

分布式数据查询策略的性能估计算法

殷蔚华, 黄载禄

(华中科技大学电信系, 武汉 430074)

摘 要: 分布式数据查询策略及其结构(DDB结构), 与传统的HLR-VLR结构有很大的不同. 本文提出了一种对DDB结构的性能估计方法, 其基础是用户的移动性和呼叫特性的概率分布. 本文对DDB结构的代价和呼叫传递时延等性能参数进行了研究, 证明了DDB结构的性能优越性.

关键词: 移动性管理; DDB; 分布式数据查询策略

中图分类号: TN91; TN929

文献标识码: A

文章编号: 0372-2112 (2001) 02-0204-04

A Methodology for the Performance Evaluation of Data Query Strategy in a DDB Structure

YIN Wei-hua, HUANG Zai-lu

(Dept. of Electronic and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: The distributed database (DDB) architectures as well as their data query strategies, are quite different from traditional HLR-VLR structure. This paper presents a methodology for evaluating hierarchical DDB architectures, based upon users' pdf of mobility and call characteristics. The main interests to DDB's performance are cost and call delivery delay, and research shows they are much better than those in a HLR-VLR structure.

Key words: mobility management; DDB; distributed data query strategy

1 引言

分布式数据查询策略^[1,2]具有层次式的树型数据库结构, 是适应移动通信发展趋势的一种策略. 文献[3]提出了一种从用户移动和呼叫活动的概率分布出发研究DDB树的性能估计方法, 其构想是很有建设性的. 但在建立模型中所做的一系列假设有明显的缺陷, 使该模型与实际情况存在较大差距, 在应用上也受到局限, 而且作者的研究仅局限于对位置更新操作时各层数据库的负载及搜索时间的估计. 本文提出的算法继承了文献[3]的思想, 避免了其缺陷. 本文的基本思想如下: 在建立合理的系统和用户模型的基础上, 推导决定DDB结构各种性能的关键的概率参数, 并进而推导DDB结构的性能参数. 本文所建立的概念可以适用于分布式数据库系统的多种不同的应用情况, 并且为比较研究分布式数据管理策略和其它位置管理策略奠定了基础.

2 基本模型

2.1 系统和用户模型

DDB系统为树状层次结构, 层数为 L , $j=1$ 指最高层(树根), $j=L$ 指树叶. 为简单起见, 假设其结构是对称的. 用户的归属 L 层节点称为HN, 所访问的 L 层节点称为VN. 每一 j 层节点下属的 L 层节点数为 N_j , $N_1=N$ 为系统中所有的 L 层节点数. 每一 j 层的分支数为 branch/j .

假设每个 L 层蜂窝是直径为 D 的圆形; 任一 j 层的覆盖范围称为 j 层蜂窝, 它可以看成由 N_j 个 L 层蜂窝环绕形成的近似圆形. 假设每层的节点设于该层蜂窝的中心. 进一步可近似假设第 j 层蜂窝也是由其所属的 $j+1$ 层蜂窝环绕而成的. 设 $m_j D$ 为 j 层蜂窝最外圈的 L 层蜂窝与圆心的距离, j 为 j 层蜂窝最外圈的 L 层蜂窝数($j < 2m_j$).

用户的运动和呼叫情况可以用概率密度函数表示, 以 $p(r)$ 表示用户位于与其HN相距 rD 的某一VN上的概率, 以 $q(r)$ 表示主叫呼叫被叫时, 主叫HN与被叫HN相距 rD 的概率, 设 $p(r)$ 、 $q(r)$ 与方向无关. DDB结构的性能参数取决于一些关键的概率参数: 用户在某一 j 层蜂窝内的概率 P_j 和 S_j , 以及呼叫局限于某一 j 层蜂窝内的概率 Q_j , 在已知 $p(r)$ 、 $q(r)$ 和DDB结构参数时可得到上述参数.

2.2 P_j 、 S_j 和 Q_j

如图1(a), 设 p_k , ($k=0 \sim m$)为当用户的HN距圆心 kD 时用户在 mD 圆内的概率, 在 rD 距离上共有 $2r$ 个 L 层蜂窝. 则

$$p_k = \begin{cases} p(0) + \sum_{r=1}^m 2r \cdot p(r), & k=0 \\ p(k) + \sum_{r=1}^m 2r \cdot p(D_{AB}), & k=1 \sim m \end{cases}$$

收稿日期: 1999-10-08; 修回日期: 2000-08-25

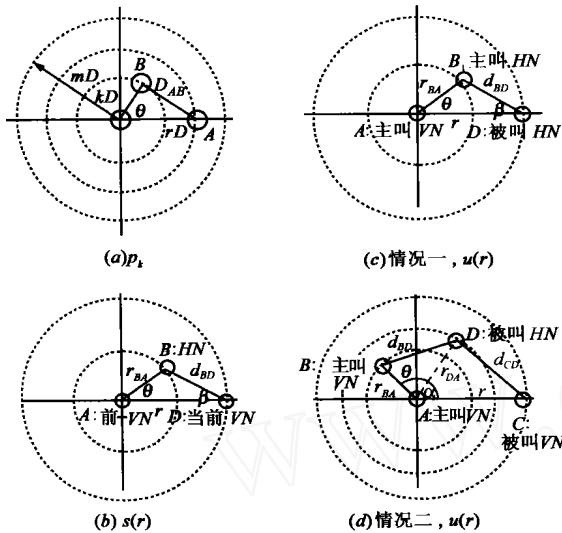


图 1 概率参数的求导

设 p_m 为用户与其 HN 同在一圆内的概率,该圆的最外圈 L 层蜂窝距中心距离为 mD . 则

$$p_m = \frac{p_{k|k=0} + \sum_{k=1}^m 2k \cdot p_k}{1 + \sum_{k=1}^m 2k}$$

则用户与其 HN 同位于第 j 层蜂窝内的概率 P_j 为:

$$P_j = p_{m_{j-1}} + (p_{m_j} - p_{m_{j-1}}) \times \frac{r_{BA}}{2m_j} \quad (1)$$

如图 1(b) 所示,用户当前位置与前一位置相距 rD 的概率为:

$$s(r) = p(r_{BA}) \times p(r_{BD}) \quad (2)$$

则同样可由 $s(r)$ 求得用户当前 VN 与前一 VN 同在 j 层下的概率 S_j .

在 DDB 结构中,用户数据,尤其是与呼叫有关的数据存储在 L 层节点的数据库中,此时有两种情况:

情况一:用户数据存于 HN 中,呼叫建立时首先将呼叫引导至用户 HN 以获取用户数据,再引导至用户当前 VN;

情况二:用户数据存于其当前 VN 中,呼叫建立时直接将呼叫引导至当前 VN.

设主叫当前所在 VN = A,主叫 HN = B,被叫当前所在的 VN = C,被叫 HN = D. 情况一时,如图 1(c) 所示, $u(r)$ 为 VN = A 的主叫呼叫 HN = D 的被叫,且 $d_{AD} = r$ 的概率,则:

$$u(r) = p(r_{BA}) \times q(r_{BD}) \quad (3)$$

情况二时,如图 1(d) 所示, $u(r)$ 为 VN = A 的主叫呼 VN = C 的被叫,且 $d_{AC} = r$ 的概率. 则:

$$u(r) = p(r_{BA}) \times q(r_{BD}) \times p(r_{DC}) \quad (4)$$

由 $u(r)$ 求得主叫与被叫同属于 j 层蜂窝的概率 Q_j ,其过程与由 $p(r)$ 求 P_j 类似.

3 代价

代价以以下公式衡量:代价 = 通讯量(消息数或比特数)

× 消息传播的距离.

3.1 位置更新代价 C_L

在位置更新时,产生两种查询过程:由当前 VN 开始的向上的位置登记,和由前一 VN 开始的向上的位置删除. 假设任一用户从 L 层节点产生的位置登记通讯量为 R_{Lr} ,产生的位置删除通讯量为 R_{Lc} . 如果某一用户在位置更新时到达的归属层顶点为 j 层,则该用户所产生的位置登记代价:

$$C_{Lr}^{(j)} = R_{Lr} \times (S_j - S_{j+1}) \times \sum_{i=L+1}^j d_{i,i+1}$$

所产生的位置删除代价:

$$C_{Lc}^{(j)} = R_{Lc} \times (S_j - S_{j+1}) \times \sum_{i=L+1}^{j+1} d_{i,i+1}$$

其中 $d_{i,i+1}$ 为 j 层与 $j+1$ 层间的平均距离. 则单个用户在一次位置更新时代价为:

$$C_L = \sum_j (C_{Lr}^{(j)} + C_{Lc}^{(j)}) = \sum_{j=L+1}^1 (S_j - S_{j+1}) \times \left(R_{Lr} \times \sum_{i=L+1}^j d_{i,i+1} + R_{Lc} \times \sum_{i=L+1}^{j+1} d_{i,i+1} \right) + (1 - S_1) \times (R_{Lr} + R_{Lc}) \times \sum_{i=L+1}^1 d_{i,i+1} \quad (5)$$

公式的第二项为移动到 1 层范围以外的用户在网内产生的位置更新代价(此时平均距离只计算网内的一段),当 S_1 较大时,可忽略.

传统 HLR-VLR 结构下单个用户单次位置更新的代价为:

$$C_{L(HLR)} = d_{HLR,VLR} \times (1 - S_{VLR}) \times (R_{Lr} + R_{Lc}) \quad (6)$$

3.2 呼叫代价 C_C

呼叫建立时,查询与呼叫传递同时进行,此时对应两种情况的呼叫代价公式不同. 情况一时,呼叫传递距离 D_j = 主叫至被叫的 HN 的距离 d_{ij} + 被叫的 HN 与 VN 间的平均距离 d_2 :

$$d_{ij} = 2 \times \sum_{i=L+1}^j d_{i,i+1}$$

$$d_2 = \sum_{k=L+1}^1 (P_k - P_{k+1}) \times 2 \times \sum_{i=L+1}^k d_{i,i+1}$$

设每个用户在任一节点产生的单次呼叫通讯量为 R_C ,则网络中单个用户的单次呼叫代价

$$C_C = C_{C1} = R_C \times \sum_j (Q_j - Q_{j+1}) \times (d_{ij} + d_2) = R_C \times \left[\sum_{j=L+1}^1 (Q_j - Q_{j+1}) \times (d_{ij} + d_2) + (1 - Q_1) \times \sum_{i=L+1}^1 d_{i,i+1} \right] \quad (7)$$

式中第二项为移动到网外的用户在网内产生的呼叫代价,其呼叫传递距离只考虑在网内的一部分,且不计 d_2 值,故第二项的值是偏小的.

情况二时呼叫代价:

$$C_C = C_{C2} = R_C \times \sum_{j=L+1}^1 (Q_j - Q_{j+1}) \times d_{ij} + (1 - Q_1) \times \sum_{i=L+1}^1 d_{i,i+1} \quad (8)$$

传统的 HLR-VLR 结构下,呼叫代价为:

$$C_{C(HLR)} = (Q_{HLR} - Q_{VLR}) \times 2 \times d_{HLR,VLR} \times R_C + (1 - Q_{HLR}) \times d_{HLR,VLR} \times R_C \quad (9)$$

3.3 平均代价 C

设用户在系统范围内是均匀分布的,即在任一 L 层节点

上都拥有相同的用户数 u 。定义平均代价为“单位时间下网络中用于位置更新或呼叫的总代价”，则有

$$C = u \times N \times L \times (C_L + C_C) \quad (10)$$

其中 C_L 为呼叫移动性比, C_C 为位置更新率。

4 呼叫传递时延 T_C

一次呼叫传递操作,假设其路径的顶点为 j 层,此时

传递时间 t_j = 传递路径上每个节点的搜索时间 t_{search} + 传递距离所消耗的时间 t_{trans} 。

t_{search} = 传递路径上所有节点的数据库容量之和 $\text{sum}l_j$ / 平均搜索每一用户的速度 V_{search}

t_{trans} = 传递距离 D_j / 平均传递速度 V_{trans}

路径 1 上的数据库容量之和为 $\text{sum}l_{1j} = L_j + 2 \times \sum_{i=L}^{j+1} L_i$

路径 2 上的数据库容量之和 $\text{sum}l_2 = \sum_{k=L-1}^1 (P_k - P_{k+1}) \times$

$$\left(L_k + 2 \times \sum_{i=L-1}^{k+1} L_i + L_L \right)$$

情况一时单个用户单次呼叫时传递时延:

$$T_C = T_{C1} = \sum_{j=L}^1 (Q_j - Q_{j+1}) \times t_j = \sum_{j=L}^1 (Q_j - Q_{j+1}) \times \left(\frac{\text{sum}l_{1j} + \text{sum}l_2}{V_{\text{search}}} + \frac{d_{1j} + d_2}{V_{\text{trans}}} \right) + (1 - Q_1)$$

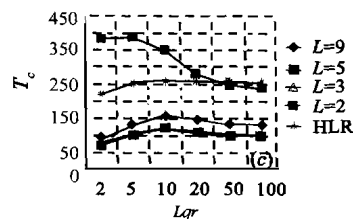
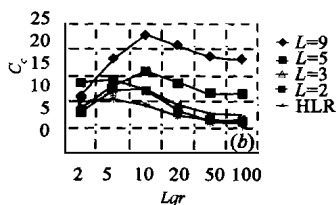
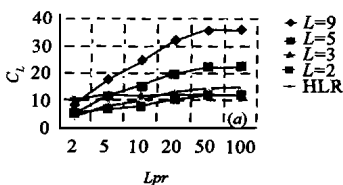


图 2 $N=256, p(r)=c, q(r)=c$ 时,不同分层下 C_L, C_C 和 T_C 的变化趋势

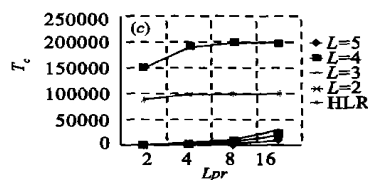
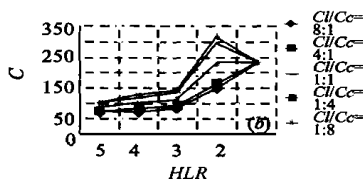
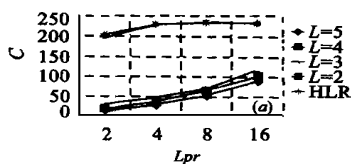


图 3 $N=100000, p(r)=c, q(r)=c, Lqr=2$ 时,不同分层或不同 L 时, C_L, C_C 和 T_C 的变化趋势

各图表中,HLR 指传统的 HLR-VLR 结构。Lpr 和 Lqr 分别为移动性和呼叫影响距离,即: $r \geq Lpr$ 时, $p(r)=0$; $r \geq Lqr$ 时, $q(r)=0$ 。假设 $R_L = R_C = R$, 位置更新代价 C_L 、呼叫代价 C_C 分别除以 R 以归一化。假设 $L + C = L(1 + \frac{C}{L})$ 为常数,则平均代价 C 除以 $u \times N \times (L + C) \times R$ 以归一化。假设网络传递信息的速率 V_{trans} 很大,则呼叫传递时延中可以近似地只考虑含 V_{search} 的部分,且 T_C 除以 $\frac{1}{V_{\text{search}}}$ 以归一化。

图 2 和图 3 分别为情况 A、B 时 DDB 结构的性能参数的变化趋势,从中可见:

(1) 一般而言,用户移动范围和移动概率的增加,或主、被叫用户距离的扩大将使 DDB 结构的代价和呼叫传递时延增

$$\times \left\{ \frac{1}{V_{\text{search}}} \left(\frac{L_i}{i=L} + \frac{d_{i,i+1}}{V_{\text{trans}}} \right) \right\} \quad (11)$$

式中 $Q_{L+1}=0$,第二项为移动到网外的用户在网内产生的单次呼叫时传递时延,其呼叫传递距离只考虑在网内的一部分, d_2 值和 $\text{sum}l_2$ 值不计,故第二项的值是偏小的。情况二时:

$$T_C = T_{C2} = \sum_{j=L}^1 (Q_j - Q_{j+1}) \times \left(\frac{\text{sum}l_{1j}}{V_{\text{search}}} + \frac{d_{1j}}{V_{\text{trans}}} \right) + (1 - Q_1) \times \left\{ \frac{1}{V_{\text{search}}} \left(\frac{L_i}{i=L} + \frac{d_{i,i+1}}{V_{\text{trans}}} \right) \right\} \quad (12)$$

传统的 HLR-VLR 结构下,呼叫传递时延为:

$$T_{C(\text{HLR})} = (Q_{\text{HLR}} - Q_{\text{VLR}}) \times \left(\frac{L_{\text{HLR}} + 2 \times L_{\text{VLR}}}{V_{\text{search}}} + \frac{2 \times d_{\text{HLR-VLR}}}{V_{\text{trans}}} \right) + (1 - Q_{\text{HLR}}) \times \left(\frac{L_{\text{HLR}} + L_{\text{VLR}}}{V_{\text{search}}} + \frac{d_{\text{HLR-VLR}}}{V_{\text{trans}}} \right) \quad (13)$$

5 数字分析

本文中分别取了 $N=256$ 和 $N=100000$ 的两套数据, $N=100000$ 的各图表反映了当用户基本在网内 ($P_1 > 0.90$),而且主叫与被叫基本都在网内 ($Q_1 > 0.90$) 条件下的情况(情况 A), $N=256$ 的图表反映了当用户部分在网外 ($P_1 < 0.90$),而且部分主叫与被叫不都在网内 ($Q_1 < 0.90$) 条件下的情况(情况 B)

大。情况 A 时,当 Lqr 较大时, C_C, T_C 随 Lqr 的增大而减小。这是因为:此时大部分呼叫传递的路径都在网外,未能计入。

(2) 情况 B 时,DDB 结构的性能大大优于传统结构,层数的增加对性能的影响不明显;而当情况 A 时,DDB 结构的性能接近或劣于传统结构,此时分层的增加将使其增大。

(3) 分层增加对 DDB 结构的代价和传递时间影响甚小,或起负面影响。因此 DDB 结构的层数在最高层负载可以承受的情况下不宜太大;

(4) 二层的 DDB 结构是不可取的,其性能远远劣于采用其它分层的 DDB 结构。图 4 表明:

(1) 当其它条件相同时,情况一时性能劣于情况二时。但

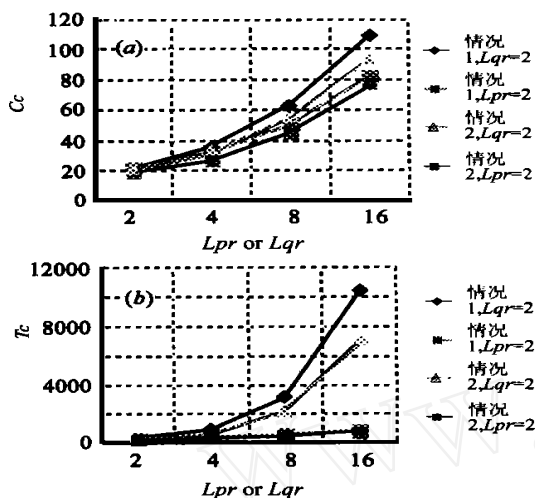


图4 $N = 100000$, $p(r) = c$, $q(r) = c$, $L = 5$ 时,
 C_c 和 T_c 的变化趋势

这里没有考虑如何将用户数据由一个 VN 传至另一个 VN , 显然通过固定网络将大大增加网络的开销, 而通过无线接口则将占用无线链路, 并且在手机一侧始终应有用户数据库的拷贝, 因此用户数据存放在何处较为有利尚有待进一步讨论.

(2) L_{pr} 的变化对性能的影响比 L_{qr} 的大. 原因是: L_{pr} 既影响 P_j , 又影响 Q_j , 而 L_{qr} 只影响 Q_j .

5 总结

本文从概率角度出发, 提供了一种研究 DDB 结构的性能参数的估计方法. 研究结果表明: DDB 结构有许多显而易见的优点, 由于充分考虑了用户移动范围的有限性和大多数呼叫的本地性特点, 使位置更新和呼叫传递的操作尽量局限于底层数据库中进行, 从而大大减少了高层数据库 (相对于传统的

HLR) 中的数据容量和信令通过高层数据库的概率, 减少了网络信令拥塞的可能性. 此外, DDB 结构从原理上讲, 允许并行查询不同的数据库分割, 这为高速数据查询和处理提供了必要的条件. 因此, DDB 结构是一种适应移动通信发展需求的结构.

参考文献:

- [1] Wang J. Z. A fully distributed location registration strategy for universal personal communication systems [J]. IEEE J. Select Areas Commun., 1993, 11(6): 850 - 860.
- [2] Maylan A. D., et al. Network architecture and signaling for wireless personal communications [J]. IEEE J. Select Areas Commun., 1993, 11(6): 830 - 841.
- [3] Eynard C., et al. A methodology for the performance evaluation of data query strategies in universal mobile telecommunication systems (UMTS) [J]. IEEE J. Select. Areas Commun., 1995, 13(5): 893 - 907.

作者简介:



殷蔚华 1965 年生, 1989 年毕业于浙江大学, 获硕士学位, 现为华中科技大学 (原华中理工大学) 电子信息与工程系博士, 从事移动通信技术方面的研究与开发工作, 主要研究方向为移动网中的数据库技术、网络管理、移动性管理技术等.

黄载祿 1937 年生, 1962 年毕业于华中工学院, 1981 ~ 1983 年在美国俄亥俄大学作访问学者, 现为华中科技大学电信系教授, 博士生导师. 从事通信、信号处理等方面的研究工作, 发表论文近百篇, 获多项科研成果奖, 所主持的 EIM-601 数字程控交换机获电子部科技进步特等奖.