

无线环境中的数据广播调度理论分析及算法研究

余 平

(华东师范大学上海数字化教育装备工程技术研究中心 上海 200062)

摘 要 数据广播是无线环境中一种独特的数据发布方式。首先系统地归纳了周期广播、on-demand 广播和混合广播 3 种广播方式的理论分析模型及其特点,然后针对单信道和多信道对各种广播方式的调度算法进行了分类、比较,并突出了理论模型对调度算法研究的指导价值。最后指出今后的研究方向。

关键词 无线环境,数据广播,调度,多信道,排队论

中图分类号 TP393 **文献标识码** A

Survey of Data Broadcast Scheduling in Wireless Environments

YU Ping

(Shanghai Engineering Research Center of Digital Education Equipment, East China Normal University, Shanghai 200062, China)

Abstract Data broadcast is an efficient way of data dissemination in wireless environments. In this paper the theoretical analysis was provided first for the three data broadcast mode (periodic broadcast, on-demand broadcast and hybrid broadcast). Then current data scheduling algorithms were classified and compared. Some typical multi-channel scheduling algorithms were analyzed in detail. Finally we pointed out some research areas for data broadcast scheduling.

Keywords Wireless environments, Data broadcast, Scheduling, Multi-channel, Queuing theory

随着无线网络技术的迅速发展和无线移动设备的日趋普及,人们可访问信息的种类越来越多,途径越来越灵活,期望访问效率也能够越来越高。但与有线网络相比,无线网络环境具有低带宽、频繁断接性、带宽多样性、网络通信的非对称性、无线设备的电源能力较低和可伸缩性等特点^[1-4],如何在无线环境中进行有效的数据管理并提供高效的数据服务,一直是学术界和工业界极为关注的研究课题。

数据广播(data broadcast)是无线环境中向大量用户提供高效数据服务的一种新的数据发送(data dissemination)方式。与点对点(point-to-point)发送方式不同,数据广播无需在收发双方建立一对一的逻辑通道,而是利用无线移动网络固有的广播机制,不断地将信息发送给广播区域中的全部用户,用户侦听广播信道并从中下载需要的信息。发送一次数据就可以同时满足多个对该数据的请求,这种数据发送方式使得系统的可伸缩性极好,用户数量的增加不会提高系统的开销,可以高效地利用有限的无线带宽资源;同时在无线通信中,接收数据比发送数据消耗的电池能量要小得多^[5],因此数据广播也是一种节省能量的数据访问方式。

本文第 1 节介绍数据广播的一般概念、特点和主要性能指标;第 2 节系统地归纳了几种数据广播调度方式的特点及其理论基础;第 3 节分别针对 3 类数据广播的主要调度方法进行了分析与比较;最后指出数据广播调度领域还存在的问题和研究方向。

1 数据广播概述

在数据广播系统中,基站服务器连续地将数据在无线信

道上发送,用户使用计算机或其他手持设备侦听信道,等待接收感兴趣的内容(见图 1)。根据数据广播过程中是否需要用户发送请求,数据广播可以分为 3 种方式^[6,7]:

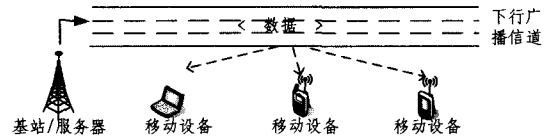


图 1 数据广播系统示意图

(1)周期广播(periodic broadcast)。服务器在下行广播信道(downlink broadcast channel)上主动对数据进行连续的、周期性的发布,用户不需要提出请求,直接侦听广播信道,从中接收需要的数据,这种方式称为基于推的广播(push-based broadcast)。由于这种广播通常具有周期性,因此也称为周期广播。

(2)on-demand 广播。用户可以通过上行信道(uplink channel)发送请求,服务器根据用户请求决定要发布的内容,这种方式称为基于拉的广播(pull-based broadcast),也称为 on-demand 广播。具体的数据发布方式可以是只对请求数据的用户发送(这种方式是点对点的请求响应,严格说并不是数据广播,本文将其一起列入数据广播范畴进行讨论),也可以将用户请求的数据通过广播方式向所有用户发送。

(3)混合广播(hybrid broadcast)。这是基于推和拉的方式组合起来的一种新的广播方式。对于用户访问频繁的数据,不需要用户提交请求,通过周期广播向所有用户发布;而用户个别请求的数据,则需要用户发出显式的请求,服务器再根据用户请求选择发送的数据。

到稿日期:2010-10-26 返修日期:2011-02-26

余 平(1972-),女,博士,助理研究员,主要研究方向为数据库与知识库、移动数据管理,E-mail:yup002002@gmail.com.

数据广播系统的主要性能衡量指标包括:

(1)访问时间(Access time, AT):从用户发出请求到得到返回结果的时间。访问时间由两部分组成:访问等待时间(access latency, AL)和数据下载时间。其中,访问等待时间是从用户发出请求到开始下载数据的时间。访问时间小,表示用于获取数据的响应时间小。

(2)调谐时间(Tune-in Time, TT):在一次数据访问的过程中,移动用户处于活动状态、从信道上接收信息(下载信息)的时间。调谐时间小,表示系统用于接收数据消耗的能量小。

由于对数据的一次广播可以同时满足对该数据的全部请求,因此数据广播最主要的优点是可以适合任意数量的用户,可伸缩性极强。但数据广播也存在不利于用户访问的因素,即广播的线性访问特性(linear access property),用户只能对广播信道中的数据进行顺序访问。当所需的数据项尚未广播时,用户需要等待,直到数据出现在广播信道上。而一旦用户错过了数据的一次广播,只能等待该数据的下一次广播才能访问到数据。线性访问特性成为影响数据广播性能的主要因素。如何获得更少的访问时间,是数据广播研究的最基本问题之一,也是数据广播调度(data broadcast scheduling)研究的核心内容。

2 数据广播调度的特点及理论分析

数据广播调度就是确定在一个广播过程中各个数据项的发送次序和频率,即服务器何时广播什么数据,使得用户访问时间最少。针对不同的数据广播方式,广播调度采用不同的理论分析模型。

2.1 周期广播调度

在周期广播中,服务器根据用户对数据的访问概率,主动地、连续地、周期性地将数据“推”到广播信道中,用户调谐到信道上进行侦听、等待,直到用户需要访问的数据在广播信道上出现,再将该数据项下载下来。服务器根据预先知道的用户访问统计信息决定广播的数据项和各数据项在广播中的次序。周期广播调度研究的目的是构造一个最优调度序列,使得平均访问时间最少。

周期广播具有以下优点:(1)可伸缩性强,可以同时将数据发布给大量用户。(2)充分地利用了有限的带宽。周期广播减少了不必要的下行数据发送(对于同一数据项的请求,发送一次就可以满足大量用户的请求)和上行数据发送(用户不必向服务器发送请求),实现了对带宽的高效利用。(3)节省能量。由于避免了上行请求的发送,因此减少了移动设备的能量消耗。

周期广播的不足:(1)不能快速响应用户请求。由于广播内容不是针对每个用户的特定请求进行发布的,因此某些用户可能需要等待比较长的时间才能从广播信道上访问到需要的数据。(2)存在一定的带宽浪费。由于广播内容不是根据当时的用户请求确定的,因此某些广播内容可能不是当前用户需要访问的,但也占用带宽发送,从这个角度来说,周期广播存在一定的带宽浪费。(3)不适用对大数据量的广播。广播数据量太大将导致平均访问时间的增大。(4)对用户访问模式的变化不能迅速响应。

假设数据项集 $D = \{d_1, d_2, \dots, d_N\}$ 表示数据库中需要广播的数据项, N 为数据项个数,数据项 d_i 的访问概率为 p_i , 长度为 l_i , 在信道上的广播频率为 f_i 。

2.1.1 单信道周期广播调度的理论分析

广播周期(broadcast cycle)是一个广播信道上所有数据

项至少广播一次的广播序列,该序列在广播信道上周期地进行广播,序列长度表示为 L 。数据项的广播实例(instance)是数据项在广播信道上的一次发送。数据项相邻两个实例之间的间隔(spacing)是从一个实例开始广播到下一个实例开始广播之间的距离。例如在图2中数据项 d_1 的两个实例之间的间隔等于 $l_1 + l_2 + l_3$ 。

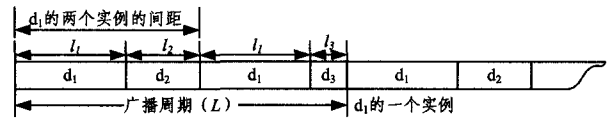


图2 广播周期的一个片段

定理1^[8,9](instance equally spacing,实例等间距规则)

当数据项的每对相邻实例之间间隔相等时,对该数据项的平均AL最小。

定理2^[8,9](square root rule,平方根规则) 给定数据项

i 的访问概率 p_i 和长度 l_i ,假设每个数据项的相邻实例都是等间距的,那么当数据项 i 的广播频率 $f_i = L \sqrt{p_i/l_i} / \sum_{j=1}^N \sqrt{p_j l_j}$ 时,平均AL取得最小值 $AL^{opt} = \frac{1}{2} (\sum_{i=1}^N \sqrt{p_i l_i})^2$ 。

定理2说明当数据项的广播频率与其访问概率的平方根成正比,且与其长度的平方根成反比时,总平均访问时间最小。定理1和定理2是单信道数据广播调度算法的重要依据。

2.1.2 多信道周期广播调度的理论分析

在多信道数据广播中,数据分布在多个信道上。假设系统中有 K 个物理广播信道集,信道 i 的带宽为 b_i ,假设 B 为系统的总信道带宽,将 N 个数据划分到 K 个信道上,信道 i 上的数据项个数为 N_i ,信道 i 上数据项的访问概率为 $p_{i1}, p_{i2}, \dots, p_{iN_i}$,数据项长度为 $l_{i1}, l_{i2}, \dots, l_{iN_i}$ 。

定理3^[10](Multi-channel square root rule,多信道平方根规则) 给定数据项 i 的访问概率 p_i 和长度 l_i ,将这 N 个数据项不重复地划分到 K 个信道上,则当每个信道 i 上的数据访问概率满足 $\sum_{j=1}^{N_i} \sqrt{p_{ij} l_{ij}} = (\sum_{i=1}^N \sqrt{p_i l_i}) b_i / B$ 时,总平均AL取得最小值 $AL_c^{opt} = (\sum_{i=1}^N \sqrt{p_i l_i})^2 / 2B$ 。

定理3说明,当每个信道上数据项的访问概率与长度乘积的平方根之和与该信道的带宽成正比时,平均AL最小。定理2和定理3是多信道广播调度算法的重要依据。

2.2 on-demand 广播调度

on-demand 广播中,服务器根据用户明确发送的请求确定要发送的数据项和次序。用户通过上行信道向服务器发送请求,请求在服务器端进行排队,服务器根据一定的策略选择要处理的请求,并决定下一个要发布的数据。服务器可以采取两种发布方式:(1)点对点方式,将数据直接发送给请求的用户,此时发送的数据只能满足单个用户的需求(这种方式严格说并不是广播方式,本文将其作为一种广义的广播进行研究);(2)广播方式,将数据在广播信道上发送,此时发送的数据可以同时满足提出相同数据请求的所有用户。on-demand 广播调度就是确定下一个要处理的请求和发布的数据,使得系统的平均访问时间最小。

on-demand 广播具有以下优点:(1)可以快速满足用户请求。服务器根据用户请求进行数据发布,比周期广播更具有针对性,可以及时满足用户请求。(2)可以节省带宽。由于发布的数据都是用户实际请求的数据,没有多余的数据在广播

信道上广播,因此节省了有限的无线信道带宽,可以发送更多的数据,满足更多用户的请求。(3)能够适应动态变化的用户访问模式。

on-demand 广播的不足:(1)不适应大量并发用户的访问。大量用户向服务器发送请求,使服务器发生阻塞,成为系统的瓶颈,大大降低了系统的访问时间。(2)需要额外的上行信道。on-demand 广播中用户需要通过上行信道发送请求,占用了宝贵的带宽资源。(3)增加了移动设备的能量消耗。移动设备发送上行请求,比被动接收数据需要消耗更多的能量。

2.2.1 点对点发布的 on-demand 广播调度理论分析

点对点发布的 on-demand 数据广播系统是一种典型的排队系统,因此通常采用排队论(Queueing Theory)^[11]对这种 on-demand 广播调度进行理论分析。

排队系统通常采用 Kendall 表示法^[11]:A/B/c 排队规则,其中 A 表示顾客到达时间间隔的分布,B 表示服务时间分布,c 表示服务器个数(即数据广播系统中的信道个数)。A 和 B 可以采用以下符号:M-指数分布,D-定长分布,G-一般分布,GI-具有独立到达时间间隔的一般分布。而队列规则是指服务器在处理队列中的请求时选择哪个请求进行服务,常用的排队规则有:FCFS(First come first served),LCFS(Last come, first served),SIRO(Service in random order),static priorities 和 dynamic priorities 等等。对于采用优先级排队规则的系统,当一个优先级更高的请求到来时,当前正在接受服务的请求可以被剥夺(preemption)或不被剥夺。排队规则缺省时采用 FCFS。例如 M/M/c 表示请求到达时间间隔和服务时间均服从指数分布,系统中有 c 个信道,采用 FCFS 规则对请求进行处理。

假设对数据项 d_i 的请求到达率(单位时间内到达的请求个数)为 λ_i ,总的请求到达率为 λ ,信道的服务率(单位时间内服务的请求个数)为 μ ,系统的负荷水平 ρ 表示请求到达率与服务率的比值。

对于 M/M/c 的 on-demand 广播系统,AL 和 AT 可以通过以下公式计算:

$$AL = \frac{\rho}{\lambda(1-\rho)} \times P_c \quad (1)$$

$$AT = AL + \frac{c\rho}{\lambda} \quad (2)$$

式中, P_c 表示到达的请求需要在队列中等待的概率,按照以下公式计算:

$$P_c = \frac{(c\rho)^c}{c! (1-\rho)} \times \left[\sum_{k=0}^{c-1} \frac{(c\rho)^k}{k!} + \frac{(c\rho)^c}{c!} \times \frac{1}{1-\rho} \right]^{-1} \quad (3)$$

由于此公式计算复杂,也可以按照以下公式估算:

$$P_c \approx \begin{cases} \frac{\rho + \rho}{2}, \rho > 0.7 \\ \rho^{\frac{c+1}{2}}, \rho \leq 0.7 \end{cases} \quad (4)$$

ρ 的取值如下:对于单信道系统, $\rho = \frac{\lambda}{\mu}$;对于等带宽的多信道系统, $\rho = \frac{\lambda}{c\mu}$;而对于不等带宽的多信道系统,假设信道 i 的服务率为 μ_i ,则 $\rho \approx \frac{\lambda}{\sum_{i=1}^c \mu_i}$ 。

对于更一般情况 GI/G/c 的 on-demand 广播系统,AL 可以通过以下公式估算,其中 c_A 是请求到达时间间隔的变异系

数(coefficient of variation), c_B 是服务时间的变异系数, P_c 按照前面的公式计算(或估算):

$$AL \approx \frac{P_c/\mu}{1-\rho} \times \frac{c_A^2 + c_B^2}{2c} \quad (\text{Allen-Cunneen formula}) \quad (5)$$

在点对点发布的 on-demand 广播中,为了防止服务器阻塞,使系统稳定运行,要求系统负荷水平 ρ 小于 1。

2.2.2 广播发布的 on-demand 广播调度理论分析

采用广播发布的 on-demand 数据广播系统,服务器对一个请求的响应可能同时满足了多个用户的请求,因此不能简单套用排队论的公式来计算平均 AL,目前尚未见针对这一类 on-demand 数据广播的理论分析。但由于服务器对一个请求的响应可以同时满足多个用户的请求,从另一个角度来看服务器的服务时间缩短了,因此单位时间内可以处理更多的请求。当系统中重复访问数据项的请求较多时,系统负荷水平 ρ 可以大于 1。

2.3 混合调度(hybrid scheduling)

鉴于周期广播和 on-demand 广播各具不足,将二者结合起来,可以弥补各自的不足。设计混合广播调度算法需要考虑 3 个问题:(1)如何分配周期广播和 on-demand 广播所需的信道带宽。(2)如何确定数据的广播模式,即确定哪些数据通过周期广播模式进行广播,哪些数据通过 on-demand 模式进行广播。以上两点是服务器需要考虑的问题。(3)对于用户来说,通过哪种方式从不同类型的信道上访问数据。

假设周期广播的带宽为 B_p , N_p 为周期广播的数据项个数,数据项 d_1, d_2, \dots, d_{N_p} 被分配到周期广播信道上,其他数据通过 on-demand 方式进行访问,那么混合调度问题就是要确定 B_p 和 N_p ,使得系统平均 AT 最小。令 AL_p 表示周期广播的平均 AL, AL_o 表示 on-demand 广播的平均 AL(二者可以通过 2.1 和 2.2 节中的公式计算),系统平均 AL 为:

$$AL = \sum_{i=1}^{N_p} p_i \times AL_p + (1 - \sum_{i=1}^{N_p} p_i) \times AL_o \quad (6)$$

其取得最小值时的 B_p 和 N_p 就是最优的带宽和数据的分配方法。当 B_p 和 N_p 确定后,再采用多信道周期广播方法将 N_p 个数据分配到多个周期广播信道。

3 数据广播调度方法

基于数据广播调度的理论分析,可以将调度方法分为两类,一类是启发式方法,另一类则完全根据调度理论(如平方根规则和排队论)而设计的方法。

3.1 周期广播调度

平坦调度(flat broadcast)是最简单的广播调度方法,所有数据周期地在广播信道上进行发送,而且在一个广播周期里每个数据项出现的频率是相同的。更多的研究在非平坦广播调度方面(non-flat scheduling 或 skewed scheduling),假设用户对于数据项的访问概率是偏斜分布的,根据数据项的访问概率确定其广播频率。

3.1.1 单信道数据广播

根据定理 2, Wong 设计了一种等长数据项的非平坦调度方法^[8]:给定广播周期长度和每个数据项的访问概率,首先计算每个数据项的广播频率,然后计算每个数据项的不同实例之间的间距,再根据数据项的广播频率和间距将数据项分配到广播信道的不同位置,尽量保证同一数据项的相邻实例等间距。这一算法完全遵循平方根规则设计,理论上可以获得最小平均访问时间,但由于存在取整的误差,不能保证实例等间距,使得该算法只能获得近似最优的平均访问时间。

Vaidya 和 Hameed 从平方根规则得出, 当 $s_i^2 \frac{p_i}{l_i}$ 为常量时, 平均访问时间最小, 并以此为出发点提出一种 on-line 算法^[9]。之后他们又发现广播调度问题类似于 packet fair queuing 问题, 因此提出采用 packet fair queuing 算法进行不等长数据项的广播调度^[12]。

Acharya 等在文献[13]中提出一种 broadcast disks 方法, 文献[14]提出的改进方法则属于启发式方法。

3.1.2 多信道数据广播

多信道数据广播调度算法中启发式方法主要包括基于“温度”的调度方法^[15]、基于数据访问概率的贪心调度算法^[16]以及基于访问概率的动态规划算法^[17]。

文献[10]等提出的多信道广播调度方法则完全基于多信道平方根规则。该算法分为两步:(1)数据划分。将 $\sqrt{p_i l_i}$ 看成是数据项的“温度”, 参考基于“温度”的调度方法将数据分配到不同信道上, 使得每个信道的温度与其信道带宽成正比。如果每个信道带宽相等, 则分配后每个信道温度相等。(2)单信道上的非平坦调度。将数据项分配到每个信道后, 再根据单信道平方根规则对每个信道上的数据项进行调度。当数据量较大或访问概率偏斜度较大时, 该方法的平均访问时间接近理论最优值^[10]。

文献[18]在数据划分后, 采用 Vaidya 和 Hameed 的调度算法, 可以对不等长数据项进行调度。文献[19, 20]在数据划分后, 在单信道上采用了类似 broadcast disks 的启发式方法, 使得每个信道上数据项的实例等间距, 这样做的代价是广播周期很长。

3.2 on-demand 广播调度

3.2.1 单信道数据广播

对于等长数据项, on-demand 广播调度的基本排队规则可以采用 FCFS, MRF (most request first), MRFL (MRF Low) 和 LWF (longest wait first)^[21] 等, 其中除了 FCFS 是经典排队论中的排队规则, 其他 3 种都考虑了对数据项重复请求的问题。除了考虑单一因素(如请求个数、等待时间等)对调度规则的影响, 一些启发式调度方法综合考虑多种因素, 使得系统在调度开销和性能方面得到平衡。

文献[22]提出了一种 $R \times W$ 方法, 同时考虑了等待数据项的请求数(R)和等待时间最长的请求已经等待的时间(W)。在每个广播时刻, 选择 $R \times W$ 值最大的数据项进行广播。为进一步减小算法开销, 文中还提出了一种参数化 $R \times W(\theta)$, 通过 θ 的变化, 在调度开销和系统性能(平均访问时间)之间进行折衷。 $R \times W$ 方法具有 FCFS 和 MRF 的优点, 并且几乎与 LWF 具有相同的性能。

文献[23]提出了一种参数化算法 priority index policies (PIP- γ), 每个广播时刻, PIP- γ 选择 $\lambda_i^{-\gamma} X_i$ 值最大的数据项进行广播, 其中 λ_i 是对数据项 i 的请求的到达率, X_i 是等待数据项 i 的请求个数, γ 是一个权重参数。文中的实验结果表明 PIP-0.5 可以获得 LWF 相似的性能, 但开销却小得多。

文献[24]综合考虑了访问时间、调谐时间和出错请求的代价, 提出了一种自适应的基于代价的有效 on-demand 调度算法。在调度时, 将计算每个数据项的代价作为调度的优先级。

对于不等长数据项调度, 文献[25]提出了一个新的性能指标 stretch(一个请求的访问时间与其服务时间的比值), 并以此为基础提出了一组基于可剥夺策略(preemptive)的调度

算法: PLWF (Preemptive Longest Wait First), SRTF (Shortest Remaining Time First), LTSF (Longest Total Stretch First) 和 MAX 算法。文中的实验结果表明, 上述算法没有一个在任何时候都表现最好, PLWF 的最坏访问时间最优, LTSF 的平均访问时间最优, SRTF 在比较平均 stretch 时最优, 而 MAX 算法在同时比较平均访问时间和 stretch 两个指标时表现更好。

3.2.2 多信道数据广播

多信道 on-demand 广播调度到目前为止尚未发现更多的启发式方法提出, 一般与周期广播结合起来进行混合调度。为了能够计算出带宽和数据项的最优划分, 目前的方法都采用经典的排队论来计算多信道 on-demand 广播的平均访问时间, 排队规则都默认为 FCFS。

文献[26]根据系统负荷采用不同的队列模型分析响应时间。假设系统中有 c 个广播信道, 当用户请求数目比较少时, 系统处于轻负荷, 采用 $M/M/c$ 队列模型; 而当用户请求数目很多时(超过服务器处理的能力), 系统处于重负荷, 则采用 $M/M/c/m$ 模型(m 为请求数据项的用户个数), 以确保响应时间不会无限增大。考虑到定长数据项广播时间是一个确定的数值, 文献[27]采用 $M/D/c$ 队列模型。当数据项不等长时, 服务时间不能确定符合什么分布规律, 文献[28]提出采用 $M/G/c$ 队列模型分析访问时间。

3.3 混合广播调度

混合广播调度由于 on-demand 广播的存在, 必然存在一个上行信道, 用户通过它向服务器发送请求, 因此都属于多信道混合调度。

多信道混合调度可以分为以下几个步骤:(1)信道分配。将系统中的 K 个信道进行功能分配, 确定哪些信道用于周期广播, 哪些信道用于 on-demand 广播。(2)数据项分配。将数据项分配到不同功能的信道上, 即确定哪些数据通过周期广播, 哪些数据采用 on-demand 方式广播。对于周期广播的数据, 还需要确定数据项在多个周期广播信道上分配的方法。

文献[29]首次提出了混合调度的解决方法, 为用户提供一条 backchannel。通过该信道发送请求, 而下行广播信道只有一条, 用来交替发送周期广播的数据项和响应用户请求的数据项(即 on-demand 数据项), 因此没有涉及到信道分配的问题。文献[26-28, 30, 31]等则讨论了多信道混合广播中信道分配和数据分配的问题。

关于信道分配和数据项分配的主要思路就是求式(6)的最小值及取得最小值时的周期广播信道个数和周期广播数据项。

在进行静态信道分配时, 采用的主要方法有穷举法^[26, 31]和有限状态自动机法^[27, 28]。

由于系统中数据访问模式是动态变化的, 因此为了更有效地利用带宽和提高系统性能, 需要根据用户访问模式的变化调整信道分配和数据分配。采用的主要方法有基于数据热点程度的动态分配^[30]、基于用户的不耐烦特性的动态分配^[27]和基于请求数阈值的动态分配^[32]。

3.4 数据广播调度方法的应用场景

周期广播调度适用于已经获取了用户对数据的需求分布的情况, 并且这一分布在调度过程中不会产生太频繁的变化。on-demand 广播调度则根据用户请求实时进行调度, 适用于用户请求实时变化的应用。混合调度对于需求变化不大的数据采取周期广播调度(大多数场合下大量数据属于这种类

型),而对于少量需求变化频繁的数据采用 on-demand 调度,以最大程度符合用户需求。

结束语 数据广播是无线移动环境中一种新的数据发送方式,可以同时为大量移动用户提供数据服务并节省系统带宽资源。数据广播可以分为周期广播、on-demand 广播和混合广播。针对这 3 类广播,本文系统地归纳了各自的理论分析模型,其中周期广播的等间距、平方根规则以及多信道平方根规则是指导各种周期广播调度的重要依据,严格按照这些规则设计的算法可以获得接近理论最优值的平均访问时间;关于 on-demand 的排队论模型经常用于在多信道混合调度中计算 on-demand 广播的平均访问时间;混合调度则基于前两个分析模型建立最优化问题模型,求解该问题就可以得到混合调度的(近似)最优方案。本文以理论分析模型为基础,分析了现有的数据广播调度方法。其中单信道和多信道的周期调度方法已经比较成熟,根据系统信道资源情况、数据情况(长度)和用户访问模式不同可以采用不同的调度策略。on-demand 方法在排队规则的选择以及广播发布的 on-demand 广播理论模型方面还有比较广阔的研究空间,尤其是后者在理论上是一个空白领域。另外关于多数据项请求的广播调度^[33]也是一个被逐步关注的问题。为了使数据广播能够获得更好的性能,需要将广播调度与其他数据广播的问题结合起来,尤其是与空中索引和广播数据缓存结合起来,这样可以进一步节省客户端能量并充分利用客户端资源,同时更好地适应无线环境。

参 考 文 献

[1] Satyanarayanan M. Pervasive computing; vision and challenges [J]. IEEE Personal Communications, 2001, 8(4): 10-17

[2] Imielinski T, Viswanathan S. Adaptive wireless information systems[C]// Proceedings of SIGDBS (Special Interest Group in Database systems) conference, Tokyo, Japan, 1994

[3] 何新贵,唐杰,李霖,等. 特种数据库技术[M]. 北京: 科学出版社, 2000

[4] 孟小峰,周龙骧,王珊. 数据库技术发展趋势. 软件学报[J]. 2004, 15(12): 1822-1836

[5] Agrawal P. Energy Efficient Protocols for Wireless Systems [C]// The Ninth IEEE International Symposium on Indoor and Mobile Radio Communications, 1998, 2: 564-569

[6] Tan K L. Data dissemination in wireless computing environments [M]. Kluwer Academic Publishers, 2000

[7] Xu J L, Lee D L, Hu Q L, et al. Data Broadcast. In Handbook of Wireless Networks and Mobile Computing [M]. John Wiley & Sons, 2002

[8] Ammar M H, Wong J W. The Design of teletext broadcast cycles [J]. Performance Evaluation, 1985, 5(4): 235-242

[9] Vaidya N H, Hameed S. Scheduling data broadcast in asymmetric communication environments [J]. ACM/Baltzer Journal of Wireless Networks (WINET), 1999, 5(3): 171-182

[10] Yu P, Sun W W, Qin Y R, et al. A Data Partition Based Near Optimal Scheduling Algorithm for Wireless Multi-channel Data Broadcast[C]// DASFAA2008, New Delli, India, 2008

[11] Bolch G, Greiner S, Meer H, et al. Queuing Networks and Markov Chains; modeling and performance evaluation with computer science applications [M]. Wiley & Sons Inc, 1998

[12] Hameed S, Vaidya N H. Efficient algorithms for scheduling data broadcast [J]. Wireless Networks, 1999, 5: 183-193

[13] Acharya S, Alonso R, Franklin M, et al. Broadcast disks: Data management for asymmetric communications environments[C]//

Proc. of ACM SIGMOD Conference on Management of Data, San Jose, CA, USA, 1995: 199-210

[14] 李霖,周兴铭. 非对称网络环境中数据广播的启发式多盘调度算法[J]. 计算机学报, 1999, 22(1): 45-50

[15] Prabhakara K, Hua K, Oh J. Multi-level Multi-channel Air Cache Design for Broadcasting in a Mobile Environment[C]// Proc. of the 16th Int'l Conf. Data Eng.(ICDE2000). 2000

[16] Yee W G, Navathe S B, Omiecinski E, et al. Efficient data allocation over multiple channels at broadcast servers [J]. IEEE Trans. on Computers, Special Issue on Mobility and Databases, 2002, 51(10): 1231-1236

[17] Arduzoni E, Bertossi A A, Pinotti M C, et al. Optimal skewed data allocation on multiple channels with flat broadcast per channel [J]. IEEE Trans. on Computer, 2005, 54(5): 558-572

[18] Zheng B, Wu X, Jin X, et al. TOSA: a near-optimal scheduling algorithm for multi-channel data broadcast[C]// Mobile Data Management, 2005: 29-37

[19] Seifert A, Hung J J. FlexSched: A Flexible Data Schedule Generator for Multi-channel Broadcast Systems[R]. 211. University of Konstanz, 2005

[20] Hung J J, Seifert A. FlexSched: A Parameterized Data Schedule Generator for Multi-channel Broadcast Systems. [C]// MDM'06. 2006

[21] Wong J W. Broadcast delivery [J]. Proceedings of the IEEE, 1988, 76(12): 1566-1577

[22] Aksoy D, Franklin M. Rx W: A scheduling approach for large-scale on-demand data broadcast [J]. IEEE/ACM Trans. on Networking, 1999, 7(6): 846-860

[23] Su C J, Tassiulas L. Broadcast scheduling for information distribution[C]// Proc. of the IEEE INFOCOM '97. Kobe, Japan, 1997: 109-113

[24] Sun W, Shi W, Shi B, et al. A Cost-efficient Scheduling Algorithm of On-Demand Broadcasts [J]. ACM Journal of Wireless Networks, 2003, 9(3)

[25] Acharya S, Muthukrishnan S. Scheduling on demand broadcasts; New metrics and algorithms[C]// Proc. of the 4th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'98). Dallas, TX, SA, 1998: 43-54

[26] Lee W C, Hu Q L, Lee D L. A study on channel allocation for data dissemination in mobile computing environments [J]. Mobile Networks and Applications, 1998: 1-20

[27] Hu C L, Chen M S. Adaptive Multichannel Data Dissemination; Support of Dynamic Traffic Awareness and Push-pull Time Balance [J]. IEEE Trans. on Vehicular Technology, 2005, 54(2)

[28] Hung H P, Chen M S. A general model of hybrid data dissemination[C]// Mobile Data Management, 2005: 220-228

[29] Acharya S, Franklin M, Zdonik S. Balancing push and pull for data broadcast[C]// Proc. of ACM SIGMOD Conference on Management of Data, Tucson AZ, USA, 1997: 183-194

[30] Stathatos K, Roussopoulos N, Baras J S. Adaptive data broadcast in hybrid networks[C]// Proc. of the 23rd VLDB Conference, Athens, Greece, 1997: 326-335

[31] Huang J L, Peng W C, Chen M S. SOM: Dynamic push-pull channel allocation framework for mobile data broadcasting [J]. IEEE trans. on mobile computing, 2006, 5(8)

[32] Saxena N, Pinotti M C. On-line balanced k-channel data allocation with hybrid schedule per channel[C]// Mobile Data Management, 2005: 239-246

[33] 张卓瑶,孙未未,余平,等. 无线环境中多数据项广播调度算法综述[J]. 计算机科学, 2009, 36(5)