

基于 DCD 协议 RSS feeds 传输模型的研究与设计

廖小荣, 范冰冰, 黄兴平

(华南师范大学 计算机学院, 广东 广州 510631)

摘要:目前 RSS feeds 的传输方式会导致大量重复传输,并且现有的 RSS 应用缺乏对 RSS 频道的有效管理的机制,文中提出了一种用 DCD 协议传输 RSS feeds 的方法,设计了利用 DCD 传输 RSS feeds 的封装格式,描述了其传输方式和传输流程,经过对典型 RSS 应用场景的模拟和理论计算,在对两种传输方式进行分析和比较后,得出了使用 DCD 协议传输的方法可很好地解决 RSS feeds 传输的及时性和传输效率问题的结论,并且提出了使用 DCD 协议频道管理交互流程来管理 RSS 频道,从而增强了 RSS 频道的可管理性。

关键词:DCD;RSS;推送

中图分类号:TP393

文献标识码:A

文章编号:1673-629X(2011)01-0129-04

Research and Design of a Delivering RSS Feeds Model over DCD

LIAO Xiao-rong, FAN Bing-bing, HUANG Xing-ping

(School of Computer, South China Normal University, Guangzhou 510631, China)

Abstract: The current solution of delivering RSS feeds results in great amount duplication and RSS channels lack efficient management in RSS application. This paper presents a distribution system of delivering RSS feeds over DCD protocol, designs the package model, and describes its delivery method and procedure. Compared with the current solution, delivering RSS feeds over DCD protocol can eliminate duplication and delivery content in time. In addition, this paper presents a method to manage RSS channels through some related transactions and messages of DCD protocol, which can enhance the RSS channels' management.

Key words: DCD; RSS; push

0 引言

RSS 技术是一种基于 XML 技术的网络内容聚合和发布方式^[1]。发布者将资讯内容生成 RSS feed 发布在网站上, RSS 阅读器向服务器轮询并下载 RSS feed,并解析其中内容。在第一次下载之后到服务器再次更新内容之前,这段时间的传输完全是浪费网络资源;即使在一个更新了的 RSS feed 中其平均的内容更新量只有 6.8%^[2],其大部分内容都在以前的轮询中下载到了客户端。在 feed 增加新条目后,客户端需等一个轮询周期之后才能得到更新^[3]。目前解决此类问题的方法有:某一客户端短时间轮询次数过多则限制其轮询次数;减少 RSS feed 的大小;利用 HTTP 协议头部限制过多的轮询;利用 RSS 2.0 存在的“TTL”元素协调服务器内容更新于客户端轮询间隔,从而达到减少轮询次数的目的^[4-6]。这些解决方案都类似于优

化措施,很难从根本上解决问题。一个更好的解决方案是当服务器有内容更新时,由服务器通知客户端获取新内容,这种“推”模式可以提高传输的及时性和传输效率。基于“推”模式的解决方案有基于 SIP 协议来实现内容的推送^[7],它可减少轮询次数,减少网络流量,缺点是需特定的网关服务器,且其频道订阅与管理实现复杂,通用性较差。

DCD(Dynamic Content Delivery)是来自开放移动联盟^[8](OMA, Open Mobile Alliance)的业务能力规范^[9]。DCD 为了增强终端用户的信息浏览体验,通过一对一或一对多、周期下载或者推送的方式,把用户个性化定制的内容更新到用户的 DCD 客户终端中。国内的相关研究文献较少^[10,11],产品开发有中国移动的快讯业务^[12],中国移动、华为公司、中兴通讯针对动态内容分发申请了相关的专利^[13]。这些专利的技术架构基本参照了 OMA 组织指定 DCD 的相关标准。文中提出了一种基于 DCD 协议传输 RSS feeds 的方法,利用 DCD 推模式的分发机制来解决传输及时性和传输效率问题,并利用 DCD 中的频道发现和频道订阅消息来增强 RSS 频道的可管理性。

收稿日期:2010-04-19;修回日期:2010-07-03

基金项目:广州市科技支撑计划(2009Z2-D261)

作者简介:廖小荣(1984-),男,江西新余人,硕士研究生,研究方向为下一代网络;范冰冰,教授,研究方向为下一代网络、网络安全。

1 RSS 传输过程

一个 RSS feed 包含多个条目及一些相关属性^[14]。RSS feed 的发布和传输过程如图 1 所示: RSS 服务器将待发布的内容按 RSS 规范生成 RSS feed, 并提供一个可以通过 HTTP 协议 GET 方法下载该 RSS feed 的 URL; 这一步通常是发布者在网页上添加新条目时由一些内容管理应用软件自动生成; RSS 阅读器通过用户的订阅操作获得标识该 RSS feed 的 URL 后, 一般会定期地或者根据用户的操作向该 URL 发出 HTTP GET 请求, 下载该 RSS feed 到客户端; RSS 阅读器解析该文档中的每个条目, 并与存储在本地的通过相同的 URL 下载的 RSS feed 进行对比, 如果文档中有新的条目, 则处理并呈现给用户, 而丢弃文档中所有旧的条目^[15]。

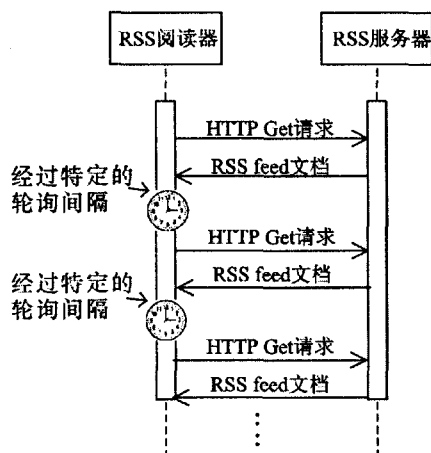


图 1 RSS feed 的传输过程

2 用 DCD 传输 RSS feeds

现设计一种基于 DCD 协议的 RSS feeds 的分发系统, 主要是利用 DCD 中“推”方式的机制来分发 RSS feeds。整个分发系统结构如图 2 所示, 即在传统的 RSS 阅读器和 RSS 生成器之间中引入 DCD 的客户端和服务端, RSS 阅读器和 DCD 客户端都宿于用户终端, RSS 阅读器负责接收来自 DCD 客户端的 RSS feeds 并将其解析、呈现给用户, DCD 客户端负责与 DCD 服务器交互, 包括用户订阅, 内容传输等。DCD 服务器负责接收 RSS feed 服务器生成的内容, 并将其按要求

分发到指定的 DCD 客户端。

DCD 客户端实体可分为三个不同的逻辑功能: 订阅和管理功能负责处理和 DCD 服务器交互的业务管理信息; 内容接收和存储管理功能负责处理接收和管理内容; 客户机应用交互接口负责与客户端应用之间的交互。DCD 服务器包括以下两个逻辑功能: 订阅和管理功能负责处理 DCD 服务器与 DCD 客户端、DCD 服务器与内容提供商之间业务管理信息的交互; 分发功能负责将内容分发给 DCD Client。

DCD 服务器可以接收各种 RSS feeds 并将其保存在服务器的数据库当中, 当 RSS 生成的 RSS feeds 有新的内容时, DCD 服务器将其封装在内容更新通告 (Content Update Notification) 消息中, 并将消息发送到相应 DCD 客户端。DCD 服务器在内容更新通告消息的消息体时, 可以剔除 RSS feeds 中的重复内容。如此, DCD 服务器还可以充当 RSS 格式转换器的角色, 将 RSS feeds 方便地转换为 DCD 客户端可以解析的格式。

2.1 用 DCD 传输 RSS feeds 的封包机制

DCD 是一个基于 XML 的分发协议, 协议栈结构可见文献[9], 其底层传输协议可以是 HTTP、WAP push、SMS 等。在 DCD 中, 其内容的封装是使用分层“信封”模型, 每个信封都包含内容的元数据和嵌套的信封, 对于每一处理层都有其对应的元数据。

DCD 协议为各种应用定义了丰富的元数据, 包括应用元数据、频道元数据和内容元数据。现将 RSS 应用作为 DCD 的频道, DCD 服务器根据 RSS 应用的特点用 DCD 定义的元数据对其进行描述, 其封包机制如图 3 所示。由于 RSS 应用比较简单, 只需用 DCD 元数据集的子集。整个 RSS feed 文档作为 DCD 频道的内容, 在整个处理过程中 RSS 内容对于 DCD 是不可见的。

2.2 用 DCD 传输 RSS feeds 的过程

使用 DCD 传输 RSS feeds 流程, 如图 4 所示。

1) 内容提供商 (RSS feed Server) 向 DCD 服务器注册频道, 定义该频道相关属性, DCD 服务器为新注册的频道指定频道标识符。

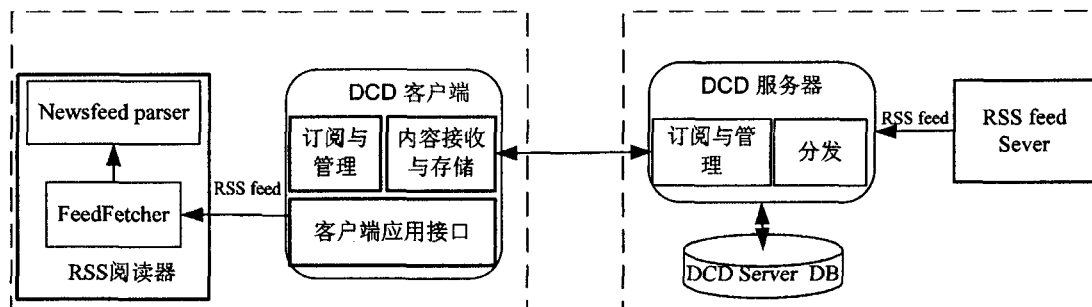


图 2 利用 DCD 协议分发 RSS feeds 的架构图

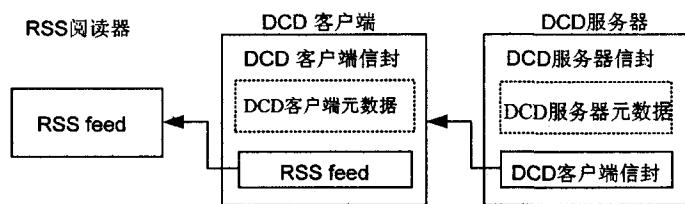


图3 利用 DCD 协议分发 RSS feeds 的封包过程

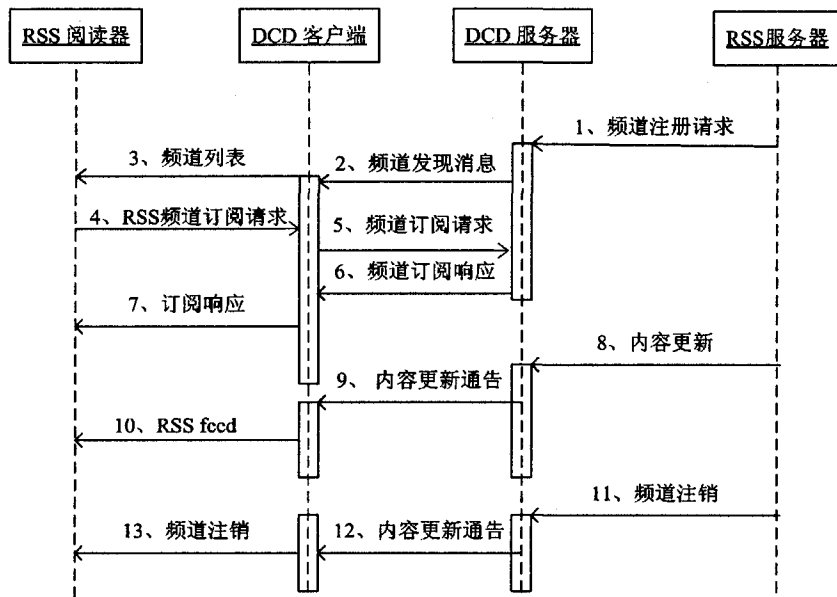


图4 利用 DCD 协议分发 RSS feeds 的流程图

2) DCD 服务器将新注册的频道信息通过频道发现消息,告知 DCD 客户端。

3) DCD 客户端将频道列表发给 RSS 阅读器。

4) RSS 阅读器订阅频道。

5) DCD 客户端将向 DCD 服务器发送频道订阅通知消息。

6) 订阅成功,DCD 服务器向 DCD 客户端发送,频道订阅通知回复消息。

7) DCD 客户端向 RSS 阅读器返回订阅成功消息。

8) RSS 生成器向 DCD 服务器传输内容 (RSS feeds)。

9) DCD 服务器将内容按订阅规则通过内容更新通告消息推送至客户端。

10) DCD 客户端将 RSS feeds 发给 RSS 阅读器。

11) 当频道停用时,RSS 生成器向 DCD 服务器注销频道。

12) DCD 服务器向 DCD 客户端发送频道停用通知。

13) DCD 客户端告知 RSS 阅读器频道已停用,RSS 阅读器解除对该频道的订阅。

定时间间隔由客户端向服务器端轮询,而基于 DCD 的方式是由 DCD 服务器主动向客户端推送更新的内容。推送方式可以及时获取更新信息,而轮询方式需经过一个轮询周期之后才能获得最新更新信息。推送方式可以避免重复下载。

其次,在 RSS 频道订阅与退订上的不同:现在

在 RSS 阅读器要订阅频道,需提供 RSS feed 的 URL 给 RSS 阅读器,RSS 阅读器将自动下载该 RSS feed 文档;要退订一个频道,用户只需在 RSS 阅读器中删除该频道即可。这种方式存在的问题有,如用户获取自己感兴趣 RSS 频道的 URL 不便;如某些频道停用,RSS 阅读器会照常轮询,除非用户手动删除频道。在 DCD 中当有频道注册到 DCD 服务器或已注册的频道废弃或过期时,DCD 服务器向客户端发送频道发现消息来告知 DCD 客户端,使其及时获取最新的频道信息,方便用户进行频道订阅。客户端要进行频道订阅时,DCD 客户端将向 DCD 服

务器发送频道订阅通知消息,该消息包含要进行订阅的各种信息。当频道订阅成功时,服务器返回频道订阅通知回复消息。通过各种元数据,用户可以对 RSS 频道进行个性化的订阅。

现设定一个场景用于比较两种传输方式的性能。考虑的主要参数^[15]有:服务器更新时间间隔 S_{ut} 、RSS feed 的平均大小 F_{sz} 、一个 RSS feed 平均的更新率 U_r 、RSS 客户端更新的间隔时间 C_{ut} 和订阅用户数 C_n 。现参考文献[2]的结论,取 $S_{ut} = 60$ 分钟, $F_{sz} = 5$ kB, $U_r = 20\%$, $C_{ut} = 30$ 分钟。现假设某一 RSS 频道在 30 天之内,按照两种不同的分发方式分发,比较如图 5。

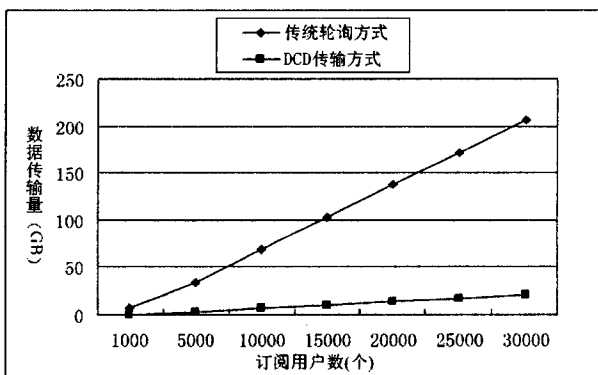


图5 两种传输方式在设定场景中数据传输量比较

设定场景的计算结果表明,服务器在为用户提供同样质量的服务情况下,利用 DCD 传输 RSS feeds 可

3 传输模型分析与比较

两种应用模型的最大不同在于,传统方式是按固

以彻底解决传统轮询方式的重复传输问题,由于 DCD 每次推给客户端的内容都是新内容,从而也避免了客户端对重复传输内容的重复解析。同时由于客户端与服务器交互次数的减少,也减少了 RSS 应用对服务器的性能压力。

4 结束语

通过对传统 RSS 应用模型与基于 DCD 的 RSS 应用模型比较分析,传统方式存在传输及时性和传输效率问题,频繁的轮询也造成了沉重的服务器负载;而使用 DCD 协议传输 RSS 则大大提高传输效率,彻底剔除重复传输。同时,通过 DCD 协议中的频道发现与频道订阅消息增强对 RSS 频道的管理。用 DCD 传输 RSS 的最大缺点是需要 DCD 协议栈的支持,由于 DCD 协议是最近才由 OMA 组织推出,在市场化和产品化上还有待进一步提高;其次需要 DCD 服务器在对 RSS feed 进行封装和 DCD 客户端对消息的解封装,从而增加了消息的处理流程。

参考文献:

- [1] 孙志国. Web2.0, 以个人为中心的互联网时代的到来[J]. 农业网络信息, 2005(12): 97-100.
- [2] Liu Hongzhou, Ramasubramanian V, Siler E G. Client Behavior and Feed Characteristics of RSS, a Publish-Subscribe System for Web Micronews[C]//Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement. Berkeley: USENIX Association, 2005: 1-6.
- [3] 杨俊, 兰宏勇. 基于 RSS 的信息推送系统的设计和实现[J]. 计算机系统应用, 2008(10): 64-67.
- [4] Hicks M. RSS comes with bandwidth price tag[EB/OL]. [2004-09-20]. <http://www.eweek.com/c/a/Messaging-and-Collaboration/RSS-Comes-with-Bandwidth-Price-Tag/>.
- [5] Robbie D S. Enhancing RSS Feeds: Eliminating overhead through Binary Encoding[C]//Proceedings of the Third International Conference on Information Technology and Applications. Los Alamitos: IEEE Computer Society, 2005: 520-525.
- [6] 卢良进, 万健, 徐向华. 基于 P2P 的 RSS 内容分发系统研究与实现[J]. 杭州电子科技大学学报, 2008, 28(2): 61-64.
- [7] Johannes K, Jukka M. Delivering RSS-feeds over SIP[C]//Proceedings of the 2008 Eighth International Workshop on Applications and Services in Wireless Networks. Aswn: IEEE Computer Society, 2008: 76-84.
- [8] 吴伟. 移动互联网业务与 OMA 技术标准[J]. 电信网技术, 2010(2): 1-7.
- [9] Open Mobile Alliance. DCD Specification[EB/OL]. [2008-12-23]. http://member.openmobilealliance.org/ftp/Public_documents/cd/Permanent_documents.
- [10] 马旭涛, 刘建新, 陈靖. 一种基于移动互联网的 DCD 业务系统模型[J]. 现代电信科技, 2008(9): 29-32.
- [11] 孙一刚, 王安生. 移动数字图书馆集成服务的体系结构与表现形式[J]. 国家图书馆学报, 2008(3): 28-32.
- [12] 中国移动通信集团公司. 移动互联网时代的快讯(DCD)业务再思考[EB/OL]. [2009-06-2]. http://labs.chinamobile.com/mblog/285_20001?wralxianrxn.
- [13] 华为技术有限公司. 动态内容分发的内容同步方法、设备及系统: 中国专利, H04L12/54[P]. 2007-06-14[2010-02-10]. <http://www.wipo.int/pctdb/ja/wo.jsp?WO=2007065331>.
- [14] Winer D. RSS 2.0 Specification[EB/OL]. [2003-07-15]. <http://cyber.law.harvard.edu/rss/rss.html>.
- [15] 于魁飞. 基于 RSS 的信息发布与订阅技术研究[D]. 北京: 北京邮电大学, 2007.

著录参考文献的方法

参考文献是指“文后参考文献”。其概念是为撰写或编辑论著而引用的有关图书资料。按规定,在各类型出版物中,凡是引用前人或其他人的观点、数据和材料等,都要对它们在文中出现的地方予以标明,并在文末列出。引文不宜过少,以 12~20 个为宜,尽量引用最新的文献和著录公开发表的文献以及《计算机技术与发展》上近期刊登的论文。

参考文献序号应按文中引用顺序排列,并

将参考文献序号标识在文内引文处。文献中责任者(作者,编者,译者等)、题名(书名,刊名,专利题名,析出题名等)、出版项(出版地,出版者,出版年、卷、期等)应齐全。外文著录人名时应姓在前、名在后,名用缩写字母,三名以上作者写出前三名,其余以“等”表示,参考文献的序号加方括号[]。文献类型请在题名后加以标识,如:报告[R],期刊[J],专著[M],论文集[C],学位论文[D],专利[P],标准[S]。