根据用户付费购买的服务将用户请求定向至相应的用户组,这种模式可以用来提供附加的资源和资源,如更好的性能和更丰富的内容。

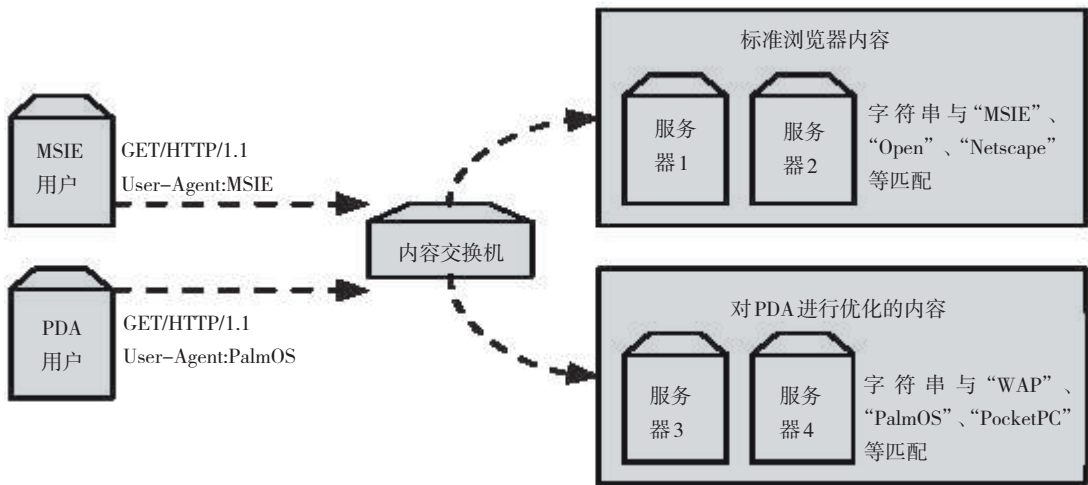


图3 使用User-Agent HTTP报头将内容发送到不同用户类型

2.4 RTSP 流媒体的负载均衡

实时或延时输入的流媒体在传输时可能会消耗大量对象服务器上的资源,还可能需要对输入的现场视频或音频进行解码和编码,因而需要对流媒体服务器进行负载均衡。可以在第四层和第七层上实施这一部署。但如果仅在第四层上进行负载均衡,面临的一个问题是客户端可以在传输视频和音频数据的 UDP 流上看到对象服务器真正的源 IP 地址,带来安全性的问题;另一个问题是如果提供服务的 ISP 或者主机服务在防火墙四周的路由器上部署了反欺骗策略,那么,真实服务器将需要位于公共的、可路由的地址空间中,这将导致较大的成

本。因而向使用 RTSP 的流媒体服务的负载均衡添加第七层智能性有许多好处。

首先,通过将 RTSP、RTP 与 RTCP 通路连接在一起,则能保证发送 RTP 内容的服务器可以位于不可路由的地址空间中,而且内容交换机可以确保这些地址无法被外部看到。

第二,在提供流媒体的服务器之间区分不同的内容类型可以更优化的配置服务器,因为不同的流类型需要不同的资源。例如:现场视频流需要消耗大量的 CPU 和内存,以便于对象服务器资源将现场视频和音频转换成正确的压缩格式,然后发送到客户端;而静态的或预先录制的内容消耗的 CPU 和内

存一般都较小,因为在内容的记录和定位过程中已经进行了压缩和编码。通过内容交换机解析RTSP流的DESCRIBE命令,如:DESCRIBE .../static/oldyella.rm(包含/static/表示静态内容);DESCRIBE .../live/bbcnews.rm(包含/live/表示动态内容),就能够区分不同的内容类型,从而将不同的请求转发到不同的对象服务器。此外,也可能有其他的要求,如区分免费内容与按需要付费的内容、已通过验证的内容与未通过验证的内容,或者使用URL来区分共享主机环境下的付费用户等等。无论是哪一种要求,解析RTSP流传输的URL的能力可以非常有效地加强

媒体流的实施。

3 结论

内容交换技术不仅在服务器负载均衡中得到了充分的应用,还可用于Web缓存重定向、防火墙和VPN负载均衡中,在网络中部署内容交换,可以更智能化的引导流量,更合理、更充分的利用网络和服务器资源,使得网络的性能得到进一步的优化。内容交换技术可以最大限度提高电子商务应用的可用性、可扩展性和安全性。我们相信,随着带宽的日益增加和内容的日益丰富,内容交换技术将迅速成为企业和ISP网络的标准。

注释及参考文献:

- [1]MATTHEW SYME, PHILIP GOLDIE.Optimizing Network Performance with Content Switching[M].于涛,罗庆华,谢文亮等译.网络性能优化——内容交换技术.北京:清华大学出版社,2004:80-110.
- [2]ANDREW S.TANENBAUM.Computer Networks[M].Fourth Edition.潘爱民译.计算机网络(第4版).北京:清华大学出版社,2004:538-557.
- [3]WADE EDWARDS, TERRY JACK, TODD LAMMLE, ROBERT PADJEN, ARTHUR PFUND, TOBY SKANDIER, CARL TIMM.CCNP Complete Study Guide[M].张波,谢琳译.CCNP四合一学习指南.北京:电子工业出版社,2005:432-441.
- [4]程薛.内容交换机智能处理网络流量[N].中国计算机报,2006 05 15.

The Application of Content Switching in Server Load Balancing

ZENG Qing

(Xichang Railway High School, Xichang, Sichuan 615000)

Abstract:By analyzing the application of content switching in server load balancing, the paper demonstrates how to use the Content Switching to optimizing server load balancing, and explaining that in order to enhance server performance and effective management and control network. It's necessary to build a network frame based on content switching.

Key words:Content switching; Server load balancing