

移动分布式实时嵌套事务提交*

刘云生⁺, 廖国琼, 李国徽, 夏家莉

(华中科技大学 计算机科学与技术学院,湖北 武汉 430074)

Commitment of Mobile Distributed Real-Time Nested Transaction

LIU Yun-Sheng⁺, LIAO Guo-Qiong, LI Guo-Hui, XIA Jia-Li

(School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China)

+ Corresponding author: Phn: 86-27-87522511, E-mail: ysliu@hust.edu.cn

<http://www.hust.edu.cn>

Received 2002-03-05; Accepted 2002-05-29

Liu YS, Liao GQ, Li GH, Xia JL. Commitment of mobile distributed real-time nested transaction. *Journal of Software*, 2003,14(1):139~145.

Abstract: For transactions' mobility and the inherence limitations of wireless network, traditional real-time transaction management mechanisms are incompetent to support the execution of mobile distributed real-time transactions in mobile distributed computing environment. In this paper, the commit mechanism for mobile real-time transactions is studied. First, a nested transaction model based on functional alternative tasks is given by analyzing the characteristics of real-time transactions in mobile distributed environment. Then a three-tier commit structure supporting the suggested model is presented. And a three-phase real-time commit protocol 3PRTC (three-phase real-time commit) is also proposed, which can guarantee the atomicity and structural correctness of the mobile real-time transactions. By performance testing, it is shown that the suggested transaction model and its commit mechanism can improve the successful ratio of real-time transactions.

Key words: mobile computing; real-time transaction; transaction processing; functional alternative; commit

摘要: 在移动分布式计算环境中,事务移动性和无线网络固有的缺陷使得传统的分布式实时事务管理机制不足以支持移动分布式实时事务的执行,故有必要为移动实时事务研究新的事务处理机制,以提高其成功率.着重研究移动实时事务的提交机制.首先,通过分析移动分布环境中实时事务的特点给出了一个基于功能替代的移动实时嵌套事务模型.然后,提出了一个基于此模型的三层提交结构以及能够保证移动实时事务原子性和结构正确性的三阶段实时提交协议 3PRTC(three-phase real-time commit).性能测试表明,所提出的事务模型及其提交机制能够提高实时事务的成功率.

关键词: 移动计算;实时事务;事务处理;功能替代;提交

* Supported by the National Natural Science Foundation of China under Grant No.60073045 (国家自然科学基金); the Defence Pre-Research Project of the 'Tenth Five-Year-Plan' of China under Grant No.413150403 (国家“十五”国防预研基金); the National Research Foundation for the Doctoral Program of Higher Education of China under Grant No.2002048706 (国家教育部博士点基金)

第一作者简介: 刘云生(1940—),男,湖南衡阳人,教授,博士生导师,主要研究领域为现代(实时、主动、内存、移动等非传统)数据库理论与技术及其集成实现,数据库与信息系统开发,实时数据工程,软件方法学与工程技术.

中图法分类号: TP393 文献标识码: A

移动分布式计算环境是为了满足人们在任意地点、任意时刻访问任意数据的需求而提出的.实时(具有定时限制)应用是该环境中的一类典型应用,如移动实时股票交易、电子商务以及交通信息的实时发布等.但由于无线网络具有不可靠、不可预测、频繁断接等特点,使得移动分布式实时事务相对于传统的分布式实时事务更难满足其截止期.因此,有必要为移动实时事务研究新的事务处理机制,以提高其成功率.

有关移动实时事务的研究还刚刚开始.已有的研究主要是在移动实时事务的执行模型^[1]、并发控制机制^[2]、调度策略^[3]等方面,可以说目前还没有有关移动实时事务提交机制的研究.2PC(two-phase commit)^[4]是分布式事务通常采用的原子提交协议,在此基础上已提出了改进的提交协议,如 PC(presumed commit)^[5]、PA(presumed commit)^[5]、O2PC(optimistic two-phase commit)^[6]等.但这些协议是非实时的,且未考虑事务的位置变化,因此都不能很好地支持移动实时事务的提交.RCP(real-time commit protocol)^[7]、PROMPT(permits reading of modified prepared-data for timeliness)^[8]等实时提交协议考虑了事务的定时限制,但它们未考虑事务的移动特性.本文研究一种支持定时限制(截止期)和移动性的三阶段提交协议.

1 移动实时事务模型

由移动主机 MH(mobile host)发出且具有定时限制的事务称为移动实时事务 MRTT(mobile real-time transaction).由于 MRTT 在执行时其位置可能发生变化,而且无线网络较大的传输延迟和频繁断接等会导致 MRTT 的执行时间相对延长,超截止期的概率也相应地增加.故有必要为移动实时事务的重要任务建立一个甚至多个功能等价任务,只要其中任何一个执行成功,则表示该任务成功.这样就提高了移动实时事务的成功率.

1.1 功能替代任务

定义 1. 设任务 tk_{ij} 与 tk_{ik} 为不同的任务,若按应用语义,它们实现同一事务 T 的一个功能,则称 tk_{ij} 和 tk_{ik} 与 T 的该功能等价,且称 tk_{ij} 和 tk_{ik} 为 T 的功能替代任务.

定义 2. 由所有与事务 T 的一个任务功能等价的任务组成的集合称为 T 的一个功能替代集,记为 FAS_i . FAS_i 中的每个任务都称为 T 的一个功能替代子事务.

通常,一个 FAS_i 中的功能替代子事务是按选优次序逐个执行的,直到某一个执行成功或该功能替代集所有子事务都夭折为止.但这种方式容易导致实时事务超截止期,而应选择多个替代子事务同时在一个或多个结点上执行.此时,只要一个替代子事务执行成功,相关的任务就成功.虽然这种并行调度方式在一定程度上增加了系统开销,但整个系统的性能将由于提高了事务的成功率而得到改善.

1.2 嵌套的移动实时事务

一个 MRTT 可能具有多个功能替代集 $FAS_i, i=1,2,\dots,n$; 每个功能替代集 FAS_i 又可能由一个或多个功能替代子事务组成,这样就形成了一个嵌套结构事务.一个移动实时事务可定义如下:

定义 3. 一个移动实时事务 MRTT 是一个四元组:

$$MRTT ::= \langle TS, R, \angle t, C \rangle.$$

其中: $TS ::= \{ \langle TK_i \rangle \mid TK_i \text{ is a task in } MRTT, i=1,2,\dots,n \};$
 $TK_i ::= FAS_i = \{ t_{ij} \mid \forall j, t_{ij} \text{ is a subtransaction functionally equivalent to task } TK_i, j=1,2,\dots,m_i \};$
 $R ::= \text{a set of resources required by tasks in } TS;$
 $\angle t ::= \text{a temporal ordering on } TS;$
 $C ::= \text{a set of constraints on } TS \text{ and } R.$

基于文献[9]中经历(history)模型的各种结构依赖关系,如提交依赖 CD(commit dependency)、夭折依赖 AD(abort dependency)、开始依赖 BD(begin dependency)等,我们根据功能替代性提出一种新的事务依赖关系——排它提交依赖 CD^X (exclusive commit dependency).

定义 4. 设 ST 为事务集, H 为 ST 的执行经历, $\forall t_i, t_j \in ST, i \neq j$, 事务 t_i 与 t_j 存在排它提交依赖, 记为 $t_i CD^x t_j$, 当且仅当

$$\neg (commit_{t_{ij}} \in H \wedge commit_{t_{ik}} \in H).$$

这样, 基于功能替代的嵌套移动实时事务的结构依赖关系有:

- (1) $\forall FAS_i \in MRTT (MRTT CD FAS_i) (i=1, 2, \dots, n).$
- (2) $\forall t_{ij} \in FAS_i (t_{ij} BD MRTT \wedge t_{ij} AD MRTT) (i=1, 2, \dots, n; j=1, 2, \dots, m_i).$
- (3) $\forall FAS_i \in MRTT (\forall t_{ij}, t_{ik} \in FAS_i (t_{ij} CD^x t_{ik})) (i=1, 2, \dots, n; j, k=1, 2, \dots, m_i; j \neq k).$

2 移动实时嵌套事务的三层提交

无线网络的不可靠和不可预测等使得移动实时事务基本上只能是软或固实时事务. 故不失一般性, 以下只讨论两层嵌套软(固)实时事务的提交. 硬实时事务的正确性须由系统来实现. 关于它的提交, 我们将另文讨论.

2.1 移动实时事务的原子性

在某种意义上, 功能替代子事务的存在需要放松对移动实时事务原子性的要求. 设 d_T 为 MRTT 的截止期, 我们有:

定义 5. 若子事务 $t_{ij} \in MRTT$ 在 d_T 内执行成功, 则称 t_{ij} 可提交或进入可提交状态; 否则称 t_{ij} 不可提交.

定义 6. 若 $\exists t_{ij} \in FAS_i$ 可提交, 则称该 FAS_i 可提交.

定义 7. 若 $\forall FAS_i \in MRTT$ 可提交, 则称 MRTT 可提交.

按照上述定义, 只要每个功能替代集中有一个子事务可提交, 则该 MRTT 可提交. 故在此, MRTT 的提交原子性是指, 当其提交时, 它的每一个功能替代集有一个也只有一个子事务提交; 若有一个 FAS_i 不可提交, 则该 MRTT 夭折. 在这种意义上, MRTT 的原子性相对于传统原子性得到了放松, 因为它只要求部分而不是所有子事务都提交. 显然, 这种放松的原子性仍然能够保证移动实时事务的正确性.

2.2 三层提交结构

为了保证分布事务的正确执行, 需要一个协调者事务来协调各子事务的执行. 通常该协调者由始发场地上的子事务担任. 但不可靠的无线网络使得发出 MRTT 的 MH 上的子事务不宜作为 MRTT 执行的协调者, 而应由 MH 所在无线单元中的固定主机 FH (fixed host) 上的子事务承担. 当 MH 迁移到新的单元时, MRTT 的协调者 t_{co} 也应迁移到新的单元, 且有关提交的上下文信息应从老的协调者传递给新的协调者. 但是, 由于 t_{co} 的位置可能发生变化, 当采用协调者和参与者两层结构提交时, 一方面, 为保持与当前协调者交换消息, 每个参与者都必须跟踪 t_{co} 位置的变化; 另一方面, 功能替代会增加 MRTT 子事务数量, 加重协调者与参与者的通信负担, 因此, 带宽有限的无线通信信道可能出现阻塞.

为此, 我们给出了一个如图 1 所示的三层提交结构. 对于一个 MRTT, t_{co} 是 MRTT 的总协调者; t_{ij} 是子事务(参与者); FAS_{coi} 是每一个功能替代集的协调者, 由任意一个在固定场地执行的子事务兼任, 负责收集其所在功能替代集中子事务的报告, 并将该功能替代集的执行结果向 t_{co} 报告. 这样, 每个 FAS_i 中只需 FAS_{coi} 跟踪 t_{co} 位置的变化, 且也只需一个子事务与 t_{co} 通信, 从而减轻了 t_{co} 的通信负担.

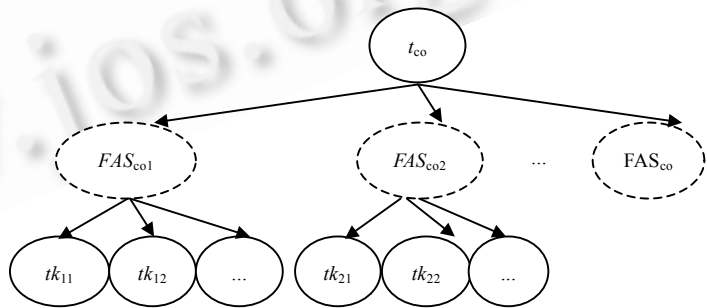


Fig.1 The three-tier commit structure
图 1 三层提交结构

3 三段实时提交协议

3.1 实时提交约束

对于实时事务,超截止期往往夭折,故提交时,可能出现下列意外情形:

- (1) 参与者报告“可提交”后,未及收到协调者的决定因超截止期而夭折;
- (2) 参与者收到来自协调者的提交决定后,来不及提交因超截止期而夭折.
- (3) 一个子事务(参与者或总协调者)未做任何报告因超截止期而夭折.

因此,为了避免因超截止期所引起的不一致性,MRTT 的提交必须增加以下约束:

约束 1. 若 t_{co} 已发出提交决定,则认为该 MRTT 已经提交.

约束 2. 若 t_{ij} 报告“可提交”后未收到 t_{co} 的决定消息,即使超截止期也不能自行夭折.

在执行提交过程中, t_{co} , FAS_{coi} 以及 t_{ij} 之间需要交换消息,不可避免地会产生通信延迟.故为避免可提交的 MRTT 因通信延迟而被夭折,我们有:

定义 8. t_{co} 接收报告截止期 $d_A = d_T + \delta$, 其中 δ 为移动网络的最大通信延迟.

δ 是动态变化的,且由 MRTT 所在的移动网络的特性决定,如通信带宽和网络的拓扑结构等.针对上述情况(3),并结合定义 8,给出以下约束:

约束 3. 若 t_{co} 到达 d_A 仍未收到所有 FAS_{coi} 的“可提交”报告,则决定夭折.

3.2 三阶段提交协议

我们给出一个考虑了事务截止期和移动性的三阶段实时提交协议 3PRTC(three-phase real-time commit). 一个 MRTT 的提交分为 3 个阶段:子事务(参与者) t_{ij} 报告阶段、功能替代集的协调者 FAS_{coi} 决定及报告阶段以及总协调者 t_{co} 决定阶段.3PRTC 协议描述如下:

```

 $t_{ij}$ :   IF  $t_{ij}$  不可提交 THEN
           向  $FAS_{coi}$  报告“FAILURE”后夭折;
           IF  $t_{ij}$  可提交 THEN
           向  $FAS_{coi}$  报告“SUCCESSFUL”后等待,直到收到并执行  $t_{co}$  的决定.
 $FAS_{coi}$ : IF 收到一个“FAILURE” 报告 THEN
           IF  $\forall t_{ij} \in FAS_i(t_{ij}$  报告“FAILURE”)THEN
           向  $t_{co}$  报告“ABORT”;
           IF 收到一个来自  $t_{ij}$  的“SUCCESSFUL”报告 THEN
           向  $t_{co}$  报告“COMMITTABLE”,且夭折  $FAS_i$  中除  $t_{ij}$  外的所有子事务.
 $t_{co}$ :   IF 收到一个“COMMITTABLE”报告 THEN
           IF  $\forall FAS_{coi} \in MRTT(FAS_{coi}$  报告“COMMITTABLE”) THEN
           决定提交,且向所有报告“SUCCESSFUL”的子事务发送“COMMIT”决定;
           IF 收到一个“ABORT”报告或到达  $d_A$  还未收到所有“COMMITTABLE”报告 THEN
           决定夭折,且向所有未报告“FAILURE”的子事务发送“ABORT”决定.

```

3.3 3PRTC的正确性

证明 3PRTC 的正确性,即要证明它能保证移动实时事务的原子性及结构的正确性.

定理 1. 3PRTC 能保证移动实时事务的原子性.

证明:令“ \exists ”表示“存在也只存在一个”.在 3PRTC 中:

- (1) $\exists t_{ij} \in FAS_i(t_{ij}$ reports SUCCESSFUL
 $\Rightarrow FAS_{coi}$ reports COMMITTABLE $\wedge \forall k \neq j, t_{ik} \in FAS_i(abort_{t_{ik}} \in H)$);
 $\forall FAS_{coi} \in MRTT(FAS_{coi}$ reports COMMITTABLE $\Rightarrow t_{co}$ makes a COMMIT decision

$\Rightarrow \exists t_{ij} \in FAS_i(t_{ij} \text{ reports SUCCESSFUL} \Rightarrow \text{commit}_{t_{ij}} \in H)$).

即若 MRTT 提交,它的每个功能替代集有且只有一个子事务提交.

- (2) $\forall t_{ij} \in FAS_i(t_{ij} \text{ reports FAILURE} \Rightarrow FAS_{coi} \text{ reports ABORT});$
 $\exists FAS_{coi} \in MRTT(FAS_{coi} \text{ reports ABORT} \Rightarrow t_{co} \text{ makes an ABORT decision} \Rightarrow \forall t_{ij} \in MRTT(\text{abort}_{t_{ij}} \in H)).$

即若 MRTT 夭折,它的所有子事务都夭折.

故 3PRTC 能保证 MRTT 的原子性. □

定理 2. 3PRTC 是放松的原子提交协议

证明:由定理 1 可知,3PRTC 只要求部分而不是所有子事务都提交,故 3PRTC 是放松原子提交协议. □

定理 3. 3PRTC 能保证移动实时事务的结构正确性.

证明:第 1.2 节描述了移动实时嵌套事务的结构依赖关系.其中 t_{ij} BD MRTT 由事务的调度策略保证;而 MRTT CD FAS_i 和 t_{ij} AD MRTT 等依赖关系在 3PRTC 中显然.故在此主要证明 3PRTC 能保证 $\forall FAS_i \in MRTT(\forall t_{ij}, t_{ik} \in FAS_i(t_{ij} \text{ CD}^x t_{ik}))(i=1,2,\dots,n; j,k=1,2,\dots,m_i; j \neq k)$ 依赖关系成立.首先:

- (1) $\forall t_{ij} \in FAS_i(t_{ij} \text{ reports FAILURE} \Rightarrow FAS_{coi} \text{ reports ABORT}).$
(2) $\exists t_{ij} \in FAS_i(t_{ij} \text{ reports SUCCESSFUL} \Rightarrow FAS_{coi} \text{ reports COMMITTABLE}).$

现在:

(1) $\exists FAS_{coi}(FAS_{coi} \text{ reports ABORT}) \Rightarrow t_{co} \text{ makes an ABORT decision} \Rightarrow \forall FAS_i \in MRTT(\forall t_{ij} \in FAS_i(\text{abort}_{t_{ij}} \in H)).$ 即 $\forall t_{ij} \in FAS_i(\text{commit}_{t_{ij}} \notin H)$.

(2) $\forall FAS_{coi}(FAS_{coi} \text{ reports COMMITTABLE}) \Rightarrow t_{co} \text{ makes a COMMIT decision} \Rightarrow \forall FAS_i \in MRTT(\exists t_{ij} \in FAS_i(t_{ij} \text{ reports SUCCESSFUL} \Rightarrow \text{commit}_{t_{ij}} \in H) \wedge \forall k \neq j, t_{ik} \in FAS_i(\text{abort}_{t_{ik}} \in H)).$ 即 $\forall FAS_i \in MRTT(\exists t_{ij} \in FAS_i(\text{commit}_{t_{ij}} \in H) \wedge \forall k \neq j, t_{ik} \in FAS_i(\text{commit}_{t_{ik}} \notin H)).$

因此, $\forall FAS_i \in MRTT, \forall j \neq k, t_{ij}, t_{ik} \in FAS_i, \neg(\text{commit}_{t_{ij}} \in H \wedge \text{commit}_{t_{ik}} \in H)$, 即, $t_{ij} \text{ CD}^x t_{ik}$. 故 3PRTC 能保证移动分布式实时事务的结构正确性. □

4 性能评价

我们在由国家自然科学基金资助、自行研制的分布式主动实时数据库原型系统 ARTs-II 上模拟移动计算环境完成了对 3PRTC 协议的性能测试.本实验采用的性能测度指标为软实时事务通常采用的事务超截止期比率 $MR = \text{NumMiss} / \text{NumTotal} \times 100\%$. 其中 NumMiss 表示超截止期事务数, NumTotal 表示事务总数.

4.1 实验模型及参数

测试系统模型如图 2 所示.移动主机事务管理器 MHTM(mobile host transaction manager)和固定主机事务管理器 FHTM(fixed host transaction manager)分别负责移动主机 MH(mobile host)和固定主机 FH(fixed host)上的事务管理及相互之间的通信.并发控制器 CC(concurrency controller)、调度管理器 SCH(scheduler)、资源管理器 RM(resource manager)分别管理各主机上的并发控制、事务调度及各种资源.为了支持实时事务,该系统采用内存数据库(memory database,简称 MDB)体系结构,即数据库的“工作版本”常驻内存,由内存数据库管理系统 MDBMS(memory database management system)管理;而磁盘数据库(secondary database,简称 SDB)只作为数据库备份,由外存数据库管理系统 SDBMS(secondary database management system)管理.为了支持移动事务,在 MH 上增加过区切换控制器(handoff controller,简称 HC)和断接处理器(disconnected processor)负责处理移动主机的过区切换和频繁断接;且在 FH 上增加位置管理器 LM(location manager),以管理位于 FH 所在无线单元内主机的位置变化.实验的主要模拟参数见表 1.

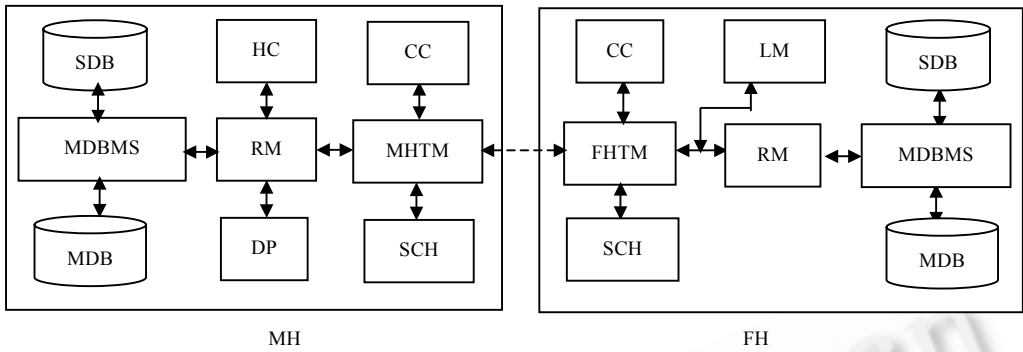


Fig.2 Simulation system model
图2 模拟系统模型

Table 1 Simulation parameters
表1 模拟参数

Parameter	Value	Parameter	Value
Think time	1s~10s	Number of functional alternative sets of each transaction	5~10
Number of mobile hosts	30	Number of sub-transactions in each functional alternative set	1~10
Number of fixed hosts	10	Number of operations in each sub-transaction	5~15
Number of wireless cells	5	Concurrency control mechanism	High priority abort (HPA)
Probability of disconnected	0.5%	Schedule police of CPU	Earlier deadline first (EDF)
Probability of handoff	0.2%	Replacement police of main memory	Least recently use (LRU)
Number of database	10	Type of transactions	Soft real-time transactions
Size of database	500	Level of nested transactions	2

4.2 实验结果分析

图3在采用3PRTC协议提交时,每个功能替代集中子事务数量的变化对MR的影响.可以看出,在每个FAS中具有多个功能替代子事务(NumST=3和5)的移动实时事务的MR比只有一个子事务(NumST=1,即任务本身)低,这是因为多个替代子事务的并发执行增加了移动实时事务成功的可靠性.然而,当NumST=7和10时,MR却上升,原因是随替代任务数量的增加,系统的开销增加,事务夭折重启的可能性也相应上升,从而导致系统性能下降.因此,替代任务的数量并不是越多越好.从测试结果来看,合适的范围应是≤5.

如图4所示为两种提交结构对MR影响的测试结果.可以看出,采用三层提交结构(包括 t_{co} , FAS_{cor} 和 t_{ij})比采用两层提交结构(包括 t_{co} 和 t_{ij})对MR的影响要低.原因是三层提交结构减轻了大多数参与者跟踪 t_{co} 的开销和无线网络的通信负担.

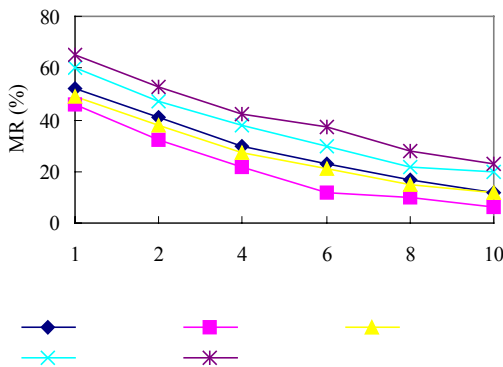


Fig.3 Impact of the number of sub-transactions in each FAS on MR

图3 每个功能替代集子事务数量对MR的影响

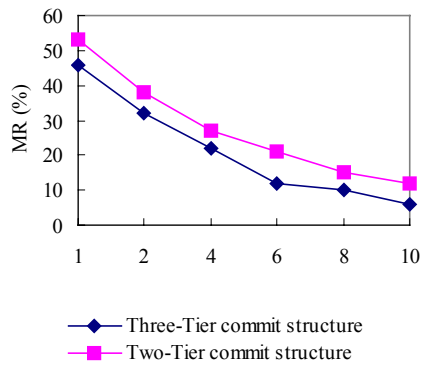


Fig.4 Impact of the commit structures on MR

图4 事务提交结构对MR的影响

5 结束语

传统的分布式实时事务管理机制不能较好地支持移动分布环境中实时事务的执行.因此,必须为移动分布式实时事务建立新的事务模型及其事务管理机制.本文在这方面做的工作概括如下:

- (1) 定义了一个基于功能替代的移动实时嵌套事务模型;
- (2) 提出了一个支持上述模型的三层提交结构;
- (3) 给出了一个考虑事务定时性以及移动性的提交协议 3PRTC,并证明了它的正确性;
- (4) 对所提出的事务模型及协议进行了性能测试,证明其可以提高实时事务的成功率.

移动实时事务是一个新的研究领域,有许多方面需要研究,如移动实时事务的调度、并发控制与恢复机制等,我们将另文专门加以讨论.

References:

- [1] Kayan E, Ulusoy O. Real-Time transaction management in mobile computing systems. In: Chen ALP, Lochovsky FH, eds. Proceedings of the 6th International Conference on Database Systems for Advanced Applications. Los Alamitos: IEEE Computer Society, 1999. 127~134.
- [2] Lam KY, Kuo TW, Tsang WH, *et al.* Concurrency control in mobile distributed real-time database systems. *Information Systems*, 2000,25(4):261~286.
- [3] Saad-Bouzeffrane S, Sadeg B, Amanton L. Soft real-time transaction scheduling in a wireless environment. In: Azzedine B, ed. Proceedings of the 4th IEEE International Symposium on Object-Oriented Real-Time Distributed Computing. Los Alamitos: IEEE Computer Society, 2001. 327~334.
- [4] Gray JN. Notes on database systems. In: Bayer R, Graham RM, Seegmuller G, eds. *Operating Systems: an Advanced Course*. Vol 60 of LNCS, Berlin: Springer-Verlag, 1978. 393~481.
- [5] Mohan C, Lindsay B, Obermarck R. Transaction management in the R^* distributed database management system. *ACM Transactions on Database Systems*, 1986,11(4):378~396.
- [6] Levy E, Korth HF, Silberschatz A. An optimistic commit protocol for distributed transaction management. In: James C, Roger K, eds. Proceedings of the ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 1991. 88~97.
- [7] Yoon Y. Transaction scheduling and commit processing for real-time distributed database systems [Ph.D. Thesis]. Korea Advanced Institute of Science and Technology, 1994.
- [8] Haritsa JR, Ramamritham K, Gupta R. The PROMPT real-time commit protocol. *IEEE Transactions on Parallel and Distributed Systems*, 2000,11(2):160~180.
- [9] Liu YS. *Advanced Database Technology*. Beijing: National Defence Industry Press, 2001 (in Chinese).

附中文参考文献:

- [9] 刘云生.现代数据库技术.北京:国防工业出版社,2001.