

移动环境下支持技术的研究

Study on Supporting Technology in Mobile Environment

吴婷婷 周兴铭

(国防科技大学计算机学院重点实验室 长沙 410073)

Abstract With the rapid development of wireless telecommunications network and mobile devices, mobile computing environment becomes more and more mature. Following the trend, mobile supporting technology is also becoming the hot research topic. Mobile supporting technology is the technology that supports people access information and do task through the network from any place. Compared with traditional distributed environment, the property of mobile environment——mobile device and resource-poor, challenges mobile supporting technology.

Keywords Mobile computing, Caching, Prefetch

1 背景

今天,我们的日常生活和工作中,各种各样的移动设备正发挥着它们各自的作用。移动电话以前所未有的速度迅速普及;手持 PC 和个人数字助理 PDA 在保证通信能力和便携性的基础上,加强了计算能力,使人们可以更加便捷地管理个人数据和进行通信;某些功能如触摸屏、声音识别、定位系统更为移动应用提供了良好的支持。在计算和存储能力方面可以与台式机相匹敌的笔记本计算机,使专业人士实现了跨越地域限制办公的可能。在无线通信系统方面,基于包交换技术正在逐步替换传统的电路交换技术,相应的标准包括对 GSM 增强的 GPRS (Generic Packet Radio System),数据传输上限大约是 100 kb/s,“第三代”移动通信系统 3G 标准正在制定和测试中。在 3G 理想的情况下,快速移动中设备的数据传输率要达到 144 kb/s;步行移动中则为 384 kb/s;静止或缓慢移动时数据传输率达到 2 Mb/s。无线局域网在其作用范围内(一般是一幢建筑内),数据传输率可以达到 2 至 10 Mb/s。新近推出的蓝牙(Bluetooth)技术实现了短距离(10 米以内)的移动 PC、移动电话和其他计算、通信设备灵活便捷地无线连接。

可以预见,移动环境和相应的支持技术将改变我们工作、学习、生活的方方面面。未来社会中,典型的移动应用包括日常生活中移动网络应用、急救应用、移动办公应用、虚拟导航等。

2 移动计算环境模型

图 1 给出了一个典型的移动计算环境模型^[3],这个模型由固定设备和移动设备组成。所有的固定设备通过高速固定网络相连(Mb/s—Gb/s)。部分固定设备具有可以与移动设备通信的无线接口,我们将这部分固定设备称为移动支持结点 MSS。每个 MSS 负责建立一个无线网络单元(如图中所示的无线广播单元或无线局域网单元)。移动设备可以从任何一个无线网络单元经由 MSS 联接到固定网络中,我们将能够提供数据广播服务的那部分 MSS 称为信息站,信息站与移动设备之间的通信带宽是不对称的。我们的模型支持小型移动设备,如 PDA、手持 PC,也支持处理功能较强的移动设备,如笔记本计算机。我们将前一类移动设备统称为手持移动设备,后一类移动设备称为移动计算机。

与基于固定网络的传统分布计算环境相比,移动计算环境具有以下几个主要特点:1)移动性:一台移动设备可以在不同的地方联通网络,而且在移动的同时也可以保持网络连接。这种计算平台的移动性可能进一步导致系统访问布局的变化和资源的移动性。2)移动设备的资源有限性:尽管移动设备在各方面能力上有了很大的发展,并且在绝对能力上还会不断提高,但与固定设备相比,它们总是资源相对有限。3)移动设备的电源能量有限性:移动设备主要依靠蓄电池供电,而电池的容量总是相对有限,一般移动设备在正常的连续使用情况下只能维持几个小时。4)移动设备的频繁

吴婷婷 博士生,研究方向为移动数据库,移动计算,分布计算。周兴铭 中科院院士,教授,博士生导师,主要研究方向为分布与并行数据库,计算机体系结构,计算机协同。

断接性:移动设备在移动过程中,由于使用方式、电源、无线通信费用、网络条件等因素的限制,一般不采用保持持续联网的工作方式,而是主动或被动地间歇性入网、断接。5)网络带宽有限性、多样性、非对称性:在大多数情况下,移动设备是通过无线连接,带宽非常有限,而移动计算机的移动性又使得不同时间可用的网络带宽与服务质量变化多端。由于物理通信媒介的限制,一般的无线网络通信都是非对称的,表现在固定服务器节点可以拥有强大的发送设备,而移动计算机的发送能力非常有限,于是下行链路(服务器到移动计算机)的通信带宽与代价和上行链路是相差很大的。6)低可靠性:无线网络与固定网络相比,可靠性较低,更容易受到干扰而出现网络故障;此外,移动计算机由于其便携性和工作环境,也带来潜在的不安全因素。

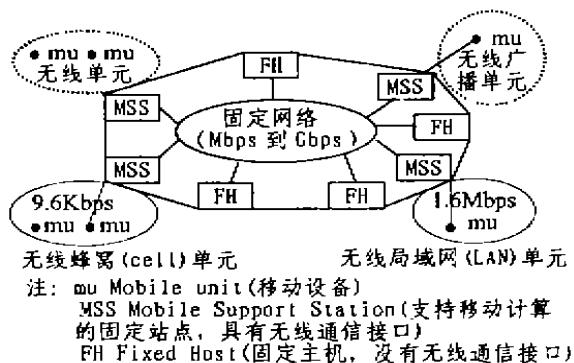


图1 移动计算环境模型

3 移动环境下的研究问题

我们的环境模型将移动设备归为两类,在下面讨论中,我们将针对它们的特点分别对待。

3.1 面向手持移动设备的研究问题

手持移动设备,因为受自身体积限制、资源有限,处理和存储能力都很弱,甚至可以被看作是固定主机的远程客户输入输出设备。这类设备所适合的场合多是需要移动设备与固定网络上主机交互,主要靠固定主机完成的应用,如日常生活中移动网络应用。出于这样的考虑,我们认为,要有效支持手持移动设备上的移动应用,关键问题就是如何针对移动计算环境这样特殊的环境,将手持移动设备与固定主机以适合的计算模型结合起来。

传统客户/服务器模型下,移动设备作为客户方直接通过网络连接向固定网络上的服务器请求服务,这样的模型存在很多问题。首先,速度慢且网络可靠性差的无线环境,会给基于消息传递的客户/服务器系统造成不可忍受的网络延迟,尤其对于手持移动设备这种依赖于服务器,网络通信频繁的客户;其次,对于基于

同步远程过程调用 RPC (Remote Procedure Call) 的客户/服务器系统,也有诸如因移动设备断接而调用阻塞等问题。最后,客户方的移动,有可能加剧网络延迟。

那么,该建立一个怎样的支持手持移动设备的移动环境计算模型,使该模型既能减少网络通信流量,加快响应时间;又能支持断接并且对于客户的移动也有良好的适应性,这是对手持移动设备这类客户提供移动支持所需重点研究的问题。

3.2 面向移动计算机的研究问题

作为有相当处理能力和存储能力的移动设备,移动计算机对固定主机的依赖性没有手持移动设备那么强,用户也希望能够用它来完成大部分工作。这类设备多为专业人事使用,典型应用如移动办公应用。出于无线网络连接带宽低、可靠性差、费用高等方面的考虑,用户通常尽量少地将移动计算机保持连线,而是通过预取和缓存,将可能需要的数据存到移动计算机上,来保证断接期间计算机的继续工作。移动计算机仅在必要时刻(如需进一步缓存或数据同步时),才与固定网络上的服务器建立连接。对于移动计算机来说,如何有效地进行预取和缓存,帮助用户充分利用移动计算机的资源,实现本地处理,才是关键问题。

移动计算环境与传统分布计算环境的缓存和预取研究的差别体现在以下几个方面:首先,传统分布式系统下的缓存和预取研究主要用于提高系统性能。但在移动计算环境下,由于缓存和预取的结果不仅是系统性能的提高,而且关系到数据的可用性,缓存失效的代价远远大于固定分布环境中的缓存失效,因此移动环境下的缓存和预取需要更高的准确性。第二,移动计算机的移动性导致网络条件的多变性,在有线与无线网络无论是带宽还是费用都有极大差别的情况下,如何选择具有适应性的缓存和预取策略,是移动环境下新的挑战。第三,缓存一致性是所有缓存研究中必然要考虑的问题,但移动计算机的断接性和移动性使服务器无法确定客户当前的位置和连接状况,传统基于回调方式的一致性办法不再适用或需要改进。第四,我们的模型中有信息站这一类固定结点,它可以支持数据广播,也就是说,在移动计算机与这类结点通信时,信道带宽不对称、下行链路(服务器到移动计算机)比上行链路的通信带宽高、代价低。如何在这样的网络条件下实现有效缓存也是移动环境下特有的问题。

综上所述,面向移动计算机,所要研究的问题包括如何智能地进行缓存和预取,以提高数据可用性,减少网络开销;采取何种策略让缓存和预取适应移动环境的多变性;如何充分利用带宽不对称性实现缓存;以及移动环境下的缓存一致性问题。

考虑到:1)通常移动计算机的磁盘容量足以装下用户要用的所有文件;2)由于文件系统的层次性组织,文件的相关性比较容易发现;3)文件系统共享冲突的

概率比较小,特别是对于使用移动计算机的用户,他们的文件通常由他们自己创建和修改。相比较而言,数据库系统的用户不太可能将整个数据库系统缓存到移动计算机上,且数据库系统中数据相关性不明显,数据库中数据更新比较多。这些特征使我们认为数据库系统的缓存和预取研究是更有意义和更具挑战性的工作,是研究重点。

4 相关工作

4.1 移动环境计算模型的研究

M. I. T. 实验室开发的 Rover 工具集^[2]中提出基于异步队列远程过程调用机制的改进客户/服务器模型,在这个模型下,移动客户发送 RPC 后, RPC 先存放在本地可靠存储的队列中,控制立即返回给客户,在合适的时刻(如下次连线时),队列中的 RPC 被取出并发送。基于异步 RPC 的客户/服务器模型在一定程度上可以支持断接,但它还是难以处理移动网络带宽低、可靠性差等问题。

最为流行的移动计算系统模型是三层客户/代理/服务器模型^[1]。在这种模型下,移动客户的请求都先送到代理,由代理处理后送到服务器;而服务器所有的应答也通过代理返回客户。代理通常是客户在固定网络上的代表。代理除了要提供消息和队列机制来支持从移动客户到代理和从代理到服务器的通信,还可能提供其他一些功能,如当事件发生时,通知客户;替客户缓存;执行过滤、压缩等。客户/代理/服务器模型通过代理实现客户在固定网络上的存在,从一定程度上减轻了带宽有限和无线连接不可靠性的影响,对于断接,移动客户将请求发送到代理,等下次连接时从代理处取结果。而所有发给断接用户的请求也可以先放在代理处,等用户连线时传送。对于弱连接,代理使用优化策略来减少数据传输量,还可以根据客户要求设定数据的优先级,由优先级决定数据发送的顺序。客户/代理/服务器模型适用于只有有限计算能力和资源的手持移动设备,因为代理可以为客户承担部分责任。这种模型的问题是对客户的移动不能有效支持,一旦客户移动到远离代理的地方,代理就无法有效发挥作用,反而白白增加开销。另一个问题是,代理需要预先配置,而针对用户个性化的配置很难预先实现。

客户/中介/服务器模型^[4]是在客户/代理/服务器模型的基础之上加入了客户方代理。客户方代理运行在移动设备上,截获移动客户请求,与固定网络上的代理交互。客户/中介/服务器模型提供了灵活的断接处理,它在客户方代理上维护一个本地缓存,用于满足断接时客户对数据的要求,以支持客户操作断接期间的继续执行。缓存失效可先存在客户方代理,等下次连线时满足。相似的,到客户的请求也可以缓存在服务方代理,等客户连接时再传送,客户方代理和服务方代理

可以通过使用多种优化策略来减少无线连接上的数据传输,提高数据的可用性,支持弱连接。客户/中介/服务器模型只适用于有足够计算能力和存储空间的移动设备。对于手持移动设备,由于本身存储和计算能力有限,很难支持客户方代理。

4.2 预取策略的研究

常见的预取方法是利用程序的空间局部性。文件系统中区缓存的 read-ahead 策略、面向对象数据库 ENCORE 中 aggressive eager prefetch 策略利用的都是空间局部性。但如果只利用程序的空间局部性,可能会预取来一些无用的信息。在移动环境下大多数预取是通过无线网络实现,如果预取来的信息没有用,浪费了宝贵的网络资源。

卡内基-梅隆大学(CMU)的 Patterson 提出 TIP (Transparent Informed Prefetching) 预取策略^[5],是利用应用层关于未来 I/O 访问的提示信息来预取数据。这样,文件系统可以较为准确地预取所需的数据。这种预取策略的问题在于完全依赖于人的参与,不能独立完成预取;而且对于那些不熟悉系统的用户,也无法提供有效的预取信息。

哥伦比亚大学 Tat 的博士论文中提出一种智能预取策略^[6],它利用应用的文件访问模式信息来预取文件的集合,但这种预取策略的问题是系统要不断检测和记录程序的行为,并不断进行匹配,一旦发现匹配就要预取,对于移动环境,受网络限制,不可能随时允许预取;而且,这样做还可能会产生额外的开销,造成系统资源的浪费。

UCLA 的 Kuenning 开发了一个面向断接操作的文件预取系统 Seer^[7],它所基于的主要思想是:系统通过观察用户行为,进一步推断文件间的语义关系,然后根据结论预取。Seer 所基于的最基本假设是用户行为的语义局部性。通过观察发现和利用这个局部性,系统可以推测各个文件间的关系。Seer 用到了语义距离 (semantic distance) 的概念。从概念上说,语义距离试图量化文件间的关系。

4.3 缓存策略的研究

缓存粒度是影响缓存性能的重要因素,传统的缓存多以页或对象为缓存粒度。针对用户对数据库的查询特征,人们提出以谓词查询为粒度的缓存。

缓存空间总是有限的,而用户用到的数据也会发生变化,因此缓存的内容需要更新,这就需要用到缓存替换策略,LRU 是最常用的替换策略。文[8]中提出了基于语义的替换策略,它根据语义距离来决定替换内容,具有一定的启示作用。

由于共享数据的多复本的存在,带来了数据一致性问题。如何保证缓存数据与最新版本数据的一致性为缓存研究的重点。移动环境下,为了节省网络开销,文[9]中缓存一致性提出“基于懒惰的拉(lazy pull-

based)策略,即由客户负责管理其缓存的有效性,并且对于无效的缓存数据也只是在下次数据被访问时才去取新的版本。每个缓存数据都有一个刷新时间(refresh time),刷新时间表示该数据假设有效的时间。当客户访问数据时,他先查看刷新时间是否到期,每次服务器给客户发送数据时,它要为该数据估算刷新时间,并一同发送给客户。使用该策略需要用户允许使用可能过期的缓存数据。

文[10]提出的伪版本(quasi-copy)概念也是基于允许缓存数据与最新版本在一定程度上有所偏差,以此来降低缓存一致性的开销。

Barbara 和 Imielinski 使用基于数据广播的使缓存无效的方法^[11]。在这个方法中,服务器定期发送无效报告来告诉客户所有改变了的数据。这样,客户无需向服务器方查询缓存数据的有效性,而只要从广播信道上听无效报告就可以。这个方法对于移动环境很有吸引力,因为服务器不再需要知道客户的位置和连接状态,客户也不再需要建立一条上行链路来判断缓存的有效性。

5 讨论

通过对现有研究成果的分析,我们看到在移动支持技术方面已有的研究还很初步。

现有的几种有代表性的移动环境计算模型大多仅仅针对移动设备的断接、弱连接的问题。一旦客户移动,网络拓扑发生变化,便不能再给予有效支持。另一个问题是这些模型大多是静态的,代理需要预先安装和设置,限制了它们满足动态移动客户需要的能力,考虑到移动代理(Mobile Agent)技术的特征,我们认为将移动代理的概念引入移动计算模型,对支持客户移动,满足移动客户的动态需要,实现系统的适应性,是一个有效的解决途径。

尽管已有的缓存和预取方面的研究比较多,但大多面向传统分布式系统和共享存储系统在移动环境下这方面的研究还很初步。尤其对于移动环境的动态变化的特征,还缺乏相应具有适应性的预取缓存策略的研究。具有适应性的预取缓存策略应该能很好地根据网络条件的变化,作出相应的变化,尽量利用有线网络下的高带宽进行预取,在无线不可靠网络下仅实现必要的缓存。

数据库方面预取的研究不多,而已有的策略也并不理想,比如利用空间局部性,或是简单通过对用户过去访问模式的匹配进行预取,这些策略的粒度多是页或对象。考虑到现实世界中用户对数据库的访问通常是基于谓词的查询,因此预取以谓词查询为单位更为合适。现有的缓存也有基于谓词查询的,但都完全由用户指定,缺乏系统支持的谓词查询预取的研究。

缓存替换策略方面,已有的研究多是基于传统

LRU 的改进,也就是根据该数据以往使用的情况决定是否被替换。基于谓词查询的缓存为缓存引入了语义信息,利用语义信息可以设计更加有效的缓存替换策略。

移动环境下缓存一致性方面的研究,大多数以对象为一致性粒度,以谓词查询为缓存一致性粒度的研究应该考虑怎样使用尽量少的网络开销,保证缓存谓词查询的完整性以及与服务方数据的一致性,因为移动环境的资源有限,客户很可能允许缓存数据与服务方数据有一定偏差。可以考虑利用这允许的偏差有效地节省网络开销,相应设计缓存一致性策略。信息站这一类服务器具有上下行带宽不对称的特征,在带宽不对称的情况下,一致性策略应尽量节省用户的网络发送,即节省上行网络链接,不惜多用下行网络链接。

结束语 传统分布计算的研究都是假设基于有线网络和固定主机,在移动计算环境中,由于移动设备和无线网络的引入,上述假设条件不再成立。移动计算环境所独具的特点:移动设备移动性,电池能量的有限性,以及网络带宽的有限性,使传统的分布处理技术不再或低效支持移动计算。移动环境给人们的研究工作提出了新的挑战。本文在提出移动支持技术的研究问题并对相关工作进行了回顾的基础上,就现有研究存在的不足,讨论了所需要进行的工作。

参考文献

- 1 Zenel B, Duchamp D. General Purpose Proxies, Solved and Unsolved Problems. In: Proc. of the Hot-OS VI, 1997
- 2 Joseph A D, et al. Rover: A Toolkit for Mobile Information Access. In: Proc. of the Fifteenth Symposium on Operating Systems, Dec. 1995
- 3 Dunham M H, Helal A. Mobile Computing and Database: Anything New? ACM SIGMOD Record, 1995, 24(4): 5~9
- 4 Housel B C, Samaras G, Lindquist D B. WebExpress: A Client/Intercept Based System for Optimizing Web Browsing in a Wireless Environment. ACM/Baltzer Mobile Networking and Applications(MONET)1997. Special Issue on Mobile Networking on the Internet
- 5 Patterson H, Gibson G, Stryanarayanan M. A Status Report on Research in Transparent Informed Prefetching. SIGOPS, Operating Systems review, 1993, 27(2): 21~34
- 6 Tait C D. A File System for Mobile Computing. [PhD thesis]. Columbia University, 1993
- 7 Kuenning G H. Seer: Predictive File Hoarding for Distributed Mobile Operation. [PhD thesis]. UCLA, 1997
- 8 Dar S, et al. Semantic Data Caching and Replacement. In: Proc. of 22nd VLDB Conf. 1996
- 9 Chan B, et al. MODEC: A Multi-Granularity Mobile Object-Oriented Database Caching Mechanism, Prototype and Performance. Distributed and Parallel Database, 1999, 7: 343~372
- 10 Alonso R, et al. Data Caching Issues in an Information Retrieval System. ACM Transactions on Database Systems, 1990, 15(3): 359~384
- 11 Barbara D, Imielinski T. Sleepers and Workaholics: Caching Strategies for Mobile Environments. In: Proc. of the ACM SIGMOD Conf. on Management of Data, 1994: 1~12