

静的二相ロックングアルゴリズムの性能解析

任 景飛、 高橋 豊、 長谷川 利治
京都大学工学部

同時実行制御アルゴリズムはデータベース管理システムが複数ユーザのデータベースへの同時アクセスを可能にするものである。本稿では静的な二相ロックングアルゴリズムを用いるデータベースにおいて、トランザクションが指数間隔で発生され、アクセス集合が一様分布に選ばれる環境での平均応答時間などの性能評価量を求める。ブロックされたトランザクションのアクセス集合が過去の履歴とは独立に選ばれるという仮定のもとで、システムの確率的挙動を二次元のマルコフ過程に帰着し、Neutsの定理を基に、マルコフ過程の厳密解を求める。数値結果とシミュレーション結果の比較から仮定の妥当性が示される。

Performance Analysis of a Concurrency Control Method with Static Locking Algorithm

Jing Fei REN Yutaka TAKAHASHI Toshiharu HASEGAWA

Department of Applied Mathematics and Physics, Kyoto University

The performance of the database system with static locking algorithm highly depends on the level of the contention for hardware as well as data resource. Transactions are generated according to a Poisson process and their access set are uniformly distributed in the database considered. Average response time is a main performance measure. Under an assumption of independent resampling, the system is modeled as a Markov process with infinite states. The matrix geometric method developed by Neuts is used in analysis. Numerical results are presented to give an insight into the trade off among system parameters.

1 まえがき

データベースシステムにおける同時実行制御法はあたかも各ユーザが単独にデータベースを利用しているかのごとく内容の首尾一貫性を保ちながら、複数のユーザのデータベースへの同時アクセスを可能にするものである。この目標を達成するためには同時にデータベースへアクセスするトランザクション間の干渉を避ける必要がある。過去十数年の間に、数多くの同時実行制御アルゴリズムが提案改良された。これらのアルゴリズムを大別すると、三つのクラスにわけられる。その一つは時刻印に基づく方法である。この方法で、トランザクションの直列可能の順序が予め決められ、トランザクションが割り当てられた時刻印に従って実行される。二番目の方法は楽観的なアルゴリズムである。このアルゴリズムでは、到着したトランザクションがとにかく実行され、データの首尾一貫性を保てない恐れが発生したときのみ関係のあるトランザクションがアボートされ、最初から再実行される。三番目の方法は今よく使われている二相ロック方法である。このアルゴリズムでは、実行の全過程において、首尾一貫性を崩さないで予め確認されたトランザクションだけが処理を開始される。その結果、必要以上に処理の開始が遅らされる可能性があるために、悲観的なアルゴリズムともいえる。

二相ロックアルゴリズムでは、トランザクションがあるデータ項目にアクセスする前にこの項目のロックを保有しなければならない。また、いったんトランザクションがロックを解放し始めると、決してロックの要求を出さない。本稿では専有ロックだけを考えることにする。一つのデータ項目が任意の時間において、高々一つのトランザクションでロックがなされる。そうすることにより、トランザクション間の干渉を避け、データベースの首尾一貫性を保つことが可である。

二相ロックアルゴリズムはロックの獲得方式により、動的なロックと静的なロックに大別できる。動的なロックでは、ロックはトランザクションの実行段階で、必要になった時点で、一つずつ要求される。動的なロックの利点はデータ資源の有効利用である。

本稿では静的な二相ロックアルゴリズムを考察する。このアルゴリズムは次ぎのように機能する。トランザクションは予めアクセスしようとする全てデータ項目を宣言する。もし、これらのデータ項目に一

つでもほかのトランザクションからのロックがなされていないければ、このトランザクションはロックの権利を獲得し、全てのアクセス項目のロックを完了して始めて、実行が開始される。もしアクセス項目の一つにでも、ほかのトランザクションからのロックがなされていれば、このトランザクションはブロックされ、ロックをかけずに、バッファでロックの解放を待つ。このようなロックメカニズムにより、アクセス項目集合に共通の部分がないトランザクション同志だけは同時に実行される。トランザクションは実行が終わると、保有していた全てのロックを解放して、ブロックされているトランザクションにロックチャンスを与える。

静的なロックアルゴリズムに基づいて運用されているデータベースシステムの確率的挙動の解析に関してはPotierとLeblanc¹はこの分野の先駆者である。彼らはトランザクションが有限個数の端末により、指数間隔で発生されるシステムを考察した。本稿で考察するのは基本的に彼らのモデルを拡張したものである。一方、MitraとWeinberger²はポアソン源からの到着したトランザクションがもしほかの実行されているトランザクションと衝突したら、直ちに廃棄されるいわゆる呼損モデルを解析した。もう一つの仮定はシステムに無限のサーバがいる仮定である。このモデルはかなり特別な場合に対応しているが、任意のアクセス集合サイズと任意のロック保有時間の分布を許すなどいくつかの面で、柔軟性を持っている。彼らはマルコフ過程の可逆性を生かして、積の形の解で、解析を厳密に行っている。また、MorrisとWong³はシステムに常に十分な数のトランザクションがある飽和システムを解析した。ロック過程をモデル化する際の工夫により、彼らはシステムのスループットの良い近似を与えている。また、Thomasian⁴はロックの要求と衝突の発生が決定的なシステムを解析した。本質的に、このモデルはいくつかの支配的なトランザクションのクラスだけを取り扱うときに使われる。

本稿では、静的なロックアルゴリズムを用いるシステムにおいてトランザクションがポアソン過程に従って生成され、ロックの要求が全項目集合に一様分布する環境での性能評価を試みる。性能評価はブロックされたトランザクションが再度ロックを要求するときに過去の要求に関して無記憶的に再び選ばれるという仮定のもとで、行われる。評価の主な尺度は平均応答時間で、最大スループットにも言及する。

本文は5節からなる。第2節にシステムとアルゴリズムを記述し、モデルを提案する。第3節にNeuts⁵の理論を使って、モデルを解析し、性能評価をする。数値結果は第4節で与えられ、第5節はまとめである。

2 システムとモデル

本文で考察するデータベースはユーザからのトランザクションでアクセスされるデータ項目の集合と考えられる。トランザクションの到着過程はポアソン過程に従い、要求されるサービス時間は指数分布に従うとする。各トランザクションがアクセスするデータ項目数は定数で、データベースに一様分布で選ばれる。もし、同時実行可能なトランザクション数に余裕があるならば、トランザクションは到着時点において、ロックを行う。ロックが保証されるトランザクションは無視できるほど短い時間に、全てのアクセスする項目にロックをかけてから、実行される。ロックが保証されないトランザクションはブロックされ、待ち行列の最後で待つ。同時実行可能なトランザクション数に余裕がない場合には、到着は単にブロックされ、待ち行列の先頭で待つ。トランザクションの処理が完了すると、持っていた全てのロックを解放するから、ブロックされたトランザクションは待ち行列の先頭から順番に再びロックの試みをする。もし、試み中、あるトランザクションがロックに失敗したら、または、実行可能なトランザクション数に余裕がなくなったら、以後のロックは中止され、次の処理完了時点まで、延ばされる。ロックに失敗したトランザクションは待ち行列の最後にもどる。上記のような複雑なシステムを厳密にモデル化するためには、ブロックされているトランザクションのアクセス集合を状態変数として記述する必要がある。しかし、通常のデータベースは膨大なデータ項目数を有することを考慮すると、定式化はほとんど不可能である。そこで、本稿では、解析を可能にするため一つの仮定を設ける。類似の仮定はほかの論文^{1,3,6,7}にでも見られる。

仮定：トランザクションがロック（再ロック）する度に、このトランザクションのアクセス集合と実行されているトランザクションのアクセス集合は過去の履歴に独立に、データベースから選ばれる。

この仮定のもとで、システムは図1のようにモデル化できる。図の中に、Fはブロックされたトランザクションにより形成される待ち行列である。実行を完了されたトランザクションの離脱時点でのみオンになる

スイッチ k_2 はブロックされたトランザクションがロックを要求するタイミングを制御している。マルチプログラミングレベル m に対応して、 m 人のサーバが同時に使える。システム内では、データ項目に対する競争のほか、サーバに対する競争も存在する。実行されているトランザクションの数がマルチプログラミングレベルと等しい時、到着したトランザクションは待ち行列の先頭に置かれ、ロック要求は次の離脱点に延ばされる。

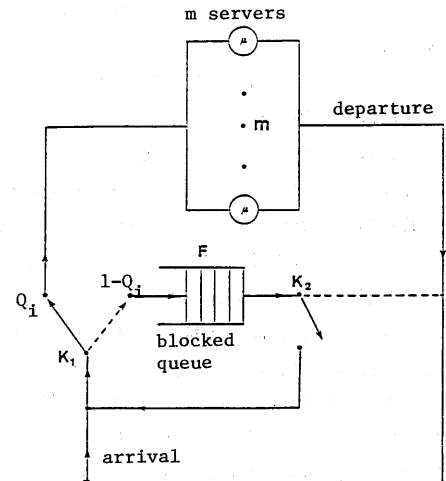


図1 システムモデル

さらに、モデルのパラメータを次のように定義する。

D: データベースにあるデータ項目総数。

S: トランザクションのアクセス集合にあるデータ項目数。

λ : トランザクションの到着率。

μ : サーバのサービス率。

m : サーバの数。

3 解析

前の仮定により、時刻 t での実行されているトランザクション数 $i(t)$ と待ち行列の長さ $j(t)$ を状態変数として選ぶとシステムの確率的挙動はマルコフ過程で記述される。平衡状態だけを考えるから、 $i(t)$ と $j(t)$ の平衡状態での値 i と j に注目する。実行されているトランザクション数が i のとき、トランザクションがロックに成功する確率は j に依存せず、次のようになる。

$$Q_i = \binom{D-iS}{D} / \binom{D}{S} \quad (0 \leq i < m)$$

$$Q_i = 0 \quad (i \geq m) \quad (1)$$

また確率 $Q^{k,i,j}$ をトランザクションの離脱の直後の状態が (i,j) である条件のもとで、 k 個のブロックされたトランザクションがロッキングに成功する確率であると定義する。(1)で定義された Q_i を用いて、 $Q^{k,i,j}$ は

$$Q^{k,i,j} = \begin{cases} Q_i Q_{i+1} \cdots Q_{i+k-1} (1-Q_{i+k}) & \text{if } i+k < m \\ & \text{and } j > k \\ Q_i Q_{i+1} \cdots Q_{i+k-1} & \text{if } i+k = m \text{ and } j \geq k \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

と書ける。

注意：全ての $j > k$ に対して、 $Q^{k,i,j}$ は $Q^{k,i,k+1}$ と等しい。この性質は後に使われる。

以上定義されたパラメータと確率 Q_i 、 $Q^{k,i,j}$ を用いて、遷移確率行列 G が求まる。ここで、無限次元の線形方程式を解く問題に直面する。標準的なアプローチは適当なサイズのところで行列 G を切り捨て、有限次元の線形方程式の解問題に直す方法である。しかし、この場合には、計算精度と計算の複雑さ間のトレードオフを考えなければならない。本稿では、標準的なアプローチの代わりに、Matrix Geometric⁵の方法を用いて、この問題を取り扱う。まず、マルチプログラミングレベル $m=3$ のとき、行列 G の例を見てみよう(図2参照)。

この例では、状態変数をスカラー化するために、状態 (i,j) は状態 n に対応させている。ただし、 $n=i+m \times j$ である。

この行列の最初の4列を別にして、同じタイプの列

は3列ごとに現れる。さらに、この性質は任意の m にもあると推論できる。したがって、行列 G はつぎのようにブロック行列の形で書ける。

| | | | | | | |
|----------|----------|-------|-------|-------|-------|-----|
| B_{00} | B_{01} | 0 | 0 | 0 | 0 | ... |
| B_{10} | B_1 | A_0 | 0 | 0 | 0 | ... |
| • | B_2 | A_1 | A_0 | 0 | 0 | ... |
| • | B_3 | A_2 | A_1 | A_0 | 0 | ... |
| • | B_4 | A_3 | A_2 | A_1 | A_0 | ... |
| • | • | • | • | • | • | • |

ここで、全ての A_k と B_k は m 次正方形行列で、 B_{10} と B_{01} はそれぞれ m 次の列と行ベクトルで、 B_{00} はスカラーである。付録1にこれらのブロックの中身が与えられている。この例からも分かるように、考察している確率過程は複雑な境界振舞いを持つ連続パラメータのGI/M/1タイプマルコフ過程である。

まず、行列 A を

$$A = \sum_{k=0}^{m-1} A_k \quad (3)$$

と定義しておく。行列 G と A_k の構造から、行列 A が既約で、なお、 $Ae=0$ という関係を満足することがわかる。ここで、 e と 0 はそれぞれ単位ベクトルとゼロベクトル

| | 00 | 10 | 20 | 30 | 11 | 21 | 31 | 12 | 22 | 32 | 13 | 23 | 33 |
|----|------------|------------------|-------------------|-------------------|---------------------|---------------------|-------------------|---------------------|---------------------|-------------------|---------------------|---------------------|-------------------|
| 00 | $-\lambda$ | λ | | | | | | | | | | | |
| 10 | μ | $-\lambda - \mu$ | λQ_1 | | $\lambda \bar{Q}_1$ | | | | | | | | |
| 20 | | 2μ | $-\lambda - 2\mu$ | λQ_2 | | $\lambda \bar{Q}_2$ | | | | | | | |
| 30 | | | 3μ | $-\lambda - 3\mu$ | | | λ | | | | | | |
| 11 | | μ | | | $-\lambda - \mu$ | λQ_1 | | $\lambda \bar{Q}_1$ | | | | | |
| 21 | | | $2\mu Q_{11}^1$ | | $2\mu Q_{11}^0$ | $-\lambda - 2\mu$ | λQ_2 | | $\lambda \bar{Q}_2$ | | | | |
| 31 | | | | $3\mu Q_{21}^1$ | | $3\mu Q_{21}^0$ | $-\lambda - 3\mu$ | | | λ | | | |
| 12 | | | μQ_{02}^2 | | μQ_{02}^1 | | $-\lambda - \mu$ | λQ_1 | | | $\lambda \bar{Q}_1$ | | |
| 22 | | | | $2\mu Q_{12}^2$ | | $2\mu Q_{12}^1$ | | $2\mu Q_{12}^0$ | $-\lambda - 2\mu$ | λQ_2 | | $\lambda \bar{Q}_2$ | |
| 32 | | | | | | $3\mu Q_{22}^2$ | | | $3\mu Q_{22}^1$ | $-\lambda - 3\mu$ | | | λ |
| 13 | | | | μQ_{03}^3 | | | μQ_{03}^2 | | μQ_{03}^1 | | $-\lambda - \mu$ | λQ_1 | |
| 23 | | | | | | $2\mu Q_{13}^3$ | | | $2\mu Q_{13}^2$ | | $2\mu Q_{13}^1$ | $-\lambda - 2\mu$ | λQ_2 |
| 33 | | | | | | | | | | $3\mu Q_{23}^3$ | | $3\mu Q_{23}^2$ | $-\lambda - 3\mu$ |

図2 $m=3$ の行列 G の例($\bar{Q}_i = 1 - Q_i$)

である。πを

$$\begin{aligned} \pi A &= 0 \\ \pi e &= 1 \end{aligned} \quad (4)$$

を満足する横ベクトルとすると、マルコフ過程の正再帰条件は

$$\pi A_{00} e < \sum_{k=0}^{m+1} (k-1) \pi A_{0k} e \quad (5)$$

となる。これはシステムの安定条件でもある。

Zを全ての成分がゼロの行列とし、(5)式が満足されるならば、行列方程式

$$\sum_{k=0}^{m+1} R^k A_k = Z \quad (6)$$

は $\text{sp}(R) < 1$ を満足する最小非負解がある。ここで、 $\text{sp}(R)$ はRのスペクトル半径で、全ての固有値のノルムの最大値と定義されている。Rが非負の場合に、対応する固有値は正の実数となる。

行列Rは $R(0)=Z$ から出発して、次の再帰的な式から計算できる。

$$R(N+1) = (A_{00} + \sum_{k=0}^{m+1} A_k R(N)^k) (-A_{11})^{-1} \quad (7)$$

再帰的な操作は

$$\text{Max}_{i,j} |R(N+1) - R(N)|_{i,j} < \epsilon \quad (8)$$

が満足されるまで続く。付録2で、収束に関する証明を示す。

状態確率をブロック行列Gに対応させて、

$$\begin{aligned} P_{00} &= \text{Prob}(i=0, j=0) \\ P_n &= \{ \text{Prob}(i=1, j=n), \text{Prob}(i=2, j=n), \dots, \text{Prob}(i=m, j=n) \} \end{aligned}$$

とする。また、行列B(R)を

$$B(R) = \begin{pmatrix} B_{00} & B_{01} & & \\ & & m+1 & \\ B_{10} & \sum_{k=1} R_{k-1} B_k & & \end{pmatrix} \quad (9)$$

と定義する。

$$\begin{aligned} (P_{00}, P_0) B(R) &= 0 \\ P_{00} + P_0 (I-R)^{-1} e &= 1 \end{aligned} \quad (10)$$

P_n は

$$P_n = P_0 R^n \quad (n \geq 1) \quad (11)$$

となる。Neutsの定理が上の結果を証明する道具になる。

確率ベクトル P_0 と行列Rを用いて、システムのいくつかの性能評価量の計算ができる。システムの最大スループットは安定条件(5)で決められる。

平均待ち行列長は

$$L = \sum_{j=1}^{\infty} j P_j e = \sum_{j=1}^{\infty} j P_0 R^j e = P_0 (I-R)^{-2} e \quad (12)$$

となる。Littleの公式により、トランザクションの平均待ち時間は

$$W = L / \lambda \quad (13)$$

となる。

平均応答時間は平均待ち時間と平均サービス時間の和になる、すなわち、

$$re = W + 1 / \mu \quad (14)$$

他のいくつかの性能評価量の計算も可能である。例えば、システムにブロックされているトランザクションがない確率は

$$NW = P_{00} + P_0 e \quad (15)$$

となり、到着したトランザクションがブロックせずに、直ちに実行される確率は

$$Nb = P_{00} + \sum_{k=0}^{\infty} P_k Q = P_{00} + P_0 (I-R)^{-1} Q \quad (16)$$

となり、実行されているトランザクションの平均数は

$$AA = \sum_{k=0}^{\infty} P_k E = P_0 (I-R)^{-1} E \quad (17)$$

となる。ここで、 $Q = (Q_1, Q_2, \dots, Q_m)^T$ 、 $E = (1, 2, \dots, m)^T$ と定義されている。

4 数値結果

前記の方法を用いて、いろいろのパラメータについて、応答時間の計算を行い、数値結果を図3~6に示す。

直感からわかるように、ロッキングに失敗する確率がゼロに近づくとき、すなわち、D/Sが非常に大きいとき、結果は通常のM/M/m待ち行列と一致する。逆に、衝突が常に存在する場合には、結果はM/M/1待ち行列と一致する。図3はこの二つの極端な場合の数値結果を示

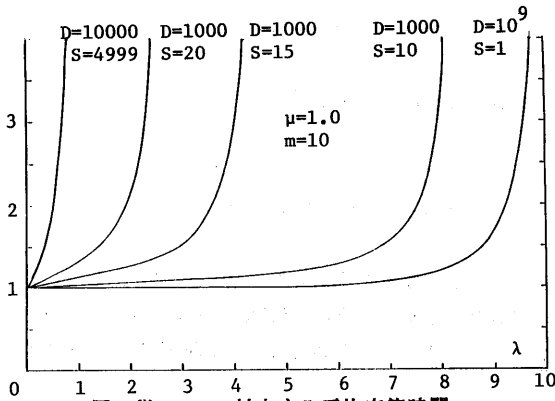


図3 様々のsに対応する平均応答時間re

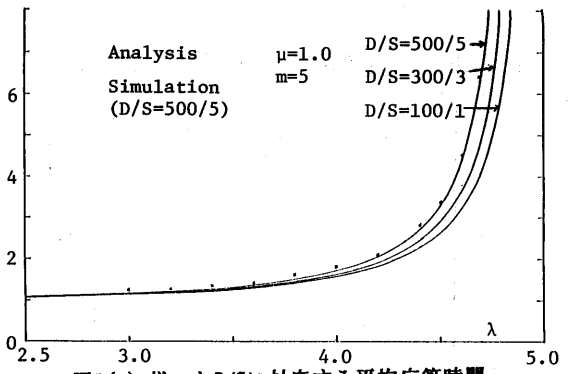


図5(a) 様々なD/Sに対応する平均応答時間

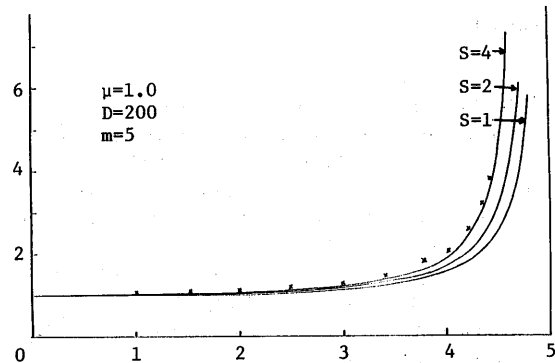


図4(a) パラメータsが小さいときの平均応答時間re

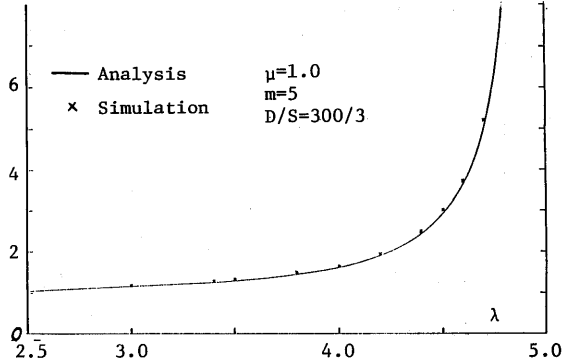


図5(b) D/S=300/3の時シミュレーションと解析の比較

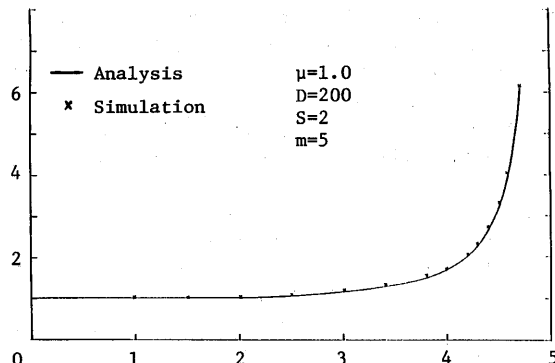


図4(b) s=2ときのシミュレーションと解析結果の比較

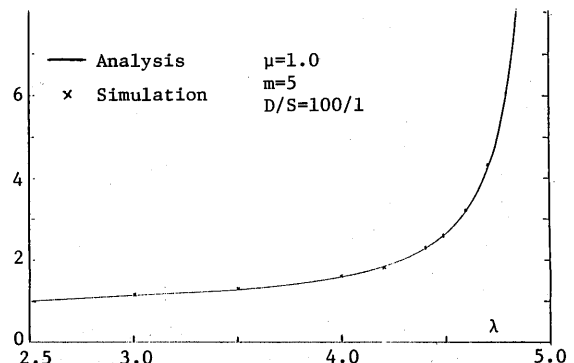


図5(c) D/S=100/1の時シミュレーションと解析の比較

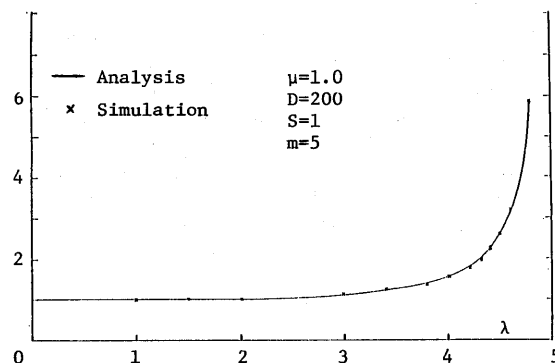


図4(c) s=1のときのシミュレーションと解析の比較

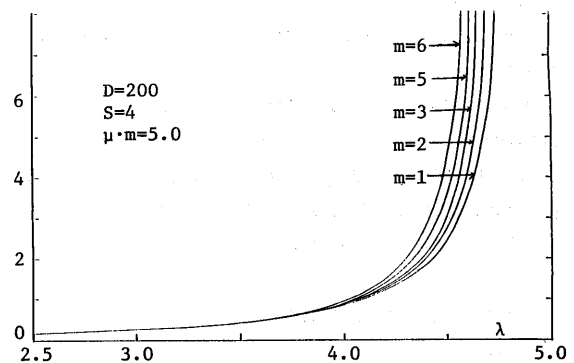


図6 総サービス率が一定の場合の平均待ち時間