オペレーティング・システム　42 － 2
（１９８９．２．２４）

カートリッジ型ＭＴにおける予測制御型負荷均衡アルゴリズムとその評価

山本　彰*　　坪井俊明**　　北嶋弘行*　　難波龍雄***　　土井　隆***
（*日立製作所システム開発研究所，　**日立マイクロコンピュータエンジニアリング，
***日立製作所小田原工場）

梗概

カートリッジ型ＭＴのための予測制御型負荷均衡制御アルゴリズムの提案とその評価結果を報告する。カートリッジ型ＭＴの制御装置においては、制御装置内にバッファメモリを設け、制御装置とＭＴの間では、複数ブロックの先読み／まとめ書きを実行する。したがって、各ＭＴを複数台の制御装置に接続する場合、１台のＭＴの先読み／まとめ書きデータが複数の制御装置に分散することをさけるため、各ＭＴをある１台の制御装置の制御下に置く方式をとる。この結果として、各制御装置に対する入出力負荷を均衡させるために、制御装置間で、各ＭＴを制御する権利（制御権）を移行させる必要が生じた。予測制御型負荷均衡アルゴリズムとは、制御装置内に組み込んだ解析モデルの性能予測結果に従って、制御権を移行すべきＭＴを決定する方式である。本講演では、以上の負荷均衡制御のための解析モデルとして、拡張漸近モデルを提案する。提案した予測制御型負荷均衡アルゴリズムの特徴を以下に示す。（1）高速アルゴリズム（2）クローズド型モデル（3）待ち時間の一部のみが漸近的に発生するという仮定を置いた点。

３ケースの実測データにより、提案アルゴリズムの評価を行ったところ、これらの評価範囲においては、提案アルゴリズムにより、最少の制御権移行回数で制御装置間の入出力負荷バランスが可能であるという結果が得られた。

A Predictive Control Algorithm for Load Balancing of Cartridge MTs and Its Performance Evaluation

Akira Yamamoto*, Tosiaki Tsuboi**, Hiroyuki Kitajima*, Tatsuo Nanba***, and Takashi Doi***
(*Systems Development Labolatory, Hitachi Ltd.****, **Hitachi Microcomputer Engineering Ltd.
***Odawara Works, Hitachi Ltd.)
****1099, Ohzenji, Asao-ku, Kawasaki, 215, Japan

Abstract

A predictive control algorithm for load balancing of cartridge MTs (Magnetic Tapes) is presented and evaluated. The controllers of cartridge MTs have beffer memory , and multiple blocks of data are collectively preloaded and written among the controllers and MT units. When each MT unit is connected with more than one controller, each MT unit is under the control of only one of the controllers to prevent distributing each MT unit's data among the controllers. Consequently, each MT unit's control right is exchanged among the controllers to balance the input/outout loads for the controllers. The predictive control algorithm makes this decision with the performance prediction using the analytic model embeded in the controllers. An extended asymptotic model is proposed as the analytic model. The presented control algorithm's characteristics are as follows: (1) high-speed algorithm (2) closed type model (3) assumption that only a part of wait time is asymptotic.

In the three cases of measurement results explained in this paper, the presented predictive control algorithm made possible to balance the input/output loads among the controllers through the minimum number of the control right's movements.

1．はじめに

ＭＴ(Magnetic Tape)の小型化、軽量化を目的として、ＭＴのカートリッジ化が進められている。カートリッジ型ＭＴでは、装置自体の小型化のために、ＭＴの高速スタート／ストップ動作には必須である真空カラムを除去している。このため、カートリッジ型ＭＴでは、制御装置内にバッファを設け、制御装置とＭＴの間で複数ブロックの先読み／まとめ書きを実行することにより、スタート／ストップ時間の増大により生ずる性能劣化を防止する方式をとる。[1),2)]

ＭＴの場合、ディスク装置などと異なり、ＣＰＵとの間でシングル・データ転送パス構成をとることも少なくないため、制御装置は１つのデータ転送パスしか持たないのが通常である。従って、高信頼化のため、マルチ・データ転送パス構成をとる場合、ＭＴを複数の制御装置と接続することになる。この時、１つのＭＴのデータを複数の制御装置のバッファに分散させてしまうことは、以下の理由で問題がある。

（１）１回の先読み／まとめ書き処理において、一括処理できる処理対象ブロック数が少なくなり、性能劣化につながる。

（２）制御装置ダウン時の障害波及度が増大する。

従って、カートリッジ型ＭＴにおいては、あるＭＴとデータ転送可能な制御装置を１台に限定するよう制御する方式を採る。（以下、あるＭＴとデータ転送可能な制御装置をそのＭＴの制御権を有する制御装置と呼ぶ。）しかし、以上のような方式をとると、どちらか一方の制御装置の制御権下にあるＭＴに入出力負荷が偏る場合がある。この入出力負荷の不均衡状態を解消するためには、適切なＭＴを選択して、制御装置間でそのＭＴの制御権を移行させる必要がある。本講演では、制御権を移行させるＭＴの選択方式として、制御装置内で収集した各ＭＴの定常的な性能特性データを用いた予測制御型負荷均衡アルゴリズムを提案する。予測制御型負荷均衡アルゴリズムとは、[3)~6)]負荷パターンの変動は負荷均衡制御周期に比べ充分長く、各ＭＴの定常的性能特性は変化しないという仮定に基づき、制御装置内に組込んだ解析モデルにより、収集した各ＭＴの性能特性データを用いて、予測計算を実行し、制御権移行後の負荷状況が最も均衡するという予測結果が得られたケースを選択する方式である。

制御権移行後の負荷状況を予測する計算は、通常の入出力処理の稼動中に実行するため、低オーバヘッドで実行できることが望ましい。さらに、この予測計算は、制御装置内で実行するため、制御装置内で収集可能なデータで計算可能でなければならない。本講演では、負荷均衡制御のために用いる予測アルゴリズムとして拡張漸近モデルを提案する。拡張漸近モデルは、[7)~10)]高速計算可能な漸近近似手法を、制御装置内で収集できる範囲のデータで解析可能な形に拡張したものである。ただし、本講演では、典型的な入出力系の構成であるチャネルとＭＴの転送速度が等しい構成をモデル化の対象とした。また、一回の選択処理で制御権移行対象とするＭＴ台数を１台とした。これは、予測計算のケース数を一定値以下とし、オーバヘッドを抑えるためである。

2．カートリッジ型ＭＴの動作

図１にカートリッジ型ＭＴを含む計算機システムの構成を示す。システム内には、チャネルＡ、制御装置Ａからなるデータ転送路、および、チャネルＢ、制御装置Ｂからなるデータ転送路、あわせて、２系列のデータ転送路が存在する。さらに、システム全体は、制御装置内のバッファを中間バッファとして、主記憶と制御装置内のバッファとの間のデータ転送処理、および、CPU内でのデータ処理を実行する上位系と制御装置内のバッファとＭＴ間のデータ転送を実行する下位系が並列に動作していることになる。この場合、制御装置内のバッファが上位系と下位系の同期をとることになる。例えば、入出力装置に対し読み取り動作を実行している場合、上位系は制御装置内のバッファにデータがなくなると、待ち状態に入ることになる。この待ち時間をバッファデータ待ち時間と呼ぶ。

各ＭＴはそれぞれ２台の制御装置に接続されている。各制御装置は、制御権を有しているＭＴとのみ、バッファとＭＴとの間の先読み／まとめ書きデータ転送処理を実行する。しかし、ＣＰＵ側からは、各ＭＴが制御装置Ａと制御装置Ｂに接続されているように見えるため、そのＭＴの制御権を持っていない方の制御装置に対しても入出力要求が発行される。以上の場合、各制御装置は以下に示すような制御方式をとる。制御装置が、チャネルからあるＭＴに対する入出力要求を受け付けた時、このＭＴの制御権を有している場合、自装置内のバッファを用いて、上位システムとの間でデータ転送処理を実行する。制御権を有していない場合、相手の制御装置内のバッファと上位システムのデータ転送処理を実行する。したがって、ある制御装置があるＭＴの要求を受け付けた時、もう一方の制御装置がこの入出力装置の制御権を有している方のバッファと上位システムとの間でデータ転送中の場合、バッファの転送路が確保できないことになり、待ち状態に入る必要が生ずる。以下、この待ち時間をバッファ転送路待ち時間と呼ぶ。ただし、各処理要求が以上述べた２種類の待ち状態に入る場合には、チャネルを離すため、他に実行可能な処理要求があれば、この入出力処理に入ることができる。

負荷均衡制御とは、各制御装置内のバッファに対して生ずる以上のバッファデータ待ち時間、バッファ転送路待ち時間、および、各ＭＴの定常的な性能特性を観測し、以上の待ち時間の差が一定値を超えたとき、それぞれの制御装置の待ち時間を均衡させるＭＴを選択し、制御装置間でＭＴの制御権を移行させるものである。本講演で提案する予測制御型負荷均衡方式とは、制御権移行後、各制御装置内のバッファに対して発生する以上の待ち時間の差を、収集した各ＭＴの性能特性データに基づき、解析モデルにより予測計算を行い、待ち時間の差が最も小さくなるという結果が得られたケースを選択する方式である。次章では、この予測方式について述べる。

３．漸近近似手法に基づく予測方式

予測計算のために用いる解析モデルの選択基準を以下の３項目とした。

選択基準１：クローズド型の解析モデル　　選択基準２：アルゴリズムの高速性

選択基準３：制御装置内で収集可能なデータで予測計算が可能であること

選択基準１はＭＴの動作特性に起因している。ＭＴは、通常１つのプログラムに占有されるため、ＭＴ１台に対応して、１つの処理要求が存在するというクローズド型のモデルを用いるのが、ＭＴの動作特性を反映できるためである。選択基準２は、この予測計算が各ＭＴに対する入出力要求と並行して実行する必要がある点から必須の基準である。選択基準３は、この計算を制御装置内で閉じて実行させるために、必要な基準である。

以上の選択基準を満たす解析モデルとして、ここでは拡張漸近モデルを提案する。拡張漸近モデルとは、講演者等が従来提案してきた漸近近似手法の拡張モデルである。漸近近似手法は、高速計算可能なクローズド型の解析モデルである。通常のクローズド型の待ち行列モデルは、正規化定数（待ち行列網モデルでは、各状態の生起確率をもとめる際、まず、各状態の生起確率の比を求め、この後、すべての生起確率の和が１になるという条件から、各状態の生起確率の比を正規化定数で割ることにより、最終解を得る。）を求める際、反復的な手続きを踏む必要があるため、計算量の点から実用的でない。ただし、ここでは、選択基準３に示したように、制御装置内で収集可能なデータで、予測計算を実行するため、モデルの拡張を行った。以下、具体的な内容について述べる。

３．１　モデルで用いる仮定

以下、提案する拡張漸近モデルで用いる仮定を、仮定１－仮定４に示す。さらに、本制御アルゴリズムでは、すでに述べたように、制御権移行後も各ＭＴの定常的な性能特性は変化しないと仮定して、予測計算を行う。この具体的な仮定内容を、仮定５、仮定６にまとめる。（仮定５、仮定６は、特に、漸近モデルを用いるために必要な仮定ではない。）

仮定１：ＩＦ（バッファがない時の上位系のスループット＜バッファがない時の下位系のスループット）

ＴＨＥＮ　システム全体のスループット＝上位系のスループット

ＥＬＳＥ　システム全体のスループット＝下位系のスループット[10]

仮定１は、中間バッファを有する系に漸近近似手法を適用する際の基本仮定[10]である。仮定１を適用すると、バッファがないとして算出したスループットが小さい方の系には、バッファデータ待ち時間は発生せず、大きい方の系のみにこの待ち時間が発生することになる。参考文献１で示した先読み／まとめ書きスケジューリング方式をカートリッジ型ＭＴに適用すると、制御装置とＭＴの間のスループットは、ＭＴの転送速度に等しい値にまで引き上げることが可能となる。従って、チャネルの転送速度とＭＴのデータ転送速度が等しい時には、CPUのデータ処理が存在する分だけ、上位系の処理時間が長くなり、上位系のスループットが下位系に比較して大きくなることはない。以上より、仮定１を前提とした場合、上位系の解析のみを行えばよいことになる。さらに、上位系には、バッファデータ待ち時間は発生しないことになるため、負荷均衡の制御としてはバッファ転送路待ち時間のみを考慮すればよいことになる。

次に、通常の漸近近似手法の基本仮定[11~9]を示す。

基本仮定：ＩＦ（平均待ち時間を０と仮定して算出した競合資源の利用率≦１）ＴＨＥＮ　競合資源の平均待ち時間＝０

ＥＬＳＥ　競合資源の平均待ち時間＝競合資源の利用率を１とするだけの待ち時間[11~9]

図２は、上記の基本仮定にしたがって算出される平均待ち時間をグラフ化したものである。この場合、待ち時間を０と仮定して算出した資源の利用率が１より小さいとき、資源の利用率を１とする待ち時間を求めると、負の値となる。待ち時間が負の値となることはないため、このような領域では待ち時間を０とするのが通常の考え方である。しかし、資源の混雑度をあらわす指標としてこの値をとらえた場合、図２に示したように、利用率を１とする待ち時間が－２０となるポイント１より、－４０となるポイント２の方がその資源の混雑度は明らかに小さいということになる。したがって、負荷均衡制御においては、資源の利用率を１とするだけの待ち時間が負の値となっても、この値をそのまま用いた方が負荷状況をより正確に把握できると考えられる。このため、ここでは以下の仮定を設けた。

仮定２：競合資源の平均待ち時間＝競合資源の利用率を１とするだけの平均待ち時間

以下、簡単のため、仮定２で示すように発生する待ちを漸近的に発生する待ちと呼ぶ。

仮定３：資源に対する各処理要求のアクセス当たりの平均待ち時間は等しい[9]。

通常、入出力系で用いるスケジューリング方式はFCFSであり、仮定３は、その資源のスケジューリング方式がFCFS(First Come First Served)である場合、しばしば用いられる仮定[9]である。

仮定４：各処理要求がＭＴに対して発行する入出力処理を実行する際、発生する待ち時間のうち、制御装置内で発生する待ち時間のみが漸近的に発生する待ち時間となる。

仮定４は、制御装置内で収集可能なデータで予測計算を可能とするために設けた仮定である。仮定４を前提としない場合、入出力処理を実行する際に発生する待ち時間の全てが漸近的に発生することを仮定していることになる。一方、仮定４を前提とした場合、待ち時間の一部が漸近的に発生することを仮定することになり、残りの待ち時間に関しては確率的に発生する面も考慮できるとも考えられる。以上より、仮定４の前提は問題ないと考えられる。

次に、仮定５、仮定６を示す。

仮定５：制御権移行には、直接関係しない各ＭＴの定常的な性能特性、具体的には、各ＭＴに対する入出力要求の発行時間間隔、一回の入出力要求当たりのデータ転送時間は、制御権移行後の系でも変化しないものとする。

仮定６：移行対象ＭＴをＭＴｊとした時、ＭＴｊを除いた制御権移行を行わないそれぞれの制御装置の制御権下にあるあわせて２つのＭＴの集合に関して、以下に示す仮定が成立するとする。すなわち、ＭＴｊを除いた制御装置Ａの制御権下にあるＭＴの集合の中の各ＭＴに、ＭＴｊの制御権移行が与える性能的影響は均等であり、制御権移行後の各ＭＴのスループット（単位時間当たり処理する入出力処理数）の増減率は等しいと考え、各ＭＴの間のスループット比は、制御権移行後も変化しないものとする。同様に、ＭＴｊを除いた制御装置Ｂの制御権下にあるＭＴの集合に関しても、各ＭＴの間のスループット比は、制御権移行後も変化しないものとする。

以上の仮定１－仮定６にしたがって、予測計算を行う。次節では、解析対象モデルについて述べる。

３．２　解析対象モデルの動作

本節では、解析対象モデルの構成、および、処理要求の動作についてのべる。表１には、本講演で用いる記号の定義をまとめる。

図３に上位系のモデルを示す。モデル内には各ＭＴに対応して処理要求が１つ存在する。処理要求はCPU上でデータ処理等を実行している時には、各処理要求対応に存在する入出力発行間隔サーバ上に滞在し、入出力処理を開始する時、まずチャネル・キューに入り、どちらか一方のチャネル側転送サーバが空くとチャネル・キューから出る。この時、すでにこの処理要求が入出力対象とするＭＴと同一の制御装置の制御権下にある別ＭＴに対する処理要求がもう一方のチャネル側転送サーバを占有して、データ転送を実行している時には、転送路キューにはいる。この後、占有されていたチャネル側転送サーバが空くとチャネル側転送サーバをつかみ、制御装置内のバッファと主記憶の間の転送処理に入ることになる。したがって、以上の解析対象モデルでは、仮定２で示した競合資源に相当する資源は、チャネル側転送サーバということになる。通常の解析モデルでは、入出力発行間隔時間、チャネル側転送時間等を入力データとし、チャネル待ち時間＋バッファ転送路待ち時間を計算結果として算出する。しかし、負荷均衡制御は制御装置内で実行するため、モデルの入力とするデータは制御装置内で計測可能でなければならない。

ここでは、入出力発行間隔時間＋チャネル待ち時間が、御装置からみた見掛け上の入出力発行間隔時間と解釈できる点に着目した。従って、図３に示したように、入出力発行間隔時間＋チャネル待ち時間を、すなわち、制御装置入出力発行間隔時間として定義し、各ＭＴごとに以下の性能特性データを観測するようにした。ただし、具体的な計算方法は、次節で述べる。

観測データ１：平均制御装置入出力発行間隔時間　　観測データ２：平均チャネル側データ転送時間

観測データ３：スループット

図４は、制御権移行後の上位系のモデル、モデルの構成、処理要求の動き、および、表１に示した記号と用語の関係を示したものである。ここでは、ＭＴｊの制御権を制御装置Ａから制御装置Ｂに移行した場合についてまとめている。図３と異なる点は、ＭＴｊの制御権の移行が行われたという点のみで、他の点については変化がない。

以下、移行対象ＭＴの具体的な選択方式について述べる。制御装置Ａの制御権下にあるＭＴの集合すべてに関して、それぞれ１台のＭＴの制御権を移行した後の、両制御装置に発生する平均バッファ転送待ち時間を算出する。この時、両制御装置に発生する平均待ち時間の差が最も小さいＭＴを移行対象ＭＴとして選択する。次節では、具体的な計算方式を示す。

３．３　予測計算方式

まず、制御権移行を行う前、すなわち、図３の構成に対応した制御装置Ａ、制御装置Ｂの平均バッファ転送待ち時間の算出方式について述べる。

通常、解析モデルを用いて評価を行う場合、処理要求の集合を評価単位とすることが多い。これは、個々の処理要求を独立して評価すると計算量が膨大になるためである。以下、評価単位とする処理要求の集合をクラスと呼ぶ。クローズド型モデルの場合、各処理要求の動作特性を平均化する際には、スループット比で重みづけを行う。制御権移行を行う前のモデルの解析に当たっては、制御装置Ａの制御権下にあるＭＴの集合、制御装置Ｂの制御権下にあるＭＴの集合をそれぞれ１つのクラスにまとめる。従って、制御装置Ａ、制御装置Ｂの平均制御装置入出力発行間隔時間、平均チャネル転送時間は、以下に示す式で表すことができる。

$$S_A = \sum_{i\in A}\{S_i * \lambda_i(A,B) / (\sum_{i\in A}\lambda_i(A,B))\} \quad (1)$$

$$X_A(A,B) = \sum_{i\in A}\{X_i(A,B) * \lambda_i(A,B) / (\sum_{i\in A}\lambda_i(A,B))\} \quad (2)$$

$$S_B = \sum_{i\in B}\{S_i * \lambda_i(A,B) / (\sum_{i\in B}\lambda_i(A,B))\} \quad (3)$$

$$X_B(A,B) = \sum_{i\in B}\{X_i(A,B) * \lambda_i(A,B) / (\sum_{i\in B}\lambda_i(A,B))\} \quad (4)$$

この時、仮定2、仮定3を適用すると、それぞれの制御装置の平均チャネル待ち時間を含む制御装置入出力発行間隔時間、平均チャネル転送時間が既知で、平均バッファ転送待ち時間が漸近的に発生することを仮定すると、それぞれの制御装置の平均バッファ転送待ち時間は、次式で表わすことができる。

$W_A(A, B)=(N_A-1)S_A-X_A(A, B)$ (5) $W_B(A, B)=(N_B-1)S_B-X_B(A, B)$ (6)

次に、MTjの制御権を制御装置Aから制御装置Bに移行した後のそれぞれの制御装置の平均バッファ待ち転送時間の算出方式について述べる。仮定5にしたがうと制御権移行後の各MTの入出力発行時間間隔、チャネル転送時間は変化しないことになる。しかし、チャネル待ち時間は、制御権移行後変化すると考えるのが妥当である。したがって、制御装置入出力間隔時間の中には、チャネル待ち時間が含まれるため、移行後の平均バッファ待ち時間の算出の際、移行前の構成で計測した制御装置入出力発行間隔時間を既知として使用するのは、無理がある。ここでは、仮定3に着目する。すなわち、チャネル待ち時間は、待ちキューが制御装置A、制御装置Bで共通であり、スケジューリングはFCFSであるため、全てのMTのチャネル待ち時間は等しいと仮定する。以上の条件は、制御権移行を行なう前の構成においても、行なった後の構成においても成立する。これより、次式が得られる。

$T_{i_1}'(A, B)=T_{i_2}'(A, B)=T_{i_3}'(A, B)$ ・・・・(7)

$T_{i_1}'(A-j, B+j)=T_{i_2}'(A-j, B+j)=T_{i_3}'(A-j, B+j)$ ・・(8)

7式、8式より、次式が展開できる。

$T_i'(A-j, B+j)=T_i'(A, B)+\alpha$ $(\forall i \in A+B)$ (9)

9式よりMTjの制御権を移行した後の制御装置入出力発行時間間隔に関しては、次式が成立する。

$X_i'(A-j, B+j)=X_i'(A, B)+\alpha$ $(\forall i \in A+B)$ (10)

以下、MTjを移行した後の各制御装置の平均バッファ転送待ち時間の算出方式について述べる。移行後の構成では、制御装置Aの制御権下にあるMTの集合は、すべて、制御権移行を行なっていないMTの集合ということになる。したがって、仮定6より、MTjの制御移行後もスループット比が変化しないとすると、制御装置Aの平均チャネル転送時間、平均制御装置入出力発行時間間隔は、次式で表わされる。

$$S_{A-j}=\sum_{i\in A-j}\{S_i'*\lambda_i'(A, B)/(\sum_{i\in A-j}\lambda_i'(A, B))\} \quad (11)$$

$$X_{A-j}(A-j, B+j)=\sum_{i\in A-j}\{X_i'(A-j, B+j)*\lambda_i'(A, B)/(\sum_{i\in A-j}\lambda_i'(A, B))\} \quad (12)$$

以上より、5式、6式と同様の展開を行うとMTjの制御権を移行した後、制御装置Aの制御権下にあるMTの集合の平均バッファ転送待ち時間は次式で表わすことができる。

$W_{A-j}(A-j, B+j)=(N_A-2)S_{A-j}-X_{A-j}(A-j, B+j)$ (13)

この時、を次式で表す変数とする。

$$X_{A-j}(A, B)=\sum_{i\in A-j}\{X_i'(A, B)*\lambda_i'(A, B)/(\sum_{i\in A-j}\lambda_i'(A, B))\} \quad (14)$$

$X_{A-j}(A, B)$を定義したを14式と$X_{A-j}(A-j, B+j)$を定義したを12式の関係、および、10式で定義した$X_i'(A, B)$と$X_i'(A-j, B+j)$の関係にしたがって、14式を13式に代入し、変形すると明らかに次式が得られる。

$W_{A-j}(A-j, B+j)+\alpha=(N_A-2)S_{A-j}-X_{A-j}(A, B)$ (15)

一方、制御装置Bの制御権下にあるMTの集合は、制御権移行を行わなかったMTの集合とMTjからなることになる。仮定6では、制御権移行を行わなかったMTの集合のスループット比が変化しないとしているため、MTjと残りのMTの平均チャネル転送時間等は平均化できないことになる。したがって、ここでは、MTjを1つのクラス、制御権移行を行わなかったMTの集合、すなわち、もともと制御装置Bの制御権下にあったMTの集合を別の1つのクラスとしてモデル化する。この時、仮定2、仮定3にしたがうと、MTjの制御権を移行した後の、制御装置Bの制御権下にあるMTの集合の平均バッファ転送待ち時間$W_{B+j}(A-j, B+j)$は、次式を成立させる値となる。

$$S_j'/(X_j'(A-j, B+j)+W_{B+j}'(A-j, B+j)+S_j')+N_BS_B/(X_B(A-j, B+j)+W_{B+j}'(A-j, B+j)+S_B)=1 \quad (16)$$

この時、11式より、明らかに次式が成立する。

$$X_B(A-j, B+j)=\sum_{i\in B}\{(X_B(A, B)+\alpha)*\lambda_i'(A, B)/(\sum_{i\in B}\lambda_i'(A, B))\}=X_B(A, B)+\alpha \quad (17)$$

17式、10式をそれぞれ16式に代入して、変形を行うと次式が得られる。

$$S_j'/(X_j'(A, B)+W_{B+j}'(A-j, B+j)+\alpha+S_j')+N_BS_B/(X_B(A, B)+W_{B+j}'(A-j, B+j)+\alpha+S_B)=1 \quad (18)$$

15式と、18式は平均バッファ転送時間以外にαという未知数を含むため、直接平均バッファ転送待ち時間を得ることはできない。しかし、ここでは、15式、18式より明かであるように、それぞれの制御装置の平均バッファ待ち時間時間にαを加えた値を求めることが可能であることに着目した。以下、この値を$W_{A-j}'(A-j, B+j)$、$W_{B+j}'(A-j, B+j)$で定義する。この時、明かに次式が成立する。

$$W'_{A\downarrow}(A-j,\ B+j)-W'_{B\uparrow\downarrow}(A-j,\ B+j)=W_{A\downarrow}(A-j,\ B+j)+\alpha-W_{B\uparrow\downarrow}(A-j,\ B+j)-\alpha$$
$$=W_{A\downarrow}(A-j,\ B+j)-W_{B\uparrow\downarrow}(A-j,\ B+j) \quad (19)$$

すでに述べたように、制御権移行対象として選択すべきMTは、制御権移行後のそれぞれの制御装置の平均バッファ待ち時間の差が最も小さくなるMTである。したがって、19式から、明かに、15式、18式を充たす$W'_{A\downarrow}(A-j,\ B+j)$、$W'_{B\uparrow\downarrow}(A-j,\ B+j)$ ($W_{A\downarrow}(A-j,\ B+j)+\alpha$、$W_{B\uparrow\downarrow}(A-j,\ B+j)+\alpha$) を算出し、それぞれの差の絶対値を最も小さくするMTを制御権移行対象として選択すればよいことになる。

4．拡張漸近モデルによる予測制御型負荷均衡制御アルゴリズムの評価

本章では、実験システムに対して、前章で示した拡張漸近モデルによる予測制御型負荷均衡アルゴリズムを適用した時の評価結果をまとめる。

実験システムの構成、システム内の処理要求の動きは、第２章で述べたとおりである。また、処理要求の動作特性に差異をもたせるため、負荷量の大きい処理要求と小さい処理要求の２種類を設けた。具体的には、転送単位となるブロック・サイズに差をもたせることにより、負荷量に差が出るようにした。処理要求の負荷量に差をもたせたのは、すべての処理要求の負荷量が均一の場合には、制御装置Ａと制御装置Ｂの制御権下に置くMTの台数を等しくすればよいため、比較的単純に負荷均衡が可能となるためである。評価を行ったケースは、３ケースである。それぞれのケースにおいて、動作させた負荷量の大きい処理要求の数、負荷量の小さい処理要求の数を以下にまとめる。

ケース１：負荷量の小さい処理要求１６個

ケース２：負荷量の大きい処理要求２個、負荷量の小さい処理要求４個

ケース３：負荷量の大きい処理要求４個、負荷量の小さい処理要求５個

ここで行った評価はいずれのケースも初期状態としては、制御装置Ａの方の入出力負荷が大きくなるようにした。すでに、述べたように本講演で提案した負荷均衡制御においては、１回に１台のMTを制御権移行対象として選択するため、それぞれの制御装置の負荷を均衡させ、制御権移行を行わなくなるという安定状態にいたるまでには、複数回の制御権移行を行う必要が生ずる。ただし、ここでの実験においては、それぞれの制御装置の平均バッファ転送路待ち時間の差がある時には、いつでも予測計算を実行し、どのMTの制御権を移行しても現状よりも負荷が均衡しないという結果が得られたときのみ制御権移行を行わないようにした。評価項目としては、以下に示す２つの項目をあげる。

評価項目１：安定状態にいたった時の２つの制御装置の平均バッファ待ち時間の差

評価項目２：安定状態にいたるまでに実行した制御権移行回数。

評価項目１の必要性は、負荷均衡制御の目的より明らかである。一方、評価項目２の意義は、制御権を移行させるためには、オーバヘッドがかかるため、移行回数が少ない方がより望ましいためである。

ここで行った評価においては、以下に示す条件の成立が最少の制御権移行回数で負荷の均衡が実現でき、安定状態に入ったことの充分条件を充たしたことになる。

（１）すべての処理要求の負荷量が均一の場合

条件１：制御装置Ａ側からＢ側に制御権下にあるMT台数が等しくなるまで制御権移行が行われる。

（２）負荷量の小さい処理要求と大きい処理要求が混在する場合

条件２：制御装置Ａ側からＢ側に、負荷量の大きい処理要求の制御権移行のみを何回か行っているか、負荷量の大きい処理要求の制御権移行以外に、負荷量の小さい処理要求の制御権を一回だけ行っているかのいずれかである。

ここで取り上げた評価ケースでは、初期状態においては制御装置Ｂの負荷量が小さいため、すべての処理要求が均一の場合、制御装置Ａの方から制御装置Ｂ側に、単純に、制御権下にあるMT台数が等しくなるまで制御権移行が行われ、安定状態に入っていれば最少の移行回数で負荷均衡状態に入ったことになる。

一方、負荷量の小さい処理要求と大きい処理要求が混在する場合には、、制御装置Ａの方から負荷量の大きいMTの制御権を移行している時、最も急速に、制御装置Ａに対する負荷量が減少し、制御装置Ｂの負荷量が増大していることになる。すでに述べたように、ここで取り上げた評価ケースでは、初期状態においては制御装置Ｂの負荷量が小さい。したがって、制御装置Ａの方から制御装置Ｂの方に、負荷量の大きい処理要求の制御権移行のみを何回か行っているか、負荷量の大きい処理要求の制御権移行以外に、負荷量の小さい処理要求の制御権を一回だけ行っているかのいずれかの制御権移行で、安定状態に入った場合、最少の制御権移行回数で安定状態に入ったことになる。例えば、負荷量の小さいMTの制御権移行を最後に２回行った場合でも、負荷量の小さいMTの負荷量が、負荷量の大きいMTの負荷量の半分未満の場合には、最少の制御権移行回数で安定状態に入った可能性もある。従って、条件２の成立は、最少の制御権移行回数で負荷の均衡が実現でき、安定状態に入ったことの充分条件となり、必要充分条件とはならないことになる。

以下、各評価ケースについてのべる。表２は、初期状態、安定状態においてそれぞれの制御装置の制御権下にあった負荷量の大きい処理要求と負荷量の小さい処理要求の数、および、安定状態におけるそれぞれの制御装置の平均バッファ転送路待ち時間の比をまとめたものである。図５は、制御権移行回数と移行後のそれぞれの制御装置の平均バッファ転送路待ち時間比をグラフ化したものである。さらに、各制御権移行において負荷量の大きい処理要求が選択されたか、小さい

処理要求が選択されたかということをまとめた。なお、本評価ケースにおいては、すべての制御権移行は、制御装置A側からB側に発生し、逆方向の制御権移行は1回も発生しなかった。評価ケース1は、すべての処理要求の負荷量が均質の場合である。評価結果においては、制御装置Aと制御装置Bの制御権下にあるMTの台数がそれぞれ8台になるまで、制御装置Aの方から制御権移行が行われている。これは、条件1を充たしており、最少の移行回数で安定状態に入ったことになる。また、安定状態に入った時のそれぞれの制御装置の平均バッファ転送路待ち時間も等しいため、負荷も均衡していることになる。

評価ケース2では、負荷量の大きい処理要求の制御権移行が1回行われた後、負荷量の小さい処理要求の制御権移行が1回行われ、安定状態に入っている。安定状態でのそれぞれの制御装置の制御権下にあるMTの集合は、負荷量の大きい処理要求が1、小さい処理要求が2と等しいため、平均バッファ転送路待ち時間も等しくなっている。また、制御権移行のパターンも条件2を充たしており、最少の制御権移行回数で安定状態に入ったことになる。

評価ケース3では、負荷量の大きい処理要求の制御権移行が1回行われた後、負荷量の小さい処理要求の制御権移行が1回行われ、安定状態に入っている。この、制御権移行のパターンも条件2を充たしており、最少の制御権移行回数で安定状態に入ったことになる。安定状態における、それぞれの制御装置の制御権下にあるMTの集合は等しくないため、平均バッファ転送路待ち時間は完全には等しくないが、その比率は1．1であるため、充分な負荷均衡状態であると考えられる。したがって、以上をまとめると、いずれのケースにおいても、本制御アルゴリズムの適用により、最少の制御権移行回数により、制御装置間の負荷均衡が実現できたことになる。

5．おわりに

カートリッジ型MTのための予測制御型負荷均衡制御アルゴリズムの提案とその評価結果を報告した。カートリッジ型MTの制御装置においては、制御装置内にバッファメモリを設け、制御装置とMTの間では、複数ブロックの先読み／まとめ書きを実行する。したがって、各MTを複数台の制御装置に接続する場合、1台のMTの先読み／まとめ書きデータが複数の制御装置に分散することをさけるため、各MTをある1台の制御装置の制御下に置く方式をとる。この結果として、各制御装置に対する入出力負荷を均衡させるために、制御装置間で、各MTを制御する権利（制御権）を移行させる必要が生じた。予測制御型負荷均衡アルゴリズムとは、制御装置内に組み込んだ解析モデルの性能予測結果に従って、制御権を移行すべきMTを決定する方式である。本講演では、以上の負荷均衡制御のための解析モデルとして、拡張漸近モデルを提案した。拡張漸近モデルは負荷均衡制御アルゴリズムが充たすべき以下の3つの条件を満足する。（1）高速アルゴリズム（2）クローズド型モデル（3）待ち時間の一部が漸近的に発生するという仮定により、制御装置内で収集可能なデータのみで性能予測を可能とした点。以上の3つの条件を充たさなければならない理由は、（1）に関しては、本制御がリアルタイムに実行されるという点、（2）に関しては、MTの動作特性を反映させるという点、（3）に関しては、本制御を制御装置内で閉じて実行させるという点に起因している。3ケースの実測データにより、提案アルゴリズムの評価を行ったところ、これらの評価範囲では、提案アルゴリズムにより、最少の制御権移行回数で、制御装置間の入出力負荷バランスが可能であるという結果が得られた。

［謝辞］拡張漸近モデルの拡張点について御討論頂いた電気通信大学亀田寿夫教授に深く感謝いたします。

参考文献　1）山本他：カートリッジ型MTにおける先読み・纏め書きスケジューリング方式の提案と解析，情報処理学会第35回全国大会，pp.219-220(1987)

2）山本他：カートリッジ型MTにおける先読み・纏め書きスケジューリング方式の評価，情報処理学会第35回全国大会，pp.221-222(1987)

3）山本他：カートリッジ型MTにおける予測制御型負荷均衡制御方式，情報処理学会第36回全国大会，pp.233-234(1988)

4）坪井他：カートリッジ型MTにおける予測制御型負荷均衡制御方式の評価，情報処理学会第36回全国大会，pp.235-236(1988)

5）山本他：バッファ付き入出力サブシステムにおける負荷均衡のための漸近近似手法，情報処理学会第37回全国大会，pp.198-199(1988)

6）坪井他：バッファ付き入出力サブシステムにおける漸近近似型負荷均衡の実験評価，情報処理学会第37回全国大会，pp.190-191(1988)

7）益田他：オペレーティングシステムの性能解析，情報処理叢書9(1982)

8）西垣，山本：資源割当て優先度のある多重プログラミングシステムのボトルネック解析，　情報処理学会論文誌，Vol.23，No.5，pp.562-569(1982)

9）山本，西垣：サービス関数による応答時間制御の下での計算機システム性能のボトルネック解析，情報処理学会論文誌，Vol.24，No.5，pp630-637(1983)

10）西垣，山本：順次アクセス入力処理におけるディスク・キャッシュ装置の効果解析，情報処理学会論文誌，Vol.25，No.2，pp630-637(1984)

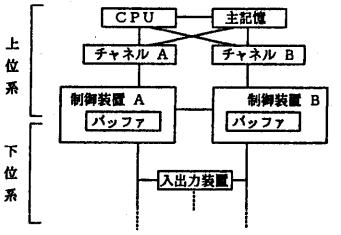

図1．システム構成

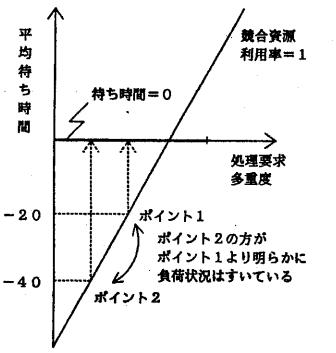

図2．漸近近似方式の基本概念

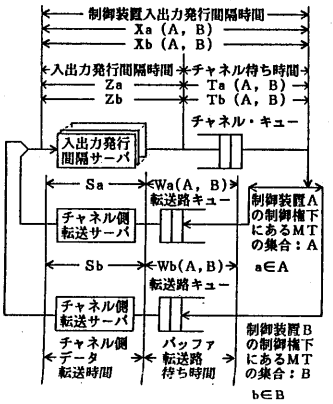

図3．上位系のモデル

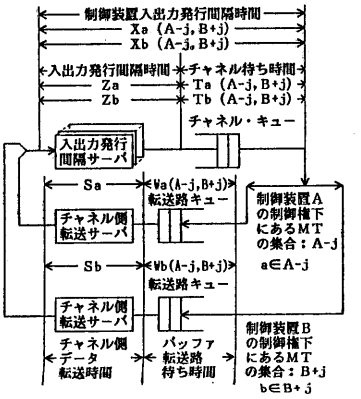

図4．MTj制御権移行後のモデル

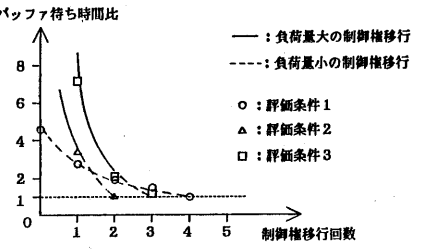

図5．バッファ待ち時間比の変化

表1．記号の定義

S_t：MT_tのチャネル側データ転送時間，　Z_t：MT_tの入出力発行時間間隔
T_t（A，B）：制御装置A，制御装置Bの制御権下にあるMTの集合がそれぞれA，Bである時のMT_tの平均チャネル待ち時間
W_t（A，B）：制御装置A，制御装置Bの制御権下にあるMTの集合がそれぞれA，Bである時のMT_tの平均バッファ転送路待ち時間
X_t（A，B）：制御装置A，制御装置Bの制御権下にあるMTの集合がそれぞれA，Bである時のMT_tの平均制御装置入出力発行時間間隔
λ_t（A，B）：制御装置A，制御装置Bの制御権下にあるMTの集合がそれぞれA，Bである時のMT_tのスループット
S_A：MTの集合Aの平均チャネル側データ転送時間，　S_B：MTの集合Bの平均チャネル側データ転送時間
N_A：MTjの制御権移行前に制御装置Aの制御権下にあるMT台数
N_B：MTjの制御権移行前に制御装置Bの制御権下にあるMT台数
X_A（A，B）：制御装置A，制御装置Bの制御権下にあるMTの集合がそれぞれA，Bである時のMTの集合Aの平均制御装置入出力発行時間間隔
X_B（A，B）：制御装置A，制御装置Bの制御権下にあるMTの集合がそれぞれA，Bである時のMTの集合Bの平均制御装置入出力発行時間間隔
W_A（A，B）：制御装置A，制御装置Bの制御権下にあるMTの集合がそれぞれA，Bである時のMTの集合Aの平均バッファ転送路待ち時間
W_B（A，B）：制御装置A，制御装置Bの制御権下にあるMTの集合がそれぞれA，Bである時のMTの集合Aの平均バッファ転送路待ち時間

表2．評価ケースのまとめ

項目＼ケース	ケース1		ケース2		ケース3	
	制御装置A	制御装置B	制御装置A	制御装置B	制御装置A	制御装置B
初期状態	負荷量小12	負荷量小4	負荷量大2 負荷量小3	負荷量 0 負荷量小1	負荷量大4 負荷量小4	負荷量大1
安定状態	負荷量小8	負荷量小8	負荷量大1 負荷量小2	負荷量大1 負荷量小2	負荷量大2 負荷量小3	負荷量大3 負荷量小1
安定状態時の平均バッファ転送路待ち時間比	1．0		1．0		1．1	