无用操作对带约束的可串行化检验的影响

刘惟一

(云南大学，昆明，650091)

THE INFLUENCE OF USELESS ACTIONS ON TESTING SERIALIZABILITY WITH CONSTRAINTS

Liu Weiyi

(Yunnan University, Kunming, 650091)

ABSTRACT

The influence of useless actions on testing serializability of a schedule is discussed in this paper. At first, we define the write-write、write-read、read-write and read-read constraints. It is shown that if there is no useless actions in the schedule, the complexity of testing serializability under the constraints is in polinomial time. We extend the problems to the multiversion systems, then some good conclusions polynomial are obtained.

摘 要

本文讨论"无用写"给调度的可串行化检验带来的影响. 首先，我们定义了写—写、写—读、读—写和读—读约束. 证明了在上述约束下，若不存在无用写，调度的可串行化检验的时间复杂性是多项式界的. 进而将上述问题推广到多版本系统，得到一系列类似的有用的结论.

§1. 引言

并发控制是数据库系统的最重要的问题之一. 一般用是否可串行化来作为衡量并发控制正确性的准则. 不幸的是，一个调度的可串行化检验往往是NP —完全问题. 因而人们进一步讨论在一定限制条件下可串行化检验的时间复杂性.

1990 年 1 月 7 日收到，1990 年 6 月 4 日定稿.

Ibaraki 等人在文献[1, 4] 中定义了写—写、写—读、读—写和读—读约束，并且利用事务输入输出图$TIO_c(h)$ 和约束图$CG_c(h)$ 的方法分别讨论了单版本和多版本系统中上述约束下的可串行化理论. Sethi 在文献[2, 5] 中讨论了无用写给“严格串行化检验”带来的影响. 本文利用$TIO_c(h)$ 方法把Sethi 的工作推进了一大步，证明了若不存在无用写，在上述约束下，调度的可串行化检验的时间复杂性是多项式界的. 进而利用$CG_c(h)$ 方法将单版本下的结果推广到多版本系统，也得到了类似的有用的结论.

§2. 基本概念

考虑到“边际效应”，我们假定事务集合T 由$\{T_o, T_i, \cdots, T_n, T_f\}$ 组成. T_o 是初始只写事务，它写所有的数据项；T_f 是终结只读事务，它读每一个数据项.

记$R_i[X]$ 为事务T_i 中的“读步”，其中$X\subseteq D$, D 为数据项集合. $R_i[X]$ 是不可分的，即在执行中不允许其它操作插入. 类似地记T_i 中的“写步”为$W_i[X]$, 它也是不可分的.

为了与多版本情况相区别，我们称单版本情况的调度为“历史”. T 上的一个历史h 是一个二元组t=(steps(T), <), 其中steps(T) 是事务集T 中所有写步和读步的集合，< 是steps(T) 上的一个全序. 对于steps(T) 中的两步A 和B, 若A<B, 则在h 中A 先于B. 对于数据项X, 若$W_j[X]$ 是h 中$R_i[X]$ 前面最邻近于$R_i[X]$ 的一个写X 操作，就说T_i 从T_j 中读X. 两个历史h 和h′, 如果对每个数据项X 和指定的i, j, 当且仅当h′ 中T_i 从T_j 读X 时，h 中有T_i 从T_j 读X, 就称h 等价于h′, 记为h≡h′. 对于h=(steps(T), <), 如果每一事务执行完后才开始执行下一事务，就说h 是串行的. 若存在一个串行的h 使h′ ≡h, 称h′ 是可串行化的.

所谓多版本系统是指一个事务对某个数据项X 施行写操作时，允许X 的旧值不被新值修改(即保留多个版本), 并且一个读X 操作可以选择分配予一个适当的版本，以利于串行化.

给定h=(steps(T), <), T 上的一个多版本调度s 是一个二元组s=(h, I), 其中I 是一个由读步集$READ(T)=\{R_i[X]|R_i[X]\in steps(T)\}$ 到写步集$WRITE(T)=\{W_j[X]|W_j[X]\in steps(T)\}$ 的映射，使得对每个$R_i[X]$ 有$I(R_i[X])=W_i[X]$. 我们称映射I 为“编译”.

其实单版本系统中，h 蕴含了一个由读步集到写步集的映射I^*, 使$I^*(R_j[X])=W_i[X]$, 其中$W_i[X]$ 是h 中$R_j[X]$ 前最邻近$R_j[X]$ 的一个写X 步. I^* 称为标准编译.

多版本系统中，调度的等价是这样定义的：设h=(steps(T), <), h′=(steps(T), <′) 是T 上的两个历史，s=(h, I), s′=(h′, I′) 是T 上的两个调度. 若I=I′ 称s′ 与s 等价，记为s≡s′. 如果一个多版本调度s′=(h′, I′) 等价于一串行调度$s=(h, I_h^*)$, 其中h 是单版本意义下的串行历史，则称s′ 是可串行化的. 在多版本意义下，若存在一个编译I, 使(h, I) 是可串行化的，则称h 是可串行化的.

为了方便，我们给每个版本一个下标，用来表示多版本调度.

例：$h'=W_0[XY]R_2[X]W_1[XY]W_2[Y]R_3[XY]W_3[Y]R_f[XY]$

$s'=(h', I')=W_0[X_0Y_0]R_2[X_0]W_1[X_1Y_2]W_2[Y_1]R_3[X_1Y_2]W_3[Y_3]R_f[X_1Y_3]$

$h=W_0[XY]R_2[X]W_2[Y]W_1[XY]R_3[XY]W_3[Y]R_f[XY]$

$s=(h, I_h^*)=W_0[X_0Y_0]R_2[X_0]W_2[Y_1]W_1[X_1Y_2]R_3[X_1Y_2]W_3[Y_3]R_f[X_1Y_3]$

可以看出$I'=I_h^*$, s≡s′. s 是一个串行调度，所以s′ 是可串行化调度，也就有h′ 在多版本意义下是可串行化的.

无论单版本或是多版本系统，若不加任何限制，h 的可串行化检验都是NP --完全问题[1]. 为此我们给出一些限制，考察在这些限制下，h 的可串行化检验是否是多项式时间的.

单版本情况下，设h=(steps(T), <) 是T 上的一个历史.

(1) 写—写限制(ww): 对某个数据项X, 若$W_i[X]<W_j[X]$, 则T_i 必须串行在T_j 之前.

(2) 写—读限制(wr): 对于某个X, 若$W_i[X]<R_j[X]$, 则T_i 必须串行在T_j 之前.

(3) 读—写限制(rw): 对于某个X, 若$R_i[X]<W_j[X]$, 则T_i 必须串行在T_j 之前.

(4) 读—读限制(rr): 对于某个X, 若$R_i[X]<R_j[X]$, 则T_i 必须串行在T_j 之前.

一个历史h 在上述(1)~(4) 限制下是可串行化的，则称h 分别属于WW, WR, RW 和RR. 历史h 在ww 和rw 限制下是可串行化的，则记h 属于WW+RW, 其余类推.

对于多版本情况，我们类似地定义上述限制下的可串行化类为MWW, MWR, MRW 和MRR, 这里M 代表多版本.

对于调度s=(h, I), $W_i[X]$ 是h 中的写X 步. 如果不存在这样的$R_j[X]$ 使$I(R_j[X])=W_i[X]$, 称$W_i[X]$ 为无用写.

§3. 单版本系统的可串行化检验

文献[1] 中定理8.1 和8.2 指出WW 和WR+RW 的成员检验(即检验h 是否属于WW 和WR+RW) 是多项式时间的，而WR, RW, RR, WR+RR 和RW+RR 的成员检验是NP —完全的. 下面讨论h 中不存在无用写时，WR, RW, ⋯ 的成员检验的时间复杂性.

定义3.1: h=(steps(T), <) 的事务描述图是一个二元组TDG(h)=(T, A). 这里T 为结点集，每个事务是一个结点；A 为边集，若T_j 从T_i 读数据项X, 则T_i, T_j 间有边，记为$T_i \rightarrow T_j \in A$, X 为边的标志. 若事务$T_k$ 有一无用写$W_k[X]$, 则增加一个哑结点$T'_k \in T'$ 和哑边$T_k \rightarrow T'_k \in A'$, h 的事务输入输出图TIO(h)=(T∪T', A∪A'). 若h 中不存在无用写，则TIO(h)=TDG(h).

定义3.2: 设c 是约束集合(例如ww, wr, ⋯). 若c 强制T_i 必须串行在T_j 之前，则TIO(h) 中加入排斥边$T_i \rightarrow T_j$. TIO(h) 加上所有排斥边的闭包，记为$TIO_c(h)$.

利用$TIO_c(h)$ 可以得到下面的定理.

定理3.1 ([1] 中引理8.1): 设C 是约束集c 限制下的一类可串行化历史. 在满足条件P 时，当且仅当$TIO_c(h)$ 无圈，则h 属于C.

条件P: 对每个数据项X 和任意的写X 事务T_i, T_j(至少有一个为有用写), 在$TIO_c(h)$ 中存在一条由T_i 到T_j 或者由T_j 到T_i 的路.

关于无用写我们有下面的定理，称之为两条排斥律.

定理3.2 ([2] 中定理3.1): 设h 中不存在无用写；T_i, T_j, T_k 是h 中三个事务，其中T_i, T_j 写数据项X, T_k 从T_i 中读X.

(1) 若T_j 必须串行在T_k 之前(记为(T_j, T_k)), 则T_j 也必须串行在T_i 之前. 也就是若(T_j, T_k), 则(T_j, T_i).

(2) 若(T_i, T_j), 则(T_k, T_j).

利用所引文献中的两个定理(即定理3.1 和3.2), 我们得到下面的结论.

定理3.3: 设h 中不存在无用写，检验h 是否属于RW 的时间复杂性是多项式界的.

证明： 设u 是一个约束，它限制h 中不存在无用写．设c 是约束rw+u．$TIO_c(h)$ 可在多项式时间内由TIO(h) 得到([2] 中定理2.1)．$TIO_c(h)$ 为有向图，有向图的非圈性可多项式时间检验[3]．利用定理3.1，只需证明条件P 对于c 成立，则定理得证．

假定$W_i[X]$, $W_j[X]$ 是h 中两个写X 步．由于不存在无用写，h 中必有$R_k[X]$, $R_l[X]$ 分别从$W_i[X]$, $W_j[X]$ 读X．所以TIO(h) 中有边$T_i \to T_k$, $T_j \to T_l$．h 中必有$R_l[X]<W_i[X]$ 或者$R_k[X]<W_j[X]$，否则，若两个写X 步在两个读X 步之前，这与假定矛盾．不失一般性，不妨设$R_k[X]<W_j[X]$．由rw 限制有(T_k, T_j)．这样$TIO_c(h)$ 中有$T_i \to T_k$, $T_k \to T_j$, T_i 到T_j 有路．证毕．

假定h=(steps(T), <) 中不存在无用写；$W_i[X]$, $W_j[X]$ 为h 中两个写X 步，h 中必有$R_k[X]$, $R_l[X]$ 分别从$W_i[X]$ 和$W_j[X]$ 中读X．

引理3.1： 在rr 约束下，若存在串行历史h′=(steps(T), <′) 与h=(steps(T), <) 等价，则h′ 中T_k 必串行在T_j 之前或T_l 必串行在T_i 之前．

证明： 不失一般性，不妨设h 中$R_k[X]<R_l[X]$．由rr 限制有(T_k, T_l)．假定T_j 串行在T_k 前，即(T_j, T_k)，由定理3.2(1) 有(T_j, T_l)．又由定理3.2(2) 有(T_l, T_i)．这样h′ 中就有$R_k[X]<R_l[X]$, $R_l[X]<W_i[X]$，这与$R_k[X]$ 从$W_i[X]$ 中读X 矛盾．所以假定(T_j, T_k) 不真，我们有(T_k, T_j)．类似地在$R_l[X]<R_k[X]$ 的前提下可得(T_l, T_i)．证毕．

定理3.4： 设h 中不存在无用写，检验h 是否属于RR 的时间复杂性是多项式界的．

证明： 设u 是一个约束，它限制h 中不存在无用写．设c 是约束rr+u．$TIO_c(h)$ 可在多项式时间内由TIO(h) 得到．有向图$TIO_c(h)$ 的非圈性可多项式时间检验．利用定理3.1，只需证条件P 对c 成立，则定理得证．

设$W_i[X]$, $W_j[X]$ 是h 中两个写X 步．由于不存在无用写，h 中必有$R_k[X]$, $R_l[X]$ 分别从$W_i[X]$, $W_j[X]$ 中读X．设h 中$R_k[X]<R_l[X]$，由引理3.1 有(T_k, T_j)．这样$TIO_c(h)$ 中有$T_i \to T_k$, $T_k \to T_j$, T_i 到T_j 有路．证毕．

文献[2] 中定理2.1 证明：若不存在无用写，“严格串行化检验”(本文称为WR 的成员检验) 是多项式时间的．不存在无用写时，由于WR, RW 和RR 的成员检验是多项式时间的，显然WR+RR 和RW+RR 的成员检验也是多项式时间的．

§4. 多版本系统的可串行化检验

文献[1] 中定理13.2 指出：MWR, MRR 和MWR+MRR 的成员检验是NP —完全的．下面讨论不存在无用写时，上述检验的时间复杂性．

定义4.1： 设h=(steps(T), <) 是T 上的一个历史，c 是一个约束集．h 的约束图是一个二元组$CG_c(h)$=(T, A)．这里T 为结点集，每个事务是一个结点；A 为边集，若c 强制T_i 必须串行在T_j 前，则$CG_c(h)$ 中有边$T_i \to T_j \in A$．

用$CG_c(h)$ 可以得到下面两个定理．

定理4.1([1] 中定理10.1)： 设c 是包含rw 的约束集．当且仅当$CG_c(h)$ 非圈，则h 属于MC．其中MC 是多版本意义下，约束集c 限制下的一类可串行化调度．

定理4.2([1] 中定理10.2)： 当且仅当$CG_{rw+ww}(h)$ 非圈，则h 属于MWW．

为了说明方便，先给出一个引理．

引理4.1： 若h 中$W_i[X]<R_k[X]$, $W_j[X]<R_l[X]$，则$W_i[X]<R_l[X]$ 或者$W_j[X]<R_k[X]$ 之一

为真.

证明：h 中$W_i[X]$和$W_j[X]$必有一个在前. $W_i[X]<W_j[X]$, 由$W_j[X]<R_l[X]$有$W_i[X]<R_l[X]$. 若$W_j[X]<W_i[X]$, 由$W_i[X]<R_k[X]$有$W_j[X]<R_k[X]$. 证毕.

定理4.3: 设h 中不存在无用写. 检验h 是否属于MWR 的时间复杂性是多项式界的.

证明：设u 是一个约束，它限制h 中不存在无用写. 设约束c=wr+u, MC 是约束c 限制下的一类可串行化调度(多版本意义). 由定理4.2 可知，当且仅当$CG_{rw+ww}(h)$非圈，则h 属于MWW. 有向图$CG_{rw+ww}(h)$的非圈性可多项式时间检验. 因此MWW 的成员检验是多项式时间的. 这里只需证MC=MWW, 则定理得证.

设I 是对于h 的任一编译，$W_i[X]$, $W_j[X]$是h 中的两个写步. 由于不存在无用写，必有$R_x[X]$和$R_l[X]$使得$I(R_k[X])=W_i[X]$, $I(R_l[X])=W_j[X]$.

首先考察$TIO_{ww}((h, I))$. 根据假定$TIO_{ww}((h, I))$中有边$T_i\rightarrow T_k$, $T_j\rightarrow T_l$, 所以$TIO_{ww}((h, I))$中包含了限制边wr 一边.

其次考察$TIO_c((h, I))$. 由wr 限制有(T_i, T_k), (T_j, T_l). 根据引理4.1, $W_i[X]<R_l[X]$与$W_j[X]<R_k[X]$必有一个为真，不妨设$W_i[X]<R_l[X]$为真. 由wr 限制有(T_i, T_l), 又由定理3.2(1)可得(T_i, T_j). 于是$TIO_c((h, I))$中有边$T_i\rightarrow T_j$, 即$TIO_c((h, I))$包含了限制边ww 一边. 我们得到MC=MWW. 证毕.

定理4.4: 设h 中不存在无用写，检验h 是否属于MRR 的时间复杂性是多项式界的.

证明：设u 是一个约束，它限制h 中不存在无用写. 设约束c=rr+u, MC 是约束c 限制下的一类可串行化调度(多版本意义). 由定理4.1 可知，当且仅当$CG_{rw+wr}(h)$非圈，则h 属于MRW+MWR. $CG_{rw+wr}(h)$的非圈性可多项式时间检验. 因此MRW+MWR 的成员检验是多项式时间的. 这里只需证明MC=MRW+MWR, 则定理得证.

设I 是对于h 的任一编译，$W_i[X]$, $W_j[X]$是h 中的两个写步. 由于不存在无用写，必有$R_k[X]$和$R_l[X]$使$I(R_k[X])=W_i[X]$, $I(R_l[X])=W_j[X]$. h 中$R_k[X]$和$R_l[X]$必有一个在前，假定$R_k[X]<R_l[X]$.

首先考察$TIO_{rw+wr}((h, I))$. 由引理4.1 可知，h 中有$W_i[X]<R_l[X]$或$W_j[X]<R_k[X]$. 根据假定$R_k[X]<R_l[X]$和$W_i[X]<R_k[X]$, 取$W_i[X]<R_l[X]$. 由wr 限制有(T_i, T_l). 于是由定理3.2(1)有(T_i, T_j), 又由定理3.2(2)有(T_k, T_j). 由wr 限制有(T_j, T_l). 于是有(T_k, T_l). $TIO_{rw+wr}((h, I))$中包含了限制边rr 一边.

其次考察$TIO_c((h, I))$. 根据假设$TIO_c((h, I))$中有$T_i\rightarrow T_k$, $T_j\rightarrow T_l$. $TIO_c((h, I))$中包含限制边wr 一边. 若(h, I) 中不存在$R_k[X]<W_j[X]$和$R_l[X]<W_i[X]$, 也就是不存在rw 约束的前提，这样$TIO_c((h, I))$满足rw+wr 限制. 若(h, I) 中有$R_k[X]<W_j[X]$或$R_l[X]<W_i[X]$. 由假定$W_i[X]<R_k[X]$, $R_k[X]<R_l[X]$, 取$R_k[X]<W_j[X]$. 根据$R_k[X]<R_l[X]$, 由引理3.1 有(T_k, T_j), $TIO_c((h, I))$中有边$T_k\rightarrow T_j$. 这样$TIO_c((h, I))$包含了限制边wr 一边和rw 一边. MC=MRW+MWR. 证毕.

若不存在无用写我们得到MWR 和MRR 的成员检验是多项式时间的，显然MWR+MRR 的成员检验也是多项式时间的.

参考文献

[1] IBARAKI, T., KAMEDA, T., MINOURA, T., "Serializability with Constraints", ACM Trans. Database Syst. 12, 3(Sep, 1987), 429–452.

[2] SETHI, R., "Useless Actions Make a Difference: Strict Serializability of Database Updates", JACM, 29,2 (Apr, 1982), 394–403.

[3] AHO, A.V., HOPCROFT, J.E. and ULLMAN, J.D., "The Design and Analysis of Computer Algorithms", Addison-Wesley, Reading, Mass, 1974.

[4] KATOH, N., IBARAKI, T., and KAMEDA, T., "Cautions Transaction Schedulers with Admission Control", ACM Trans. Database Syst. 10,2 (June, 1985), 205–229.

[5] SETHI, R., "A Model of Concurrent Database Transactions", 22nd Annual Symposium on Foundations of Computer Science. Nashville, Tenn., (Oct, 1981), 175–184.

全国模式识别与人工智能学术会议(PRAI90) 在哈尔滨召开

中国自动化学会模式识别与机器智能专业委员会和中国计算机学会人工智能与模式识别专业委员会于1991年7月24至26日在哈尔滨工业大学召开了中国自动化学会第八届模式识别与机器智能和中国计算机学会第三届模式识别联合学术会议.

全国40余所高校、科研和生产部门的120余名专家、教授和研究人员，克服洪水阻隔的重重困难，汇聚一堂. 会议由程序委员会主席戴汝为教授致开幕词，宣国荣教授和戴汝为教授分别就"模式识别进展"和"计算机智能接口与人工神经网"作了综述报告. 会议代表按图象处理、文字识别、语言识别与信号处理、人工神经网络、模式识别等专题进行了认真交流与讨论. 会议期间还参观了哈尔滨工业大学机器人研究所和有关实验室.

本次会议是在全国七五科技攻关取得丰硕成果的基础上召开的. 会议论文集收录了175篇论文. 它们反映了我国模式识别与人工智能领域在七五期间的研究成果，跟踪了国际高技术领域的进展，并有广泛的应用背景和国民经济意义. 宣国荣教授在闭幕词中代表程序委员会高度评价了由本届会议论文集中反映出来的我国模式识别与人工智能领域在理论和实践上取得的迅速进展，并对哈尔滨工业大学筹办这次会议表示感谢.

会议期间自动化学会模式识别与机器智能专业委员会根据自动化学会要求，还就专业委员会换届工作进行了酝酿.

经过二学会专业委员会到会委员讨论，下届联合学术会议初步定于1993年春在成都召开.