

論 文

マルチアクセス通信のためのグループ化2次元ラウンドロビン
スケジューラの提案高城 衛^{†*} 木村 成伴^{††} 海老原義彦^{††}A Proposal of Grouped Two-Dimensional Round-Robin Schedulers for
Multiaccess CommunicationsMamoru TAKAJO^{†*}, Shigetomo KIMURA^{††}, and Yoshihiko EBIHARA^{††}

あらまし マルチアクセス通信におけるパケットスイッチのスケジューリングを行う機構の一つに、各入力ポートに出力ポート別の FIFO 方式の待ち行列をもつ 2DRR スケジューラがある。この 2DRR スケジューラは、高いスループット、フェアネスを達成するが、 N 入力 N 出力スイッチに対して、公平な割当機会を与えるためには $N+1$ が素数でなければならないという制約がある。また、スケジューリングに要する計算量と記憶容量は $O(N^3)$ である。そこで本論文では、 N 個の入出力ポートを $a_0 \times \dots \times a_{m-1}$ ($m \geq 2$) に階層化するグループ化 2DRR スケジューラを提案する。そして、本スケジューラにより上述の制約が緩和され、また、 $a_{m-1} \leq N^{1/2}$ としたとき、本スケジューラの計算量が $O(N^{5/2})$ に、記憶容量が $O(N^2)$ に抑えられることを示す。最後に、計算機シミュレーション実験によって、 a_{m-1} は 4 以上に設定すべきであり、その場合、フェアネスなどといった 2DRR スケジューラの長所を失うことなく、2DRR 方式と同様の遅延特性をもつことを示す。

キーワード 2DRR スケジューラ、マルチアクセス通信、マルチアクセスチャネル、パケットスイッチ、フェアネス

1. ま え が き

近年のネットワークの広域化、多様化に伴い、通信チャネルを共有し、一つのノードへ複数のノードからアクセスが生じるマルチアクセス通信の要求がますます高まってきている。このようなシステムとして代表的なものに衛星システム、マルチドロップ電話回線、マルチタップバスシステムなどがある。一般的に、複数のノードからアクセスがあると、これらの信号は混信し、正常なアクセスはできなくなる [1], [2].

この問題を解決するために従来から多くの手法が開発されてきたが、ノード数が比較的小さく、パケットの到着率が大きい場合は、スケジューリングを行い、チャネル利用の順番を一定手順に従って決定（予約）

する方式がとられている [1], [2]. 予約方式としては、衝突解決を用いるものと、TDM やラウンドロビン [3] を用いるものなどがある。このうち、ラウンドロビンによる予約方式には、マルチアクセスメディアごとにそれら特有の性質を使った多くのバリエーションがある。例えば、衛星通信網で使われる予約システムとしては文献 [4]~[6] などが、LAN では文献 [7], [8] などがあげられる。

これらの予約システムは、 N 入力 N 出力システムのスケジューリングを規定時間内に行うシステムに一般化される。ラウンドロビン方式におけるこのようなシステムの一つに、2DRR スケジューラ (Two-Dimensional Round-Robin Scheduler) [9] がある。この方式は、 N 入力 N 出力システムの完全スケジューリング (ランダム性のないアルゴリズム手順に沿ったスケジューリング) を実現するものであり、高いスループット (多くの要求に対応すること) と、フェアネス [10] (各入力からの出力要求に公平な出力機会を与えること) を達成している。また単純なアルゴリズムで構成されているために、ハードウェアへの効果的な

[†] 筑波大学大学院理工学研究科, つくば市

Master's Program in Science and Engineering, University of Tsukuba, Tsukuba-shi, 305-8573 Japan

^{††} 筑波大学電子・情報工学系, つくば市

Institute of Electronics and Information Sciences, University of Tsukuba, Tsukuba-shi, 305-8573 Japan

* 現在, 日本電気(株) 交換移動通信用事業本部

実装が可能であり、高速なスイッチングが期待される。

ところで、この2DRR方式において、 N 入力 N 出力システムの各入出力に対して公平な割当機会を与えるためには $N+1$ が素数でなければならないという制約がある。また、スケジューリングに要する計算量と記憶容量は $O(N^3)$ である。そこで本論文では、前者の制約を緩和し、後者を軽減するため、 $N = a_0 \times a_1 \times \dots \times a_{m-1}$ のとき、入力ポートを m レベルに階層化するグループ化2DRRスケジューラを提案する。そして、本方式において、各 $a_i + 1$ ($0 \leq i \leq m - 1$)が素数であるときに、システムの各入出力に対して公平な割当機会を与え、更に、 $a_{m-1} \leq N^{1/2}$ ならば、計算量を $O(N^{5/2})$ に、記憶容量を $O(N^2)$ に抑えることができることを示す。

本論文の構成は以下のとおりである。まず、2.で2DRRスケジューラについて簡単に触れた後、3.でグループ化2DRRスケジューラの提案を行う。4.で、計算機シミュレーションを行い、本方式と従来の2DRR方式とのスループットとフェアネスを比較する。その結果、いずれも同等な処理能力をもつことを示す。最後に5.で、結論と今後の課題について述べる。

2. 2DRR スケジューラ

2.1 2DRR スケジューラによるスケジューリング

2DRR方式では、 N 入力 N 出力パケットスイッチの完全スケジューリングを実現するため、各入力ポートにそれぞれの出力ポートのためのFIFO方式の待ち行列をもつ。各入力ポートに到着したパケットは出力ポート別の待ち行列に入れられるが、そのなかから同一タイムスロットで出力できるパケットは一つのみである。また、各入力ポートは各出力ポートと全結線されており、出力先が競合しない限り他の入力ポートからの出力と同時にパケットを出力することができる。図1に4入力4出力スイッチを示す。

2DRR方式では、スケジューリングを行うために次に示す4種類のマトリクスを使用する。

(1) リクエストマトリクス (RM)

R 番目の入力ポートに C 番目の出力ポートへの出力パケットがあるとき、 $RM[R, C]$ には1が示され、そうでなければ0が示される。ここで、 $RM[R, C]$ とは行列 RM の R 行 C 列の要素を表す。後述の PM 、 SM 、 AM に対しても同様の記法を用いる。

図1の4入力4出力スイッチにおいて、各入力ポートの待ち行列内の黒い部分は送信をもつパケットを示

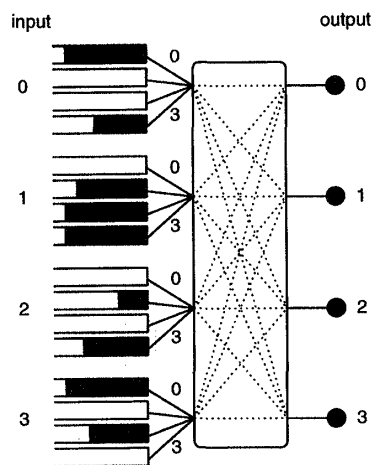


図1 4入力4出力スイッチの構成
Fig.1 The construction of 4 × 4 switch.

		output				timeslot				
		0	1	2	3					
input	0	1	0	0	1	0	0	1	2	3
	1	0	1	1	1	1	3	0	2	
	2	0	1	0	1	2	0	3	1	
	3	1	0	1	0	3	2	1	0	

Request Matrix (RM) Pattern Matrix (PM)

図2 リクエストマトリクスとパターンマトリクス
Fig.2 Request and pattern matrices.

しているとするとき、このとき RM は図2左となる。

(2) パターンマトリクス (PM)

$PM[I, J] = (I + IJ + J) \text{ mod } (N + 1)$ ($0 \leq I, J \leq N - 1$)により定義され、 SM の作成や選択に用いられるほか、パケットのスケジューリングのために用いる。その際、各要素の散らばりが等しくなるように配置される (PM の任意の行にある隣接した2数の並びが一度しか存在しない。列についても同様のことが成り立つ)ことが望ましいが、これは $N + 1$ が素数のときにのみ可能である[9]。散らばりが等しくない場合には、スケジューリングに偏りが生じる。図1の場合の PM は、図2右のようになる。

(3) スケジューリングマトリクス (SM)

SM には段階0から段階 $N - 1$ までの N 種類がある。段階 P の SM は、 $PM[R, P]$ 行 R 列 ($0 \leq R \leq N - 1$)に0が置かれ、その0の位置から同じ行で列番号が増加する方向に順に $1, 2, \dots, N - 1$ が配置される。ただし、列番号が $N - 1$ になった場合には、列番号0から配置を続ける。

図1の場合、段階0から段階3が作成される(図3

step 0 (timeslot 0-3)				step 2 (timeslot 8-11)			
0	1	2	3	2	3	0	1
3	0	1	2	0	1	2	3
2	3	0	1	1	2	3	0
1	2	3	0	3	0	1	2

step 1 (timeslot 4-7)				step 3 (timeslot 12-15)			
3	0	1	2	1	2	3	0
1	2	3	0	2	3	0	1
0	1	2	3	3	0	1	2
2	3	0	1	0	1	2	3

図3 スケジューリングマトリクス
Fig. 3 Scheduling matrixes.

参照). 例えば, 段階1の SM の第0列は $PM[1,0] = 1$ であることから, SM の第0行第1列を0に, そして, そこから右へ順に1, 2, 3を割り当てる. ただし, 最後の3は SM の右枠からはみ出すことから, いちばん左の列に戻る. 同様に, 第1行は第3列が, 第2行は第0列が, 第3行は第2列が0となり, そこから右に1, 2, 3が割り当てられる.

(4) アロケーションマトリクス (AM)

どの入力ポートからどの出力ポートへの出力を許可するかを記録するために用いる. R 番目の入力ポートから C 番目の出力ポートへ出力が許可されたときに $AM[R,C]$ に1が, そうでなければ0が示される.

ここまでで作成した RM と SM から, AM を定める. 以下で, タイムスロット0における AM の決定方法を説明する. なお, タイムスロットについては後述する.

AM は四つのステップを経て決定される. これを, パターン0, 1, 2, 3と呼ぶ. 初期化として, AM の要素すべてを0にしておく. パターン0では, まず, SM の要素の内0がある位置を調べる. 現在, SM は段階0を使用しているとすると, $SM[0,0]$, $SM[1,1]$, $SM[2,2]$, $SM[3,3]$ がこれに該当する. つぎに, RM におけるこれらの位置の要素が1かどうかを調べ, 該当位置の AM の要素を1とする. これにより, 図2の RM では, 入力ポート0から出力ポート0への出力と, 入力ポート1から出力ポート1への出力が許可されることになる. パターン1, 2, 3についても同様に, 順にそれぞれ SM の要素で1, 2, 3がある位置を調べ, RM との比較を行う. ここで, スイッチにおける仮定より, 同一の入力ポートから複数のパケットを出力することや, 同一の出力ポートに複数のパケッ

トを出力することはできない. したがって, パターン1では $RM[2,3]$ が選択され, パターン2ではいずれの入出力対も選択されず, パターン3では $RM[3,2]$ が選択される.

ところで, このパターンの順番はタイムスロットの番号に該当する PM の列によって決定される. すなわち, タイムスロット0では図2の PM の第0列を参照し, これが第0行から順に0, 1, 2, 3となるため, パターンが0, 1, 2, 3の順に進められた. 同様に, タイムスロット1では1, 3, 0, 2の順に進められ, 以後 N タイムスロットごとに巡回する. また, N タイムスロットごとに SM の段階が1増加し, N 段階で段階0に戻る. つまり, 段階は N^2 タイムスロットで巡回する.

2.2 2DRR スケジューラの特徴

N 入力 N 出力パケットスイッチにおいて, 各パターンにおけるパターンの番号と同じ要素の位置を SM 上の N^2 個の要素から探し出し, それらの組について, RM から出力の要求があるかどうかを N 回 (パターンの個数) 調べ, 出力があれば (最大 N 回), その組がスイッチにおける仮定のうで送出可能かどうか, AM に対し入出力それぞれについて $N-1$ 回比較を行う. これが N パターンシーケンスで行われることから, スケジューリングには $(N^2 + N + 2(N-1)N) \times N = 3N^3 - N^2$ の計算量が必要とされる. また, 2DRR方式ではスケジューリングに $N \times N$ の SM が N 個, RM , AM , PM がそれぞれ1個ずつ必要となる. これによりスケジューリングに要する記憶容量は $N^2 \times (N+3) = N^3 + 3N^2$ となる.

3. グループ化 2DRR スケジューラ

前章で示したように, 2DRR方式は, N 入力 N 出力システムの各入出力に対して公平な割当機会を与えるためには $N+1$ が素数であるという制約があり, スケジューリングに要する計算量と記憶容量は $O(N^3)$ であった. これらを解決するため, 本章では, 2DRR スケジューラの入出力ポートをグループ化したグループ化 2DRR スケジューラを提案する.

3.1 グループ化 2DRR スケジューラの提案

図4に16入力16出力のスイッチにおけるグループ化 2DRR スケジューラの構成図を示す.

グループ化 2DRR スケジューラにおいてグループ化を行うには, 入出力ポートの個数 N が $N = a_0 \times a_1 \times \dots \times a_{m-1}$ ($m \geq 2$, 各 a_i は2以上の整

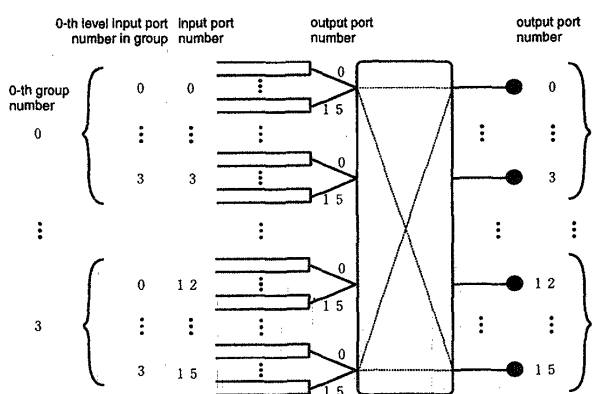


図4 グループ化2DRRスケジューラの構成

Fig. 4 The construction of a grouped 2DRR scheduler.

数)と表される必要がある。ただし、2DRRと同様に、スケジューリングにおいて各ポートの選択機会の偏りをなくすためには、 $a_0 + 1, a_1 + 1, \dots, a_{m-1} + 1$ は素数でなければならない。この制約は2DRRスケジューラのそれよりも緩和されており、これにより、2DRRよりも多くの種類のスイッチサイズで偏りが無いスケジューリングが可能になる。

ここで、 $N = N_0, N_i = a_i \times N_{i+1} (0 \leq i \leq m-1), N_m = 1$ とおく。

グループ化は第 $m-2$ レベルから第 0 レベルまで段階的に行われる。第 i レベル ($0 \leq i \leq m-2$) では、 N 個の入力ポートは N_i 個ずつに分け、その各々を a_i 個に分割し、更にこれを $a_{m(i+1)} = \min(a_{i+1}, a_{i+2}, \dots, a_{m-1})$ 個に分割する。これらの入力ポートのまとまりを順にクラス、グループ、入力ポート番号割当単位と呼び、それぞれに、第 i レベルのクラス番号 ($0 \sim N/N_i - 1$)、グループ番号 $0 \sim a_i - 1$ 、グループ内入力ポート番号 $0 \sim a_{m(i+1)} - 1$ を付ける (図 5 参照)。

各レベルは次のように定義される。まず、第 $m-2$ レベルでは入力ポート番号順に入力ポートを並べ、これを機械的に分割する。すなわち、クラス番号 c のクラスには入力ポート番号 $c \times N_{m-2} \sim c \times N_{m-2} + c - 1$ の入力ポートを、クラス番号 c 、グループ番号 g のグループには、 $c \times N_{m-2} + g \times N_{m-1} \sim c \times N_{m-2} + g \times N_{m-1} + g - 1$ の入力ポートを割り当てる。更に、グループ内の各入力ポート一つずつに対して、グループ内入力ポート番号 $0 \sim a_{m(m-1)} - 1$ を入力ポート番号の小さい順に割り当てる。例えば、16 入力 16 出力スイッチにおいて、 $16 = 4 \times 4$ と分割した場合、 $m = 2$ であり、第 $m-2$

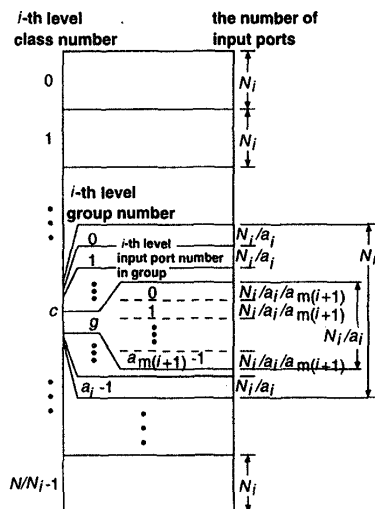


図5 第 i レベルのグループ化

Fig. 5 The i -th level grouping.

レベルは第 0 レベルである。このとき、第 0 レベルのクラスは一つしかなく、グループ番号とグループ内入力ポート番号は図 4 に示したようになる。

つぎに、第 i レベル ($1 \leq i \leq m-2$) から第 $i-1$ レベルの構成方法を述べる。最初に、第 i レベルのクラス a_{i-1} 個 (第 i レベルのクラス番号 $a_{i-1} \times c \sim a_{i-1} \times c + a_{i-1} - 1$) を連結し、第 $i-1$ レベルクラス番号 c のクラスとする。第 i レベルのクラスを連結した様子と、これに対応する第 $i-1$ レベルのクラスを図 6 の左枠及び右枠にそれぞれ示す。更に、図に示したように、第 i レベルの各クラスはクラス番号順に第 $i-1$ レベルのグループ (第 $i-1$ レベルのグループ番号 $0 \sim a_{i-1} - 1$) になる。ただし、連結前に第 i レベルの各クラス内のグループと各グループ内の入力ポート番号割当単位の順番を次の手順で入れ換える。

まず、 $a_i \times a_i$ の PM を用い、第 i レベルのグループ番号 g のグループをグループ番号 $PM[g, (t + s_i) \bmod a_i]$ にする。つぎに、段階 r_i の $a_{m(i)} \times a_{m(i)}$ の SM を用いて第 i レベルのグループ内の入力ポート番号割当単位を入れ換えるが、 $a_i = a_{m(i)}$ であるかどうかによって処理が異なる。なお、 t, s_i, r_i はその時点のタイムスロット、第 i レベルのスロットシーケンス及び段階を表し、これについては後述する。

$a_i = a_{m(i)}$ の場合、グループ番号 g 、グループ内入力ポート番号 k の入力ポート番号割当単位を、グループ番号 $g' = SM[k \bmod a_{m(i)}, g]$ のグループ内入力ポート番号 k にする。図 6 では、グループ内入力ポート番号 0 と k の入れ換えの様子が示されているが、変

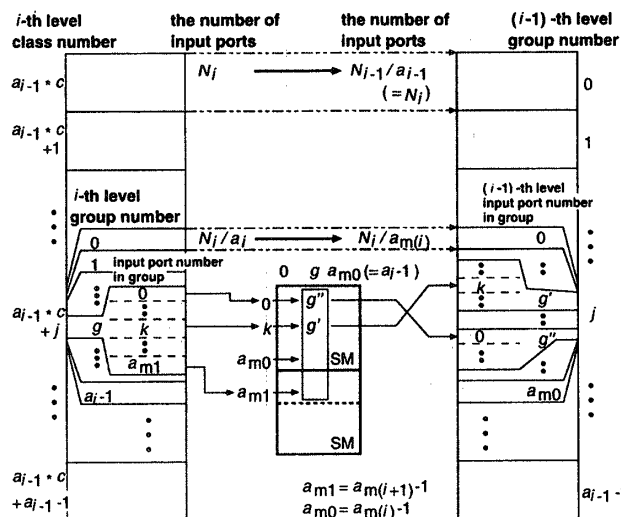


図6 $a_i = a_{m(i)}$ のときのグループの割当方法
Fig. 6 The group assignment at $a_i = a_{m(i)}$.

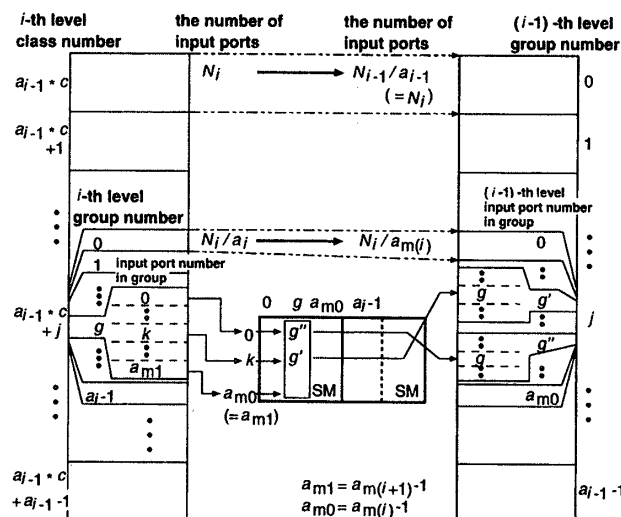


図7 $a_i \neq a_{m(i)}$ のときのグループの割当方法
Fig. 7 The group assignment at $a_i \neq a_{m(i)}$.

換後の結果が第 $i-1$ レベルに移行されているので注意されたい。SM の構成より、この割当ては一意に定まり、異なる入力ポート番号割当単位が同じグループの同じグループ内入力ポート番号に重複して割り当てられることはない。入れ換え後、第 i レベルのグループは第 $i-1$ レベルの入力ポート番号割当単位になる。

$a_i \neq a_{m(i)}$ の場合、グループ番号 g 、グループ内入力ポート番号 k の入力ポート番号割当単位を、グループ番号 $g' = SM[k, g \bmod a_{m(i)}]$ 、グループ内入力ポート番号 g にする。図 7 では、グループ内入力ポート番号 0 と k の入れ換えの様子が示されており、その変換結果が第 $i-1$ レベルに移行されているのは図 6 と同様である。これにより、新しいグループ番号の範囲は $0 \sim a_{m(i+1)} - 1$ に、グループ内入力ポート番号は $0 \sim a_i - 1$ になっている。仮定より、 $a_{m(i+1)} = a_{m(i)}$ であるから、この変換により、第 i レベルのグループを第 $i-1$ レベルの入力ポート番号割当単位とすることができる。

第 0 レベルのクラスが完成したら、第 0 レベルに対してもグループと各グループ内入力ポート番号割当単位の順番を上述の手順で行う。そして、各入力ポートをグループ番号、グループ内入力ポート番号の順に並べ直した後、再び入力ポートを a_{m-1} 個ごとのグループに分ける。そして、このグループに対してグループ番号 $0 \sim N/a_{m-1} - 1$ を付ける。最後に、グループ o ($0 \leq o \leq N/a_{m-1} - 1$) に対して出力ポート番号 $a_{m-1} \times o \sim a_{m-1} \times o + a_{m-1} - 1$ の a_{m-1} 個の出力ポートを割り当てる。

このようにしてできた N/a_{m-1} 個の $a_{m-1} \times a_{m-1}$ の入出力ポート対に、前節で述べた 2DRR スケジューラによりスケジューリングを行う。以上の作業をすべてのスロットシーケンス s_{m-2}, \dots, s_0 ($0 \leq s_i \leq a_i - 1$) の値の組合せに対して順に行い、それらの結果の合成が最終的な結果となる。したがって、スケジューリングは $a_{m-2} \times \dots \times a_0 = N/a_{m-1}$ 回繰り返すことになる。組合せの順序は固定されていれば問わない。タイムスロット t は、組合せがすべて終わるまで固定され、この期間に行われる 2DRR スケジューラのタイムスロットもすべてこの値が用いられる。全組合せの終了後、 t が 1 増加される。これが a_{m-1} に達すると t を 0 に戻し、2DRR で用いる段階数を 1 増やす。この段階数が $a_{m(m-1)}$ に達して 0 に戻ったら、第 $m-2$ レベルで用いる段階数 r_{m-2} を 1 増加させる。この値が $a_{m(m-2)}$ に達し、0 に戻ったら第 $m-3$ レベルの r_{m-3} を 1 増やし、以下同様に $r_{m-4} \sim r_0$ が増加する。

したがって、段階数の組合せ数は $\prod_{i=0}^{m-1} a_{m(i)}$ である。

さて、16 入力 16 出力スイッチを 4×4 にグループ化する例において、第 0 レベルの構成方法は既に示した。以下では、 $t + s_0 = 1, r_0 = 0$ としたときのグループ化構築の続きを示す。ここで、 $a_0 = a_1 = 4$ であるので、ここで用いる PM と SM は 4×4 のマトリクス (図 2, 図 3 参照) のみ、 RM と AM は 2DRR 方式と同じ 16×16 のマトリクスである。

まず、第 0 レベルのグループ (0~3) を PM により入れ換える。 PM の第 $(t + s_0)$ 列から、グループ

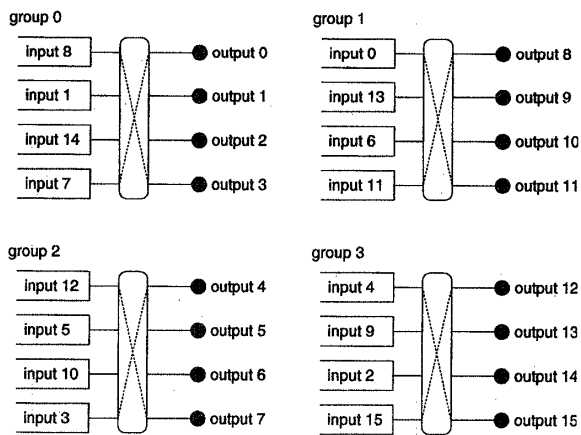


図8 4入力4出力のスイッチへの分割
Fig.8 Dividing to 4 × 4 switches.

0, 1, 2, 3はそれぞれ1, 3, 0, 2になる。つぎに, SM を使って各グループ内の入力ポート番号割当単位の入れ換えを行う。例えば, グループ0の入力ポート番号割当単位0~3は, 段階 r_0 の SM の第0列を参照して, 0番は(グループ番号, グループ内入力ポート番号) = (0, 0) に, 1~3番はそれぞれ (3, 1), (2, 2), (1, 3) に移る。グループの入れ換えによってグループ0には入力ポート番号8~11が割り当てられていることから, 実際には, これらの入力ポートが上述の場所に置かれることになる(図8参照)。最後に, グループ0~3に図8のように出力ポートを四つずつ割り当て, 各々で4入力4出力の2DRRスケジューラによるスケジューリングが行われる。

このスケジューリングは $s_0 = 0 \sim 3$ まで行われ, すなわち, $(t + s_0) \bmod 4 = 2, 3, 0$ に対しても, 同様なスケジューリングが行われる。 s_0 が4に達したら, s_0 は0に戻され, t が1増やされる。

3.2 グループ化2DRRスケジューラの特徴

2.において2DRR方式では N 入力 N 出力の場合, 計算量が $3N^3 - N^2$, 記憶容量が $N^3 + 3N^2$ 必要とされることを示した。これに対してグループ化2DRR方式では, $N = a_0 \times a_1 \times \dots \times a_{m-1}$ のとき, これらは以下のように求められる。

まず計算量は, グループ化に要するものとスケジューリングに要するものに分けられる。前者は, 各スロットシーケンスにおいて, 第 i レベル ($0 \leq i \leq m-2$) の各クラス (N/N_i 個) でグループ (a_i 個) の入れ換えを行い, つぎにその各グループ内の入力ポート番号割当単位 ($a_{m(i)}$ 個) を入れ換える必要がある。した

がって, これに要する計算量は, $a_i(1 + a_{m(i)})N/N_i$ である。ここで, $2 \leq a_{m(i)} \leq a_{m-1}$ であり, $i \leq m-2$ から $N_i = a_i \times N_{i+1} \geq a_i \times a_{m-1}$ であることに注意すれば, $a_i(1 + a_{m(i)})/N_i \leq 3/2$ が成り立つことがわかる。したがって, 全レベルでの総計算量は $N \sum_{i=0}^{m-2} a_i(1 + a_{m(i)})/N_i < (3/2)N \cdot m \leq (3/2)N \log N$ である。

後者は, 各スロットシーケンスにおいて $a_{m-1} \times a_{m-1}$ の2DRRスケジューラを N/a_{m-1} 回行うので, その総計算量は $a_{m-1}(3a_{m-1} - 1)N$ である。

スケジューリングが完了するまでに, 上記の操作を1タイムスロット分 (N/a_{m-1} スロットシーケンス) 行うので, 総計算量は $((3/2)(\log N)/a_{m-1} + 3a_{m-1} - 1)N^2$ となる。このとき, $a_{m-1} \leq N^{1/2}$ を満たすように選べば, 総計算量は $O(N^{5/2})$ となる。また, 各スロットシーケンスでの $a_{m-1} \times a_{m-1}$ の2DRRスケジューリングは並列に処理することが可能なので, この場合は $O(N^2)$ となる。

つぎに記憶容量であるが, グループ化の際に, 第 i レベル ($0 \leq i \leq m-2$) で $a_i \times a_i$ の PM が1枚と $a_{m(i)} \times a_{m(i)}$ サイズの SM が $a_{m(i)}$ 段階分必要となる。スケジューリングの際には, $a_{m-1} \times a_{m-1}$ の2DRRスケジューラが行われるため, $a_{m-1} \times a_{m-1}$ の PM が1枚と $a_{m-1} \times a_{m-1}$ サイズの SM が a_{m-1} 段階分, これに加えて $N \times N$ サイズの RM と AM が1枚ずつ必要である。 $a_{m-1} = a_{m(m-1)}$ であるから, 総記憶容量は $(\sum_{i=0}^{m-1} (a_{m(i)}^3 + a_i^2)) + 2N^2$ となる。また, $m \geq 2$ より $a_i/N \leq 1/2^{m-1} \leq 1/m$ であることから, 総記憶容量は $N^3/(\log N)^2 + N^2/(\log N) + 2N^2$ よりも小さいことがわかる。更に, $a_{m-1} \leq N^{1/2}$ のときは, $0 \leq i \leq m-1$ に対して $a_{m(i)} \leq N^{1/2}$ となることから, $\sum_{i=0}^{m-1} a_{m(i)}^3 \leq (\log N)N^{3/2}$ となる。したがって, 総記憶容量は $(\log N)N^{3/2} + N^2/(\log N) + 2N^2$, すなわち, $O(N^2)$ になることがわかる。また, $a_0 = \dots = a_{m-1}$ の場合は PM と SM は $a_{m-1} \times a_{m-1}$ のマトリクスだけですみ, 総記憶容量は最小の $N^{3/m} + N^{2/m} + 2N^2$ となる。

ところで, 2DRRスケジューラでは, 各入出力の組は $2N-1$ タイムスロットに最低1回の送信機会と, N タイムスロットに平均1回の送信機会が存在する[9]。同様な議論から, グループ化2DRRスケジューラでも

同じ結果を得ることができる。すなわち、 $m-1$ レベルにおいて、グループ化により選択された a_{m-1} 個の入出力からなる $a_{m-1} \times a_{m-1}$ の 2DRR スケジューラで、ある入出力の組がパターンシーケンス 0 で選ばれるのは、各段階で 1 回だけであり、 $0 \leq i \leq m-2$ レベルにおいて、あるグループがスロットシーケンス 0 で選ばれるのは各段階で 1 回だけである。したがって、ある入出力の組が送信機会を得てから、次の送信機会を得るまでの最長の時間は、まずすべてのレベルで、0 番目のパターンにある入出力の組が選ばれ、すべてのレベル i で次の段階に移り、すべてのレベルで a_i 番目のパターンにある入出力の組が選ばれる場合である。この場合、送信されるまでの待ち時間は最長の $2N-1$ になる。したがって、各入出力の組の送信機会は $2N-1$ タイムスロットに最低 1 回は存在する。

また、グループ化 2DRR スケジューラも 2DRR スケジューラと同様 N^2 タイムスロットで一巡し、 N^2 タイムスロットの間に少なくとも N 回の送信機会が得られることから、平均 N タイムスロットに最低 1 回の送信機会を得ることができる。

4. シミュレーションによる評価

本章では、本論文で提案したグループ化 2DRR スケジューラの評価を計算機シミュレーションによって行う [11]。

4.1 シミュレーションモデル

モデルとして 16 入力 16 出力と 48 入力 48 出力のスイッチを用いる。シミュレーション条件は以下のとおりである。

- パケット長は一定とする。
- パケットは送信が許可されたタイムスロット内で送信を完了する。
- 入力キューのバッファ長は無限とし、パケットの廃棄はないものとする。
- 各入力ポートへのパケットの到着間隔
 - 到着間隔は指数分布に従う。
 - 各入力ポートへの平均到着間隔は等しく、それぞれ独立な乱数発生系を用いる。
 - 到着パケットの出力先は等確率に割り当てる。

なお、精度検定はバッチサイズ 100,000、信頼区間 95% のバッチ平均法により行い、精度は 5% 以内とした。

4.2 シミュレーション結果と考察

図 9 に、16 入力 16 出力のスイッチでの高優先度パ

ケットの比率が 0.5 の場合の平均到着率と平均待ち時間の関係を示す。ここで、図中の“2DRR”は従来の 2DRR 方式の結果を、“Grouped 2DRR”は 16 個の入出力ポートを 4×4 にグループ化を行ったときの結果を表す。図から、グループ化を行った場合と行わなかった場合を比較すると、グラフがほぼ重なっており、遅延特性は、ほぼ同等であるといえる。

つぎに、特定の入力ポートにのみトラヒックが到着した場合、グループ化により入力ポートの割当てに偏りが生じないかどうかを調べた。図 10 に、16 入力 16 出力のスイッチを 4×4 にグループ化した場合とグループ化しない場合について、入力ポート 5~9 にのみトラヒックが到着した場合の各ポートのサービス率を示す。グループ化した場合は、対象入力ポートは第 0 レベルのグループ 1 のグループ内ポート番号 1~3 と、グループ 2 のグループ内入力ポート番号 0~1 に相当する。ここで、一つのグループ内の入力ポートだけを選択すると、それぞれの入力ポートが同じグループにグループ分けされることがないため、最初のパターンシーケンスにおいてパケットの衝突は起きない。こ

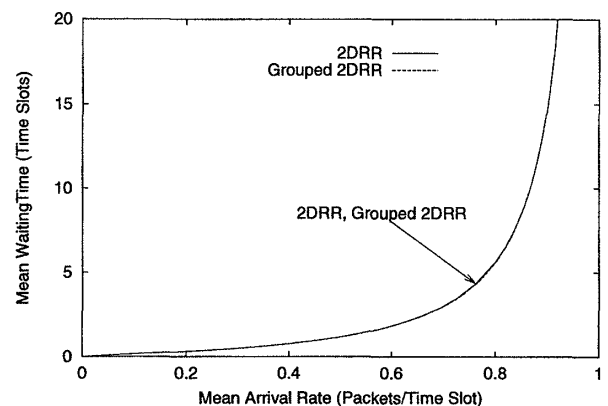


図 9 平均到着率に対する平均待ち時間

Fig. 9 The mean waiting times for the mean arriving rates.

	2DRR	grouped 2DRR
node 5	0.200943	0.20072
node 6	0.201391	0.201555
node 7	0.198394	0.198215
node 8	0.199195	0.200059
node 9	0.200077	0.19945

図 10 パケットが特定の五つの入力ポートにだけ到着する場合のサービス率

Fig. 10 The service rates when packets arrives at the fixed five input ports.

では複数のグループにまたがって入力ポートを選択しており、グループ化によって入力ポートの選択に偏りがある場合は、到着率の増減が現れるはずである。これにより、フェアネス（公平性）への影響を検証することができる。

さて、ここで入力のある入力ポートの数は五つであるので、フェアネスが達成されている場合、理論上それぞれのサービス率は0.2になる。図10から、グループ化しない場合は0.198394~0.201391, グループ化を行った場合は0.198215~0.201555であり、グループ化した場合、しない場合にかかわらずフェアネスが保たれていることが確かめられた。

以上の結果から、グループ化2DRR方式は2DRR方式と同等の性能を保持しながら、各ポートの選択機会の偏りをなくすためのスイッチサイズの制約を2DRR方式よりも緩和し、更に、計算量、記憶容量を軽減していることが示された。ただし、グループ化のレベル数が増えるほどグループ化2DRR方式オーバーヘッドは増えることから、レベル数は極力抑えることが望ましい。

つぎに、グループ化の分割方法による性能への影響を調べるため、48入力48出力のスイッチを、グループ化しない場合と、 4×12 , 12×4 , $2 \times 4 \times 6$, $6 \times 4 \times 2$, $12 \times 2 \times 2$ にグループ化した場合についてシミュレーションを行った。このときの平均到着率と平均待ち時間の関係を図11に示す。図中の“48”はグループ化しない場合を示す。“ 4×12 ”は 4×12 にグループ化したグループ化2DRRスケジューラの結果を表し、それ以下についても同様の結果を意味する。これより、 4×12 , 12×4 , $2 \times 4 \times 6$ にグループ化した結果、及び $6 \times 4 \times 2$, $12 \times 2 \times 2$ にグループ化した結果がそれぞれ重複しており、各々が同様の遅延特性をもつことがわかった。また、前者の結果と比べ、後者のほうが遅延が大きいことがわかる。

前者と後者の違いは、 $N = a_0 \times \dots \times a_{m-1}$ と表したときの a_{m-1} が2であるか否かということである。 $a_{m-1} = 2$ の場合、実際にスケジューリングを行う2DRRスケジューラのサイズは 2×2 であるため、同じグループに割り当てられる入力ポートのバリエーションが少ない。このために、競合する入力ポートが数個に限定されることから、 $a_{m-1} \neq 2$ の場合よりも平均待ち時間が大きくなると思われる。このことは、 a_{m-1} をいくつに設定するかが、 a_0, \dots, a_{m-2} をいくつに設定するかよりも非常に重要な意味をもつこと

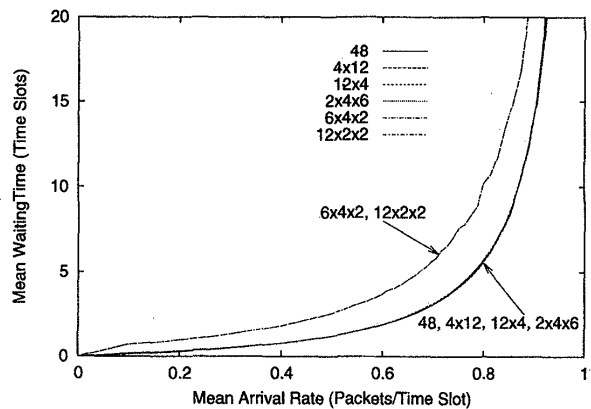


図11 グループ分けによる平均待ち時間の比較

Fig. 11 The comparison for the mean waiting times of grouping sorts.

を示している。 4×12 , 12×4 , $2 \times 4 \times 6$ にグループ化した場合はいずれも同様な結果が得られていることから、 a_{m-1} は4以上に設定すれば十分であるといえる。

また、グループ化しない場合と a_{m-1} が4以上でグループ化した場合の平均待ち時間はほぼ一致しており、入出力のサイズが48の場合においても、遅延特性がほぼ同等であることが示された。

5. むすび

2DRRスケジューラは、高いスループットの達成が可能であり、単純なアルゴリズムを用いているため、ハードウェアへの実装が容易であることから高速なスイッチングが可能であることが知られている。本論文では、2DRRスケジューラのこれらの特徴を損なうことなく、より多くのサイズのスイッチを構成可能にするグループ化2DRRスケジューラを提案した。そして、 N 入力 N 出力のスイッチを $N = a_0 \times \dots \times a_{m-1}$ とグループ化した場合、 $a_{m-1} \leq N^{1/2}$ ならば、2DRR方式では $O(N^3)$ であった計算量を $O(N^{5/2})$ に抑え、記憶容量は $O(N^3)$ から $O(N^2)$ に抑えることを示した。また、2DRR方式と同様にして、ある入出力の組について、 $2N - 1$ タイムスロットに確実に1回はパケットを送信する機会をもち、平均すると N タイムスロットに最低1回は送信する機会が得られることを示した。これは、ラウンドロビンスケジューリングにより得られる特質であり、フェアネスの実現のための重要な要素である。

なお、現在販売されているスイッチでの N は最大

数千程度であり、このようなスイッチに本スケジューラを適用した場合、2DRRを適用した場合と比較して、単純計算で計算量が $\sqrt{N} \sim 60 \sim 70$ 倍程度、記憶容量が数千倍程度の差が期待される。

つぎに、提案する方式を計算機シミュレーションにより性能評価を行い、グループ化した場合の遅延特性が2DRRと同等であることがわかった。更に、 N 入力 N 出力のスイッチを $N = a_0 \times \dots \times a_{m-1}$ とグループ化した場合、 a_{m-1} は4以上が適切であり、ある入力ポートにパケットの到着が集中した場合について各ポートのフェアネスが達成されていることを確かめた。それ以外の値であれば、 $a_i + 1$ が素数であるという制約を除けば同様な結果が得られるが、記憶容量の観点からすれば、各 a_i の最大値はなるべく小さいほうが好ましい。また、 N の因数に同じ値がある場合は、マトリクスが共用できるため、更に記憶容量が少なくなる。しかし、グループ化の数が多くなればなるほど計算が複雑になり、本スケジューラの特徴であるスケジューリングの単純さが失われるおそれがある。以上の観点からすると、例えば、3階層程度に抑えるならば、 $N = 48$ の場合は 4×12 や $2 \times 4 \times 6$ が適当である。

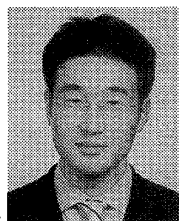
最後に、本論文ではパケットの到着過程としてポアソン過程を用いたが、現実のネットワークでのトラヒックはバースティな場合がほとんどである。提案するモデルの基本的な特性を解析することが本論文での目的であり、ポアソン過程による計算機シミュレーションを行うことで、提案したパケットスイッチの特性が十分把握できたが、今後は現実的なトラヒックによる解析を行い、実際のネットワークにより適応したスケジューラを提案することが課題である。

文 献

- [1] D. Bertsekas and R.G. Gallager, "Data Networks," pp.241-337, Prentice-Hall, Inc., 1987.
- [2] C. Partridge, "Gigabit Networking," pp.43-151, Addison-Wesley Publishing Company, 1994.
- [3] E.L. Hahne, "Round-Robin Scheduling for Max-Min Fairness in Data Networks," IEEE J. Selected Areas in Commun., vol.9, no.7, pp.1024-1039, 1991.
- [4] M.A. Bonuccelli, I. Gopal, and C.K. Wong, "Incremental Time-Slot Assignment in SS/TDMA Satellite Systems," IEEE Trans. Commun., vol.39, no.7, pp.1147-1156, 1991.
- [5] W.T. Chen, H.J. Liu, and Y.T. Tsay, "High-Throughput Cell Scheduling for Broadband Switching Systems," IEEE J. Selected Areas in Commun., vol.9, no.9, pp.1510-1523, 1991.

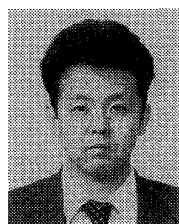
- [6] T. Inukai, "An Efficient SS/TDMA Time Slot Assignment Algorithm," IEEE Trans. Commun., vol.COM-27, no.10, pp.1449-1455, 1979.
- [7] M.G. Hluchyj and M.J. Karol, "Queueing in High-Performance Packet Switching," IEEE J. Selected Areas in Commun., vol.6, no.9, pp.1587-1597, 1988.
- [8] H. Matsunaga and H. Uematsu, "A 1.5 Gb/s 8×8 Cross-Connect Switch using a Time Reservation Algorithm," IEEE J. Selected Areas in Commun., vol.9, no.9, pp.1308-1317, 1991.
- [9] R.O. LaMaire, "Two-Dimensional Round-Robin Schedulers for Packet Switches with Multiple Input Queues," IEEE/ACM Trans. Networking, vol.2, no.5, pp.471-482, 1994.
- [10] J. Wong, J.P. Sauve, and J.S. Field, "A Study of Fairness in Packet Switching Networks," IEEE Trans. Commun., vol.COM-30, no.2, pp.346-353, 1982.
- [11] M.H. MacDougall, "Simulating Computer Systems: Techniques and Tools," pp.122-167, The Massachusetts Institute of Technology, 1987.

(平成10年6月5日受付, 11月16日再受付)



高城 衛 (正員)

平8筑波大・情報・情報科学卒。平10同大学院修士課程了。同年日本電気(株)に入社、現在に至る。マルチアクセスチャネルに関する研究などに従事。



木村 成伴 (正員)

平2東北大・工・情報卒。平7同大学院博士後期課程了。同年筑波大・電子・情報工学系講師、現在に至る。情報科博。代数プロセスの合成、フェアネスの導入などの研究に従事。情報処理学会、ソフトウェア学会各会員。



海老原義彦 (正員)

昭45東北大・工・電子卒。昭50同大学院博士課程単位取得退学。同年同大助手。現在、筑波大・電子・情報工学系教授。平10より同大・術情報処理センター長。工博。コンピュータネットワークアーキテクチャ、デジタル通信システムの性能評価、及び知的通信システムの研究などに従事。情報処理学会会員。