

一种域名式信息推送系统的研究

牛 颢¹, 朱 敏¹, 张 明², 王 舒¹, 黄 建¹, 赵 瑜¹

(1 四川大学 计算机学院, 四川 成都 610000; 2 四川省计算机研究院 网络信息研究室, 四川 成都 610041)

摘 要: 对基于互联网的信息推送技术进行了简要介绍, 指出了网络信息服务采用推送技术的必要性. 通过对比传统的信息推送技术所采取的“广存网”技术, 提出了一种新的模型来实现基于域名式的网络信息推送, 即通过对现有网络创建分路镜像的方法来实现, 并通过 ActiveX 控件来解决该模型下存在的问题. 实验证明, 对比传统的信息推送方式, 该模型在推送信息准确率、网络稳定性、系统安全性以及服务器负载等四个方面有着明显的改进.

关键词: 信息推送; 广存网; 分路镜像; 服务器负载

中图分类号: TP391

文献标识码: A

文章编号: 1000-7180(2010)05-0095-04

Research on Information Push System Based on Domain Name

NIU Hao¹, ZHU Min¹, ZHANG Ming², WANG Shu¹, HUANG Jian¹, ZHAO Yu¹

(1 Institute of Computer Science and Technology, Sichuan University, Chengdu 610000, China;

2 Network and Information Security Research Room, Sichuan Research Center of Computer Applications, Chengdu 610041, China)

Abstract: A kind of technology about Information Push based on domain name is introduced in this paper, it is necessary to use this Push Technology in Web Information Service nowadays. By contrast the traditional Information Push Technology which used Broadcasting & Storing Technology, a new model is designed to implement the Internet Information Push based on domain name through making the Image Splitter, and the problem of this model could be solved by creating the ActiveX control. The experimental results show that, in contrast with the traditional Information Push Technology, this model has improved perceptibly in accuracy of information, network stabilization, system security and server load.

Key words: information push; broadcasting & storing technology; image splitter; server load

1 引言

信息推送技术 (Information Push Technology) 是为了在最大程度上满足用户在网络上及时查找有用信息而应运而生, 并成为第三代浏览器的核心技术. 以信息拉取技术 (Information Pull Technology) 为核心的第二代浏览器在互联网信息激增的情况下, 无论是在信息检索的时间、效率还是在检索质量上, 都已经越来越无法使广大用户感到满意. 而推送技术正在以其提供服务的主动性, 返回信息的新颖性、及时性, 以及它在个性化信息服务方面的优势日益得到人们的重视. 信息推送的技术特点决定了该技术必将被网络内容服务商 (ICP) 广泛采用, 并成

为其开展更有效的信息服务所依赖的一种技术手段.

2 传统的信息推送技术存在问题

信息推送技术事实上是媒介“Push”信息方式的回归, 这种方式符合人们长期以来形成的获取信息的习惯. 但是该技术的实施, 还存在一定的问题.

2.1 网络拥塞问题

与人们预期的相反, 传统的信息推送技术不但没有缓解网络线路的紧张程度, 还过多地占据了带宽, 造成了网络拥塞. 这是由于当前网络技术的不完善所造成的, 现今的 Internet 上普遍采取的通信协议是 TCP/IP 协议. 该协议假定不同用户的信息需

求是彼此独立的,只考虑如何实现端口间的最佳连通,而不顾网络中传输信息的内容是否有关联,有什么关联,在实现端口间的最佳连通是以信息的“点对点”的个性化连接为代价的.例如:一条信息被 N 个用户取用,这条信息就要在互联网上取用 N 次, N 值越大,传输次数也就越多;在面向个性化连接的分组交换网络中,信息条目的共享度越大,对信道传输的压力也就越大.在当前的互联结构下进行信息推送,势必会造成网络信道的拥塞以及降低信息整体的推送量.

2.2 不完善的信息个性化服务

从资源管理的角度来看,信息推送技术所代表的是一种信息资源管理方式,表示的是特定的 Web 页面由浏览器自动检索以使用户以后阅读,因此我们需要实现基于特定查询信息推送技术.对用户来说,就是实现定制信息的个性化服务.目前频道技术在很多网站都得到了应用,但是在这些网站上,几乎都采用了静态页面的方式,定义静态模式的频道.静态频道的特点是,频道的内容是网站设计者预先设定的,然后将这些频道相应内容链接在这些频道之上.这种方式对于更为专业的个性化服务而言,还有很大差距,因为可能有很多用户的实际需求与网站设计者预先设定的频道相差很大.这就要求实现动态频道的推送技术.而目前动态频道技术的实现还很少,也就是说用户的个性化信息需求还无法得到完全满足.现阶段的网上推送技术没有真正达到提高效率的目的,没实现真正的个性化信息服务.

3 基于分路镜像的信息推送系统模型

3.1 推送系统总体结构

对于传统的信息推送系统的不足,文中提出了一种新的信息推送模型,即通过在现有网络上创建分路镜像的方法来实现定制信息的推送,如图 1 所示,该模型主要分为两部分:(1)基于以太网点对点协议的拨号网络(PPPOE);(2)以分路镜像的方式创建一个信息推送服务器.在(1)中,需要构建一个基于 PPPOE 的拨号网络,需要有安装 PPPOE 协议的多台主机、一个 BRAS 服务器、一个 RADIUS 服务器,以及认证数据库.不同的主机通过指定的用户名和密码发出拨号请求,经过网络传达到 BRAS 服务器(宽带接入服务器),BRAS 服务器接到请求后向 RADIUS 服务器(远程用户拨号认证系统)发出 ACCESS REQUEST 请求包,其中含有用户的用户名、密码、端口类型等,经 RADIUS 服务器与认证数

据库比对后,向 BRAS 回送 ACCESS RESPONSE 响应包,其中包含用户的合法性和一些设置,如用户 IP 地址、掩码、网关、域名、用户可使用的带宽等.用户接收到这些信息后就可以接入网络,联网期间 BRAS 不断向 RADIUS 发送信息,这些信息包括用户的上网时间、用户流量、用户下网时间等,可以方便网络的维护和管理.该模型的第二部分则是以分路镜像的方式创建一个信息推送服务器接入到 BRAS 服务器,用于截取主机发出的 HTTP 请求,进行分析处理并将结果返回给主机.

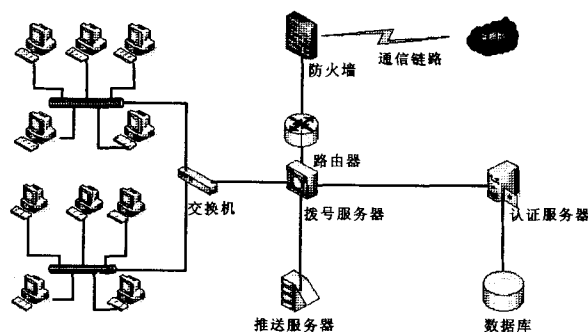


图 1 基于分路镜像的信息推送系统模型

3.2 推送系统实现原理

如图 2 所示,一个完整的信息推送流程是首先由主机通过 URL 向目标服务器发出 HTTP 请求报文,该请求报文经过交换机分别转发于信息推送服务器以及目标服务器.当信息推送服务器获取到该 HTTP 报文之后,便对该报文进行分析并提取出其中相关的关键词进行信息处理.分析完成后,结果将由推送服务器送出返回给用户主机并由主机的浏览器下载.与此同时,目标 Web 服务器同时也获得到主机所发出的 HTTP 请求报文,收到请求报文之后目标 Web 主机经过确认向其发出响应报文,两个服务器所发出的响应报文同时返回给主机的浏览器并下载显示给用户.这样的信息推送方式既不影响当前用户所浏览的网页也可以达到对用户的所需的个性化信息的推送.

3.3 该推送系统模型的优点

同传统的信息推送方式相比,基于分路镜像的信息推送系统模型有着十分明显的优点:(1)由图 1 可以明显看出,实现该模型不需要对现有的网络结构进行改造或重新设计,只需要在原有网络的宽带接入端创建分路镜像作为推送服务器即可,也不需要用户的主机安装专门的客户端,实现起来非常方便,也极大地节约了建设成本.(2)由该模型的构成原理可以看到,经过分路的推送服务器处理,我们

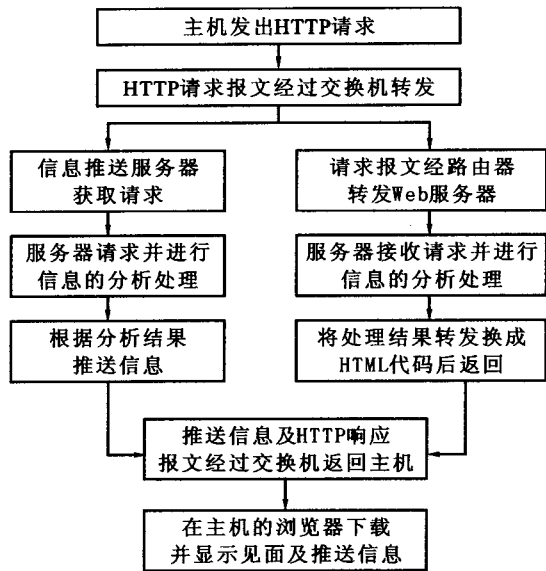


图 2 信息推送流程图

可以实现信息的动态推送.通过分析用户发出的 HTTP 请求报文,可以提取出其中的关键字作为用户所需求的重点,根据关键字将不同的信息推送给用户,这种方式可以更好地满足用户的个性化需求,也可以使得用户自己定制所需要的信息.静态信息推送技术可用于传播大众化的信息,信息推送方通过对所收集到的大众信息进行了整理,按主题加以组织,设定成静态频道供用户选择.静态推送技术利用广存网进行传播,动态推送技术则利用互联网进行传送.二者的结合既能有效缓解通信线路的紧张,又能实现完全个性化的信息服务.

3.4 信息推送过程中出现的问题以及解决办法

对于不同的服务器返回响应数据包的时间不同我们有如下定义:

- (1) 由 Web 服务器接收到主机 HTTP 请求至主机接收到 Web 服务器 HTTP 响应报文的时间为 $W(t)$;
- (2) 由信息推送服务器接收到主机 HTTP 请求至主机接收到信息推送服务器 HTTP 响应报文时间为 $P(t)$;
- (3) 设 $I = W(t)/P(t)$,表示响应时间比;

在信息推送的实验环境中,会出现用户无法访问某个目标网页的现象,即网页中显示错误信息,经过分析得出如下原因:在用户主机发送 HTTP 请求时,会分别发送请求报文向 Web 服务器和推送服务器,此时,由于网络传输上面的延迟或各个服务器性能以及配置上的差别,响应报文的时间也不尽相同,如表 1 所示.

当 $I > 1$ 时,即表示从推送服务器所返回的数

表 1 信息推送时出现的问题	
响应时间比	信息推送时出现的问题
$I > 1$	目标网页无法打开,只显示推送页面.
$I < 1$	目标网页可以正常打开.
$I = 1$	网页及推送页面无法打开,页面报错.

据包先于 Web 服务器到达用户主机.那么此时用户就无法正常浏览其所需的网页而只能在浏览器中看到推送页面.当用户主机再次发送 HTTP 请求时,由于浏览器缓存已经保存了先前的推送页面信息,浏览器会直接从缓存中读出刚刚显示的推送页面,导致用户持续的无法访问目标页面.

当 $I < 1$ 时,即表示从 Web 服务器所返回的数据包先于推送服务器而到达用户主机,此时用户可以正常访问网页,推送信息也会显示而不会影响用户.

当 $I = 1$ 时,则表示从推送服务器和 Web 服务器返回的数据包是同时到达用户主机的,在这个过程当中,会出现网络上的“抢包”现象,即浏览器不停的接收来自两个服务器的数据包,造成浏览器无法解析,页面出错.

对于 $I > 1$ 和 $I = 1$ 所出现的情况,我们需要清空客户端的浏览器缓存,而现今基于安全的考虑,绝大部分浏览器都禁用了 JavaScript 脚本,所以是无法通过脚本语言来操作客户端.基于以上问题的考虑,可以通过编写浏览器 ActiveX 控件来实现清空缓存的目的.ActiveX 控件能够直接嵌入到浏览器和其他应用程序中,并具有和应用程序一样强大的数据处理,磁盘操作等能力.

如图 3 所示,当浏览器发出请求时,信息推送服务器向用户浏览器回传内嵌 ActiveX 控件的页面,由浏览器负责解释.在解释过程中首先用该控件的在页面中注明的 id 值先在本地的注册表内进行查询,若已经存在,则说明该控件已经在本地安装,然后通过注册表中的相关信息直接使用该控件;否则就要根据页面中所提示的该控件所在的服务器上的路径到推送服务器上去下载并且自动完成在本地的

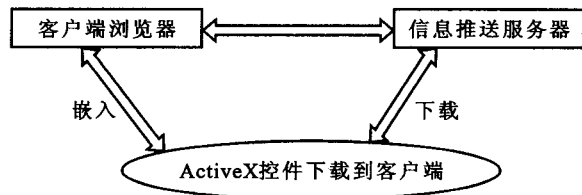


图 3 信息推送流程图

安装注册,使该控件成为本地资源,供以后使用.在控件的安全性方面,通过 C# 语言来实现 IObject-Safety 接口,将该 ActiveX 控件标记为安全的 ActiveX 控件,在最大程度上降低了风险.代码如下(其中 Guid 是不能修改):

```
[ Guid (" CB5BDC81 - 93C1 - 11CF - 8F20 - 00805F2CD064"), InterfaceType ( ComInterfaceType. InterfacesIsUnknown)]
```

```
public interface IObjectSafety
{
    // 接口实现方法
    void GetInterfaceSafyOptions(
        System. Int32 riid,
        out System. Int32 pdwSupportedOptions,
        out System. Int32 pdwEnabledOptions);
    void SetInterfaceSafetyOptions(
        System. Int32 riid,
        System. Int32 dwOptionsSetMask,
        System. Int32 dwEnabledOptions);
}
```

4 试验结果对比分析

在实验中,将信息推送程序以及目标访问页面放于一台服务器中用于模拟传统的信息推送方式.在通过与传统的信息推送方式相对比,基于分路镜像的信息推送模型在推送信息准确率、网络稳定性、系统安全性,以及服务器负载等四个方面有着明显的改进.

在信息推送准确率方面,连续不间断的向 Web 服务器发送 100 个 HTTP 请求,在传统的信息推送模型下,正确获得 85 个推送响应,正确率为 85%.在基于分路镜像的信息推送模型下,正确获得 92 个推送响应,正确率为 92%.由于基于分路镜像的推送模型是分别处理主机的 Web 请求和推送请求,其推送准确率明显高于传统的推送模型.

在网络稳定性方面,随着推送请求的不断增加,在传统的信息推送模型下,主机发送数据至目标网页的传输时延不断增加,网络的带宽占有量也不断增加,稳定性逐渐变差.而在基于分路镜像的推送模型下,由于推送服务器是置于局域网端,其传输时延并不受推送请求的增加而影响.

在系统安全性方面,由网络结构可以清楚看出,传统的信息推送模型依赖现行的网络环境,而基于分路镜像的信息推送模型是对 Web 请求以及推送请求分别处理,相互不受干扰,即当推送服务器出现

问题时,并不影响用户访问网络.

在服务器负载方面,随着 Web 请求和推送请求的不断增加,传统的信息推送模型的服务器端负载不断增加,在连续不间断的向服务器发送 100 个 HTTP 请求时,其 CPU 占用率最高到达 30%,而基于分路镜像的信息推送模型的推送服务器端 CPU 占用率最高为 16%,由此可见,采用分路镜像的模型可以很好的减轻服务器端的负载压力.

5 结束语

文中提出了一种基于分路镜像的信息推送系统架构,并分析了该系统在信息推送过程中的优势以及所出现的问题,提出了一种相应的解决方案.在网络信息资源飞速增长的情况下,推送技术将成为一种对信息资源进行有效开发、允许资源共享的重要手段.

参考文献:

- [1] 彭路,胡越明. Push 技术及其在 Internet 上的应用[J]. 计算机应用研究,1999(9):42-59.
- [2] 张健,袁春风. Internet Push 技术的分析与展望[J]. 计算机应用研究,2000(12):4-7.
- [3] 李秀. 推送技术——完善网络信息服务的关键技术[J]. 现代图书情报技术,2001,89(5):56-58.
- [4] 杨桂芝. 一种基于信息推送的搜索引擎模型[J]. 现代电子技术,2007,247(8):81-83.
- [5] James F Kurose, Keith W Ross. 计算机网络[M]. 陈鸣,译. 北京:机械工业出版社,2004.
- [6] Keith Cheverst, Keith Mitchell, Nigel Davies. Investigating context-aware information push vs. information pull to tourists [C]// Consumer Communications and Networking Conference. USA, IEEE, 2006:437-441.
- [7] Keith Cheverst, Gareth Smith. Exploring the notion of information push and pull with respect to the user intention and disruption[C]// Consumer Communications and Networking Conference. USA, IEEE, 2006:437-441.
- [8] George Cybenko, Brian Brewington. The foundations of information push and pull. Automatic generation of intelligent agent programs[J]. IEEE Intelligent Systems, 1997, 12(1):3-4.

作者简介:

牛 颢 男,(1983-),硕士研究生.研究方向为计算机和信息系统.

朱 敏 女,(1971-),教授.研究方向为智能交通、计算机网络.

张 明 男,(1978-),工程师.研究方向为计算机网络.