

WAP 网关中 Cache 的实现研究

The Realize of Cache in WAP Gateway

许小刚 夏 勤 顾冠群

(东南大学 WAP 研究小组 东南大学网络研究室 南京210096)

Abstract This Paper introduces the function and location in network of cache;summrizes the framework of WAP Gateway,realized by WAP research group of Southeast University;the paper also discusses how to place the cache in WAP Gateway and its replacement,then analyses those performance and feasibility. Last,the paper proposes the related works in the future.

Keywords WAP,Cache,Proxy

一、介绍

caching 通过节省网络带宽(具体地说就是减少网络交互双方交互的次数,和收发数据的时间)减少终端用户的延迟和服务端的负载来提高网络和系统的性能,因此在网络中放入 cache 已经被绝大多数网络运行程序所采纳。但是 cache 存放位置和其替代算法对 cache 是否能够充分地发挥其作用起着至关重要的作用。而选择 cache 存放的位置和其替代算法,是与具体使用 cache 的程序的结构和功能,以及 cache 内容的特性而言的,而没有通用的方式来使用 cache 的。

本文主要讨论的是东南大学 WAP 研究小组的 WAP 网关中 cache 的应用。本文概述了 WWW 中 cache 使用的位置和所起到的作用。综述了 WAP 网关的体系结构,和其使用 cache 的必要性并定位了 cache 在本网关中要起到的作用。讨论了网关中 cache 的具体实现方法和 cache 位置的存放。简要地分析了 cache 对网关性能的提高。最后提出了今后的相关工作和 cache 实现的建议。

二、Cache 综述

网络中减少数据获取时间最通用的技术是 caching。WWW 中的 caching 可分为三类:客户端 caching、网络 caching(例如 proxy caching)、服务器端的 caching。我们将简要地描述这些 WWW caching。

客户端 caching(Client-side caching) 是指在网络中接收数据的一方对接收的数据进行缓存。这对于那些特殊类型的用户和频繁访问某一特殊内容的数据很有用。客户端的 caching 又有会话级的、主机级的和 LAN 级的 caching。实验表明 LAN 级的是最有效的策略,因为会话级的和主机级的要在每个客户端保存文

档的拷贝,因此浪费了 cache 的空间。但这三种级别的 cache 适用于不同的应用中,不能够绝对地比较谁优谁劣。

网络 caching(Proxy Caching) 就是指在网络数据传输中的某些网络节点对数据进行缓存。这种 cache 主要的用处有两种:一是对不同的服务流的数据的 cache,二是网关代理级的 cache。网络 cache 中主要的问题是 cache 的位置问题(文[2]主要讨论了在网络中如何放置给定数量的 cache,从而更有效地减少网络的负载和提高平均延迟)。网关代理级的 cache 主要是指那些在网络服务或者网关级的网络节点中的 caching。网络服务节点的 cache 对提高该节点服务的性能和质量起着很重要的作用。网关级的 cache 主要是在网络的边缘对特殊内容的 cache。WAP 网关中的 cache 就属于这一种。

服务器端 caching(Server-side caching) 是指在网络数据源对客户端请求的数据进行缓存,其主要目的是提高服务器访问服务器中资源的次数从而减少服务器的负担,提高服务器的效率。服务器端的 caching 中有文件级的和内存级的 cache 存放,但绝大多数都是文件级的。

客户端和代理 cache 主要目的是减少由于网络传输而造成的内容的读取延迟。服务端的 cache 主要目的是减少服务器端的响应时间和提高服务器的吞吐量。由于 cache 的目的不同,以及客户端、代理、服务器的功能的不同导致了它们对 cache 的大小、方式、阶段都存在一定的差异。客户端的 cache 一般情况下不很大,替代算法、刷新策略都比较简单,因此效率不是很高。代理 cache 一般面向的是某一类特殊的用户,这些用户访问的内容有很多是相同的,而且访问也很频繁,因此代理 cache 的好坏可以使某些网络流量大大地减

少,从而导致了服务器端的负载和客户端的延迟。服务器端的 cache 的主要目的是为了减少服务器的响应时间和服务器的吞吐量。

三、WAP 网关的体系结构

首先要明确的是 WAP 网关是边缘网关,它起的作用是协议转换的作用,因此为了减少网络流量和用户延迟,就必须在网关中加入 cache。WAP 网关的体系结构如图1所示。

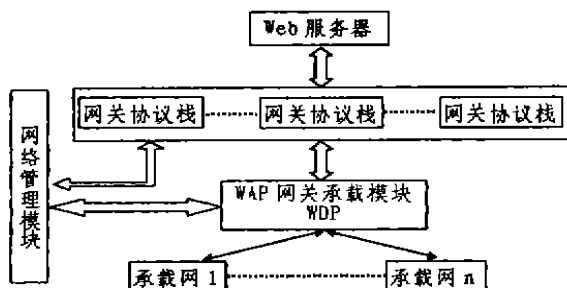


图1 WAP 网关体系结构

图1中,WAP 网关中的承载模块即 WDP 所起的作用是:1)匹配不同的承载网络包括 SMS, USSD, GSM 等;2)为上层转发报文,如果是支持 IP 的承载网络直接转发 IP 如:GPRS。

WAP 的网关协议栈的功能是完成 WTLS, WTP, WSP 等协议的功能,最后向 Web 服务器发送 HTTP 请求,获得数据,编码,返回给客户端。

从图1中可以看出网关协议栈可以有多个,这是指网关协议栈可以和底层运行于同一台主机,也可以独立于 WDP 而运行于不同的主机上,由 WDP 为不同的协议栈进行选路。这种设计思想是基于集成服务的思想的。这样可以在客户端很多的情况下动态地增加协议栈从而可以增加整个网关的运行效率,减少某一协议栈的负载,同样在客户端较少的情况下,可以通过网管动态地减少协议栈的个数以节省系统资源。

这种集群化服务的设计思想对于整个网关系统的

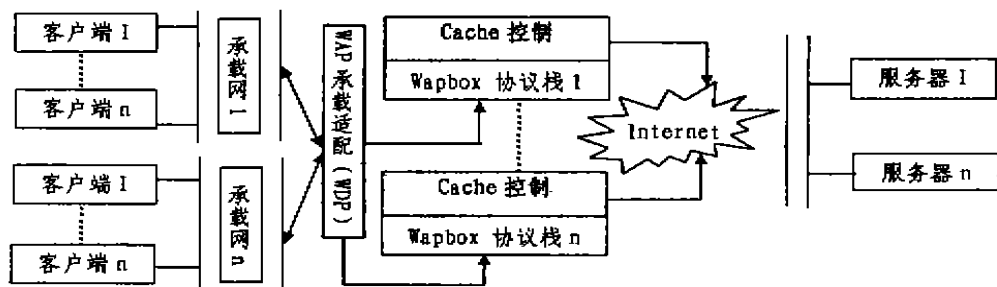


图3 分散式 cache 实现方案

运行有很好的可扩展性,但是由于运行多个协议栈,而且它们是相互独立的,因此如果没有 cache 或者 cache 处理的不当,不但使系统的效率降低,而且会增加不必要的请求,增加了网络负载。

四、WAP 网关中 Cache 的实现

WAP 网关是连接无线网络和有线网络的协议转换网关,是属于代理网关,所以在 WAP 网关实现的过程中我们可以借鉴代理网关中通用的 cache 实现方法,然后结合 WAP 网关自身的特点来设计 cache 的实现。

通用代理网关 cache 的实现(图2) 这种设计思想是客户端访问 Internet 时首先与 WWW 代理建立连接,当客户端的请求在 cache 中时,代表服务器根据一定的规则判断是将数据直接返回给客户端还是向远端服务器请求,如果请求不在 cache 中,代理即向远端的 WWW 服务器请求数据,在将数据返回给客户端的同时,代理按照一定的规则决定是否将数据写到 cache 中去。

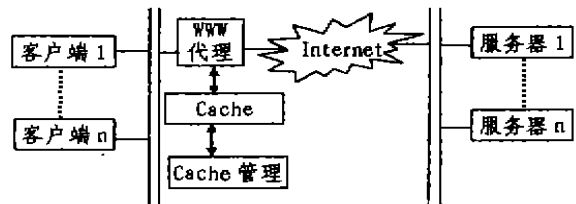


图2 通用代理网关 cache 实现

以上这种通用的 cache 实现方法对于用户数比较少,访问不是很频繁的请求下效果比较好。

WAP 网关中 cache 的实现 WAP 网关是代理网关,所以在 cache 的实现中也参照了上面的实现模式。但本网关采用的是集群化的思想,所以在实现时要注意到不同主机之间的协调问题,cache 具体实现有下面两种方式。

一种是在运行 WAP 协议栈的每台主机上分别放置 cache,不同主机上的 cache 互不相关,具体如图3。

图3的方案是在运行 wapbox 协议栈的各个主机上分别有 cache 控制,不同主机上的 cache 互不影响,这是一种分散的设计思想,这种方案中 cache 的实现简单,因为各个 cache 之间没有关系,实现时就不会考虑 cache 内容的重复,cache 的替代算法也可以相互独立。但是由于 WDP 不会解析哪一次请求的 url 在那个 wapbox 上已经存在 cache 中,它的分配是随机的,因

此 wapbox 处理的内容也是随机的,不会和 url 有关系,这就会导致 wapbox 中 cache 内容几率减少,相同的 url 会请求多次,网络流量会增加。

为克服分散设计思想中的 cache 内容几率减少, url 请求多次的问题,我们设计了集中式的设计方案,如图4所示。

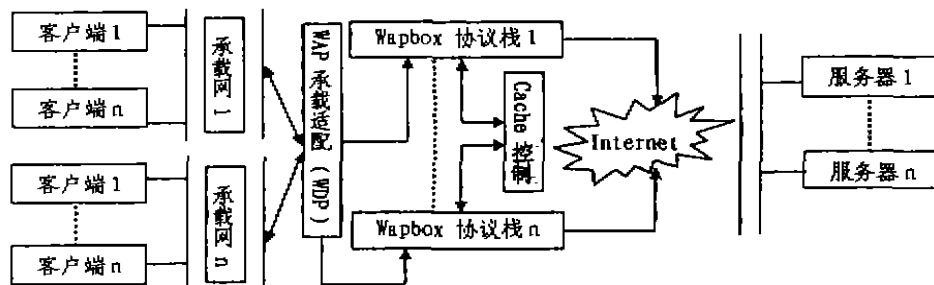


图4 集中式 cache 处理方案

集中式设计是指当每个协议栈要发送 url 请求时首先访问公共的 cache 管理,看 url 是否已经在 cache 中,如果没有再发送 url 请求到 Internet。这种方案解决了分散方案中的问题,但是实现更复杂。

五、分析

上述两种方案设计各有各的优点和缺点,但是通常情况下采用集中式的设计方案,因为这种设计方案效率很高,虽然这种方案增加了各个协议栈和 cache 控制的交互,但是 cache 的位置和协议栈是在一个局域网中,这种交互带来的延迟要比到 Internet 上取数据的延迟要少得多。

然而,由于目前 WAP 的用户不是很多,访问量不很大,我们针对这种情况设计了单机版的 WAP 网关,这种网关是图3的一种特例,cache 采用的是固定 cache 内容的最大生命期,替代算法采用的是 LRU,cache 内容的查找采用的是顺序查找的算法。实验表明对于目前 WAP 服务单一,WAP 网站少的情况下运行是良好的。

但随着 WAP 的应用的丰富、用户的增多,WAP 网关还是要采用图4的方案。

下一步工作是:图4方案中 cache 的具体实现以及 cache 和各个协议栈的交互等还有待进一步的研究。其性能还要在以后用户和服务都有相当数量的基础上,作详细的分析。

总之,cache 的研究有很多内容,本文提出的想法和建议还要通过实践来检验和改善。

参考文献

- 1 Chang Da-Wei, et al. Adaptive-level memory caches on World Wide Web servers. Computer Networks
- 2 吴东升,等. WWW 代理现状与改进方案. 计算机应用, 1999,12
- 3 周军,等. Proxy Server 中 cache 的管理和使用. 现代计算机
- 4 WAP Architecture Specification. WAP Forum. URL: <http://www.wapforum.org/>
- 5 Wireless Application Environment Overview. WAP Forum. URL: <http://www.wapforum.org/>
- 6 Wireless Application Environment Specification. WAP Forum. URL: <http://www.wapforum.org/>
- 7 东南大学 WAP 研究小组相关文献