社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2007－ＳＬＤＭ－129／2007－ＥＭＢ－4（７）
2007／3／15

Responsive Linkの追い越し用バッファによる待ち時間を有効利用した実時間通信

片山 陽平† 加藤 真平†† 山崎 信行†,††

†慶應義塾大学理工学部情報工学科
††慶應義塾大学大学院理工学研究科開放環境科学専攻
〒223-8522 横浜市港北区日吉3-14-1
E-mail: †{katayama,shinpei,yamasaki}@ny.ics.keio.ac.jp

あらまし 本論文では，分散リアルタイム通信規格 Responsive Link を用いたマルチホップな周期通信の実時間通信手法について述べる．各ノード通信に対して，周期，ホップ数，転送時間に基づいてデッドラインを決定し，それに基づいてパケットの優先度決定とスケジュール可能性判定を行なう通信手法を提案する．その際，パケット退避機能によって生じる遅延を考慮したスケジュール可能性解析を用いることで，より厳密にリアルタイム性を保証する．シミュレーションの結果，パケット退避機能を考慮したスケジューリング可能性解析を用いた評価において，コネクションの受け入れ率が最大15%，ネットワーク全体のスループットが最大55%向上したことを示す．

キーワード リアルタイム通信，パケットスケジューリング，追い越し用バッファ，Responsive Link

Real-Time Communication with Efficient Use of Waiting Time in the Overtaking Buffer of Responsive Link

Yohei KATAYAMA†, Shinpei KATO††, and Nobuyuki YAMASAKI†,††

† Department of Information and Computer Science, Faculty of Science and Technology, Keio University
†† Keio University
3-14-1 Hiyoshi, Kouhoku-ku, Yokohama, Kanagawa 223-8522 Japan
E-mail: †{katayama,shinpei,yamasaki}@ny.ics.keio.ac.jp

Abstract In this paper, we deal with the method of real time communication for multi hop periodic communication that uses Responsive Link, which is designed for distributed real-time system. We propose the method of communication which dicides deadline for multi hop communication on each node based on period, the number of hops and transferring time, decides priority of communication and tests the schedulability of the communication. Additionaly, we propose the schedulability analysis for latency caused by the mechanism of buffer overflow prevention for overtaking buffer to guarantee strict real-time behavior. The result of simulation shows that acceptance ratio has improved by 15% and throughput has improved by 55% in the maximum on the evaluation which uses the schedulability analysis proposed by us.

Key words Real-Time Communication, Packet scheduling, Overtaking Buffer, Responsive Link

1. 序 論

近年，様々な分野において分散リアルタイムシステムの必要性が高まっている．分散リアルタイムシステムはノード内の処理だけでなく，ノード間の通信においても，リアルタイム性が求められる．

Responsive Link は分散リアルタイム制御用に設計および実装された通信規格であり，スループットの高いマルチホップなリアルタイム通信の実現をサポートする．Responsive Link は現在，国内では情報処理学会試行標準 IPSJ-TS 0006:2003 として標準化され，国際的には ISO/IEC SC25 WG4 において標準化作業が行なわれており，今後，分散リアルタイムシステムに広く用いられる可能性が高い通信規格である．

Responsive Link の優先度制御はソフトウェア依存である．そのため，高いスループットの実現とリアルタイム性の保証を行なうにはソフトウェアによる適切な通信管理が必要になる．

本論文の構成は以下の通りである．2. 章では，本研究の前提となる通信規格 Responsive Link の特徴と，Responsive Link を

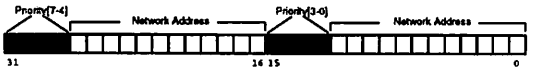

図 1　Responsive Link のヘッダフォーマット

用いたリアルタイム通信について述べる．3. 章では，マルチホップな周期通信のリアルタイム性の保証とコネクション受け入れ率，スループットの向上を実現するパケットスケジューリング手法について述べる．4. 章では，追い越し用バッファのパケット退避機能によって生じる遅延の影響を考慮したスケジュール可能性解析について述べる．5. 章では，本提案手法の評価を示す．最後に，6. 章で結論と今後の課題を述べる．

2. 背　　景

2.1 Responsive Link

Responsive Link [4] は，柔軟なリアルタイム通信を実現するために，優先度によるパケットの追い越し機構，優先度による経路制御，各ノードにおけるパケットの優先度付替，トポロジーフリーな結合形態等の様々な機能を実現した，分散リアルタイム通信規格であり，我々が提案し，我々が研究開発を行っている Responsive Multithreaded Processor (RMT Processor) [2] に実装されている．

図 1 に Responsive Link におけるパケットのヘッダフォーマットを示す．優先度を用いた追い越し機構を実装するために，ネットワークアドレスに 8bit の優先度を付加している．通信パケットは固定長となっている．

Responsive Link では，各通信ノードが持つルーティングテーブルを用いてハードウェアにより静的な経路制御を行う．ルーティングテーブルはローカルなノードからソフトウェアによって読み書きされる．通信パケットが入力されると，パケットのヘッダ部に付加されたネットワークアドレスと優先度を元にルーティングテーブルを参照し，出力ポート番号を決定する．その際に，パケットの優先度を付け替えることが可能である．

Responsive Link では，追い越し用バッファと退避用外部記憶を有したネットワークスイッチによって，優先度を用いたパケットの追い越し機構を実現している．

図 2 は 5 入力 5 出力で 1 つの入力部あたり追い越し用バッファが 4 パケット分あるネットワークスイッチの一部を簡略に示したものであり，入力ポート (In1,In2) に対する入力部とその周辺部である．入力ポート (In) と出力ポート (Out) において最後の数字はポート番号を示している．

Responsive Link の追い越し機構では，異なる入力ポートから入力された通信パケットが衝突した場合，低優先度のパケットを追い越し用バッファに貯めて出力を待たせ，高優先度の通信パケットを先に出力させることにより，パケットの追い越しを実現している．また，同じ入力ポートにおいて既に出力待ちを行なっている通信が存在する場合において，利用していない追い越し用バッファが存在し，かつ，すぐに出力できる通信パケットが入力された場合には，同一経路上の先行する低優先度パケットを追い越すことが可能である．

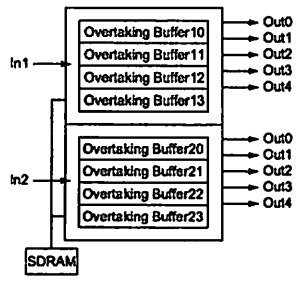

図 2　簡略に示した Responsive Link のネットワークスイッチの一部

一方，同じ入力ポートにおいて送信待ちを行なう通信パケットが連続して入力された場合，追い越し用バッファを溢れさせないために少なくとも 1 つ以上の空バッファを常に用意する．空バッファが少なくなって行き，残り 1 本になった場合，次からの入力パケットは空バッファを利用して退避用外部記憶 (SDRAM 等) に退避する．出力が進み，空バッファが残り 2 本以上になると，退避用外部記憶に退避されていた入力パケットを優先度に従って追い越し用バッファに書き戻すことにより，出力を継続する．退避されたパケットが全て書き戻されるまで，入力パケットは空バッファに退避される．

2.2 Responsive Link を用いたリアルタイム通信

Responsive Link は，ソフトウェアによってルーティングテーブルを設定することによって，パケットの優先度付替，経路制御等を行なう．また，パケットの優先度付加ポリシもソフトウェアによって決定する．このため，ソフトウェアによって適切に優先度制御を行ない，スケジュール可能性判定と受け付け制御を行なうことで通信のリアルタイム性を保証する必要がある．

マルチホップなリアルタイム通信では，ホップ数が多くなるほどデッドラインミスしやすくなる．このため，ホップ数に応じた優先度決定を行なうことで，コネクションの受け入れ率が向上する．

例として，ノード N_0 をソースノード，ノード N_3 をデスティネーションノードとし，ノード N_1，ノード N_2 を経由するコネクション C_i を考える．コネクション C_i は周期とデッドラインが同じ長さであるとする．

コネクション C_i の周期 P_i を各通信路における通信のデッドラインとして割り当ててスケジュールした場合の各ノードにおけるスケジューリング例を図 3 に示す．

図 3 の四角形で表される部分は転送された通信パケットの集合を表す．$C_i^{[n]}$ はコネクション C_i が n 番目に経由する通信路におけるコネクション C_i の通信を表し，$d_i^{[n]}$ は $C_i^{[n]}$ の絶対デッドラインを表す．図 3 の灰色部は Responsive Link の追い越し機構によってコネクション C_i より高優先度のコネクションの通信が行なわれている状態を表す．

図 3 では，各通信路のデッドラインが長く設定されているために，通信全体のデッドラインを守れない．

[3] は，経路上の各ノードにおいて周期をホップ数 h_i で等分したデッドライン (式 1) を設定し，その値に基づいて優先度決定とスケジュール可能性判定を行なうことで，通信全体のリア

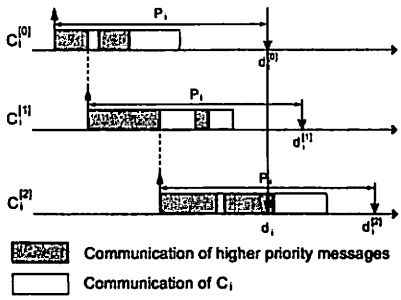

図3　各通信路に対して全体と同じデッドラインを割り当てた場合のスケジュール例

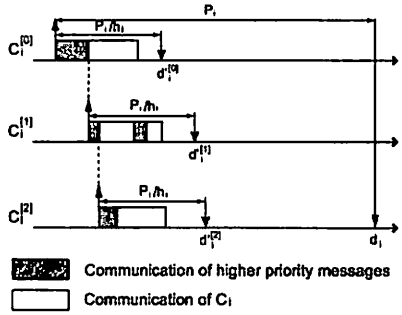

図4　[3] の手法を用いたときのスケジューリング例

ルタイム性を保証する方法を提案，実装した．この手法を用いたときのスケジュール例を図4に示す．

$$d_i^{[n]} = \frac{P_i}{h_i} \quad for \quad \forall n \tag{1}$$

[3] のパケットスケジューリングはデッドラインに基づいた優先度決定を行なうため、Deadline Monotonic [5] である．

3. パケットスケジューリング

従来手法 [3] はパケットの転送時間と各ノードにおけるパケットの到着時刻を考慮していないため，全体のスループットが低下する．つまり，複数ノードを経由する通信の優先度が必要以上に高く設定されているために，他の通信の資源割り当てが阻害される．また，デッドラインが短く設定されているために，その通信自体のスケジュール可能性も低下する．

そこで，パケットの転送時間と各ノードにおけるパケットの到着時刻を考慮したデッドライン設定方法を提案する．

Responsive Link を用いてマルチホップに通信を行なう場合，各ノードに到着したパケットは，到着したものから順に次のノードに転送される．そのため，経路上の前のノードにおける転送時間と次のノードにおける転送時間をオーバーラップして考えることができる (図5)．これにより，式1の各ノードにおける相対デッドライン $D_i^{[n]}$ は式2に改良することができる．ただし，コネクション C_i の各周期のデータ転送量を T_i とした．

$$D_i'^{[n]} = \frac{P_i + (h_i - 1)T_i}{h_i} \quad for \quad \forall n \tag{2}$$

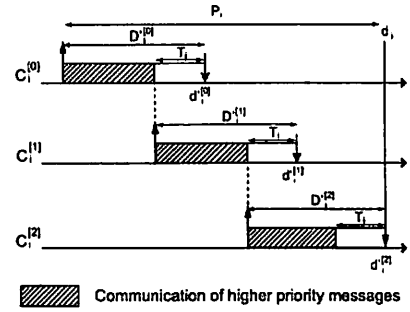

図5　転送時間を考慮したデッドライン設定

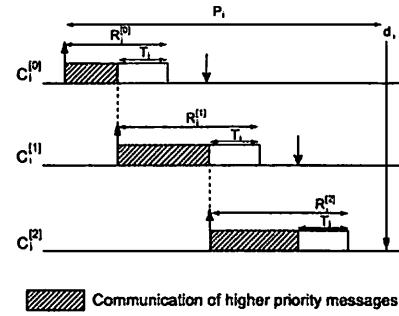

図6　最悪応答時間に基づくデッドライン設定

次に，各ノードにおける到着時刻を考慮したデッドライン割り当て方法について述べる．

各ノードでは，新しく追加するコネクションのスケジュール可能性判定を行なう場合，Response Time Analysis (RTA) [1] を行なう．RTA では，最悪応答時間とデッドラインの大小関係によってスケジュール可能性が判定される．応答時間とは，通信の到着から通信が完了するまでの時間を指し，最悪応答時間は，解析上最悪の状況を仮定した場合の応答時間である．

最悪応答時間により，経路上の直前のノードにおいて，設定されたデッドラインよりも最悪でどの程度早く転送が完了するのかがわかる．図6のように，早く転送が完了した分の時間だけ，以降のノードのデッドラインを長く設定することで，図5の場合よりもスケジュール可能性を高めることができる．

コネクション C_i の n 番目に経由する通信路における通信 $C_i^{[n]}$ は，その通進路までの経路上の通信 $C^{[k]}(0 \le k < n)$ のデッドラインと最悪応答時間の差分をデッドラインに加えることができる．新しく設定されるデッドライン $D_i''^{[n]}$ は，式3で表される．

$$D_i''^{[n]} = D_i'^{[n]} + \sum_{k=0}^{n-1} (D_i'^{[k]} - R_i^{[k]}) \tag{3}$$

この方法を用いる場合，通常のスケジュール可能性判定に加えて，新しいコネクションの追加によって最悪応答時間が変化した通信に対してもスケジュール可能性判定を行なう必要がある．

D'' を優先度決定に用いた場合，連鎖的に再計算が必要になるため，優先度決定には D' を用い，D'' はスケジュール可能性

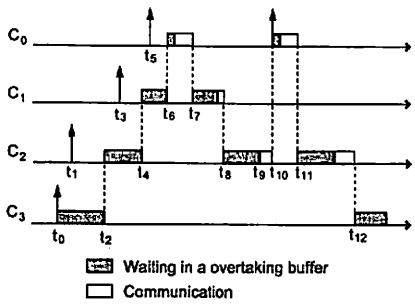

図7 最悪状態を考慮したスケジュール例

判定にのみ用いる．

4. スケジュール可能性解析

Responsive Link の追い越し機構では，先に到着したパケットの出力待ちによって追い越し用バッファが占有されると，パケットの退避機能により，出力待ちを行なっているパケットが出力されるまでの間，後から入力されたパケットが優先度に関わらず退避用外部記憶において待たされる現象が生じる．出力ポートにおける衝突と異なり，出力待ちに関しては高優先度パケットによるプリエンプションはできない．先に到着した低優先度のパケットが長時間出力待ちを行なう可能性があるため，パケット退避機能による遅延を無視すると厳密にリアルタイム性を保証できない．

一周期のデータ転送量が数パケット程度しかない場合，パケット退避機能による遅延の影響はあまりないと考えられる．一方，一周期のデータ転送量がある程度大きい場合，一度通信が出力待ちを行なうと，連続的に同じ通信のパケットが入力されるため，パケット退避機能による遅延が生じる可能性が大きい．

4.1 解析手法

以下，各通信のデータ転送量は十分大きく，転送待ちを行なった場合，必ず追い越し用バッファを占有するものと仮定する．

パケットの退避機構を考慮した場合の最悪応答時間を考えるために，図7を用いる．以降，パケットの退避機構を考慮した場合の最悪応答時間を**実最悪応答時間**と呼び，考慮しない場合のものと区別する．図7は同じ入力部におけるパケットスケジューリング例である．図7の上矢印は通信の到着を，灰色の四角形は追い越し用バッファにおける出力待ち状態を表し，各通信 C_i の優先度は添字 i の小さいものほど高優先度とする．

例えば，図7の時刻 t_1 において，通信 C_2 が到着するが，先に到着した通信 C_3 が既に出力待ちを行なっているため，SDRAM に退避される．時刻 t_2 で通信 C_3 が出力を再開すると，SDRAM に退避された通信 C_2 のパケットが優先度に従って追い越し用バッファに書き戻される．通信 C_3 のパケットは，出力待ちを行なっていたもの以外は SDRAM に退避されており，高優先度の通信が全て出力されるまで出力されない．追い越し用バッファに書き戻された通信 C_2 のパケットは出力ポートで衝突が起きたために時刻 t_2 から出力待ちを行なっている．時刻 t_3 から時刻 t_8 の間のように，通信 C_2 が出力を行なう前に高優先度通信が到着すると，通信 C_2 の出力は進まない．時刻 t_8 から時刻 t_{10} のように出力の再開後まで高優先度の通信が同じ入力ポートから入力されなかった場合に時刻 t_9 から時刻 t_{10} のように連続的な出力が行なわれる．

入力ポートのスケジューリングにおいて，各通信は，各周期に最大1回，先に到着した低優先度通信の出力待ちによって待たされる可能性がある．そのため，ある通信の最悪状態が発生する状況は，以下の二つの条件を満たす場合である．

- 低優先度通信のうち最も出力待ちの長いものが当該通信の到着直前に到着し，かつ，最大の出力待ちを行なう．
- 当該通信の到着後に到着した高優先度通信が，その高優先度通信よりも優先度が低い通信のうち，最も出力待ちが長いものによって，最大の遅延を被る．

ここで，各通信は各周期に最大で一回しかプリエンプションを行なわないことに着目して，プリエンプションを行なう通信と，プリエンプションされた通信のプリエンプション直前の出力待ち時間を一組みにして考える．これにより，扱いの難しいプリエンプション不可能な低優先度通信の出力待ちを，プリエンプションを行なった高優先度通信の転送時間の一部に置き換えて考えることができる．

入力ポートが同じ通信の集合を $\{C_m | 0 \le m < M\}$ とおく．M は要素の個数を表し，添字 m が小さいものほど高優先度とする．このとき C_m の実最悪応答時間 R'_m は，式4で表される．

$$R'_m = T'_m + \sum_{k=0}^{m-1} \left\lceil \frac{R'_m}{P_k} \right\rceil T'_k \tag{4}$$

$$T'_k = \begin{cases} \max\limits_{(j|k<j\le m)} \left\{ R_j - T_j \right\} + R_k & for \quad k < m \\ \max\limits_{(j|m<j)} \left\{ R_j - T_j \right\} + R_m & for \quad k = m \end{cases} \tag{5}$$

T'_k はその通信自体の最悪応答時間 R_k に，その通信にプリエンプションされた通信のプリエンプション直前の出力待ち時間の最悪値を加えたものである．式4は RTA の表現式と同形であり，漸化式によって再帰的に解くことができる．

5. 評　価

5.1 評価方法

評価として，確立するコネクション数を 25,75,...,925,975 と変化させたときのコネクション受け入れ率とネットワーク全体のスループットに関する比較を行なった．シミュレーションは 4×4 の 2D トーラス (図8) と木構造 (図9) の2種類のネットワークトポロジーについて行なった．発生させるコネクションは次の2種類のパラメータを用いた．

表1 コネクションのパラメータ

	パラメータ 1	パラメータ 2
周期	変動幅が ±5% の一様分布	変動幅が ±5% の一様分布
帯域幅利用率	[0.01,0.1] の一様分布	[0.01,0.5] の一様分布

パラメータ1は発生するコネクションの帯域幅利用率が比較

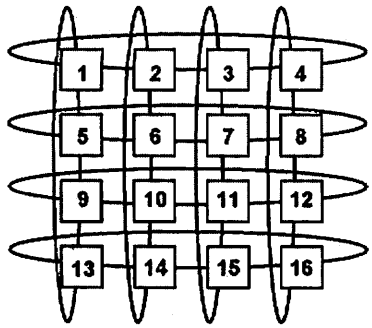

図 8　シミュレーションに用いたネットワークトポロジー (4×4 の 2D トーラス)

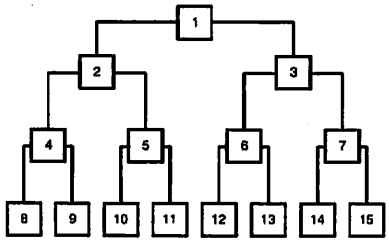

図 9　シミュレーションに用いたネットワークトポロジー (木構造)

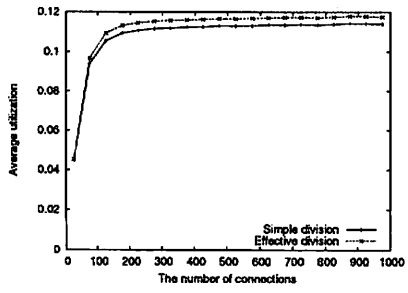

図 10　スループット (2D トーラス，帯域幅利用率小)

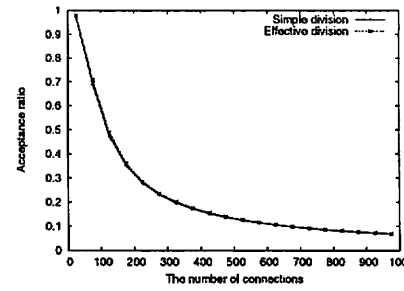

図 11　受け入れ率 (2D トーラス，帯域幅利用率小)

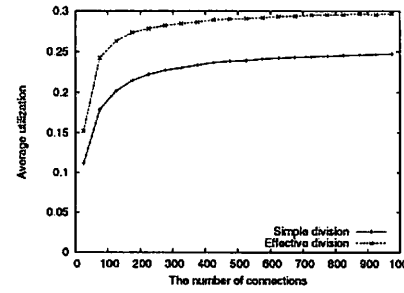

図 12　スループット (2D トーラス，帯域幅利用率大)

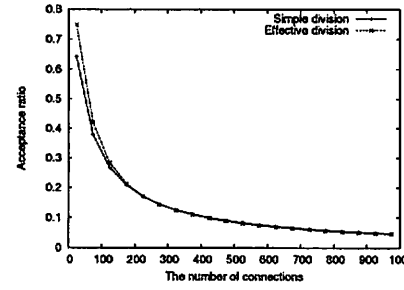

図 13　受け入れ率 (2D トーラス，帯域幅利用率大)

的小さく，パラメータ 2 は発生するコネクションの帯域幅利用率が比較的大きい．スケジュール可能性判定においては，周期の大きさではなく，周期の大きさに対する変動幅が意味を持つので，変動幅のみを指定している．

ソースノードとディスティネーションノードは無作為に決定した．

スケジュール可能性判定においては，パケットの退避機能による遅延を考慮したスケジュール可能性解析を用いた．

5.2　評価結果

5.2.1　4×4 の 2D トーラスでの評価

パラメータ 1(帯域幅利用率小) を用いた場合の，従来手法 [3](Simple division)，提案手法 (Effective division) の評価結果を図 10 と図 11 に示す．

選択可能な経路が多いネットワーク構造であることと，各コネクションの帯域幅利用率が小さいことから，手法間の違いはあまり見られない．

次にパラメータ 2(帯域幅利用率大) を用いた場合の各手法の評価結果を図 12 と図 13 に示す．

パラメータ 2 を用いた場合には，従来手法と比較して最大で，受け入れ率が 17%，スループットが 37%向上している．これは，各コネクションの帯域幅利用率が大きいほど提案手法が有効であるためである．

図 13 では，コネクション発生数が 200 以上になると手法間で受け入れ率の違いが見られないが，これは，提案手法のほうがより帯域幅の大きいコネクションを受け入れているためである．これはコネクション発生数の少ない辺りにおける受け入れ率と図 12 のスループットから確認できる．

図 10 と図 12 を比較すると，パラメータ 2 を用いた方がスループットが高いことがわかる．また，図 11 と図 13 を比較すると，全体のコネクション確立数はパラメータ 1 を用いた方が高いことがわかる．つまり，パラメータ 1 を用いた場合には，帯域幅利用率の小さなコネクションが多く確立されるが，帯域

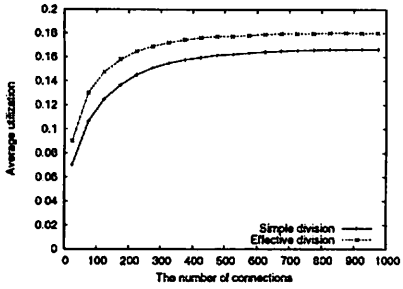

図 14　スループット (木構造，帯域幅利用率小)

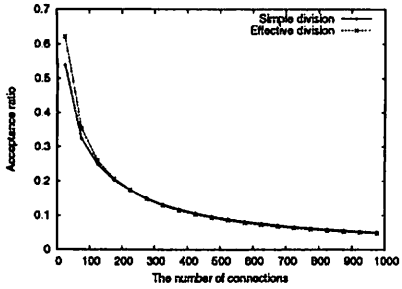

図 15　受け入れ率 (木構造，帯域幅利用率小)

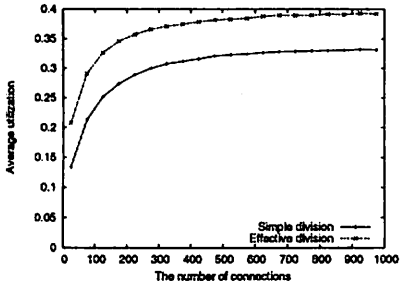

図 16　スループット (木構造，帯域幅利用率大)

利用率の大きな通信を用いた場合に比べてスループットは出ないと言える．これは，パケット退避の影響を考慮しているためである．多くの通信が発生した場合，出力待ちを行なう通信が増えるため，パケット退避による遅延が大きくなり，スループットが低下する．

5.2.2　木構造での評価

パラメータ 1(帯域幅利用率小) を用いた場合の各手法の評価結果を図 14 と図 15 に，パラメータ 2(帯域幅利用率大) を用いた場合の各手法の評価結果を図 16 と図 17 に，それぞれ示す．

木構造による評価結果と 2D トーラスによる評価結果を比較した場合，木構造を用いた場合の方が，2D トーラスを用いた場合に比べて全体的にスループットが向上している．これは，木構造は負荷が偏りやすい構造であり，最悪応答時間を考慮した場合には，負荷の低いノードにおける余裕時間を負荷の高いノードに当てることが可能なためである．

図 14 と図 15 の評価において，提案手法は従来手法に比べて最大で，受け入れ率が 16%，スループットが 28%向上している．

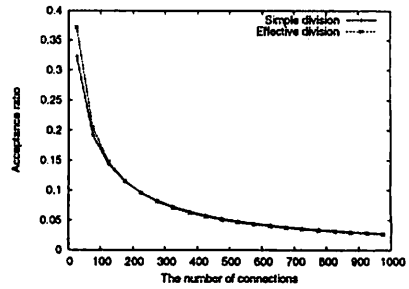

図 17　受け入れ率 (木構造，帯域幅利用率大)

また，図 16 と図 17 の評価において，提案手法は従来手法に比べて最大で，受け入れ率が 15%，スループットが 55%向上している．

6. 結　　論

本研究では，Responsive Link の追い越し機構を考慮したパケットスケジューリング手法とスケジュール可能性解析手法を提案した．

パケットスケジューリング手法では，経路上のノードに設定するデッドラインを適切に設定することで，通信全体のリアルタイム性保証とスケジュール可能性向上を同時に実現した．また，スケジュール可能性解析手法では，Responsive Link の追い越し機構によるパケットの退避を考慮したスケジュール可能性解析を行ない，提案のパケットスケジューリング手法と合わせて用いることで，厳密なリアルタイム性保証を実現した．

これらの提案手法に対してシミュレーションによる評価を行ない，その有効性を評価した．評価の結果，提案手法では従来の手法に比べ最大で，コネクション受け入れ率が 15%，スループットが 55%向上した．

今後の課題として，本研究ではノード間の通信に対してのみスケジュール可能性判定を行なったが，各ノードがローカルなプロセッサとの間でパケットを読み書きする場合についてもノード間通信と同様の機構を用いているため，同様のスケジュール可能性判定をローカルなプロセッサとの通信部分に対しても行なう必要がある．

謝辞　本研究は，科学技術振興機構 CREST の支援による．

文　　献

[1] N. Audsley, A. Burns, M. Richardson, K. Tindell, and A. J. Wellings. Applying New Scheduling Theory to Static Priority Pre-emptive Scheduling. *Software Engineering Journal*, 8(5):285–292, 1993.

[2] N. Yamasaki. Responsive Multithreaded Processor for Distributed Real-Time Systems. *Journal of Robotics and Mechatronics*, 17(2):130–141, 2005.

[3] 佐々木 貴宏. Responsive Link を用いたリアルタイム通信管理機構. 修士論文 慶應義塾大学理工学部, 2006.

[4] 山崎 信行. 分散制御用リアルタイム通信 Responsive Link の設計および実装. 情報処理学会論文誌コンピューティングシステム, 45(SIG 3 (ACS 5)):50–63, 2004.

[5] N. C. Audsley. Deadline Monotonic Scheduling. *YCS 146, Department of Computer Science, University of York*, 1990.