移动环境下数据广播中的混合索引技术*

李国徽，唐向红
(华中科技大学 计算机科学与技术学院，湖北 武汉 430074)

摘　要：为了提高移动客户机的电能使用效率，提出结合两种常用的索引技术——树索引和哈希索引的复合性索引模型。该模型既减少了移动客户机的电量消耗，延长了使用时间，同时又保证了访问时间没有大幅增加。最后用 JavaSIM 仿真工具对混合索引模型、树索引模型和哈希索引模型进行了仿真研究。仿真结果表明，在访问时间、调谐时间、索引效率等方面，混合索引模型具有较好的性能和实用价值，达到了设计目的。

关键词：移动计算；索引；谐调时间；访问时间；仿真
中图分类号：TP311.13　　**文献标志码**：A　　**文章编号**：1001-3695(2007)06-0315-03

Hybrid Indexing Technique for Data Broadcast under Mobile Enviroment

LIU Guo-hui, TANG Xiang-hong
(College of Computer Science & Technology, Huazhong University of Science & Technology, Wuhan Hubei 430074, China)

Abstract: This paper proposed a new hybrid data indexing model taking advantages of tree index and Hash index, which would improve the efficiency of power in mobile clients. This model not only declined the consumption of power in mobile clients, but also didn't improve access time. The result of simulation shows that the model has high efficiency and practical value at aspects of access time, tune-in time and indexing efficiency, compared to tree index and Hash index.

Key words: mobile computing; indexing; tune-in time; access time; simulation

0　引言

随着移动设备和通信设备的发展，移动计算越来越普及。人们可以在移动设备上如手提电脑、手机、PDA 等通过无线网络来存储数据。然而，无线网络的带宽有限且具有非对称性(下行带宽远大于上行带宽)。数据广播技术是针对移动环境中无线网络特点的一种有效数据访问方式，以较小的代价来支持大量移动设备并行访问数据库[1]。

从广播信道中有效地访问数据有两个关键的要求，即节约电能和尽量减少移动客户端的等待时间。在基于数据广播的系统中，移动客户端必须等待到服务器广播它所需要的数据时才能获得所需要的数据；而客户端的等待时间取决于广播数据的长度。广播数据的长度又叫做广播周期；广播周期越长，数据访问时间也就越长。

在无线环境中，移动设备所配备的电量是有限的。研究表明，移动设备的电量大部分是消耗在监听无线信道和检查所接收的数据包上[2]。在文献[3]中，有选择地监听概念的提出就是降低基于数据广播系统的电量消耗。通过有选择性地监听，移动客户端可以在大部分时间处于休眠状态，仅仅当它想要的数据来到时才被激活。索引技术就是在无线环境下用来实现选择性监听的。索引与数据一起广播帮助移动客户端定位请求的数据。移动客户端大部分时间处于休眠状态，因此极大地降低了电量消耗。同时，带索引的数据广播增加了广播周期的长度，这导致了移动客户端的等待时间的延长。在实际应用中，不同的应用对客户端的等待时间和电量消耗的要求各不一样。

近年来，为了改善数据的访问性能，提出了不少数据访问方法。大部分方法都是基于下列三种技术[3]：树索引、签名索引[4]和哈希索引。一些混合数据访问方式同时被提出。评价数据访问方式的优劣有如下三个标准：

(1) 访问时间(Access Time)。它是指从客户提出访问请求，到结果返回所需的时间。访问时间决定了用户访问数据的响应时间。

(2) 谐调时间(Tune-in Time)。它是指在完成一个访问请求期间，移动设备保持接听广播的总时间。调谐时间决定了移动设备的电源消耗。因为在不接听广播的时间里，移动设备可以转入休眠模式；此时移动设备消耗的电源相对于启动状态可以忽略不计。

(3) 索引效率(Indexing Efficiency)。它是指每增加一单位的访问时间能够节约多少谐调时间。它与访问时间和谐调时间都相关。它是用来刻画在维持可接受访问时间增加的情况下把谐调时间减低到最小。

总之，一种节约电能的索引技术必须平衡索引代价(访问时间的增加)和谐调时间的节约，以达到索引效率的最大化。

1　常见的索引技术

数据广播中一般采用简单的树索引结构和哈希索引结构。树索引结构能大幅减少谐调时间但同时也增加了数据的访问

收稿日期：2006-04-04；修返日期：2006-06-28　　基金项目：国家自然科学基金资助项目(60203017)；教育部留学回国基金资助项目

作者简介：李国徽(1973-)，男，湖南人，教授，博导，博士，主要研究方向为现代数据库技术、实时计算、移动计算；唐向红(1979-)，男，湖南人，硕士研究生，主要研究方向为现代数据库技术(txhwuhan@163.com).

时间;而且数据广播的组织形式对其性能影响较大。哈希索引结构虽然减少谐调时间的性能没有树索引好,但它几乎不增加数据的访问时间;同时它的适应能力强,其性能受数据广播的组织形式比较少。

1.1 树索引技术[5]

树索引是一种比较传统的数据索引技术。树索引技术在无线数据广播中,数据帧到达时间代替了数据记录的物理存储位置。访问采用树索引的数据广播分为下面几步:

(1) 初始探测。客户机监听广播信道,获得广播中下一个树索引的到达时间。

(2) 搜索。客户机顺着一系列指针找到想要帧的到达时间。遍历的指针数目等于索引树的高度。

(3) 检索。客户机监听信道下载其请求的所有数据帧。

图 1 描述了在一个广播周期中具有 81 个节点的索引树例子。最下层的方块代表几个数据帧的集合。它分为两部分,一部分为可重复部分,一部分为非重复部分。

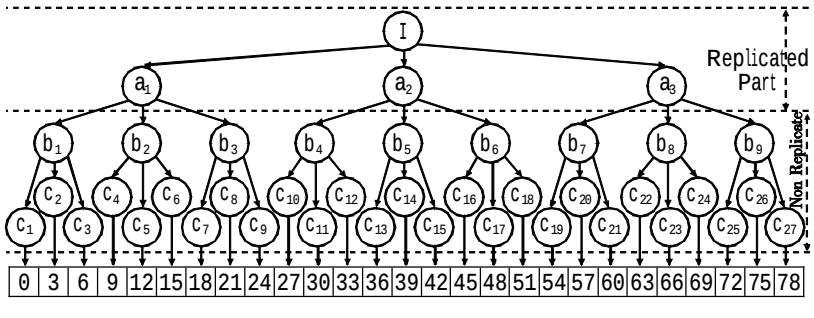

图 1 索引树结构

1.2 哈希(Hash)索引技术

哈希方法广泛用于信息检索。一个数据帧的哈希值是数据帧的关键字通过一个特定的哈希函数运算得来的。哈希函数的选择对哈希索引技术性能来说至关重要。哈希索引技术一般包括以下步骤:

(1) 初始监听。客户机监听广播信道,获得第一个哈希值。

(2) 过滤。客户机访问连续的哈希值和数据帧,找到想要的数据。一般来说,它要花广播的半个周期去找到它想要的第一个数据帧。

(3) 检索。客户机监听信道得到连续想要的数据帧。

哈希值的长短与哈希函数的平均失效率对谐调时间和访问时间都有影响。

2 混合索引的设计

树索引技术和哈希索引技术都有其优缺点。比如,树索引技术适合随机数据访问;哈希索引技术适合顺序结构数据,类似广播信道。树索引技术对簇集的数据广播非常有效;但簇集对哈希索引技术性能影响不大。哈希索引技术特别适合多属性的数据索引;树索引技术提供了一种基于索引值较准确和完整的全局视图,客户机能快速地在树索引上找到想得到的数据的到达时间,这样,谐调时间自然就缩短了。由于哈希索引不包含数据帧的全局信息,它仅仅只能对客户机判定当前数据帧是否与查询有关提供帮助。其过滤的有效性在很大程度上取决于哈希索引的平均失效率。

下面提出一种称之为混合索引的新型索引方法。该方法结合了树索引和哈希索引两种索引技术的优点。广播索引结构如图 2 所示。

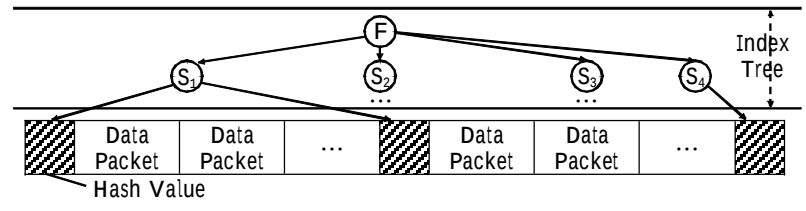

图 2 混合索引的逻辑结构图

这个结构的上层是一个索引树可重复部分的结构。最下层是一个哈希索引结构。

客户机能通过搜索索引树获取想要数据帧的大概位置信息。索引树高度一般不高,因此检索索引树的代价不高。由于混合索引技术是建立在哈希索引技术之上的,它保留了哈希索引技术的所有优势;同时,索引树提供数据的全局信息,进一步减少了谐调时间。这种新的索引技术访问数据分以下几步:

(1) 初始探测。客户机监听信道判定下一棵索引树的到达时间。

(2) 搜索。根据接收到的索引树,客户机得到自己想得到的数据帧的最近监听时间。

(3) 过滤。在最近的位置,通过哈希值进行过滤,直到找到想要的数据帧。

(4) 检索。客户机监听信道下载所需要的数据帧。

2.1 混合索引模型代价分析

为了分析这种新型的索引技术,定义了表 1 所示的参数表。

表 1 分析模型的参数表

参数	含 义
D	数据广播中数据帧的数量
N	整个数据广播中数据包的数量
P	一个数据帧中包含的数据包数量
A	访问时间
T	谐调时间
B	哈希函数过滤的失效率
Tp	访问一个数据包所需的时间
H	索引树的高度
K	索引树的扇出度
S	所要检索数据包的数量
Q	被哈希值所占的数据包数量

假定一个数据帧中包含 P 个数据包,一个索引树节点只占用一个数据包。同时在每个数据帧头部都有一个哈希值,所检索的数据包均聚集在一起。设 X 为索引树的全部节点数,根据混合索引的索引结构可知:

$$D = N/P$$

$$H = \lceil \ln D/\ln K$$

$$X = K^{H+1} - 1$$

谐调时间包括检索索引树的时间和哈希过滤的时间。其中检索索引树的时间为$(H+1) \times T_p$;哈希过滤时间为 $1/2B \times P \times T_p + (1-B) \times Q \times T_p$。所以

$$T = (H+1) \times T_p + 1/2B \times P \times T_p + (1-B) \times Q \times T_p$$

检索 S 个连续数据包的访问时间包括谐调时间、等待 S 个数据包中第一个数据包到达的时间和从数据广播中获取连续 S 个数据包的时间。其中等待 S 个数据包中第一个数据包到达的时间为 $(N/2) \times T_p$;从数据广播中获取连续 S 个数据包的时间为 $S \times T_p$。所以 $A = T + (N/2 + S) \times T_p$。

2.2　仿真实验

为进一步测试新型混合索引的性能, 本文用 Java 语言的一个仿真软件包 JavaSIM[6, 7] 进行了仿真实验。仿真参数如表 2 所示。

表 2　模拟比较环境参数

参　　数	含　　义
数据源	database
数据类型	文本
数据包数量	5 000 ～5 000 000
数据包的长度	800 Bytes
哈希值所占数据包数量	2
请求数量	> 40 000
数据帧中的数据包数量	5
数据访问方式	Zipf

对在这种环境下的几种数据广播索引技术进行仿真比较。仿真实验结果如图 3 所示。

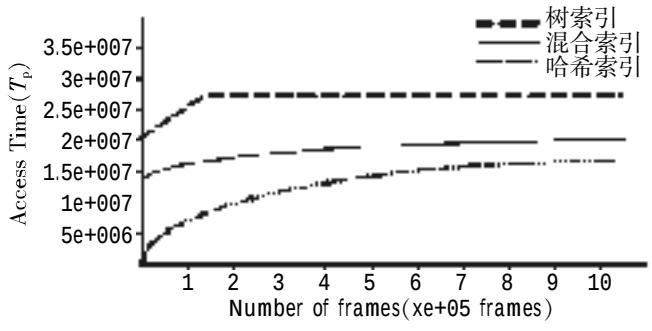

图 3　访问时间比较

从图 3 的访问时间看, 随着数据帧数量的增大, 每一种索引技术的访问时间都会增加。在增加的幅度上, 树索引的增加最快, 混合索引居中。与以前的分析很拟合, 在访问时间上哈希索引的性能最好, 混合索引居中, 树索引性能最差。

从图 4 可知, 随着数据帧数量的增加, 树索引和混合索引在谐调时间上趋于稳定, 哈希索引有所增加; 在谐调时间上哈希索引的性能最差, 混合索引居中, 树索引性能最好。

从图 5 可知, 在数据帧数量较小时, 哈希索引的效率最高。随着数据帧的增加, 索引树和混合索引的效率逐渐提高, 哈希索引效率趋于稳定。在数据帧数量进一步增大时, 混合索引显示出了它的优势。

3　结束语

本文提出的混合索引方法达到了设计目的, 在一定程度上减少了谐调时间, 具有很高的索引效率和一定的实用性。但该索引方法还有改进的地方, 如能否降低索引树的高度、在哈希方法时采用多级哈希函数。这些新的思路还要有待研究。

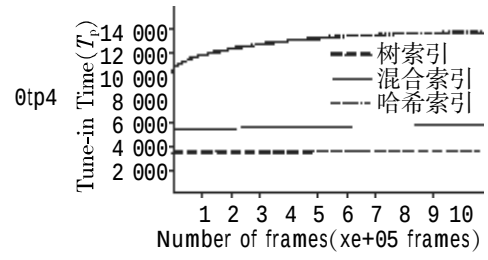

图 4　谐调时间比较

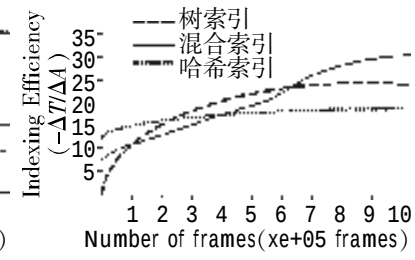

图 5　索引效率的比较

参考文献:

[1] SAI Chengfa, TSAI Chunwei. A new approach for solving large traveling salesman problem using evolutionary ant rules: proc. of International Joint Conference on Neural Network[C]. New Jersey: IEEE Press, 2002: 1540-1545.

[2] LEE D, HU Q, LEE W C. Power conserving and access efficient indexs for wireless computing[M]. Boston: Kluwer Academic Publisher, 2000: 249-264.

[3] IMIELINSKI T, VISWANATHAN S, BADRINATH B R. Energy efficient indexing on air: proc. of the 4th International Conference on Extending Database Technology[C]. [S. l.]: [s. n.], 1994: 245-258.

[4] LEE W C, LEE D L. Using signature techniques for information filtering in wireless and mobile environments[J]. Distributed and Parallel Database Journal, 1996, 4(3): 205-227.

[5] IMIELINSKI T, VISWANATHAN S, BADRINATH B R. Data on the air organization and access[J]. IEEE Transactions of Data and Knowledge Engineering, 1997, 9(3): 353-372.

[6] PARRINGTON G D, SHRIVASTAVA S K, WHEATER S M, *et al*. The design and implementation of arjuna[J]. Computing Systmes, 1995, 8(2): 255-308.

[7] LITTLE M C, MCCUE D L. Construction and use of a simulation package in C++[R]. UK: University of Newcastle upon Type, 1993: 437-454.

(上接第 314 页)

4　结束语

视频监控管理平台主要为适应目前监控网络的不断扩大及其日益复杂的功能需求而设计。其重点在于对视频监控系统的有效管理。

本文结合视频监控系统的特点和管理平台的功能需求, 提出了一套完整的视频监控系统管理平台的设计方案, 较好地实现了这些要求。随着网络的发展, 带宽资源越来越丰富, 通过网络进行高清晰视频已经成为可能。因此采用先进视频编码技术的高清晰视频监控就成为未来监控系统的需求。但是对于基于 Internet 等低带宽环境下的视频监控需求, 需要系统提供配套的低带宽监控方式, 以满足不同带宽的要求。将高清晰视频和低带宽的视频监控相结合将是视频监控管理平台改进的一个方向。

参考文献:

[1] 万加富, 张文斐, 张占松. 网络监控系统原理与应用[M]. 北京: 机械工业出版社, 2003: 15-211.

[2] 赵慧玲, 叶华. 以软交换为核心的下一代网络技术[M]. 北京: 人民邮电出版社, 2002: 36-76.

[3] 谢俊元, 刘样, 徐曼. IP 组播技术在远程视频监控系统中的应用[J]. 计算机应用研究, 2002, 19(2): 108-110.

[4] STEVENS W R. UNIX 网络编程[M]. 第 1 卷. 北京: 清华大学出版社, 1999: 112-373.

[5] IETF. RFC 3261, SIP: session initiation protoco1[S]. [S. l.]: [s. n.], 2002.

[6] MOIZARD A. oSIP2 document[EB/OL]. http://www.gnu.org/software/osip/doc/html/index.html.

[7] 毋河海. 地图数据库系统[M]. 北京: 测绘出版社, 1996: 56-167.

[8] 陈元琰, 邓宗明, 张睿哲, 等. Visual C++ 编程实用技术与案例[M]. 北京: 清华大学出版社, 2001: 75-342.