

# 用 SQUID 架构本地信息代理服务器

罗忠石 刘虹霞

(鸡西大学计算机系 鸡西158100) (河北师大数信学院 石家庄050016)

**摘要** 仿照电视广播里的插播模式,我们设想了一个活动的代理中心结构作为支持本地信息高效传送的手段,当网页通过与内容服务器协作的动态代理服务器被取回的时候,本地信息将基于需要灵活地被插入到网页中。本文介绍了用 Squid-based 架构的代理服务器上的信息传送的设计和各種功能,这种方案的操作对于网页客户与内容提供商来说完全透明的,是切实可行的。

**关键词** 因特网,万维网,本地信息,代理转换

## The Local Information Based on the Squid-Based Proxy Server

LUO Zhong-Shi LIU Hong-Xia

(Department of Computer Technology, Jixi University, Jixi 158100)

(College of Mathematics and Information Technology, Hebei Normal University, Shijiazhuang 050016)

**Abstract** According to the advertisement model in television broadcasting, we conceive an effective transmitted mean that is based on an active proxy. The locale information should be inserted into the Web page on-demand, as soon as a Web page is retrieved through an active proxy server that is in collaboration with the content server. This page introduces a mean and other function based on the Squid-based proxy server, and this project is practical because its operation can be made completely transparent to Web clients and content providers.

**Keywords** Internet, WWW, Locale information, Proxy transformation

通常,一个香港的网民从位于美国的一个网站上看到了一则关于纽约的某大商场让利销售的信息,他不会愿意为这次让利销售就专程跑到纽约。这样,网上冲浪者对插播信息的兴趣就变得很小。由于“WORLD WIDE WAIT”的问题,他甚至拒绝该网页的主要内容。

我们设想了一个动态的代理中心框架去支持当地网页的传播。这个框架基本工作模式与电视广播中插播方式很相似,协议的信息在因特网内容服务器与代理服务器之间被制作。当一个网页通过协议好的代理服务器从内容服务器上被恢复时,代理服务器将会插入到当地信息的网页中。由于代理商与客户物理上很接近,对客户有比因特网内容供应商更好的理解,因此,通过代理商上传的信息对网上冲浪者来说,将会在内容上更相关、更有吸引力。

### 1 基本代理中心系统体系结构

代理服务器是解决“WORLD-WIDE-WAIT”问题和涉及管理与政策问题的最有效的手段。事实上,在近一、二年内对

网络系统和其它信息资源存储、查询、管理的系统,并以此为基础实现用户身份的统一管理和各类应用系统的集中授权。

### 3.7 目录服务系统与企业数据仓库等知识管理系统结合

大型企业建立的数据仓库系统主要是解决专业领域的决策问题和企业整体决策问题。单一的数据仓库借助目录服务的已有信息资源可方便地与其他基础数据库结合,与决策应用系统结合,而且可与其他数据仓库结合。知识管理系统中强调人的管理,这正是目录服务系统的优势,两者结合可以使知识管理系统与企业人员密切结合。

**结束语** 企业网络信息系统引入目录服务系统将带来巨大变化,使集中管理成为现实。目录服务系统可以与企业的多

高效代理服务器的需求有极大的增长,代理商对于那些通过代理服务器上网的网上冲浪者的访问行为有着更好的认识。网页服务器对网页客户的访问行为的完全了解是在访问内容上,而不是他所访问的网页。代理服务器是个发生网络申请的成本高效的地方,它能为众多客户和服务器提供一个一对多的解决方案。代理服务器与因特网内容服务器之间应当建成完全的协作,两者应就下面的问题达成协议:

代理服务器不应该要求因特网服务器在它的网页上增加任何标记来帮助信息插入的进程,因此,内容服务器会从改变它们的网页或维护它的站点的工作中解脱出来。随着代理服务器信息插入的实际进程的认可,它应该具有如下特点:

- 服务器透明处理:本地信息的插入进程相对因特网服务器透明这点是很重要的。这将减轻内容服务器的工作负担还给他们更多的与代理服务器协作的动力。

- 实时处理:本地信息的插入应该作为在网页上实时数据恢复处理的一部分被实时执行。这也可确保它与 HTTP 协议完全兼容。网页访问的自然流应该被维护,以至客户觉察时间

种应用结合,充当系统间交互的可靠桥梁和纽带。目录服务系统纽带作用的发挥,在提升信息应用系统整体安全性和可管理性的同时,将极大促进网络信息应用系统的融合并向更高层次发展,使网络信息系统更好地服务于企业的生产经营。

### 参考文献

- 1 [美]Adams C, Lloyd S. 公开密钥基础设施——概念、标准及实施. 北京:人民邮电出版社,2001
- 2 关振胜. 公开密钥基础设施 PKI 与认证机构 CA. 电子工业出版社,2002
- 3 [美]Silberschatz A, Korth H F, Sudarshan S, 等(杨冬青 唐世渭等译). 数据库系统概念. 北京:机械工业出版社,2000

没有被影响。

·用户化的处理:被插入到网页的本地信息的选择是个动态的进程。它能够根据时间、访问者的描述、访问历史和代理服务器上的信息的优先权进行变化。这可以确保系统有足够的灵活性去控制各种信息需求。

·高效的处理:与对象恢复时间相比,插入信息到恢复的数据中的时间很低。这也确保信息处理不会显著增加访问的等待时间。

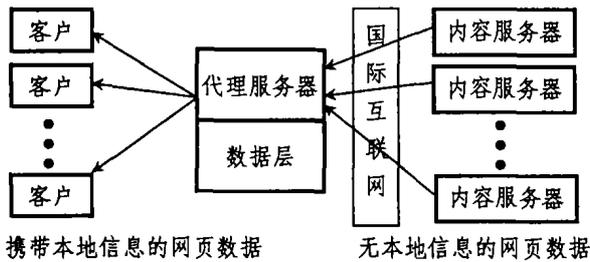


图1 通过代理服务器插入本地信息的系统结构体系

图1中显示了我们所设想的在互联网上插入本地信息的代理服务器的体系结构,也展示了上面所提及的那些令人向往的特点。内容服务器上的数据按它的原始格式被存贮,而不包含插入信息。这减轻了内容服务器为支持在它的网页上插入信息的进程而承受的负担。代理服务器就变成发生这种信息插入网页行为的地方。当代理服务器接收到来自于客户的HTTP请求时,它会作出把信息插入到客户所请求的网页上的决定。通过把客户请求的网页的URL(或IP)地址与协议好的网页列表上的URL名相比较,代理服务器将决定是否进行信息的插入。对于那些能够被插入信息的网页,信息横幅的选择将基于以下信息去快速地进行。

客户的个人描述;客户对当前网页的访问历史;客户想要访问的域名;因特网内容服务器在协议中所认可的优先选择的信息形式;代理服务器所拥有的本地信息服务商。

然后,当通过代理服务器把回复的数据传递给网上冲浪者时,被选中的本地信息的标记将会被插入网页的HTML源程序中。

为了使得这种选择插入到被请求的网页上的信息容易处理,我们在动态代理服务器上设置了一个数据层。在这个数据层里存有三种基本类型的数据库,每一个与内容服务器、信息服务商或网上冲浪者有关。

第一个数据库与有协议的内容服务器有联系。对那些每一个内容服务器都乐于与代理服务器签协议的网页下有联系的信息被记录在数据库中的输入属性字段中。

与代理服务器为信息插播有协议的网页的URL与IP地址;本地信息在网页中插入点;网页的内容域;网页所喜欢的本地信息的内容域。

第二个数据库与本地信息的横幅有关。下面的信息是为每一个与代理服务器有联系的信息横幅而被记录的。

信息横幅的URL地址标记;被插入的URL地址标记的大小;信息横幅的内容域;信息横幅所喜欢的网页的内容域;代理人服务器的信息横幅的优先权。

第三个数据库有在代理服务器控制下的网上冲浪者有关。依赖于代理服务器所喜欢的分析的程度,这个数据库可以包含通过对于网页冲浪者对网页的详细访问历史进行分析所得到的个人首选项的信息。

利用来自三个数据库的信息,代理服务器能够找到最适

当的本地信息,把它以飞行的方式插入到网上冲浪者所请求访问的网页上。

## 2 在SQUID代理服务器上本地信息的插入

我们已经在上一部分讨论了本地信息的基本的中心代理系统的体系结构。为了研究设想的可行性与实用性,我们使用这样的本地信息系统,并且把它集成为一个全功能的代理服务器。通过执行演习,可以更好地理解大多数有关系统设计的细节。

在我们的演习研究中,最基本的工作环境是这样设想的:客户/代理服务器 $\leftrightarrow$ 代理服务器 $\leftrightarrow$ 网页内容服务器

在因特网内容服务器上存贮的多媒体数据不包含本地信息。只要网页通过代理服务器被恢复,它就会被进行本地信息的插入检查。这些插入处理将飞速执行,而那些不需要当地信息的网页将会绕过代理服务器的插入模块。

我们选择SQUID代理服务器2.3版作为服务器原型和实验的平台。选择这种平台是因为它是已经在因特网上得到广泛使用的一个高效的代理服务器软件,而且它的软件代码是可用的。在SQUID中,数据以模块的方式进行传送。只要代理服务器接收到从它的上面的网络层级传来的数据块,它就会立即把数据推向下一层级的客户/代理服务器,而不需要等待完全的数据传送完。

## 3 基于SQUID的信息代理服务器的数据流动

我们希望把在SQUID代理服务器上的一个网页请求的实际工作流程与本地信息的插入联系起来。这不仅仅可以帮助我们更好地理解基于SQUID的信息代理服务器的完全的系统设计,还可以帮助我们去欣赏这种设计以及如何把它们结合到一个单一的系统。

当代理服务器接收到客户的请求,它将把请求的URL地址与被贮存在协议好的内容服务器的数据层中的网页的URL地址列表相比较。比较工作由访问控制阶段的ClientAccessCheck()功能完成。假如找到匹配项,Ad-Insert-Flag变量将被设置并通报回应处理产生回应信息,说明被请求的HTML对象需要被插入本地信息。

在系统里的启发式和预先定义的规则被用来决定哪一个信息条幅可以被用在插入处理中。关于这个被选取的条幅的信息,如信息的插入点、大小、URL标记的内容将会被更新到HttpStateDate数据结构中的Ad-Insert-Location, Ad-Tag-Length和Ad-URL-Tag等字段中。

当代理服务器接收到一个回应,它将会检查HttpParseReplayHeader()功能中的Ad-Insert-Flag标记与含“200 OK”回应代码。如果两种情况都符合,意味着被恢复的网页是合法,可以在其中插入本地信息。含本地信息的对象新的内容长度将被通过使用原始的Content-Length字段与Ad-Tag-Length字段计算出来。存在New-Content-Length中计算的结果将用来在ClientBuildReply()功能中创建回应的报头。用ClientSendMoreData()功能定位插入点,以便把被插入的本地信息的URL地址标记插入到网页中。这个处理的完成需要使用下面的内容:

- 在HttpStateDate结构中的Ad-Insert-Location字段
- 为详细的插入标志对流出数据作解析

在这两种情况下,在插入点被找到后,信息的URL地址

(下转第126页)

## 参 考 文 献

- 1 Object Management Group (OMG). Common Object Request Broker Architecture (CORBA), v3.0. Dec. 2002
- 2 Cukier M, et al. AQUA: An adaptive architecture that provides dependable distributed objects. In: Proc. of the IEEE 17th Symposium on Reliable Distributed Systems, West Lafayette, IN, Oct. 1998. 245~253
- 3 Landis S, Maffies S. Building reliable distributed systems with CORBA. Theory and Practice of Object Systems, 1997, 3(1): 31~43
- 4 Felber P, Guerraoui R, Schiper A. The implementation of a CORBA object group service. Theory and Practice of Object Systems, 1998, 4(2): 93~105
- 5 Marchetti C, Mecella M, Virgillito A, Baldoni R. An interoperable replication logic for CORBA systems. In: Proc. of the International Symposium on Distributed Objects and Applications, Antwerp, Belgium, Sep. 2000. 7~16
- 6 Natarajan B, Gokhale A, Yajnik S, Schmidt D C. DOORS: Towards high-performance fault-tolerant CORBA. In: Proc. of the Intl. Symposium on Distributed Objects and Applications,

- Antwerp, Belgium, Sep. 2000. 39~48
- 7 Moser L E, Melliar-Smith P M, Narasimhan P. Consistent object replication in the Eternal system. Theory and Practice of Object Systems, 1998, 4(2): 81~92
- 8 Narasimhan P. Transparent Fault Tolerance for CORBA: [PhD thesis]. Department of Electrical and Computer Engineering, University of California, Santa Barbara, Dec. 1999
- 9 Narasimhan P, Moser L E, Melliar-Smith P M. Strong replica consistency for fault-tolerant CORBA applications. In: Proc. of the IEEE 6th Workshop on Object-Oriented Real-Time Dependable Systems, Rome, Italy, Jan. 2001
- 10 Nam D, Lee D, Youn H Y, Yu C. Group communication support for CORBA using OCI. In: Proc. 12th IASTED Intl. Conf. on Parallel and Distributed Computing and Systems, Las Vegas, NV, Nov. 2000. 106~111
- 11 IONA Technologies, Inc. 2001. ORBACUS. <http://www.orbacus.com/ob/>.
- 12 Mena S, Schiper A, Wojciechowski P. A Step Towards a New Generation of Group Communication Systems: [Technical Report IC/2003/01]

(上接第102页)

标记会在当数据在客户 socket 上被写时插入到流出数据中。说明一点是,这个插入处理不会改变在贮存入口的数据内容。确保在代理服务器中被隐藏的数据拷贝不包含本地信息。

## 4 执行

我们通过实时的、基于代理服务器的 HTML 对象的传输来研究了我们为本地信息所设想的解决方案。依靠在更高的执行效果、全功能的代理服务器上的方案的执行,显示了方案的系统体系结构和设计的可行性和实用性。

第一个有影响的因素是相对 HTML 对象的访问延时的插入处理的运行时间。这个执行的测验是在基于 SQUID2.3 版本的代理服务器上进行的。它设置在一个双 CPU 的 Pentium III 1GHz, 512M 内存, 128M 交换空间和 4G SCSI 硬盘计算机上。操作系统为 Linux RedHat 6.0。所用的代理服务器是不含本地信息的相同的 SQUID 转置代理服务器,两者在同一台机器上测试。我们观察到对 HTML 对象的平均访问时间为 1~2 秒,在信息插入之上的额外的运行时间少于 1 毫秒,这是足够小以至可以被忽略的时间。

第二个有影响的因素是在 HTML 网页的访问延时之上的插入处理的运行时间。我们发现额外的延时由于下面二个原因事实上是不会被观察到的。

·每一个网页通常都包含多个嵌入的对象。由于信息标记的插入仅仅作用于 HTML 对象,它对整个网页恢复时间的作用效果远远小于 HTML 对象。另外,HTML 对象的访问延时往往比在网页中嵌入图像对象用时要小得多。

·在网页上的对象的恢复是运行在受限的并行模式下。因此,对一个对象访问的信息插入延时通常隐藏在数据传送通道中。

第三个影响因素是恢复信息条幅和呈现它给网页冲浪者

的延时,这个延时依赖条幅图像的大小,条幅服务器的工作量以及在网页条幅服务器和网上冲浪者之间的可用的带宽。

**结论** 仿照电视广播里的插播模式,我们设想了这个动态的代理中心结构作为支持本地信息高效传送的手段,当网页通过与内容服务器协作的动态代理服务器被取回的时候,当地信息将基于需要灵活地被插入到网页中。这个网页信息通过代理服务器时在其中插入本地信息的代理服务器-中心的解决方案对于网页客户与内容提供商来说是完全透明的,切实可行的。

## 参 考 文 献

- 1 Abrams M, Standridge C R, Abdulla G, Williams S. Caching Proxies: Limitations and Potentials. In: Proc. of the 4th Intl. World Wide Web Conf. Dec. 1995. <http://www.w3.org/Conferences/WWW4/Papers/155>
- 2 Bennett D. How Interactive Ads are Delivered and Measurement Implication. White Paper, Audit Bureau of Circulation. <http://www.accessabvs.com/webaudit/admeasurement.html>.
- 3 Berthon P, Pitt L, Prendergast G. Visits, Hits, Caching, and Counting on the World Wide Web: Old Wine in New Bottles? Internet Research: Electronic Networking Applications and Policy, 1997, 7 (1) <http://www.emerald-library.com/brev/17207aa1.htm>.
- 4 Broadvision Inc. <http://www.broadvision.com>
- 5 Cassidy P F. How New Auditing Tools Legitimize Web Advertising. Feb. 1997. <http://www.netscapeworld.com/netscapeworld/nw-02-1997/nw-02-webaudit.html>
- 6 Chi C-H, Li X. Content Transformation Model for Pervasive Internet Access: [Internal Report]. School of Computing, National University of Singapore, 2000. Also submitted to IEEE Multimedia and Expo, 2000