

动态网页分发加速技术研究

张铃丽, 石磊

(郑州大学 河南省信息网络重点开放实验室, 河南 郑州 450052;
郑州大学 信息工程学院, 河南 郑州 450052)

摘要:随着 Web 技术的迅速发展, 动态和个性化网页的比重日益增加, 而传统缓存一般只适用于静态内容, 难以减少获取动态网页所需的流量和延时代价。为了更有效地分发动态网页, 人们提出了各种动态内容加速方案。文中研究了典型的动态网页分发加速方法, 并对相关的加速技术进行了分析和比较。针对 ESI 和 CDE 这两种技术的优缺点, 提出基于共享片段的动态网页分发加速模型。实验结果表明, 与 ESI 和 CDE 相比, 该模型可以节省更多的带宽, 减少更多的延时。

关键词: Web 缓存; 动态网页; 网页分发; Web 性能; 共享片段

中图分类号: TP393

文献标识码: A

文章编号: 1673-629X(2008)08-0130-04

Research on Techniques for Efficient Delivery of Dynamic Web Pages

ZHANG Ling-li, SHI Lei

(Henan Provincial Key Lab. on Information Network, Zhengzhou University, Zhengzhou 450052, China;
School of Information Engineering, Zhengzhou University, Zhengzhou 450052, China)

Abstract: With the rapid development of the Web, dynamic and personalized Web pages increase gradually. However, traditional Web caching is commonly applicable to static content, and is hard to reduce the traffic and latency in retrieving dynamic Web pages. In order to deliver dynamic Web pages more efficiently, researchers proposed some kinds of schemes for accelerating dynamic content. Classical techniques for efficient delivery of dynamic Web pages are studied in this paper, then the accelerating techniques are analyzed and compared. In view of the advantages and disadvantages of ESI and CDE, a new model of dynamic Web pages for efficient delivery based on shared fragments is presented and discussed. Experimental results show that the model can provide more bandwidth savings and latency reduction than existing solutions ESI and CDE.

Key words: Web caching; dynamic Web page; Web page delivery; Web performance; shared fragments

0 引言

近年来, 虽然 Internet 的网络基础设施有了很大的发展, 但网络设施的发展远跟不上网络应用和网络用户的快速增长, 由于接入 Internet 的用户数量剧增以及 Web 服务和网络固有的延迟, 使得网络越来越拥挤, 用户的服务质量得不到很好的保证^[1]。如何缩减 Web 访问延迟、改善 WWW 服务质量已经成为改善 Internet 应用环境的主要目标。

目前, 越来越多的网站提供个性化和动态内容, 但随着 WWW 上动态内容和个性化服务的比重日益增加, 传统缓存带来的性能改善已不再显著^[2]。另外, 生

成动态内容的代价比较大, 为大量用户同时动态生成个性化的页面将导致传输延时和数据传输失败。传统的网站缓存方案在提高静态 Web 内容的传输速度方面发挥了很大的作用, 但动态内容和个性化服务比重渐增, 单一的静态缓存技术和传统的 CDN 技术对网络性能的改善并不明显。为了满足用户的个性化需求及减少网络带宽需求和网络延时, 人们提出了各种动态网页分发加速技术, 从而来改善 WWW 服务质量。

1 动态网页分发加速技术

动态网页分发加速技术可以提高 Web 页面的交付速度, 缩短响应时间。动态网页分发加速技术可以分为两大类: 其一, 传输加速技术, 目的是减少从原服务器到客户之间传输动态网页所需的带宽和延时, 主要包括: ESI (Edge Side Includes) 规范和 CDE (Class-based Delta Encoding)^[3]; 其二, 动态网页生成加速技术, 用以提高服务器端产生动态网页的性能, 主要包括

收稿日期: 2007-11-20

基金项目: 国家自然科学基金(60472044)

作者简介: 张铃丽(1982-), 女, 河南周口人, 硕士研究生, 主要研究领域为动态缓存技术; 石磊, 博士, 教授, 主要研究领域为高性能计算、Web 预取、Web 挖掘与中间件技术。

预取技术^[4]、DUP(Data Update Propagation)^[5]和 DC-CP(Dynamic Content Caching Protocol)^[6]。

1.1 ESI

片段缓存是加速 Web 应用的一种有效技术,并且一些通用的应用服务器也支持片段缓存^[7]。ESI(Edge Side Includes)对片段缓存进行了规范,已提交给 W3C 作为标准。ESI 是一种基于 XML 的标记语言,用于定义包含片段、变量和其他控制指令的模板,使得客户、代理和 CDN 能够在片段的级别缓存文档。它描述了可缓存和不可缓存的网页组件,支持页面片段的动态装配,允许高速缓存静态页面元素,缩短网页的整体传输时间。ESI 允许将网页分解成具有不同缓存特性的片段,每个片段作为一个独立的元素保存在缓存服务器上。当响应用户请求时,缓存服务器只需要获取那些不可缓存或者过期及没有命中的片段,减少了响应时间和原服务器的负担。

传统的内容生成和发布网络以及 ESI 分离内容传输和内容生成分别如图 1 和图 2 所示。

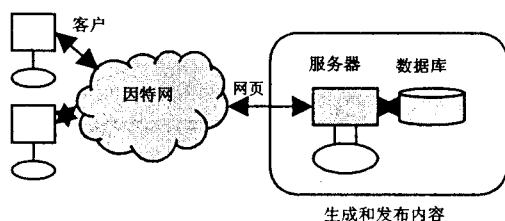


图 1 传统的内容生成和发布网络

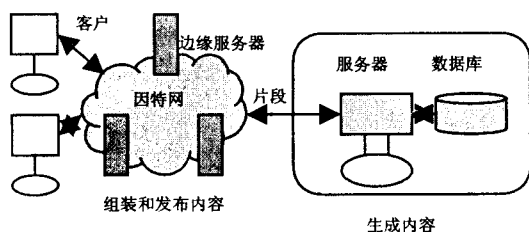


图 2 ESI 分离内容传输和内容生成

1.2 CDE

DE(Delta Encoding)^[3]指的是以一个或多个通常相似的文件作为基文件对目标文件进行编码的过程,编码后的结果称为 delta(增量)。该方法是在一个慢速链接的两端分别存储一个文档的相同副本,称该副本为基文件。慢速链接的服务器端从服务器取回文档的当前副本,并计算当前副本与存储副本的差异(delta),然后再把 delta 发送到慢速链接的另一端。当客户机收到 delta 以后,就把 delta 与基文件结合起来生成当前副本。因此,响应一个对动态网页的请求时,只需要传输编码得到的 delta。客户通过旧版本和 delta 进行解码来恢复原来的响应(如图 3 所示)。

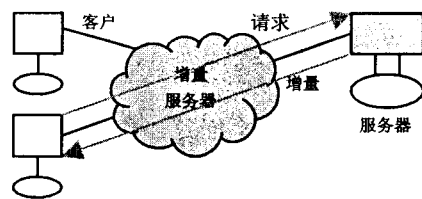


图 3 使用增量编码的网络模型

DE(增量编码)为开发动态流量中的时间相关性提供了一种有效的方法,但是它不能解决服务器端的可扩展性问题,动态文档的数量在不断增加,服务器端存储容量的需求越来越大。针对这个问题,文献[3]提出了一种 CDE(Class-based Delta Encoding)方法,对多个文档进行分类,每类只保存一个基文件。与前面方法不同的是,当服务器端对当前响应进行 DE 的时候,所用的基文件可能客户没有保存,编码后的响应中包含了该基文件的 URL,客户收到响应后,如果发现没有该基文件就先请求该文件。这样,在第一次请求该网站的网页时,需要请求两次才能获得该网页。但是,这个方法需要保存的基准文件比较少并且可以被很多文档共享。

1.3 预取

预取也被称为主动缓存技术,Cao 等人提出主动缓存方案(active cache scheme)^[4],它通过一种 Cache Applet 的方式把部分用户请求处理的任务转移到代理上。Cache Applet 是由原服务器提供给代理的代码,Cache Applet 可以通过其 URL 获取,当缓存网页时,代理会从原服务器上获取相应的 Cache Applet。当用户请求命中缓存的动态网页时,代理把用户请求作为参数传递给已下载到本地的 Applet,Applet 要么通过计算返回一个新的页面,要么返回对象的拷贝,要么要求代理转发请求到原服务器。

预取是根据用户当前的访问请求和历史访问记录,预测用户将来可能发出的访问请求,在用户浏览当前 Web 页面时将预测的内容取到本地高速缓存(cache)中^[2]。用户在真正要访问这些页面时只需从本地高速缓存下载,从而在很大程度上达到减小用户访问延迟的效果。

1.4 DUP

DUP^[5](Data Update Propagation)可以确定底层数据的改变对缓存对象产生的影响。它使用对象依赖图(ODG)建立底层数据和缓存内容的精确依赖关系并计算和监视这种依赖关系,触发缓存内容更新^[8]。当底层数据变化的时候,系统遍历依赖图更新或者使缓存对象失效。缓存内容可能是整个 HTML 网页也可能是 HTML 网页中的片段。当系统感知到底层数据改

变时,就使用图遍历算法确定受到影响的缓存对象。一旦发现缓存对象过期就让它失效或者进行更新。

1.5 DCCP

Ben Smith 等人于 1999 年提出了一个新协议 DC-CP^[6](Dynamic Content Caching Protocol),此协议允许个性化内容生成程序(如 CGI 应用软件,Java servlets 等)指定不同 GET 请求之间的等价或者部分等价。Web 缓存可以依据该信息来使用这些内容的其他形式。所以,可以使用缓存的结果来响应相同的或者等价的请求。对于部分等价的请求,当生成和传送实际内容的时候,可以把事先缓存的内容作为一个近似的结果传送给客户。客户在等待更多精确信息时就可以迅速地先浏览部分或者相似的结果。

2 基于共享片段的分发加速模型

在动态网页的分发加速技术中,ESI 和 CDE 最具代表性。ESI 可以在片段级别达到巨大的重用,但是不能重用相似的内容且需要对源代码进行修改。CDE 可以重用相似内容,不需要修改代码和目前的 WWW 架构,容易在现有的基础上部署,但 CDE 是在完整网页的级别重用相似网页,由于目前动态和个性化内容的比重日益增加,CDE 已不能充分减少冗余流量。针对两者的优缺点提出基于共享片段的分发加速模型(Model of Efficient Delivery Based on Shared Fragments, MEDBSF)。它对动态网页分类,不同类型的网页使用不同方法,具有 ESI 和 CDE 的优点,实现了片段级别的增量编码。

为对页面进行分类,代理为缓存中的每个页面记录两个参数:请求频率(requestfrequency)和更新频率(updatefrequency)。设定参数 α 为两者比率,即 $\alpha = \frac{\text{requestfrequency}}{\text{updatefrequency}}$ 。

定义 1 如果某个页面的请求频率小于或等于更新频率,即 $\alpha \leq 1$,称该页面为积极页面。

定义 2 如果某个页面的请求频率大于更新频率,即 $\alpha > 1$,称该页面为消极页面。

对于积极页面和消极页面,应该采用不同的处理办法。积极页面由于更新频率较快,采用主动更新较为合适;消极页面,由于请求频率占优,采取被动更新为宜。

定义 3 页面更新定义为四元组 $U < F, D, R, H$,其中 F 表示片段, D 表示 delta, R 表示片段和网页之间的映射关系, H 表示网页。由 F 、 D 和 R 可以完成对页面 H 的更新。

该方法包括以下三个步骤:

Step1:分类,动态网页分为积极页面和消极页面。

Step2:当积极页面发生变化时,就立即对缓存中的网页进行更新,保持缓存中的页面一直有效。对于消极页面,只有当用户请求时才进行更新。

Step3:采用 ESI 和 CDE 相结合的方法对页面进行更新。

Step4:更改相应页面的更新频率和请求频率。

系统结构如图 4 所示。可以看出,修改了 delta 服务器,修改后的 delta 服务器在空闲的时候对网页分类,分析积极页面并导出片段和页面之间的映射关系。分析积极页面也可以由专门的分析器来完成以避免 delta 服务器超载。页面和片段之间的映射关系是多对多的,一个页面由多个片段组成,同时一个片段也可能被多个页面所包含。当 delta 服务器收到用户的请求时,先判定它所请求页面的类型,如果是积极页面,则直接响应;如果是消极页面,则转发给原服务器,以确保强一致性,在收到原服务器的响应后,对响应进行片段级别的增量编码,然后将修改后的响应发给客户,同时更新缓存的消极页面。对于每个页面,如果发生一次更新或者被用户请求一次,它的更新频率或请求频率也会被修改一次。

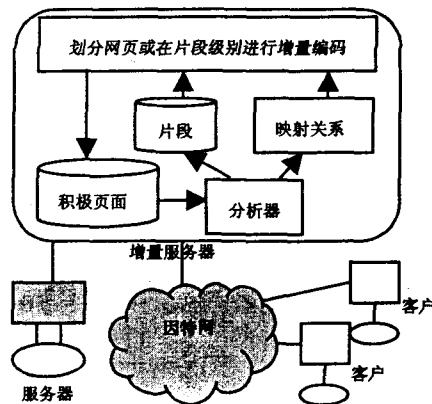


图 4 基于共享片段的系统结构图

该模型结合了 ESI 和 CDE 的优点,在片段级别进行编码以重用相似的片段,从而可以最大程度地节省带宽。一个片段存储一次就可以被多个网页使用,缓存的共享片段具有较高的重用性和较低的存储开销。经常使用的片段可以缓存在本地或者代理服务器中,因此可以减少延时。可以在不修改程序代码、WWW 架构和 HTTP 协议的情况下进行部署。缓存片段和缓存静态网页是一样的,代理可以缓存片段,该模型可以利用代理分发动态网页而不需要对他们进行修改。

图 5 显示了三种方案的相对带宽需求。对于每种方案,相对带宽需求就是实施该方案后所需带宽与修改系统前所需带宽的比值。从图中可以看出,MEDB-

SF 所需带宽在所有网站中都是最少的。然而,对一些网站(如 51job 和 bokee)CDE 的带宽需求和 MEDBSF 很接近,这是因为,对于这些网站,CDE 基准文件的命中率和 MEDBSF 片段的命中率差不多。图 6 显示了三种方案的加速比。对于每种方案,加速比是在实施该方案后的延时与修改系统前的延时的比值。从图中可以看出,MEDBSF 的性能在所有网站中都是最高的。和带宽需求结果类似,对于一些网站,CDE 的性能和 MEDBSF 接近。

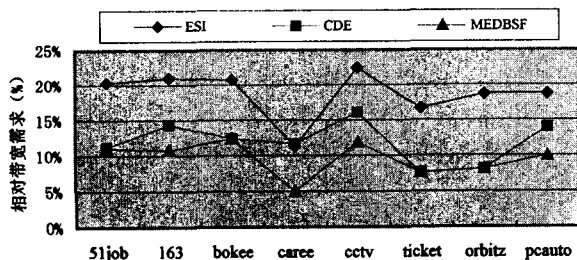


图 5 8 个网页集的相对带宽需求

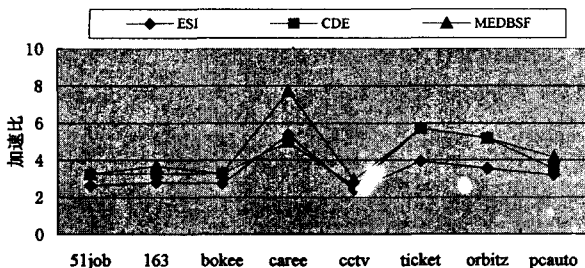


图 6 8 个网页集的加速比

3 结束语

讨论了目前几种典型动态网页分发加速技术,表 1 中给出相关技术的分类及比较,可以看出整体上可以分为两类:传输加速技术和生成加速技术。其中 CDE 和 ESI 最具代表性。CDE 可以有效地减少响应时间和带宽需求,但在许多个性化站点中的效果并不显著。ESI 在片段的粒度达到了巨大的内容重用,也没有引入额外的计算代价,但是它需要 Web 开发者对网页进行手工分片。

文中提出的基于共享片段的分发加速模型结合了 CDE 和 ESI 的优点,它在片段级别进行增量编码,可以重用相似片段,最大限度地减少了数据传输量,节省带宽。该方案可以在不修改程序代码、WWW 架构和 HTTP 协议的情况下部署。但是它也有自己的局限性,主要是需要 DE(增量编码)计算,但由于 DE 是 CPU 密集型计算,所以会较明显地降低服务器端的吞吐量^[3]。然而,CPU 技术发展迅速且价格降低,计算成本相对便宜。

动态网页分发加速技术今后的发展方向主要是传

输加速技术的改进。因为对于生成加速技术,比较容易在服务器端用增加硬件和部署新技术来提高动态网页的生成速度。传输加速技术的本质是尽量避免传输冗余信息。ESI 可以有效减少冗余信息的传输,但需要修改源代码,CDE 虽不能有效减少冗余信息的传输但不需要修改源代码。文中提出的模型具有 ESI 和 CDE 的优点,不需要修改源代码且可以充分减少冗余信息的传输,理论上可以节省更多的带宽,减少更多的延时。

表 1 相关技术的分类及比较

传输加速技术	ESI	对片段缓存进行了规范,提高了网页的可缓存性,可以实现片段级别的重用,但不能重用相似内容而且需要对源代码进行修改。目前 ESI 只受到 Java 类开发平台的支持,其它如 ASP、NET、PHP 尚无支持 ESI
	CDE	可以有效地节省带宽减少延时,可重用相似内容,不需修改代码和 WWW 架构,易于部署,但是由于 CDE 是在完整网页而非片段级别重用,不能充分减少冗余流量,在个性化站点中效果并不显著
	MEDBSF	它对动态网页分类,不同类型的网页使用不同方法,具有 ESI 和 CDE 的优点,实现了片段级别重用,可以重用相似的内容,可以在不修改程序代码、WWW 架构和 HTTP 协议的情况下部署
生成加速技术	预取	Web 预取可以对用户未请求过的页面进行缓冲,是一种主动的高速缓存技术,可以大大缩减 Web 访问延迟,但是如果控制不当将引起较大的额外带宽的消耗
	DUP	使用对象依赖图(ODG)来表示对象和基本数据之间的依赖关系,只更新那些受到影响的对象。可保持缓存内容的强一致性,但需要开发者指定缓存对象和底层数据的依赖关系
	DDCP	对于采用 GET 方法的查询请求,原服务器可以使用该协议指定等价或部分等价的请求,从而达到动态内容的重用。可以重用缓存的等价或者相似的结果,但需要修改协议,也只适用于 GET 请求

参考文献:

- [1] Wang Jia. A survey of Web caching schemes for the Internet [J]. Computer Communication Review, 1999, 29(5): 36-46.
- [2] Shi Lei, Han Yingjie, Ding Xiaoguang, et al. An SPN based Integrated Model for Web Prefetching and Caching [J]. Journal of Computer Science and Technology, 2006, 21(4): 482-489.
- [3] Psounis K. Class - Based Delta - Encoding: A Scalable Scheme for Caching Dynamic Web Content [C]//22nd International Conference on Distributed Computing Systems Workshops (ICDCS 2002). Vienna, Austria: [s. n.], 2002: 799-805.
- [4] Cao Pei, Zhang Jin, Beach K. Active cache: Caching Dynamic Contents on the Web [C]//In Proceedings of Middleware '98. Lake District, UK: [s. n.], 1998: 373-388.
- [5] Challenger J, Iyengar A, Dantzig P. A Scalable System for Consistently Caching Dynamic Web Data [C]//In Proceedings of the IEEE INFOCOM 1999. New York: [s. n.], 1999: 294-303.
- [6] Smith B, Acharya A, Yang Tao, et al. Exploiting Result Equivalence in Caching Dynamic Web Content [C]//Proceedings of Second USENIX Symposium on Internet Technologies

(下转第 147 页)

些问题。虽然文献[6]认为 AES-128 加密算法对大部分商业应用来说是足够安全的,而且 NIST 也预计 AES-128 加密至少用到 2036 年是安全的^[7],但是文献[8]认为单一的对称加密算法在数据加密和密钥交换中可能带来安全隐患,研究非对称加密在密钥分配中应用具有前景。文献[9]建议将椭圆曲线加密方法作为公钥非对称计划在 ZigBee 技术中应用。在 ZigBee 组网方面,基于 ZigBee 技术组成的网状网只适合数据传输较低的应用,例如工业控制领域,而不适合数据传输量多的应用,因此需进一步加强 ZigBee 组网的研究,使之应用领域更为广泛。对于无线传感网络中数据安全交换方面,ZigBee 联盟只是在理论上对网络层安全协议进行描述,并没有对不同应用应采取具体安全级别有具体的研究,因此加强针对不同应用的具体安全措施还有待进一步发展,同时对数据完整性和认证技术研究以及根据不同的应用情况,进行安全属性的灵活配置研究也很重要。由于 IPv6 拥有巨大的地址空间,能为每一个 ZigBee 节点分配一个全球唯一的网络地址,同时还能提供很好的 QoS 和安全的通信保障,因此,IPv6 和 ZigBee 结合是未来发展的一个亮点,目前国内相应的产品也已经问世。总之,为 ZigBee 技术有更加广阔的应用空间,ZigBee 技术的安全技术,如密钥分配协议的需求与性能指标、密钥管理的方案等,还需进一步深入研究。

表 1 ZigBee 安全级别

安全级别 Security level	安全属性 Security attribute	数据加密 Data encryption	帧完整性 Frame integrity	完整性代码 Integrity code
0	没有	没有	没有	0 bit
1	MIC-32	没有	有	32 bit
2	MIC-64	没有	有	64 bit
3	MIC-128	没有	有	128 bit
4	ENC	有	没有	0 bit
5	ENC+MIC-32	有	有	32 bit
6	ENC+MIC-64	有	有	64 bit
7	ENC+MIC-128	有	有	128 bit

6 结束语

ZigBee 是一种新兴的无线网络通信技术,它因优

越的特性在众多技术中脱颖而出,被业界认为是最适合无线传感网络的新技术。在安全方面,ZigBee 技术对协议栈各层加强安全保护,采用 AES 加密算法对数据加密,同时提供数据完整性检查和鉴权措施,还建立信任中心机制对安全钥匙管理,这些安全措施采用使无线网络通信具有良好安全保护机制。通过对 ZigBee 技术安全分析,对 ZigBee 安全优势和不足有一定了解,但是随着许多应用对安全需求的提升,进一步加强安全研究是必要的。

参考文献:

- [1] ZigBee Alliance document[EB/OL]. 2004. <http://www.zigbee.org>.
- [2] Paolo B, Prashant P, Vince W, et al. Wireless Sensor networks: A survey on the state of the art and the 802.15.4 and ZigBee standards[J]. Computer Communication, 2007, 30(7):1655-1695.
- [3] 任秀丽,于海斌. 基于 ZigBee 技术的无线传感网的安全分析[J]. 计算机科学, 2006, 33(10):111-113.
- [4] Moazzam K, Fereshteh A, Jelena M. Key Exchange in 802.15 Networks and Its Performance Implications[C]//In Proc of 2nd International Conference on Mobile Ad-hoc and Sensor Networks (MSN 2006), LNCS, Vol. 4325. [s. l.]: Springer-Verlag, 2006:497-508.
- [5] ZigBee Alliance. ZigBee Security Specification Overview[EB/OL]. 2005. http://www.zigbee.org/en/events/documents/December2005_Open_House_Presentations/Zigbee_Security_Layer_Technical_Overview.pdf.
- [6] Ferguson N, Schneier B. Practical cryptography[M]. New York: John Wiley and Sons, 2003.
- [7] Krasner J. Using Elliptic Curve Cryptography(ECC) for Enhanced Embedded Security[EB/OL]. 2004. <http://www.embedded-forecast.com/EMF-ECC-FINAL1204.pdf>.
- [8] Ondrej H, Peter K, Petr F, et al. On security of PAN wireless systems[C]//In Proc of 6th International Workshop on Embedded Computer Systems(SAMOS 2006), LNCS, Vol. 4017. [s. l.]:Springer-Verlag, 2006:178-185.
- [9] Pereira R. ZigBee and ECC Secure Wireless Networks[EB/OL]. 2004. <http://www.elecdesign.com>.

(上接第 133 页)

- and Systems. Boulder, Colorado, USA: [s. n.], 1999:209-220.
- [7] Yuan Chun, Cheng Yu, Zhang Zheng. Evaluation of Edge Caching/Offloading for Dynamic Content Delivery[J]. IEEE

Transactions on Knowledge and Data Engineering, 2004, 16(11):1411-1423.

- [8] 范国闯,钟华,黄涛,等. Web 应用服务器研究综述[J]. 软件学报, 2003, 14(10):1728-1739.