

基于 动态代理服务器的地方广告

罗忠石 崔学文 赵炜平

(黑龙江省鸡西大学计算机系 158100)

摘要: 本文主要介绍了基于Squid-based代理服务器的地方广告的设计思想与各种功能的执行细节。这种网络广告的模式对“因特网”来说意义重大, 它为涉及到的各方提供了一个双赢的局面。

关键词: 因特网 万维网 信息商业 电子商务 因特网广告 代理转换

Local Web Advertisement Through Dynamic Active Proxy

1 引言

为了开发因特网广告的市场潜力, 我们设想了一个动态的代理中心框架去支持地方网页广告。这个框架基本工作模式与在电视广播中播放的广告方式很相似。广告在因特网内容服务器与代理服务器之间制作。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地方广告到网页中。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。通过代理商上传的广告对网上浏览者来说, 将会在内容上更相关、更有吸引力。在本文中, 我们描述了基于SQUID的代理服务器

上的地方广告的设计思想和实现细节。这种模式对于因特网广告很重要, 它为所涉及到的三方提供了一个双赢的局面。

(1) 内容供应商既可以提高收入又可以不用化费额外的精力去查询和维护自己站点上的“远程”广告;

(2) 代理服务商(主要指“因特网”服务访问供应商, ISAP)可以利用出售通过他们服务器的网页广告而获得一种新的获利的商机;

(3) “网民”则可以接收到与自己更相关的、更感兴趣的地方广告。此外, 由于这种解决方案的操作对于网页客户与内容提供商来说完全透

明, 所以是切实可行的。

2 在SQUID代理服务器上的地方广告的插入

我们选择SQUID代理服务器2.1版作为服务器原型和实验的平台, 选择这种平台是因为它已经是一个在因特网上得到广泛使用的一个高效的代理服务器软件。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快, 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。

2.1 执行考虑

在SQUID代理服务器上的地方广告的执行,几个设计思想得到很好的解决。

2.1.1 被请求的数据对象的大小

在广告插入处理中,关于地方广告条幅的URL地址标记需要被嵌入到被请求的网页的HTML代码中,这显然增加了HTML对象的大小。这可能使因特网上的数据传送成为一个问题。因为在把广告的实际地址标记被插入到回复数据中之前,有关包含对象的原始长度内容的网页请求的回复报头将会被推向客户。此外,没有时间去把回复报头传到客户端去更新这些有新内容、新长度的信息,在因特网上的对象传送时的错误匹配,可能会导致数据丢失和依赖使用HTTP工具版本的因特网连接的超前断开。

然而,与其他的成熟的网络技术如流水线与数据压缩不一样,新HTML对象的文件大小可以在广告选择的处理中预先计算出来,在广告横幅被选中后,HTML对象的新尺寸可能通过条幅的URL地址标记大小字段(Ad_Tag_length)和被请求对象的当前内容尺寸字段(Content_length)的和计算出来。用这种方法,确保插入广告前后的HTML对象的内容长度就是最新的长度。

广告插入后的HTML对象尺寸的预先计算

是非常重要的,它可以允许广告插入处理飞速执行而不需要回复数据的外在缓冲器。否则,携带地方广告的HTML对象仅仅在广告插入处理完全完成之后,才能被传送给网上浏览者。这会严重地降低代理服务器的执行,因为数据传送通道被破坏,还导致了更多的磁盘操作。

2.1.2 数据结构

在SQUID中,访问所涉及的请求与回复数据通过Httpstatedata这个数据结构连接在一起,它记录了请求与回复的所有必要的信息,这个结构是在代理服务器处理请求时候产生的,被应用在完全请求和回复处理的全过程中。因此,在我们的执行中,在Httpstatedata中有一个位置用来存储有关被插入的广告条幅的信息,即Ad_Insert_Flag,还有四个字段分别为,详细描述HTML对象在适应广告插入后的文件大小的增长的字段AD_Tag_Length;描述被插入的广告的URL标记的字段,Ad_URL_Tag;详细描述允许广告URL标记插入的HTML对象的内部位置的字段Ad_Insert_Location和用来记录广告插入后的对象内容的新尺寸的字段New_Content_Length。

2.1.3 高速缓冲存储器的版本

根据在代理服务器上的广告插入,通过代理服务器的每一个HTML对象可能有两个版本,不

含地方广告标记的原始版和含地方广告标记的最终版。这会由于两个版本共享同一个对象名而产生问题。为降低地方广告代理服务器的系统复杂性,提高客户插入的最终广告处理的灵活性,我们采取了以下方法。在内容服务器与代理服务器上的所有对象都不含地方广告。只要对象一通过代理服务器(包括那些从代理服务器的高速缓冲存储器恢复和远程内容服务器取过来的对象),广告将飞速地插入到HTML对象中。由于插入广告对象标记到HTML对象中的用时少得可以忽略,这种方法不会对代理服务器的执行产生延时。

2.1.4 广告信息与选取

我们这种解决方案的特色之一就是广告处理的用户化。只有在内容服务器上被选中的对象才被插入地方广告;其余的则绕过代理服务器的广告驱动。另外,被插入广告的选取是一个动态的处理过程,它基于客户优先选择权和访问方式的历史记录,代理服务器的广告条幅和有协议内容服务器的需求。

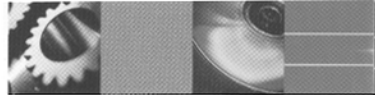
为了在我们的系统框架中支持这点,我们把有关广告选取的规则存储在代理服务器的数据层中,此外,保存在数据层中的信息会在系统启动时被加载到内存中以备快速访问。只要有请求在受访问控制的代理服务器上通过,代理服务器就会使用在数据层中的信息去更新与广告插入有联系的请求的HttpStateDate数据结构中的所有字段,包括Ad_Insert_Flag,Ad_Tag_Length,Ad_URL_Tag和Ad_Insert_Location。

3 基于SQUID的广告代理服务器的数据流动

在这部分,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地方广告的插入联系起来,这不仅仅可以帮助我们更好地理解基于SQUID的广告代理服务器的完全的系统设计,还可以帮助我们如何把它们结合到一个单一的系统。

3.1 在SQUID上的需求处理





在对 SQUID 的代理服务器中的广告插入处理进行描述前, 先来理解一下在 SQUID 上处理客户请求的两个基本步骤是非常有帮助的, 它们是请求与回复。

3.1.1 请求处理

请求处理可以粗略地分为下面七步。

(1) 客户连接: 当代理服务器接收到来自客户的请求时, 它会通过建立与客户的连接来回应。

(2) 请求解码, 被接收的 HTTP 请求将会在代理服务器中被读取并解释。

(3) 访问控制, 代理服务器的访问控制包括被请求的 IP 地址和域名同先前已定义的 URL 地址列表的检查, 如果他们不在 ip_cache 中, 可以使用 dnsserver 程序去获取 DNS 结果, 当访问控制检查完成后, 代理服务器的客户边将建立一个 ACL 状态数据结构并为通知注册一个回叫功能。

(4) 高速缓冲存储器 碰撞/错过, 在访问控制通过后, 代理服务器将在它的本地高速缓冲存储器搜索被请求的对象, 假如有高速缓冲存储器碰撞, 代理服务器的客户一边将在存储入口注册这个信息, 否则, SQUID 将把请求推向下一个网络层级的网页服务器/代理服务器。

(5) ICP 探测, 被向前推进的请求处理用 ProtoDispatch() 开始, 它的功能等同于一个选取过程, 包括传送 ICP 需求和接收 ICP 回应, 在 ICP 回应处理后, ProtoStart() 将启动, 它的功能是调用一个适当明确的协议功能去向前推进请求。

(6) HTTP 连接, 这个 HTTP 模块首先将打开一个与起始服务器或等同的下一级高速缓冲存储器的连接, 然后一个新的连接请求将随着回应功能被传给网络通信模块, Comm.c 例行程序将试图建立连接, 并尝试多次连接直到放弃。

(7) HTTP 通信, 在 TCP 连接建立成功后, HTTP 将建立一个请求缓冲器并使它服从在 socket 之上写操作, 然后再注册一个读管理去接收和处理 HTTP 的回复。

3.1.2 回应处理

回应处理可以被简单地分为 6 个处理步骤:

(1) 回应设置, 当回应被代理服务器接收时, 它的 HTTP 回应报头将会被解析并放入回应数据结构。

(2) 数据接收, 当回应的数据被代理服务器读出来时, 它将会附加到存储入口, 代理服务器的客户-边将通过回叫功能被通报新数据。

(3) 客户数据流, 在代理服务器的客户-边被通报新数据后, 它会从存储入口复制数据并使它可以在客户 socket 之上进行写操作, 回应报头也同样被建立。

(4) 本地数据储存, 当接收到的数据被附加到存储入口并流向客户时, 这些数据可以服从代理服务器在磁盘上的写操作。

(5) 数据流的终止, 当 HTTP 模块结束从上端的服务器读回应数据时, 它会把存储入口标记成“完成”, 然后服务器的 socket 将被关闭。

(6) 请求的断开, 当代理服务器的客户一边已经结束对全部对象在本地磁盘上的写操作时, 它将从存储入口注销自己并会释放所有可用的资源, 同时它也会关闭客户的连接。

3.2 广告的上传

图 1 显示了有地方广告插入的 SQUID 的工作

流程图, 在 SQUID 执行中, 广告插入处理由两个处理阶段构成, 第一阶段是广告的检查, 在代理服务器的请求处理中执行, 第二阶段是广告标记的插入, 在代理服务器回应处理中执行。

3.2.1 阶段一: 广告检查

当代理服务器接收到客户的请求, 它将把请求的 URL 地址与被储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_Inset_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 对象需要被插入地方广告。

在系统里的启发式和预先定义的规则被用来决定哪一个广告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, 如广告的插入点, 大小, 广告 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。

3.2.2 阶段二: 广告标记的插入

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, 可以在其中插入地方广告, 含地方广告的对象的内容长

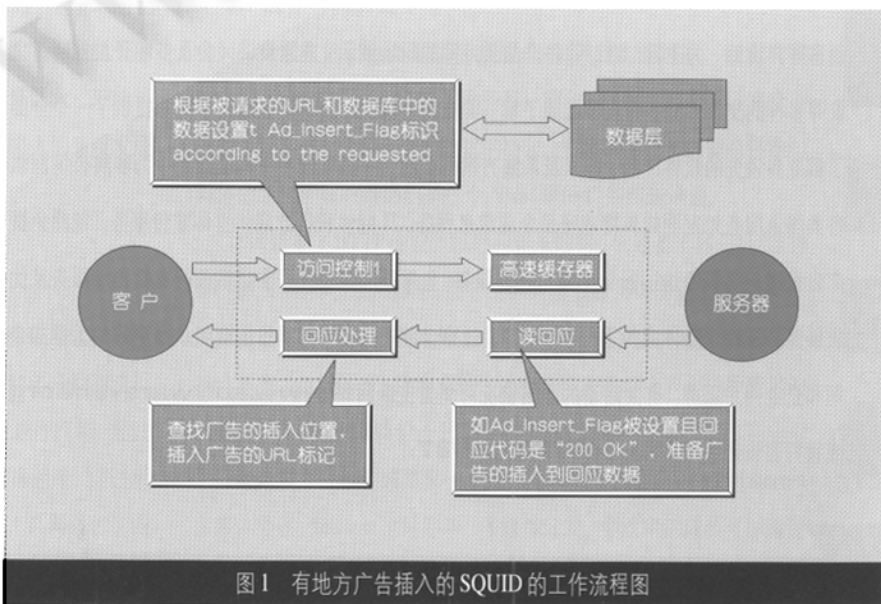


图 1 有地方广告插入的 SQUID 的工作流程图

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。

代理服务器将用 ClientSendMoreData() 功能定位插入点, 以便把被插入地方广告的 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:

(1) 在 _HttpStateData 结构中的 Ad_Insert_Location 字段

(2) 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, 广告的 URL 地址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: 这个插入处理不会改变在储存入口的数据内容。这会确保在代理服务器中被隐藏的数据拷贝不包含地方广告。

4 结论

本文介绍了有关在全功能的、高效运行的代理服务器上地方广告系统的设计思想与执行的主要技术细节。这个方案是建立在可行性和实用性基础上的。这个方案会因为内容提供商、代理服务提供商和网页客户之间建立双赢的局面, 而给因特网广告创造更多的商机。 ■

引导企业需求, 创造客户价值

企业应用软件盛会将于今年 6 月在沪举行

由上海市人民政府信息化办公室、上海软件行业协会主办及多家 IT 媒体承办的“2003 年上海国际软件行业年会”(CISE2003) 以及相关专业展览和论坛将于今年 6 月下旬在上海举行。年会针对不同的行业应用特点举办不同主题的行业活动, 其中包括 2003 年制造业信息技术应用展览及会议、2003 年供应链及物流应用技术展览和会议、2003 年电子政务应用技术展览和会议、2003 年企业应用软件及开发技术展览和会议四大专题活动, 做为今年亚太城市信息化高级论坛 (APCITY2003) 的系列活动之一, 已连续五届成功举办的软件年会活动将继续以“引导企业需求, 创造客户价值”为主旨, 通过软件产品及方案的现场展示, 主题演讲大会及多场开放式研讨会等多样的交流形式, 为众多渴望了解、实施应用解决方案的企业及行业用户提供了一个全面了解业界领先的软件产品、技术及系统方案的良机。主办单位将在本次活动中为参展公司提供各类增值服务如定期软件需求安装企业信息服务, IT 媒体和行业媒体立体宣传服务 (包括免费广告报道与高级专访), 行业信息与咨询服务, 全部参加企业代表数据库和行业数据库服务及优质展览会及论坛整体宣传等, 同时希望通过整合政府及各协办单位的资源从而更好的为参加公司和企业用户服务, 有关活动的详细情况可浏览主题网站 www.softwarevent.org 或拨打咨询热线 021-64477681/85/87。



倪鸣

上海软件行业协会会务部, 上海霖科

TEL: 021 - 64477680/81/82

FAX: 021 - 64477680

EMAIL: marknee@softwarevent.org

Web: www.softwarevent.org

2003 年上海国际软件行业年会

引导企业需求, 创造客户价值。

2003 年 6 月

引导企业需求, 创造客户价值! 更多精彩内容, 尽在 www.softwarevent.org