

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4995101号
(P4995101)

(45) 発行日 平成24年8月8日(2012.8.8)

(24) 登録日 平成24年5月18日(2012.5.18)

(51) Int.Cl.

H04L 12/56 (2006.01)

F 1

H04L 12/56 200Z

請求項の数 20 (全 28 頁)

(21) 出願番号 特願2007-557329 (P2007-557329)
 (86) (22) 出願日 平成18年2月28日 (2006.2.28)
 (65) 公表番号 特表2008-532145 (P2008-532145A)
 (43) 公表日 平成20年8月14日 (2008.8.14)
 (86) 國際出願番号 PCT/DK2006/000119
 (87) 國際公開番号 WO2006/089560
 (87) 國際公開日 平成18年8月31日 (2006.8.31)
 審査請求日 平成21年2月26日 (2009.2.26)
 (31) 優先権主張番号 PA200500304
 (32) 優先日 平成17年2月28日 (2005.2.28)
 (33) 優先権主張国 デンマーク (DK)
 (31) 優先権主張番号 60/656,377
 (32) 優先日 平成17年2月28日 (2005.2.28)
 (33) 優先権主張国 米国 (US)

(73) 特許権者 507289117
 テクラテック・アクティーゼルスカブ
 T E K L A T E C H A / S
 デンマーク、デーコー-2800リングビ
 ー、ビグニング377、ディプロムヴァイ
 (74) 代理人 100101454
 弁理士 山田 阜二
 (74) 代理人 100081422
 弁理士 田中 光雄
 (74) 代理人 100125874
 弁理士 川端 純市
 (72) 発明者 トビアス・ビエルレゴー
 デンマーク、デーコー-2000フレデリ
 クスペアウ・セー、2チル・ヴェンストレ
 、ダルガス・ハーウェ34番
 最終頁に続く

(54) 【発明の名称】共有リソースへのアクセスを制御する方法及びシステム

(57) 【特許請求の範囲】

【請求項 1】

共有リソースへのアクセスを制御する方法であって、

- 互いに異なる3つ以上の所定の優先順位のうちの1つの優先順位を既にそれぞれ割り当てられているデータ項目であって、送信されるべき少なくとも1つのデータ項目及び当該データ項目にそれぞれ割り当てられた少なくとも1つの優先順位のうちの少なくとも一方をデータプロバイダから受信するステップと、

- 複数のデータ項目を上記リソースに、

・ 第1の時点において、最も高い非待ち受けの優先順位を有する1つの送信されるべきデータ項目を上記共有リソースに送信し又は当該送信されるべきデータ項目の識別を上記データプロバイダに送信するステップと、

・ その優先順位が、引き続き、上記第1の時点において非待ち受けでありかつ空でなかった低い各優先順位からの1つのデータ項目の送信を待ち受けるステップとによって、繰り返し送信することを含む共有リソースへのアクセスを制御する方法。

【請求項 2】

上記複数の優先順位のうちの1つの優先順位にそれぞれ関係する複数の待ち行列を提供するステップをさらに含み、

- 上記受信するステップは、1つの待ち行列内の各データ項目又は優先順位を受信することを含み、上記待ち行列は上記データ項目に割り当てられた優先順位に関係し、

- 上記送信するステップは、上記最も高い優先順位を有する1つの非待ち受けの待ち行

列から上記データ項目又は識別を送信することを含み、

- 上記待ち受けるステップは、その待ち行列が、その後、低い優先順位を有する全ての空でない非待ち受けの待ち行列からの1つのデータ項目又は識別の送信を待ち受けることを含む請求項1記載の方法。

【請求項3】

上記待ち受けるステップは、非待ち受けの低い各優先順位からのデータ項目又は識別が上記データプロバイダから送信されたとき、上記待ち行列からの次のデータ項目又は優先順位をメモリへ転送することを含み、

上記送信するステップは、上記メモリ内の最も高い優先順位を有するそのデータ項目又は当該データ項目の識別を送信することを含む請求項2記載の方法。

10

【請求項4】

複数の上記データプロバイダを提供するステップをさらに含み、

上記複数のデータプロバイダはそれぞれ、

- 上記複数の優先順位のうちの1つの優先順位をすでにそれぞれ割り当てられている複数のデータ項目を提供するように、かつ、

- 上記データプロバイダから上記共有リソースへの送信準備が整っている少なくとも1つのデータ項目及び当該データ項目にそれぞれ割り当てられた少なくとも1つの優先順位のうちの少なくとも一方を提供するように適合化されており、

- 上記送信するステップは、1つのデータプロバイダが上記データ項目を上記リソースに送信することを含み、

20

- 上記待ち受けるステップは、低い各非待ち受けの優先順位からのデータ項目がデータプロバイダから上記共有リソースに送信されるまで、上記待ち受けの優先順位を有する送信準備の整ったデータ項目の任意のデータプロバイダからの送信を防止することを含む請求項1乃至3のうちのいずれか1つの請求項記載の方法。

【請求項5】

上記複数のデータプロバイダは、指示されるとそれに従って、データ項目を転送するように適合化されており、

- 上記送信するステップは、最も高い非待ち受けの優先順位の送信準備の整ったデータ項目を有する1つのデータプロバイダに上記データ項目を上記共有リソースに転送するように指示することを含み、

30

- 上記待ち受けるステップは、待ち受けの優先順位の送信準備の整ったデータ項目を有するデータプロバイダに、上記待ち受けの優先順位の任意のデータ項目を送信するように指示しないことを含む請求項4記載の方法。

【請求項6】

上記受信するステップは、上記複数のデータ項目を受信しつつ上記複数のデータ項目を記憶装置又はメモリに提供することを含み、

上記送信するステップは、上記データ項目を上記記憶装置又はメモリから送信することを含む請求項1乃至5のうちのいずれか1つの請求項記載の方法。

【請求項7】

上記共有リソースは、リンク、メモリ、プロセッサ、集積回路又はクロスバーである請求項1乃至6のうちのいずれか1つの請求項記載の方法。

40

【請求項8】

1つの所定の優先順位のための上記待ち受けるステップは、上記所定の優先順位が複数のデータ項目を送信したときのみ、上記少なくとも1つのデータ項目の少なくとも1回の送信を待ち受けることを含む請求項1乃至7のうちのいずれか1つの請求項記載の方法。

【請求項9】

共有リソースへのアクセスを制御する方法であって、

- 互いに異なる3つ以上の所定の優先順位のうちの1つの優先順位を既にそれぞれ割り当てられているデータ項目であって、送信されるべき少なくとも1つのデータ項目及び当該データ項目にそれぞれ割り当てられた少なくとも1つの優先順位のうちの少なくとも一

50

方をデータプロバイダから受信することと、

- 各データ項目は、当該データ項目に関する優先順位の1つの送信プロセスを割り当てられており、優先順位毎に少なくとも1つの送信プロセスを定義することと、
- 複数のデータ項目を上記リソースへ、

・第1の時点において、最も高い優先順位のために定義される非待ち受けの送信プロセスを割り当てられた送信されるべきデータ項目を上記共有リソースに送信し又は当該送信されるべきデータ項目の識別を上記データプロバイダに送信することと、

・引き続き、最も高い非待ち受けの優先順位のために定義される上記非待ち受けの送信プロセスが、そのうちの少なくとも1つの送信プロセスが上記第1の時点において非待ち受けでありかつ空でない低い各優先順位からのデータ項目の送信を待ち受けることによって、繰返し送信することを含む共有リソースへのアクセスを制御する方法。
10

【請求項10】

上記送信プロセスは、上記時点において送信された上記データ項目又は当該データ項目の識別の送信の時点においてデータ項目又は当該データ項目に割り当てられた優先順位のみが受信されていた、データ項目及び当該データ項目の識別のうちの少なくとも一方の送信を待ち受ける請求項9記載の方法。

【請求項11】

共有リソースへのアクセスを制御するためのシステムであって、

- 互いに異なる3つ以上の所定の優先順位のうちの1つの優先順位を既にそれぞれ割り当てられているデータ項目であって、送信されるべき少なくとも1つのデータ項目及び当該データ項目にそれぞれ割り当てられた少なくとも1つの優先順位のうちの少なくとも一方をデータプロバイダから受信するように適合化された受信手段と、
20

- 送信機手段とを備え、

上記送信機手段は、

・送信されるべき最も高い非待ち受けの優先順位を有する1つのデータ項目を上記共有リソースに送信し又は当該送信されるべきデータ項目の優先順位の識別を上記データプロバイダに送信することと、

・引き続き、その優先順位からの1つのデータ項目の送信を、上記データ項目又は当該データ項目の識別の送信の時点で非待ち受けでありかつ空でなかった低い各優先順位によって1つのデータ項目が送信されるまで待機させることによって、複数のデータ項目を上記リソースに繰返し提供させるように適合化されたシステム。
30

【請求項12】

上記複数の優先順位のうちの1つの優先順位にそれぞれ関連する複数の待ち行列をさらに備え、

- 上記受信手段は、1つの待ち行列内に各データ項目又は優先順位を提供するように適合化されており、上記待ち行列は上記データ項目に割り当てられた優先順位に関係し、

- 上記送信機手段は、最も高い非待ち受けの優先順位を有する上記データ項目又は識別を上記最も高い優先順位を有する任意の非待ち受けの待ち行列のデータ項目から送信し、かつ、引き続き、その行列に低い優先順位に関連する全ての非待ち受けの待ち行列からのデータ項目又は優先順位の送信を待ち受けさせるように適合化される請求項11記載のシステム。
40

【請求項13】

上記送信機手段は、非待ち受けの低い各優先順位からのデータ項目又は識別が1つの待ち行列から送信されたとき、次のデータ項目又は識別を上記待ち行列からメモリへ転送し、かつメモリ内の上記最も高い優先順位を有するそのデータ項目又は識別を上記リソースに送信するように適合化される請求項12記載のシステム。

【請求項14】

複数の上記データプロバイダをさらに備え、

上記複数のデータプロバイダはそれぞれ、

- 上記複数の優先順位のうちの1つの優先順位をそれぞれ割り当てられている複数のデ
50

ータ項目を提供するように、かつ、

- 上記データプロバイダから上記共有リソースへの送信準備が整っている少なくとも 1 つのデータ項目及び当該データ項目にそれぞれ割り当てられた少なくとも 1 つの優先順位のうちの少なくとも一方を提供するように適合化されており、

上記送信機手段は、上記最も高い非待ち受けの優先順位を有する上記データ項目を上記リソースに送信するように 1 つのデータプロバイダに対して指示し、その後、上記最も高い非待ち受けの優先順位より低い各非待ち受けの優先順位からのデータ項目がデータプロバイダから上記共有リソースに送信されるまで、上記最も高い非待ち受けの優先順位を有しあつ送信準備の整っているデータ項目が任意のデータプロバイダから送信されることを防止するように適合化される請求項 1 1 乃至 1 3 のうちのいずれか 1 つの請求項記載のシステム。

10

【請求項 1 5】

上記複数のデータプロバイダは、指示されるとそれに従って、データ項目を転送するように適合化されており、

上記送信機手段は、上記最も高い非待ち受けの優先順位の送信準備の整っているデータ項目を有する 1 つのデータプロバイダに上記データ項目を上記共有リソースに転送するように指示し、かつ、引き続き、上記待ち受けの優先順位の送信準備の整っているデータ項目を有する複数のデータプロバイダに、上記データプロバイダが低い各非待ち受けの優先順位からのデータ項目を送信するように指示されるまで、上記待ち受けの優先順位の任意のデータ項目を送信するように指示しないように適合化される請求項 1 4 記載のシステム。

20

【請求項 1 6】

上記受信手段は、上記複数のデータ項目を受信しあつ上記複数のデータ項目を記憶装置又はメモリに提供するように適合化されており、

上記送信機手段は、上記複数のデータ項目を上記記憶装置又はメモリから送信するように適合化される請求項 1 1 乃至 1 5 のうちのいずれか 1 つの請求項記載のシステム。

【請求項 1 7】

上記共有リソースは、リンク、メモリ、プロセッサ、集積回路又はクロスバーである請求項 1 1 乃至 1 6 のうちのいずれか 1 つの請求項記載のシステム。

30

【請求項 1 8】

上記送信機手段は、ある優先順位からの複数のデータ項目が上記リソースに送信された後にのみ、上記優先順位を待ち受けさせるように適合化される請求項 1 1 乃至 1 7 のうちのいずれか 1 つの請求項記載のシステム。

【請求項 1 9】

共有リソースへのアクセスを制御するためのシステムであって、上記システムは、

- 互いに異なる 3 つ以上の所定の優先順位のうちの 1 つの優先順位を既にそれぞれ割り当てられているデータ項目であって、送信されるべき少なくとも 1 つのデータ項目及び当該データ項目にそれぞれ割り当てられた少なくとも 1 つの優先順位のうちの少なくとも一方をデータプロバイダから受信するように適合化された受信手段と、

- 各データ項目が、当該データ項目に関係する優先順位の 1 つの送信プロセスを割り当てられている、優先順位毎の少なくとも 1 つの送信プロセスと、

- 複数のデータ項目を上記リソースへ、

・ 第 1 の時点において、最も高い優先順位の非待ち受けの送信プロセスを割り当てられた送信されるべきデータ項目を上記共有リソースに送信し又は当該送信されるべきデータ項目の識別を上記データプロバイダに送信することと、

40

・ 引き続き、最も高い非待ち受けの優先順位の上記非待ち受けの送信プロセスに、そのうちの少なくとも 1 つの送信プロセスが上記第 1 の時点において非待ち受けでありかつ空でない低い各優先順位からのデータ項目の送信を待ち受けさせることによって繰返し提供するように適合化された送信機手段とを備えた共有リソースへのアクセスを制御するためのシステム。

50

【請求項 20】

上記送信機手段は、上記送信プロセスに、上記時点において送信された上記データ項目又は当該データ項目の識別の送信の時点においてデータ項目又は当該データ項目に割り当てられた優先順位のみが受信されていた、データ項目及び当該データ項目の識別のうちの少なくとも一方の送信を待ち受けさせるように適合化される請求項 19 記載のシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、共有リソースへのアクセスの制御に関し、特に、共有リソースへのデータ転送に所定の保証を提供するために制御することに関する。

10

【背景技術】

【0002】

このタイプの技術は、特許文献 1 及び非特許文献 1 乃至 4 に記述されている。

【0003】

【特許文献 1】米国特許出願公開第 2004 / 001502 号明細書。

【非特許文献 1】Felicijan, T. et al., 「An asynchronous on-chip network router with quality-of-service (QoS) support」, SOC conference, 2004, proceedings. IEEE, International Santa Clara, CA, USA, Sept. 12-15, 2004, Piscataway, NJ, USA, IEEE, pp 274-77.

20

【非特許文献 2】Felicijan, T. et al., 「An asynchronous low latency arbiter for quality of service (qos) applications」, microelectronics, 2003, ICM 2003, Proceedings of the 15th international conference on Cairo, Egypt, Dec.9-11, 2003, Piscataway, NJ, USA, IEEE, pp 123-26.

【非特許文献 3】Bjerregaard, T. et al., 「Virtual channel designs for guaranteeing bandwidth in asynchronous network-on-chip」, Norchip conference, 2004, proceedings, Oslo, Norway, 8-9 Nov. 2004, Piscataway, NJ, USA, IEEE, pp 269-272.

【非特許文献 4】Zhang, H. et al., 「Rate-controlled static-priority queueing」, Networking: foundation for the future, San Francisco, March 28 April 1, 1993, Proceedings of the annual joint conference, on the computer and communications societies (INFOCOM), Los Alamitos, IEEE Comp. Soc. Press, US, vol. 2, conf. 12, pp 227-236.

30

【発明の開示】

【発明が解決しようとする課題】

【0004】

多くのデータ転送アプリケーションに見受けられる問題は、ある量のデータがリンク又はメモリ等の共有リソースへの転送を予定されていて、しかもこのデータの全てを同時に送信することができないことがある。従って、どのデータを最初に送信し、どのデータが待機しなければならないかについて決定を下さなければならない。当然ながら、この決定は、待ち時間等の共有リソースへ向けたデータの転送に影響を与え、ひいては、リソースが関与しているアプリケーションのパフォーマンスに影響を与える。本発明は、共有リソースに向けられるデータを、待ち時間及び帯域幅に関する保証が得られるようにスケジューリングする新規な方法を提供する。

40

【課題を解決するための手段】

【0005】

第 1 の態様において、本発明は共有リソースへのアクセスを制御する方法に関し、

- 複数の異なる所定の優先順位のうちの 1 つを既にそれぞれ割り付けられている、送信されるべき少なくとも 1 つのデータ項目に関係する情報を受信することと、

- 複数の項目を上記リソースに、

・最も高い非待ち受けの優先順位を有する 1 つの送信されるべきデータ項目を上記リソースに送信することと、

50

・その優先順位が、引き続き、送信されるべきデータ項目に関する情報が既に受信されている非待ち受けの低い各優先順位からの1つのデータ項目の送信を待ち受けることによって、繰り返し送信することを含む。

【0006】

一般に、データ項目は、イーサネットのパケット、フリット（フロー制御ユニット）又はこれらの一一部等の任意タイプのデータ項目であってもよい。各データ項目は単独で処理されかつ優先順位を付けられてもよく、又はより多い量のデータが優先順位を付けられ、次に複数のデータ項目に分割されてもよい。

【0007】

データの優先順位は予め決定されてもよく、そのデータ項目又はそのデータ項目が引き出されたより多い量のデータのコンテンツから決定されてもよい。あるいは、優先順位は、そのソース（発信者）又はその受信者から決定されてもよい。

【0008】

本コンテキストにおいては、優先順位は任意の望ましい方法で表現されてもよい。本発明は通常、電気回路において実現されているので、通常は、整数等の優先順位を表す数が使用される。しなしながら、電気回路であってもこれは必要条件ではない。

【0009】

一般に、優先順位の順番は、任意の優先順位ペアからどちらが高く、どちらが低いかを決定できるようにして決定される。これは、優先順位の実際の表現方法とは無関係である。本コンテキストでは、優先順位は任意の方法で表現されてもよく、その順位（どれが他より高いか）は任意の方法で決定されてもよい。

【0010】

受信ステップは、データだけを受信してもよく（その後、その優先順位を決定する）、優先順位だけ、又はデータ及び割り当てられた優先順位の両方を受信してもよい。以下、これらの情報を受信する方法の異なる様態についてさらに説明する。

【0011】

現在、共有リソースは、データを受信するように適合化された、メモリ、リンク、プロセッサ及び／又はクロスバー等の任意タイプのリソースであってもよい。

【0012】

本発明は、データ項目の共有媒体への送信を制御する方法に関し、個々のデータ項目に焦点を合わせるのではなく、優先順位に基づいてデータ項目の流れを制御する。

【0013】

優先順位は、待ち受けであってもなくてもよい。待ち受けの優先順位とは、その優先順位を有するデータはどれも、今は共有リソースに送信され得ないことを意味する。その優先順位を有するデータ項目に共有リソースへの送信を望むものがないとしても、その優先順位は待ち受けであると言える（即ち、このようなデータ項目の送信準備が整っていることを示す情報が受信されている）。

【0014】

好ましくは、特定の優先順位だけ待ち受けしなければならない非待ち受けの低い優先順位は、関係する優先順位を有するデータ項目の送信の時点で非待ち受けである優先順位である。

【0015】

任意の時点で、次にリソースに送信されるべきデータ項目は、下記のようなデータ項目である。

- 送信の準備が整っており（即ち、それに関する情報が既に受信されており）、
- 待ち受けでない優先順位を有しており、かつ
- データ項目の準備が整っている非待ち受けの優先順位のうちで最も高い優先順位を有するもの。

【0016】

そのデータ項目が一旦送信されると、その優先順位は、待ち受けでなくかつそのための

10

20

30

40

50

データの送信の準備が整っている低い優先順位が全てデータ項目を共有リソースに送信する機会を得るまで、待ち受けしている。

【0017】

ある優先順位が待ち受けである間に、この待ち受けの優先順位より低い他の優先順位に関するデータ項目の送信の準備が整うことになる点は注目すべきである。待ち受けの優先順位は、このような優先順位からの送信をも待機する必要はない。

【0018】

現在の優先順位が再び送信するまでに、どの低い優先順位が送信しなければならないかに関して、優先順位毎に記録をとることが所望されてもよい。

【0019】

最も高い優先順位は、全ての低い（他の全ての）非待ち受けの優先順位がデータ項目を送信する度に、好ましくは少なくとも一度送信を許可されるだけであるとしても、待ち時間及び取得可能な帯域幅の両方に関して、全ての優先順位に所定の保証が得られる点は注目される。さらに、データ項目の共有リソースへの送信は、実際にどの優先順位がデータ項目を送信するかに関わらず、最適（リンクの全能力等）であり得る点も注目される。これは、ある優先順位が待ち受けであり得るのは、他の優先順位を割り当てられたデータ項目の送信の準備が整っている場合に限られることから分かる。1つの優先順位のみが、送信されるべき複数のデータ項目を有するときは、この優先順位は待ち受けにはならず、そのデータ項目は1つずつ共有リソースに送信されてもよい。

【0020】

複数のデータ項目を提供する1つの方法は、
上記複数の優先順位の1つにそれぞれ関係する複数の待ち行列を提供するステップをさらに含み、

- 上記受信するステップは、1つの待ち行列内の各データ項目を受信することを含み、上記待ち行列は上記データ項目に割り当てられた優先順位に関係し、

- 上記送信するステップは、上記最も高い優先順位を有する1つの非待ち受けの待ち行列から上記データ項目を送信することを含み、

- 上記待ち受けるステップは、その待ち行列が、その後、低い優先順位を有する全ての空でない非待ち受けの待ち行列からの1つのデータ項目の送信を待ち受けることを含む方法である。

【0021】

従って、データ項目は受信され、かつ関係する優先順位に関する待ち行列に提供される。

【0022】

これに関しては、上記待ち受けるステップは、非待ち受けの低い各優先順位からのデータ項目がデータプロバイダから上記共有リソースに送信されたとき、上記待ち行列からの次のデータ項目をメモリへ転送することを含むかもしれません、上記送信するステップは、上記メモリ内の最も高い優先順位を有するそのデータ項目を送信することを含む。

【0023】

これは、待ち受けの状態と、どのデータ項目を送信するかの決定とがさらに分離される点で、プロセスを単純にする。この実施形態では、メモリは単に非待ち受けである空でない待ち行列毎にデータ項目を保持するだけである。従って、これらのデータ項目のうちのどれが送信されるべきかの決定が簡単である。

【0024】

実際に、どの優先順位が待ち受けであるかを可視化し、かつ送信するために適合化される次の優先順位の容易な決定を促進するためには、この同じメモリ構造が、実際のデータ項目のためではなく、データ項目に関連する情報のために使用されてもよい。

【0025】

メモリにデータ項目を持たない空でない待ち行列が待ち受けの待ち行列になるように、待ち受けの待ち行列は、データ項目をメモリに送信することを防止されてもよい。

10

20

30

40

50

【0026】

別の実施形態では、本方法は、複数のデータプロバイダを提供するステップをさらに含み、

上記複数のデータプロバイダはそれぞれ、

- 上記複数の優先順位のうちの1つをすでにそれぞれ割り当てられている複数のデータ項目を提供するように、かつ、

- 上記データプロバイダから上記共有リソースへの送信準備が整っている少なくとも1つのデータ項目に関する情報を提供するように適合化されており、

- 上記送信するステップは、1つのデータプロバイダが上記データ項目を上記リソースに送信することを含み、

- 上記待ち受けるステップは、低い各非待ち受けの優先順位からのデータ項目がデータプロバイダから上記共有リソースに送信されるまで、上記待ち受けの優先順位を有する送信準備の整ったデータ項目の任意のデータプロバイダからの送信を防止することを含む。

【0027】

この状況において、本方法又は本方法を実行するシステムはデータを保有又は受信しなくてもよいが、データ項目をいつ送信するか（及び、好ましくはプロバイダが幾つかのデータ項目のうちのどれを送信するかを決定するためのその優先順位の識別）を（データ項目の送信方法を知らせる）データプロバイダに指示するように適合化される。

【0028】

1つの方法では、待ち受けるステップは、防止信号を送信する等により、データプロバイダが待ち受けの優先順位を有するデータ項目を送信することを能動的に防止する。あるいは、プロバイダは、送信を指示されるまではデータ項目を送信しないように適合化される。これは、上記複数のデータプロバイダは、指示されるとそれに従って、データ項目を転送するように適合化されているときであってもよく、

- 上記送信するステップは、最も高い非待ち受けの優先順位の送信準備の整ったデータ項目を有する1つのデータプロバイダに上記データ項目を上記共有リソースに転送するように指示することを含み、

- 上記待ち受けるステップは、待ち受けの優先順位の送信準備の整ったデータ項目を有するデータプロバイダに、上記待ち受けの優先順位の任意のデータ項目を送信するように指示しないことを含む。

【0029】

一般に、受信するステップは、好ましくは、上記複数のデータ項目を受信しかつ上記複数のデータ項目を記憶装置又はメモリに提供することを含み、上記送信するステップは、上記データ項目を上記記憶装置又はメモリから送信することを含む。

【0030】

ある実施形態では、1つの所定の優先順位のための上記待ち受けるステップは、上記所定の優先順位が複数のデータ項目を送信したときのみ、上記（少なくとも1つの）データ項目の（少なくとも1回の）送信を待ち受けることを含む。この方法では、この所定の優先順位は、上記優先順位が待ち受けの段階になる前にデータ項目が毎回1つしか送信されない場合よりも、リソースへの大きい帯域幅を与えられる。

【0031】

この状況においては、非待ち受けの待ち行列についてリソースに送信されるべきデータ項目を保持するメモリはどれも、所定の優先順位のための複数のデータ項目のためのスペースを有し、他の優先順位のためにはそれより少ないスペースを有してもよい。

【0032】

他の態様において、本発明は共有リソースへのアクセスを制御する方法に關し、

- 複数の異なる所定の優先順位のうちの1つを既にそれぞれ割り付けられている、送信されるべき少なくとも1つのデータ項目に關係する情報を受信することと、

- データ項目に関する情報の各要素は上記關係する優先順位の1つの送信プロセスを割り当てられており、優先順位毎に少なくとも1つの送信プロセスを定義することと、

10

20

30

40

50

- 複数の項目を上記リソースへ、

・ 最も高い優先順位のために定義される非待ち受けの送信プロセスに、その関連情報が既に割り当てられている 1 つのデータ項目を上記リソースに送信することと、

・ 引き続き、その送信プロセスが、送信されるべきデータ項目に関する情報を既に割り当てられかつそのうちの少なくとも 1 つの送信プロセスが非待ち受けである低い各優先順位からのデータ項目の送信を待ち受けることによって繰返し送信することを含む。

【 0 0 3 3 】

この態様は、第 1 の態様に密接に関連するものであるが、この態様においては、少なくとも 1 つの優先順位に関して、第 1 の態様の送信の特徴に従う幾つかの送信機手段が定義されてもよい。この方法では、この優先順位は、他の優先順位に対して決定可能かつ現実的なサービス保証を提供しながら、より大きい帯域幅を得ることができる。 10

【 0 0 3 4 】

この場合もやはり、受信される情報をどの送信プロセスに割り当てるか、及びデータ項目の送信にどの送信プロセスが選択されるかを制御することによって、多種多様な振る舞いを得ることができる。また、受信される情報を送信プロセスに割り当てるタイミングは、システムの振る舞いを決定する。好ましくは、情報は、最も早い時点で送信される関係するデータ項目を有することになる送信プロセスに割り当てる。

【 0 0 3 5 】

ある実施形態では、この送信プロセスは、引き続き、実際に、送信されるべきデータ項目に関する情報が割り当てられている、低い優先順位の各非待ち受けの送信プロセスからのデータ項目の送信を待ち受ける。 20

【 0 0 3 6 】

一般に、上記第 1 及び第 2 の態様においては、上記優先順位 / 待ち行列 / 送信プロセスは、好ましくは、時間に間に合って送信された上記データ項目の送信時にはその情報が既に受信されているデータ項目のみの送信を待ち受ける。従って、これらの態様においては、どの待ち行列、送信プロセス又は優先順位が非待ち受けでありかつ空でないかの決定は、好ましくは、そのデータ項目の送信の時点で実行される。ある方法では、各優先順位にメモリが割り当てられ、このメモリには、この優先順位が再び非待ち受けになる前にデータ項目を送信すべき、空でない待ち受けの低い優先順位の情報が与えられてもよい。次に、このメモリは、データ項目の送信の度に更新されてもよい。 30

【 0 0 3 7 】

本発明の別の態様は、共有リソースへのアクセスを制御するためのシステムに関し、

- 複数の異なる所定の優先順位のうちの 1 つを既にそれぞれ割り付けられている、送信されるべき少なくとも 1 つのデータ項目に関係する情報を受信するように適合化された受信手段と、

- 複数のデータ項目を上記リソースに、

・ 1 つのデータ項目を送信させ、かつ最も高い非待ち受けの優先順位を上記リソースに送信させることと、

・ 引き続き、その優先順位からの 1 つのデータ項目の送信を、送信されるべきデータ項目に関する情報が既に受信されている非待ち受けの低い各優先順位によって 1 つのデータ項目が送信されるまで待機されることによって、繰返し提供せんように適合化された送信機手段とを備える。 40

【 0 0 3 8 】

上述したように、上記受信手段は実際のデータ項目を受信しつつ関係する優先順位をそれ自身が決定してもよく、データ項目及び優先順位の両方を受信してもよく、又は優先順位のみを受信してもよい。

【 0 0 3 9 】

ある総合的な実施形態では、上記複数の優先順位のうちの 1 つにそれぞれ関連する複数の待ち行列をさらに備え、

- 上記受信手段は、1 つの待ち行列内に各データ項目を提供するように適合化されてお 50

り、上記待ち行列は上記データ項目に割り当てられた優先順位に関係し、

- 上記送信機手段は、最も高い非待ち受けの優先順位を有する上記データ項目を上記最も高い優先順位を有する任意の非待ち受けの待ち行列のデータ項目から送信し、かつ、引き続き、その行列に低い優先順位に関連する全ての非待ち受けの待ち行列からのデータ項目の送信を待ち受けさせるように適合化される。

【0040】

次に、上記送信機手段は、非待ち受けの低い各優先順位からのデータ項目が1つの待ち行列から上記共有リソースに送信されたとき、次のデータ項目を上記待ち行列からメモリへ転送し、かつメモリ内の上記最も高い優先順位を有するそのデータ項目を上記リソースに送信するように適合化されてもよい。

10

【0041】

上述したように、このタイプのメモリは、データ項目を送る次の優先順位の迅速な決定を促進するために、非待ち受けの優先順位のための情報のみがメモリにおいて受信されるような方法で受信されるその情報のために、追加的に又は代わりに使用されてもよい。

【0042】

この同じ、又は他の実施形態において、複数のデータプロバイダをさらに備え、上記複数のデータプロバイダはそれぞれ、

- 上記複数の優先順位の1つをそれぞれ割り当てられている複数のデータ項目を提供するように、かつ、

- 上記データプロバイダから上記共有リソースへの送信準備が整っている少なくとも1つのデータ項目に関する上記情報を提供するように適合化されてもよく、

20

上記送信機手段は、1つのデータプロバイダに、上記最も高い非待ち受けの優先順位を有する上記データ項目を上記リソースに送信するように指示しつつ、引き続き、低い各非待ち受けの優先順位からのデータ項目がデータプロバイダから上記共有リソースに送信されるまで、任意のデータプロバイダからの上記待ち受け優先順位を有する送信準備の整っているデータ項目の送信を防止するように指示するように適合化される。

【0043】

この点に関して、上記複数のデータプロバイダは、指示されるとそれに従って、データ項目を転送するように適合化されてもよく、上記送信機手段は、上記最も高い非待ち受けの優先順位の送信準備の整っているデータ項目を有する1つのデータプロバイダに上記データ項目を上記共有リソースに転送するように指示し、かつ、引き続き、上記待ち受けの優先順位の送信準備の整っているデータ項目を有する複数のデータプロバイダに、上記データプロバイダが低い各非待ち受けの優先順位からのデータ項目を送信するように指示されるまで、上記待ち受けの優先順位の任意のデータ項目を送信するように指示しないように適合化される。

30

【0044】

一般に、上記受信手段は、好ましくは、上記複数のデータ項目を受信しかつ上記複数のデータ項目を記憶装置又はメモリに提供するように適合化されており、上記送信機手段は、上記複数のデータ項目を上記記憶装置又はメモリから送信するように適合化される。

【0045】

本コンテキストにおいては、上記共有リソースは、リンク、メモリ、プロセッサ、集積回路又はクロスバーであってもよい。

40

【0046】

上述したように、上記送信機手段が、ある優先順位からの複数のデータ項目が上記リソースに送信された後にのみ、上記優先順位を待ち受けさせるように適合化されるときに、より広い帯域が所定の優先順位に提供されてもよい。

【0047】

第4の態様においては、本発明は、共有リソースへのアクセスを制御するためのシステムであって、上記方法は、

- 複数の異なる所定の優先順位のうちの1つを既にそれぞれ割り付けられている、送信

50

されるべき少なくとも 1 つのデータ項目に関係する情報を受信するように適合化された受信手段と、

- データ項目に関する情報の各要素は上記関係する優先順位の 1 つの送信プロセスを割り当てられている、優先順位毎の少なくとも 1 つの送信プロセスと、

- 複数のデータ項目を上記リソースへ、

・最も高い優先順位のために定義される非待ち受けの送信プロセスに、その関連情報が既に割り当てられている 1 つのデータ項目を上記リソースに送信させることと、

・引き続き、その送信プロセスに、送信されるべきデータ項目に関する情報を既に割り当てられかつそのうちの少なくとも 1 つの送信プロセスが非待ち受けである低い各優先順位からのデータ項目の送信を待ち受けさせることによって繰返し提供させるように適合化された送信機手段とを備える。

【 0 0 4 8 】

上述したように、これは、データ項目のフローを制御しつつ優先順位に決定可能かつ制御可能な送信保証を与える追加的な方法を提供する。

【 0 0 4 9 】

ある実施形態では、上記送信機手段は、引き続き、この送信プロセスに、送信されるべきデータ項目に関する情報を既に割り当てられている、低い優先順位の各非待ち受けの送信プロセスからのデータ項目の送信を待ち受けさせるように適合化される。

【 0 0 5 0 】

また、上記第 3 及び第 4 の態様において、上記送信機手段は、好ましくは、上記優先順位 / 待ち行列 / 送信プロセスに、時間に間に合って送信された上記データ項目の送信時にはその情報が既に受信されているデータ項目のみの送信を待ち受けさせるように適合化される。その決定の時点が確定されるときは、これにより上記プロセスの制御はより容易になる。

【発明の効果】

【 0 0 5 1 】

本発明は、共有リソースに向けられるデータを、待ち時間及び帯域幅に関する保証が得られるようにスケジューリングする新規な方法を提供する。

【発明を実施するための最良の形態】

【 0 0 5 2 】

以下、図面を参照して、本発明の好適な実施形態について説明する。

【 0 0 5 3 】

本実施形態は、本発明（以下、ALG（A s y n c h r o n o u s L a t e n c y G u a r a n t e e）- 非同期の待ち時間保証と呼ぶ。）を非同期のオンチップネットワーク（ネットワーク・オン・チップ（Network-on-Chip）、NoC）におけるリンクスケジューラとして使用し得る方法を示す。ALGにより提供される複数の保証されるサービス（GS（Guaranteed Service））は互いに逆依存しておらず、従って、ALGは、時分割多重（TDM）に基づく帯域幅（BW）割当ての方法の限界を克服し、かつ異なる複数のGSの必要条件によって特徴づけられる広範なトラフィックタイプをサポートする。GSのスペクトル範囲の反対側の両端において、ALGは臨界待ち時間及び割込み等の低いBWトラフィックの両方をサポートし、また、厳しい待ち時間の要請はないがBWに関してGSを要求するストリーミングデータもサポートする。さらに、ALGは、完全な非同期環境において動作する。以下、ローエリア 0 . 1 2 μ m のCMOS 規格のセル実装によってこれを実証する。

【 0 0 5 4 】

本明細書の以後の編成は、次の通りである。セクション 1 では、GS の背景を考察し、現在の解決方法を述べ、NoC における最適な解決方法の必要条件を定義する。セクション 2 では、ALG スケジューリングの概念を説明してその機能性の証拠を呈示し、セクション 4 では、セクション 3 で述べた証拠を拡大して検討してバッファの限界についても説明する。セクション 4 は、この方法における帯域幅の割当てに関する。セクション 5 では

10

20

30

40

50

、オンチップALGリンクの実装について述べ、セクション6では、シミュレーションの結果を提供する。

【0055】

1. 保証されるサービス

以下、複数のネットワークパフォーマンスパラメータについて検討し、分類を確立する。次に、接続指向のルーティングの必要性について論じ、現在のNoC及びマクロネットワークにおいて使用されるGSの技術について検討し、最後にNoCにおけるGSの一連の必要条件を提案する。

【0056】

1.1. パフォーマンスパラメータ

サービス保証は、少なくとも1つのパフォーマンスパラメータによって定量化される。サービスの限界を適切に指定するためには、BW及び待ち時間の両方が示されなければならない。維持され得るスループットが小さすぎれば、待ち時間保証は役に立たず、同様に、負わされる待ち時間に限界がなければBW保証は役に立たない。

【0057】

複数フリットの1つのストリームのBWの限界はそのパスにおけるボトルネックによって決定されるが、ネットワークにおける1つのフリットの合計の待ち時間は、発生する待ち時間の総和によって特徴づけられる。これらは、必要とされるネットワーク接続がその間にアクセスされるネットワークのアドミッション待ち時間 t_{admit} 及び多くのホップ待ち時間を含む。ここで、1つのホップは、1つのルーティングノード内のバッファからリンクを通じて隣のルーティングノード内のバッファへ至るフリットの移動である。ホップ待ち時間は、フリットが例えばリンクである共有のルーティングリソースへのアクセスが許可されるのに要する時間であるアクセス待ち時間 t_{access} と、アクセスが一旦許可されているときに、フリットを次のルーティングノード内のバッファに送信するために要する時間である送信待ち時間 t_{link} とから成る。従って、Xホップの長さであるパスをトラバースして移動する1つのフリットの合計の待ち時間は、 $t_{total} = t_{admit} + t_{access} + t_{link} + \dots + t_{access}X + t_{link}X$ となる。

【0058】

1.2. 接続指向のGS

確かなサービス保証を提供するためには、接続指向のルーティングが絶対に不可欠である。接続なしのルーティングでは、全てのデータが同じ論理ネットワーク上を移動し、どの送信も別の送信をストールする可能性を有する。GSトラフィックは、ネットワーク内の他のトラフィックとは論理的に独立していかなければならない。サービス保証に関する確かな限界は、モジュールの設計フローを促進することにおいて、システムレベルの視点から有益である。このような保証がなければ、システムの変更は、大規模なトップレベルの再検証を必要とすることがある。従って、NoCにおけるGSは、大規模SoC設計のターンアラウンド・タイムを短縮する可能性を含んでいる。また、BETルーティングネットワークのパフォーマンスの形式的な検証は可能でない場合が多いが、臨界的なリアルタイムシステムでは望ましいように、GSはこれを可能にする。

【0059】

1.3. GSの技術

BW保証を提供する基本的な解決方法は、フェア・フリュイド・キューイング (FFQ) (fair fluid queuing) に基づく。FFQは、接続毎に別々の複数の待ち行列を実現するヘッド・オブ・ライン・プロセッサ・シェアリング (HOL-PS) (head-of-line processor sharing) の一般的な形式である。待ち行列の先頭は、フェアシェアなアクセスを例えばリンクである共有媒体へ提供するような方法でサービス処理がされる。

【0060】

非同期のNoCにおいては、FFQ型のアクセス方法には、極めて長い最悪のケースの

10

20

30

40

50

待ち時間という不快な欠点がある。ツリーアービタ（ツリーの調停装置）は、FFQを近似することができるが、複数の入力の互いに対するタイミングに関しては何も言えないの10で、所定のチャンネルに到達するパケットはどれも、サービス処理がされるまでに他の全ての入力を待機せざるを得ないという可能性がある。従って、このリンクアクセス方法を実現する非同期NOCにおいて、複数のリンクのシーケンスを通じて累積される最悪のケースの待ち時間は極めて長い。また、アクセス時間はBWの予約に反比例する。短い待ち時間を達成するためには、BWの大部分が予約されなければならない。全体的に同期のNOCでは、ネットワークを通じた待ち時間は1ホップ当たり1クロックサイクルに保証され得るが、ソースにおける接続へアクセスする待ち時間は依然として、予約されるBWに反比例する。また、複数の接続は独立してはバッファされないので、このように短いホップ当たりの待ち時間を実現するためには、明示的なエンド・ツー・エンドのフロー制御が要求される。BW保証から分離された、待ち時間のよりよい限界を提供するためには、異なる方法が必要である。

【0061】

広域ネットワークでは、複数のネットワークノード間においてクロックレベルでの同期を実現することは明らかに不可能であるので、マクロネットワークは全体的には非同期の性質を有する。これにより、マクロネットワークは幾分、非同期NOCに類似する。GSに対するFFQ型の解決方法を実現するに当たっては、非同期NOCに関して上述したような待ち時間に関する問題点は周知の欠点である。これらの欠点を克服するためには、レートで制御される静的優先順位（RCSP（rate controlled static priority））スケジューリングが使用されることが多い。RCSPでは、アドミッションコントローラは、着信する全てのパケットに適格な送信時刻を割り当てる。この時刻になると、パケットは静的優先順位付き待ち行列（SPQ）に入れられる。この方法は、BWの保証だけでなく、待ち時間の保証も、互いに独立して提供することができる。しかしながら、アドミッション制御は、ノードがローカルな時間の概念を有することを必要とする。よって、アドミッション制御は非同期NOCにおける実現には不適となる。本方法の別の欠点は、それが仕事量非保存型（nonwork-conserving）であることにあり、これは、チャンネル待ち行列内にその適格な送信時刻を待っているパケットが存在していても、ルータはアイドル状態であってもよいことを意味する。これは、利用可能なネットワーキングリソースを使用する効率を低下させ、待ち時間の限界が尊重されるとても、平均の接続待ち時間及びリンク利用度は低減する。

【0062】

1.4. NOCにおけるGSのための必要条件

NOCにおけるGSのための解決方法に関する出願人による必要条件の案は、本解決方法が（i）高い動作速度及び低いハードウェアオーバーヘッドを容易にするために単純であり、（ii）ネットワーキングリソースを効率的に利用するために仕事量保存型（work conserving）であり、（iii）待ち時間及びBWに関して分離される、又は少なくとも互いに逆依存しない限界を提供できる、というものである。さらに、非同期NOCにおけるGSのための解決方法に関する必要条件として、（iv）本解決方法は、ローカルであれグローバルであれ、時間の概念を必要としないということがある。ALGは、これら全ての必要条件に適合し、よって同期システム及び非同期システムの両方においてGSを提供する有効な解決方法である。

【0063】

セクション2では、ALGスケジューリングの規律について説明し、共有リンク上の待ち時間及びBW保証を提供する際のその使用を実証する。但し、ALGベースのアクセスは任意の共有媒体に適用できる点に注意する。また、本明細書における実装は非同期の回路に基づいているが、ALGはこのようなものに限定されない。しかしながら、ALGは、時間の概念を必要としないという事実により、特に、非同期システムに適している。

【0064】

2. 基本的なALGスケジューリング

10

20

30

40

50

本セクションでは、基本的な A L G スケジューリングの規律について説明する。まずは、その動作の直観的な理解について述べ、続いて、その動作を正式に証明する。以下の説明における時間表示は全て、時間単位であるフリット時間 (f l i t - t i m e) を使用して量子化される。1 フリット時間は、物理リンク上で 1 つのハンドシェークサイクルを完了するのに要する時間として定義される。V C 制御手段は、リンク上でストールするフリットがないことを保証し、よって、このようなハンドシェークの時間は十分に定義される。複数の回路は非同期であるため、当然ながら、フリット時間はネットワーク全体を通じて一定ではない。しかしながら、ここでは、フリット時間はかなり均一であることを仮定している。

【 0 0 6 5 】

10

図 3 は、完全な A L G リンクを示す。A L G アドミッション制御及び静的優先順位付き待ち行列 (S P Q) は、A L G スケジューラを実現する。V C 制御は、これらを包み込む。これらの 3 つのサブシステムがどのように同時に動作して、1 つの物理リンクを共有する複数の V C に渡る待ち時間及び B W 保証を提供するかは、以下の説明で明らかとなるであろう。A L G スケジューリングの原理は、完全に最もよく理解される。S P Q は、複数の V C に優先順位を付け、それ相応に待ち時間保証を提供するが、それは所定の複数の条件下に限定される。アドミッション制御は、これらの条件が満たされることを確実にする。V C 制御機構は、受信側に空のバッファスペースが存在する場合にのみフリットが共有リンク上で送信されることを保証し、こうして、フリットがリンク上でストールしかつ待ち時間及び B W 保証を無効にすることを防止する。

20

【 0 0 6 6 】

2 . 1 . 優先順位付けされたチャンネル

待ち時間保証を提供するためには、リンクアクセス時間に関する限界を提供する必要がある。図 2 を見て、フリットがチャンネル A から D へ無作為に、但し大きい間隔で到着することを想定されたい。次に、これらのチャンネルは優先順位によってサービス処理されており、A が最も高い優先順位を有すると考える。A に到着するフリットは常に、ただちにサービス処理がされ、よって、最大のリンクアクセス時間は 1 フリット時間、即ち、進行中である可能性のある送信を完了するために要すると思われる時間、であることが保証される。この時点では、A に到着する複数のフリット間に大きい間隔が存在するという単純な仮定を行っているので、B に到着するフリットは、サービス処理がされるまでに A を 1 回だけ待機することになる。従って、B 上のフリットは、最大で、進行中の送信の完了及び A 上の送信を待つので、最大 2 フリット時間だけ遅延される。同様に、C は最大で 3 フリット時間だけ待つ、等々となる。結果的に、最大のリンクアクセス時間はチャンネルの優先順位に比例する。これが、図 3 において S P Q により実現されている機能性である。

30

【 0 0 6 7 】

2 . 2 . アドミッション制御

上述の規律は、大きいフリット間隔を必要とする。これを保証することは、特に分散されたルーティング制御を有する非同期ネットワークにおいては必ずしも可能ではない。仮にソースにおいて特定のフリット間隔が提供されるとしても、ネットワークはデータストリームにジッタをもたらし、その結果、間隔への必要条件がネットワーク内部のどこかで無効にされることがある。これにより、S P Q へのアドミッションを調整するアドミッション制御ステージが必要になる。図 3 では、A L G アドミッション制御は S P Q の前段の複数のボックスとして示されている。これは、アドミッション制御ステージ及び S P Q も実装する点で、マクロネットワークに使用される R C S P に幾分、類似している。しかしながら、R C S P の場合、アドミッションはチャンネルのローカルタイミングに基づいている（優先順位毎に適格な送信時刻がスケジュールされる）。これは、時間の概念を全く持たない完全な非同期システムでは不可能である。

40

【 0 0 6 8 】

S P Q の待ち時間の限界が無効にされないように A L G アドミッション制御により実現

50

されるべき条件は、所定の（高い優先順位の）VC上のフリットが別の（低い優先順位の）VCにおけるフリットを1回だけストールできるというものである。これは、1つのフリットが所定のVC上へのアクセスを競っているときに、SPQ内で待機している複数のフリットに注目することによって達成され得る。低い優先順位の複数のVC上にあるフリットの待ち時間保証を無効にしないために、所定のVC上にある先行するフリットがSPQ内で順位を与えられている間に待機し続けている全てのフリットは、新たなフリットがアドミッションを得る前にサービス処理がされなければならない。これは、フリットがリンク上で送信されているときにSPQの占有をサンプリングすることによって保証される。同じVC上の新たなフリットは、その時点でSPQ内で待機している全てのフリットがサービス処理がされると、アドミッションを得ることができる。従って、低い優先順位のVC上にあるフリットは、高い優先順位の各VC上のフリットにより最大で1フリット時間だけストールされる。所定のフリットがリンクへのアクセスが許可される場合、SPQ機能の定義により、高い優先順位のVC上にあるフリットは全て既にサービス処理がされていることになり、存在するフリットは低い優先順位のVC上で待機しているものだけになる点に留意されたい。

【0069】

図4は、ALGを例示したものであり、B及びCの待ち行列の待ち時間保証が如何にして満たされるかが分かる。Aの待ち行列には、多すぎるフリットが到着し、よって、アドミッション制御により抑制が行われている。Aのフリットは先行するリンク上へ極めて迅速に（保証されている待ち時間の限界より高速で）送信されているので、Aにおけるバーストの理由は、ネットワークにおけるより早いポイントにおいて見出すことができるものと思われる。

【0070】

2.3. ALGの待ち時間及び帯域幅保証

このセクションでは、基本的なALG接続により提供される待ち時間及びBW保証について述べる。結果は、セクション3.4において正式に演繹されるであろう。ALGのサービス保証は、接続のために予約される各VCの優先順位のレベル及び各リンク上のVCの合計数によって特徴づけられる。優先順位のレベルQ1, Q2, ..., QXを有する複数のVCがALGリンク1, 2, ..., Xのシーケンス上に予約されている接続について考察する。各リンクは、複数N個のVCを実装している。これらのリンクは、リンクアクセス時間に関してQ1, Q2, ..., QXフリット時間の限界を提供している。これは、ソースにおいてフリット間隔、即ち $t_{interval} = N + Q_{max} - 1$ フリット時間が尊重されるという条件下のことである。ここで、 Q_{max} はVCシーケンス上の最大のQ値である。これは、セクション4.4において導出される、いわゆる間隔条件である。間隔条件は、接続のアクセスレート保証でもあり、よって、接続のBW保証を全リンク容量の一部： $BW_{min} = BW_{min}[Q_{max}] = BW_{link} / (N + Q_{max} - 1)$ として特徴づける。

【0071】

重要な点は、ALGにより提供される待ち時間保証はBW保証から分離されるという事実に注目することである。リンク上のVCの数であるNが増加すると、BW保証は任意に少なくなる可能性があるが、リンクアクセス時間は依然として1フリット時間への低下に維持される。従って、例えば割込みである低いBWニーズを有する待ち時間臨界接続は、BWの過剰割当ての必要なしにサポートされる。

【0072】

TDM型のBW割当てに基づくNOCの既存の（同期）GS規律は、1ホップ当たり1フリット時間の待ち時間を実現するが、最初の接続アクセス待ち時間は、依然として合計のエンド・ツー・エンドの待ち時間をBW保証に対して反比例させる。ALGは、間隔条件が満たされる限り、GS接続への即時のアクセスを提供する。また、非同期ネットワークにおける1ステージ当たりの転送待ち時間はごく短くされてもよく、同様の同期回路のクロックサイクルより遙かに短くされてもよい点に留意することもできる。このように、

10

20

30

40

50

ALGは待ち時間に関して限界を保証するが、同期NOCも潜在的に遙かに低い最小待ち時間を有する。同期回路ではなく非同期回路を使用するNOCを実装する主たる優位点は、このことにある。

【0073】

2.4. 証明

以下、ソースにおいて間隔条件が尊重されれば、エンド・ツー・エンドの接続待ち時間に関する限界を定め得ることを証明する。アドミッション制御はフリットを保持するかもしれないが、そのフリットがそのグローバルなスケジュールより進んでいる場合に限られ、ローカルに観察されるフリット間隔は短縮されることになる。証明は、2つの部分から成る。第1部では、ALG規律が単一のリンクに作用することを証明する。はじめに、接続上に送信される第1のフリットが待ち時間の必要条件に適合する、又はデッドラインに間に合うことを示す。次に、デッドラインに間に合ったフリットに続く、間隔条件を固守する任意のフリットもまたそのデッドラインに間に合うことを示し、これから間隔の値に到達する。間隔条件を固守する全てのフリットは、帰納法によりそのデッドラインに間に合う。証明の第2部では、ALGリンクに関して、ソースにおいて間隔条件が尊重されれば、フリットは各リンクでそのデッドラインに間に合うことを示す。従って、接続のエンド・ツー・エンド待ち時間は、ネットワーク内部で間隔条件が無効にされることに関わらず、各リンク上の待ち時間保証の総和によって限界を付けられる。

10

【0074】

単一リンクの定理：複数N個のVCを実装するALGリンク上では、VC Q上の全てのフリットは、フリット間隔条件 $t_{interval} \leq N + Q - 1$ フリット時間の下で、最大のリンクアクセス時間であるQフリット時間を保証される。

20

【0075】

証明：それぞれがSPQにおける優先順位のレベル1, 2, 3, ..., Nに対応する複数N個のVCを実装する所定の1つのリンクをとり挙げる。所定のVC Q {1, ..., N}上に到着する第1のフリットは、SPQへのアクセスがただちに許可される。SPQにおいて、これは、リンクへのアクセスが許可されるまでに最大でQフリット時間だけ待機する。従ってこれは、Qフリット時間の最大のリンクアクセス時間により限界を付けられるそのデッドラインに間に合う。

30

【0076】

次に、Q上の、SPQへのアクセスがただちに許可され、よってそのデッドラインに間に合うフリットAと、フリットAに続く、同じくQ上のフリットBについて考察されたい。フリットBは、フリットAから $t_{interval}$ フリット時間だけ後に到着する。フリットAは、SPQへアクセスするために0フリット時間だけ待機し、かつSPQ内で最大でQフリット時間だけ待機する。フリットAがリンクへのアクセスが許可されたとき、SPQ内では、Qより低い優先順位にあるVCの数である最大でN-Q個のフリットが待機していた。ALG規律によれば、これらは全て、Q上の次のフリットがSPQへのアクセスが許可される前に送信されなければならない。最悪の場合のシナリオでは、Qより高い優先順位のVCの数である最大でQ-1個のフリットが、Qのアドミッション制御がフリットBによるSPQへのアクセスを許可する前に送信されなければならないN-Q個のフリットより高い優先順位をとりうる。これらの部分遅延の合計は、Q上の1つのフリットとSPQへのアドミッションが与えられる次のフリットとの間で経過し得る最大時間を示し、 $Q + (N - Q) + (Q - 1) = N + Q - 1$ になる。これは、 $t_{interval} \leq N + Q - 1$ フリット時間であれば、フリットBは確実にSPQへのただちにアクセスが許可されてSPQ内において最大でQフリット時間だけ待機し、これもまたそのデッドラインに間に合うことを意味する。従って、間隔条件下では、そのデッドラインに間に合ったフリットに続く任意のフリットは、それ自体もそのデッドラインに間に合い、かつ第1のフリットはそのデッドラインに間に合うので、帰納法により、全てのフリットはそのデッドラインに間に合う。

40

【0077】

50

次に、ALGリンクのシーケンスに関して、間隔条件がネットワークに導入されるジッタに起因して局地的に無効にされるとしても、ALGは、全てのフリットが各リンクでそのデッドラインに間に合うことを保証する点を示す。従って、エンド・ツー・エンド待ち時間限界は、各リンクにおける待ち時間限界の総和である。この時点では、依然として、ノードには常に十分なバッファスペースが存在することを仮定している。セクション5では、バッファの必要条件を計算してこの証明を補強する。

【0078】

リンクのシーケンスの系 (Corollary) : 常に十分なバッファスペースが存在するという仮定の下では、それぞれが複数N個のVCを実装する複数X個のALGリンクのシーケンス上に予約されたVCであるQ1, Q2, …, QXを有する所定の接続の場合、待ち時間の限界は、ソースにおいてフリット間隔 $t_{interval} = N + Q_{max} - 1$ フリット時間が尊重されるという条件下で、各リンクにおける待ち時間限界の総和である。ここで、 Q_{max} は {Q1, Q2, …, QX} の最大値である。

【0079】

証明：問題の接続上にあり、 $VC = Q = \{Q1, Q2, …, QX\}$ が予約されているリンク、及び上記リンク上でそのデッドラインに間に合った上記接続上のフリットAについて考察されたい。このフリットはそのデッドラインに間に合ったので、上述の単一リンクの定理の証明に従って、アドミッション制御は、同じVC上で、フリットAがSPQへのアクセスが許可された後、最大で $N + Q - 1$ フリット時間の後、進行中のフリットBのSPQへのアドミッションのために開く。フリットAは、そのデッドラインに間に合ったので、スケジュール通りであったか、スケジュールより進んでいた。フリットBは、スケジュールについてフリットAよりさらに先に進んでいれば、フリットAがSPQへのアクセスが許可された後に、 $N + Q - 1$ フリット時間より少ない間隔で到着し、アドミッション制御は、そのSPQへのアクセスをただちには認めない可能性もある。フリットAがアクセスを許可された後、最も遅い $N + Q - 1$ フリット時間で、フリットBは確実にアクセスが許可される。ソースにおいて、これらの分離間隔は $N + Q_{max} - 1$ フリット時間であった。この時点では、この分離間隔は同じであるか短くなり、よって、フリットBのスケジュールは少なくともフリットAと同程度に先に進むことになる。従って、フリットBもまたそのデッドラインに間に合う。反対に、ネットワーク内のより早いステージにおける輻輳に起因してフリットBのスケジュールがフリットAより遅れていれば、フリットBは、フリットAがSPQへのアクセスが許可されて以降、 $N + Q_{max} - 1$ フリット時間より後に到着する。従って、これはただちにアクセスが許可され、そのデッドラインに間に合う。

【0080】

接続上に送信される第1のフリットは、どのリンクのアドミッション制御においてもストールされないので、そのデッドラインに間に合う。そのデッドラインに間に合うフリットに続くフリットは何れも、それ自体もそのデッドラインに間に合うので、帰納法により、上記接続上の全てのフリットは全てのリンクにおいてそのデッドラインに間に合う。

【0081】

維持できる最小のBWは、これから生じる：

【0082】

最小帯域幅の系：それが合計帯域幅 BW_{link} を提供する複数X個のALGリンクのシーケンス上に留保された複数個のVCであるQ1, Q2, …, QXを有する所定の接続上では、維持される最小帯域幅は、 $BW_{min} = BW_{link} / (N + Q_{max} - 1)$ になる。ここで、 Q_{max} は {Q1, Q2, …, QX} の最大値である。

【0083】

証明：上記リンクのシーケンスの系に従って、1つのALG接続上にある、間隔条件 $t_{interval} = N + Q_{max} - 1$ フリット時間を固守する全てのフリットは、限界を付けられた待ち時間を有する。従って、複数のフリットのストリームは、リンクによりサポートされる合計フリットレートの少なくとも $1 / (N + Q_{max} - 1)$ のフリットレー

10

20

30

40

50

トで輻輳の発生なく送信されてもよい。これから直接、維持できる帯域幅は少なくとも、 $BW_{min} = BW_{Link} / (N + Q_{max} - 1)$ であることが得られる。

【0084】

3. バッファ

前のセクションでは、フリットが、ALGリンクのアクセススケジューリング規律のみによって制約されてネットワーク内を自由に流れると仮定していた。この作業は、フリットが絶対にドロップされない損失なしのネットワークを目標にしているので、各リンクは、受信側が自由なバッファ空間を有している場合にのみVC上で1つのフリットが送信され得ることを保証するバック・プレッシャ・フロー制御も実装しなければならない。これは、アドミッション制御の余分な層、即ち図3に示すVC制御を導入する。VC制御はALGアドミッション制御及びSPQを包み込み、受信側のVCバッファが空のスペースを有することを示している場合にのみフリットを通させる。フリットは、リンクの受信側へ自由に移動できる場合にのみ、ALGアドミッション制御へ呈示されなければならない。そうでなければ、ALG規律により提供される待ち時間保証は、リンク上でストールするフリットによって無効にされることがある。逆に、フリットはVC制御により、デッドラインに間に合わず、やはりALG待ち時間保証を無効にするほど過分に遅延されではならない。

【0085】

この作業では、共有ベースのVC制御を使用する。図6に示す方法は、VC当たり1本のワイヤを使用して、例えばリンクである共有媒体への非ブロッキングのアクセスを実現している。共有ボックスは、1つのフリットにアドミッションを与えた後にロックし、別のフリットが通ることを許可しない。上記1つのフリットは、向こう側の非共有ボックスへこの媒体を通って進む。非共有ボックスはラッチを実装し、フリットはラッチへと受け入れられる。フリットが非共有ボックスを出るときは、アンロック制御ワイヤがトグルする。これにより共有ボックスはアンロックされ、別のフリットに媒体へのアドミッションが与えられる。媒体がデッドロックしない限り、フリットは媒体内でストールしない。

【0086】

図5に示すように、接続を複数のALGリンクのシーケンスとしてモデル化し、入力ポートと上記接続のために予約されたVCバッファとの間に直接の回線が存在する。この仮定は、非ブロッキングのスイッチングを実装するルーターアーキテクチャに関して有効である。図中のVCバッファは、その入力に非共有ボックスを実装しつつ出力に共有ボックスを実装している。関連する待ち時間は、1つのフリットがリンクへのアクセスが許可されるのに要する時間であるリンクアクセス待ち時間 t_{access} と、リンクアクセスが許可されるとルータからリンクを介して次のVCバッファへ至る1つのフリットの待ち時間であるリンク転送待ち時間 t_{link} と、アンロック信号が前のVCバッファ内の共有ボックスへ戻るのに要する時間であって、別のフリットへリンクへのアクセスを許可してもよいことを示すアンロック待ち時間 t_{unlock} である。輻輳が発生しないので、リンクアクセス待ち時間以外の待ち時間は全て一定である。それぞれが複数N個のVCを実装する複数X個のALGリンクのシーケンスより成る接続の、エンド・ツー・エンドの待ち時間限界は、セクション3で述べた t_{total} 、即ち、 $t_{end2end} = t_{access} + t_{link} + t_{access} + t_{link} + \dots + t_{access} + t_{link}$ と同様である。単純化のために、本明細書では、全リンク上でNを同じであるものとしている。リンクアクセス時間は、各リンクにおいて予約されているVCの優先順位、 Q_1, Q_2, \dots, Q_X により決定され、 $t_{access} = Q_1$ フリット時間、 $t_{access} = Q_2$ フリット時間、等々となる。接続上の最大のQである Q_{max} は、パスのボトルネック、即ち $BW_{min} = BW_{Link} / (N + Q_{max} - 1)$ であるので、セクション4.3に従って接続のBW保証の必要条件である。

【0087】

ここで、フリットがSPQへアクセスできるほど機が熟したとき、即ち、フリットがそのスケジュールより0時間進んでいるときには、共有ボックスが常にアンロックされると

いう必要条件を決定する必要がある。そうであれば、フリットは A L G アドミッション制御へ提示され、A L G 規律に従って、そのデッドラインに間に合う。以下、フリット間隔条件及び $t_{link} + t_{unlock} < N - 1$ フリット時間というリンクサイクル条件下で A L G スケジューリング規律を正しく機能させるには、単一エレメントの V C バッファで足りることを証明する。

【0088】

単一バッファの定理：フリット間隔条件 $t_{interval} \leq N + Q_{max} - 1$ フリット時間及びリンクサイクル条件 $t_{link} + t_{unlock} < N - 1$ フリット時間の下では、リンクのシーケンスの系の有効性を保証するためには、各ノードにおける各 V C につき単一エレメントのフリットバッファで足りる。

10

【0089】

証明：図 5 に示すように、それぞれが複数 N 個の V C を実装する複数の A L G リンクのシーケンス上に予約された複数の V C (\dots, Q_i, Q_j, \dots) を有する接続の 1 つのセクションについて考察されたい。V C バッファである V C buf_i 及び V C buf_j はそれぞれ、1 フリット分のバッファスペースを有する。リセット時は、これらは空であるので、接続上で送信される第 1 のフリットは V C 制御によって制限されず、A L G 規律に従ってそのデッドラインに間に合う。ここで、そのデッドラインに間に合うように進行中のフリット A に続く、フリット B について考察されたい。フリット A は、そのデッドラインに間に合いつつあるので、遅くとも、それがそのスケジュールより 0 時間進んでいることに対応する時間 0 において S P Q_j へのアクセスを取得する。この時点で、V C buf_j は V C buf_i に別のフリットを受け入れる準備が整っていることを信号で伝える。従って、V C buf_i は、時間 t_{unlock} の後に、次のフリットであるフリット B のためにその出力を開放する。フリット A は、遅くとも時間 $0 - t_{link}$ には S P Q_i を出発しているので、adm1 は $0 - t_{link} + N - 1 = N - 1 - t_{link}$ フリット時間以内にはフリット B が S P Q_i へ入ることを許可する。この時間が、V C buf_i がフリット B を通過させた後であれば、V C buf_i はそのフローにとっての制限ファクタとならない。従って、V C 制御が制限ファクタにならないための必要条件は、 $N - 1 - t_{link} > t_{unlock}$ 且 $t_{link} + t_{unlock} < N - 1$ フリット時間となる。これは、リンクサイクル条件を構成する。このリンクサイクル条件が成立すれば、フリット B は、時間 : $Q_1 + t_{link} + N - 1 - t_{link} = Q_1 + N - 1$ フリット時間に V C buf_j に到着する。これは、必要とされるフリット間隔である $Q_{max} + N - 1$ フリット時間以下である。よって、フリット B はそのスケジュールに間に合って adm_j に到着した。

20

【0090】

ソースにおける、最小のフリット間隔が $Q_{max} + N - 1$ フリット時間という間隔条件の下、及び $t_{link} + t_{unlock} < N - 1$ フリット時間というリンクサイクル条件の下では、そのデッドラインに間に合ったフリットに続く任意のフリットもまたそのデッドラインに間に合う。第 1 のフリットがそのデッドラインに間に合うので、帰納法により、その接続上の全てのフリットはそのデッドラインに間に合う。

30

【0091】

4. 帯域幅の割当て

40

これまでに述べたように、A L G リンクは、その基本形式においてフレキシブルな待ち時間保証を提供する。V C は、達成されるべき接続の待ち時間の必要条件にその優先順位が対応するものが選択されてもよい。しかしながら、帯域幅保証は、選択される V C の優先順位に従って確定される。また、異なる優先順位の中でも、B W はさほど異ならず、最も高い優先順位の 1 / N から、最も低い優先順位の 1 / (2 N - 1) までの範囲である。以下、基本的な A L G 構成のフレキシブルな待ち時間保証をなお享受しつつ接続へのフレキシブルな B W 割当てを達成するための 3 つの方法について説明する。

【0092】

第 1 の方法（マルチチャンネル方法）は、単一の接続への各リンクに幾つかの V C が割り当てられる自明な方法である。これは事実上、実際には幾つかの「並行する」接続であ

50

って増大されたデータ容量を有する接続を生成する。その使用は、各リンクで局所的に扱われてもよく、ネットワークアダプタにおいてエンド・ツー・エンド・ベースで扱われてもよい。BWのこの割当て方法における優位点は、理解及び実装が容易であることがある。また、フリットはバーストで到着する（間隔はほとんどない）。最初のX個のフリットは、独立してバッファされる別々のVCシーケンス上で送信される。ここで、Xは各リンクにおいてその接続のために割り当てられるVCの数である。従って、これら全ての待ち時間は、高い優先順位のVCを選択することによって極めて短くされてもよい。パケット全体に極めて低い合計待ち時間の保証が与えられるので、これは、小さいパケットの主要な優位点である。OCPインターフェース等のバスタイプ（メモリがマップされた）のアクセスソケットを有するネットワークでは、単純なOCPコマンドを運ぶ大部分のパケットは、32ビットのネットワークでは、2つから3つのフリットより成る。このようなパケットにおけるフリットの数に対応する幾つかのVCシーケンス又は集合を割り当てるにより、これらのコマンドの転送待ち時間は、割り当てられるVC集合のうちで最も遅いVC集合のALG待ち時間にまで短縮される。この方法の欠点は、リンクのエリアの必要条件が多かれ少なかれBWの粒度に対して線形に増大し、上記エリアは、主としてフリットバッファによって決定されていて、各VCが1つのフリットバッファを必要とするので、BWの粒度はVCの数に対してほぼ非線形である、という事実にある。また、広範なSPQも必要とされる。これは、SPQのパフォーマンスを低下させ、本実施形態ではSPQはリンクのボトルネックであるので、これによりフリット時間が短縮する。

【0093】

10

第2及び第3の方法は、より多くのバッファスペースを実装することなく、基本的なALGアドミッション制御方法により、1つの優先順位が通常許可されるより多いアクセスを許可するという基本概念を軸に展開する。従って、1つのVCはSPQへのアクセスがより頻繁に許可されるので、このVCのスループットは増大する。しかしながら、複数のフリットが独立して始動することはできないために、接続上の複数フリットのパケットの合計待ち時間は第1の方法に比較して長い。複数のフリットはネットワークを介して同じバッファシーケンスを使用するので、1つのフリットは直前のフリットが接続の直前のノードにおけるバッファを去るのを待機しなければならないという事実に起因して、複数のフリットは接続上において、無視できない分離を有する。

【0094】

20

第2の方法（マルチアドミッション制御方法）によれば、各VCは幾つかのアドミッション制御を有する。これらのうちの任意の1つが非待ち受けであれば、フリットはアドミッションを与えられてもよい。次に、上記フリットにアドミッションを与えたアドミッション制御（より多くが非待ち受けであれば、これらのうちの任意の1つが上記フリットにアドミッションを与えた「その1つ」であるとして選ばれてもよい）は、それがアドミッションを与えたフリットの送信時に、基本的なALGスケジューリングの方法に関して上述した方法に従って、SPQの低い優先順位の占有をサンプリングする。各VCは、それが実装する（又はアクティブである）各アドミッション制御ステージにつき一度、低い優先順位のVCの送信を潜在的にブロックすることができるので、これらの待ち時間保証が適宜低減されることが分かる。従って、VCの待ち時間保証は、高い優先順位のVCの数ではなく、高い優先順位のVCのアクティブなアドミッション制御ステージの合計数に関連する。本方法は、基本的なALGリンクにおける2つ以上の異なる優先順位の使用へのアクセスを有しかつ單にアドミッションに対してオープンであれば何れをも選択するフリットバッファに対応する。しかしながら、複数の優先順位は絶対に同期的には使用されないので（フリットバッファが1つしか存在しないため）、これらは等しい待ち時間保証を経験し、2つの優先順位のうちの高い方に従って、絶対に互いにブロックし合わない。

【0095】

40

各リンクでは、フリットは、次のノードにおけるフリットバッファが空である場合にのみ送信されてもよい。即ち、フリットは、先行するフリットが先行するリンク上で送信された後でなければ送信され得ない。これは、分離条件を発生させる。フリットは、次のノ

50

ードではアドミッション制御によってブロックされない、即ち上記フリットは、そのスケジュールより先に進んでいないこと、即ち、 $t_{\text{分離}} = t_{\text{リンク}} + Q$ (先行するリンク) + $t_{\text{アンロック}}$ 、を仮定する。先に示した間隔条件とは異なり、分離条件はNとは独立である点に留意されたい。間隔条件は依然として各アドミッション制御ステージを求めるが、接続上のフリットのストリームに適用される分離条件とは独立している。従って、VCが2つのアドミッション制御ステージを実装すれば、即ち、 $t_{\text{間隔}} = 2 \times t_{\text{分離}}$ であれば、所定のVCについて2倍のBWが得られる。これは、図5のリンクサイクル条件に対応する。 $t_{\text{間隔}} < 2 \times t_{\text{分離}}$ であれば、フリットは、間隔条件よりも分離条件によってより厳しい制限を受けるので、2倍未満のBWが保証される。従って、達成可能な最大帯域幅は、リンクサイクル条件及びこの条件の倍数の間隔条件による「バランス」によって決定される。

【0096】

ここで言うQは、SPQの優先順位と同等のものではなく、高い優先順位のVCのアクティブなアドミッション制御ステージの合計数に1を加算して計算されたQであることに留意されたい。

【0097】

第3の方法(アドミッション制御計数法)によれば、アドミッション制御ステージによる占有(及び、低い優先順位のVCで待機中のものが存在すれば、後続の近接アクセスも)のサンプリングは、その優先順位でフリットが送信される毎にではなく、X回目毎に行われるだけである。実際には、フリットのバーストは、アドミッション制御が開放する(非待ち受けになる)度に送信されてもよい。バーストの最後のフリットの後、アドミッション制御は閉止し、低い優先順位のVCの占有をサンプリングし、送信を待機中のフリットにより占有されているものを待機する。従って、低い優先順位VCの各フリットは、潜在的に、この特定の(高い優先順位の)VCをX回待機しなければならない。従って、その待ち時間保証は低減される。マルチアドミッション制御方法の場合と同様に、VCが利用可能な帯域幅の増大は、低い優先順位のVCの待ち時間を犠牲にして達成される。

【0098】

5. 実装

図3によれば、基本的なALGリンクは、VC制御、アドミッション制御及びSPQという3つの基本的なサブシステムから成る。

【0099】

5.1. VC制御

採用されているVC制御方法の機能性については、セクション4で説明した。図7は、1つのVCの共有ボックス及び非共有ボックスの実装を示す略図である。1本のワイヤのアンロック信号は、2相の肯定応答として機能する。`pull_se_gen`により発生されるパルスは、Cエレメントの`c_lock`をリセットするに足る長さでなければならない。共有ボックスの出力における`output_protocol`回路は、VCから共有媒体を分離する。従って、個々のVCが遅いことに関わらず、フリットの自由な流れが保証される。

【0100】

5.2. アドミッション制御

ALGスケジューリングの新規性はアドミッション制御ステージに見いだされ、アドミッション制御ステージはSPQが適切な待ち時間限界を規定することを許容しながら、フリットの流れを制御する。

【0101】

アドミッション制御の各チャンネルは、低い優先順位のチャンネル毎に1ビットのステータスレジスタを実装する。所定のチャンネルの少なくとも1つのステータスピットが設定されると、アドミッション制御は、そのチャンネル上でのフリットのSPQへのアドミッションを停止する。そのチャンネル上のフリットがリンクへのアクセスが許可されると、SPQの占有のスナップショットに従ってステータスピットが設定される。占有は、所

10

20

30

40

50

定のチャンネルが（優先順位を付けられている）リンクへのアクセスが許可される間に、どのチャンネルが S P Q 内で待機中のフリットを有するかを指定する。ステータスピットは、これらの待機中のフリットがリンクへのアクセスが許可されるときに、引き続いてリセットされる。全てのフリットが送信されると、ステータスピットは全てクリアになり、アドミッション制御は上記所定のチャンネル上の別のフリットにアドミッションを与える。

【 0 1 0 2 】

図 8 は、チャンネル n のアドミッション制御を示す略図である。低い優先順位のチャンネルのそれぞれには、1つのステータスピットレジスタ [$n - 1 \dots 0$] が R S ラッチとして実装される。チャンネル n は、所定の時点でリンクへのアクセスを競う最も高い優先順位のチャンネルであると考える。S P Q は占有ベクトルを生成し、 n の肯定応答がハイである間は、その値は安定している。ステータスピットレジスタのセット入力は、 n の肯定応答及び占有ベクトルの論理和である。このようにして、適切なステータスピットは、上記チャンネルがリンクへのアクセスが許可される時点（その上昇する肯定応答により示される）の S P Q の占有に従って設定される。ステータスピットレジスタのリセット入力は、単に対応するチャンネルの肯定応答信号へ接続される。1つのチャンネルがリンクへのアクセスが許可されると、その肯定応答はハイになり、これにより、このチャンネルに対応するステータスピットは、高い優先順位のチャンネルのそれぞれのアドミッション制御においてリセットされる。全てのステータスピットがローであれば、入力要求は出力へ伝搬することを許容される。ローカルな肯定応答はステータスピットを設定させるので、要求パスにおいては、A N D ゲートではなく C エレメントが必要とされる。これは、入力要求がローにされるまで出力要求はローにされないことを保証する。

【 0 1 0 3 】

所定の時間においてリンクへのアクセスを得ることのできるチャンネルは1つのみであるので、ステータスピットレジスタのセット及びリセットは互いに排他的であることに留意されたい。

【 0 1 0 4 】

5 . 3 . 静的優先順位付き待ち行列

図 10 は、出願人による S P Q アービタ及び付随するマージデータパスの非同期の実装を示す。主要な制御信号は、「イネーブル」である。リセット時には、「イネーブル」はローである。「イネーブル」はアービタのアクティブフェーズを示すので、「イネーブル」がハイである間は、入力における非対称の C エレメントは、内部の要求信号がローにされないことを保証する。ミューテックス (m u t e x) は、いわゆるロックレジスタを実装する。少なくとも1つの入力要求がこれらの出力を介して伝搬すると、これは、「a n y _ r e q」を発生する O R ゲートによって検出され、イネーブルはハイレベルにされる。次に、これによりロックレジスタがロックされ、「イネーブル」がハイである限り、新たに到着する入力要求はロックされる。これ以降、出力の C エレメントのうちの正確に1つがファイアし、適切なチャンネルが肯定応答され、上記特定のチャンネルのデータが出力ラッチにおいて捕捉される。この時点で、「イネーブル」はローにされ、優先順位付けのフェーズの終わりが示される。「イネーブル」がローである限り、出力の C エレメントの何れもファイアすることができないが、肯定応答されたチャンネルがその要求を取り下げる、かつリンクが出力要求を認識すると、対応する C エレメントはリセットし、「イネーブル」に別の優先順位付けのフェーズを開始させる。

【 0 1 0 5 】

ミューテックスの「g 1」出力は「g 2」出力より甚だしくは遅くないものと仮定すると、高速 R T Z 相が実現する。R T Z に要求されることは、優先順位が付けられたチャンネルの「r 2」入力が「g 2」を介して出力 C エレメントへ伝搬することが全てである。ローにされる「g 2」が出力の C エレメントを介して伝搬して入力へ戻り、イネーブルを再度ファイアする前に、「g 1」がローにされる限り、この仮定は安全である。「g 1」及び「g 2」を含む両方の R T Z タイミングパスは、「イネーブル」がローになるのに伴

10

20

30

40

50

って始動する。S P Q のサイクルは、V C の数のあらゆるダブリングにおいて 3 つの 2 入力ゲート深さ (0 . 1 2 マイクロメートル技術で約 3 0 ~ 4 0 p s) 分だけ、即ち非対称の C エレメントにおける 1 つと、完了を検出する O R ゲートにおける 2 つ分が増大するだけである。

【 0 1 0 6 】

アドミッション制御によりそのステータスピットの設定に使用される占有ベクトルは、イネーブル及びロックされる入力要求の論理和を構成する。すぐ前のセクションで指摘したように、このベクトルは、肯定応答がハイである間は安定していることが要求される。ロックされる要求ベクトルをイネーブルと論理和演算することにより、占有ベクトルのビットは、肯定応答がハイであるときに安定しているかローであることが保証される。ステータスピットレジスタはハイである占有ビットによって設定されるので、これで十分である。10

【 0 1 0 7 】

6 . 結果

市販の 0 . 1 2 μ m の C M O S 規格セルを使用して、1 6 ビットの、8 個の V C の A L G リンクを実装した。典型的なタイミングパラメータを適用し、1 . 4 ナノ秒のフリット時間に対応する 7 0 2 M D I / 秒の測度で設計をシミュレートした。共有の物理リンクは、3 段パイプラインを実装している。リンク全体のセル面積、即ちプレレイアウトは 0 . 0 1 2 mm^2 であり、A L G スケジューラのコア（アドミッション制御及び S P Q ）は僅か 0 . 0 0 4 mm^2 にした。これは、A L G の恩恵が面積に関しては全く高価でないことを示している。20

【 0 1 0 8 】

3 つの A L G リンクのシーケンス上の接続をエミュレートする試験セットアップをシミュレートした。接続は、高い優先順位の V C を予約する高速の接続及び低い優先順位の V C を予約する低速の接続の 2 つを観察した。他の全ての V C 上でランダムなバックグラウンドトラフィックが誘導される間に、フリットの待ち時間を記録した。図 9 は、1 0 0 0 0 個を超えるフリットについて記録された異なるネットワーク負荷に対するフリット待ち時間の分布を示す。ネットワーク負荷が 1 0 0 % であっても、接続上のフリットがそのデッドラインに間に合っている様子が分かる。ネットワーク負荷の増大に伴って、待ち時間の分布図は待ち時間限界へと押し上がるが、絶対に限界を超えない。転送待ち時間限界は、3 . 6 ナノ秒 / ホップから 1 . 4 ナノ秒 (1 フリット時間) の増分で上方へと取得可能である。これは、A L G アクセス待ち時間と、リンクに渡る一定の転送待ち時間（共有ボックス、マージ、パイプライン、スプリット及び非共有ボックス - 約 2 . 2 ナノ秒）とを含む。B W 保証は、高速の接続上で $1 / 8 \times 7 0 2 \text{ M D I} / \text{秒} = 8 8 \text{ M D I} / \text{秒}$ あり、低速の接続上で $1 / 1 5 \times 7 0 2 \text{ M D I} / \text{秒} = 4 7 \text{ M D I} / \text{秒}$ であった。30

【 0 1 0 9 】

表 1 は、A L G の保証を、N o C に使用される既存のスケジューリング方法のそれと比較したものである。表において、N は、各リンク上の V C の数であり、h は所定の接続によりスパンされたホップの数である。T D M は同期 N o C において使用され、ある種のエンド・ツー・エンド・フロー制御が実装されることを考慮すれば、接続待ち時間の限界を N + h に規定している。この実装がなされなければ、待ち時間限界は、非同期フェアシェアのレベルまで、即ち (N + 1) × h まで低減される。本表は、A L G が待ち時間に関してより質の高い限界を規定すること、及びインスタンス化され得る様々なタイプの接続に関して A L G が一般に、よりフレキシブルであることを示している。40

【 0 1 1 0 】

【表1】

	同期	非同期		
	T D M	フェアシェア	A L Gの高速パス	A L Gの低速パス
t_{admit}	N	0	0	0
t_{access}	1	N	1	N
t_{link}	1	1	1	1
待ち時間	$N + h$	$(N + 1) \times h$	h	$(N + 1) \times h$
帯域幅	$1/N$	$1/N$	$1/N$	$1/(2N - 1)$

異なるG Sの技術の待ち時間及び帯域幅保証。

【図面の簡単な説明】

【0 1 1 1】

【図1】独立してクロック同期された複数のコア同士を接続する非同期ネットワークが大規模S o C設計におけるモジュール性を促進することを示す。

【図2】仮想チャネルA乃至Dが1つの物理リンクを共用する基本的なリンクを示す。

【図3】静的優先順位付き待ち行列(S P Q)がリンクへのアクセスに優先順位を付けて待ち時間保証を提供し、アドミッション制御はフリットのフローがS P Qにより要求される条件を固守することを確実にし、V C制御は非ブロッキングの振る舞いを保証する、好適な実施形態に係る完全リンクを示す。

【図4】図3のリンクのオペレーション例を示す。

【図5】接続により予約されるV Cのシーケンスの1つのセクションのモデルである。

【図6】共有ベースのV C制御を示す。

【図7】共有ボックス及び非共有ボックス概略的に示す。

【図8】アドミッション制御を概略的に示す。

【図9】(a)は低速パスのフリット待ち時間の分布とネットワーク負荷との関係を示し、(b)は高速パスのフリット待ち時間の分布とネットワーク負荷との関係を示す。

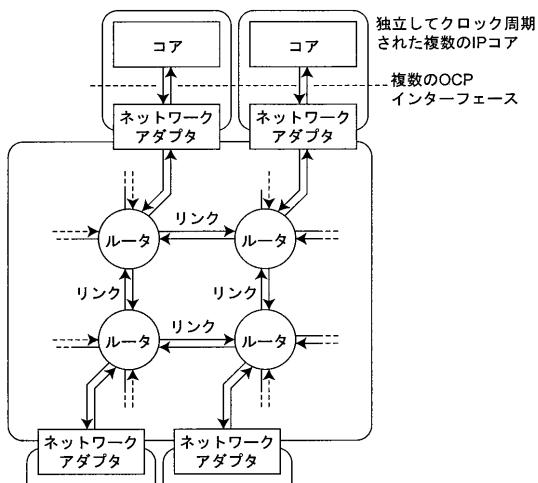
【図10】S P Qアービタ及び付随するマージデータパスの好適な非同期実装を示す。

10

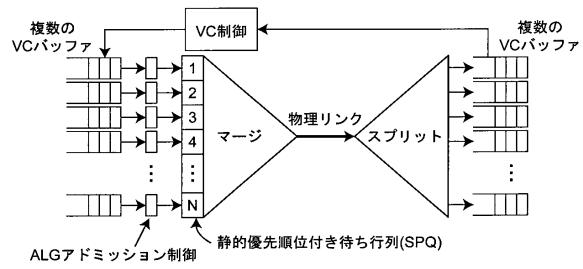
20

30

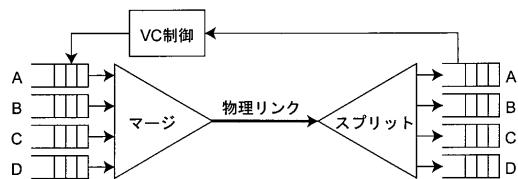
【図1】



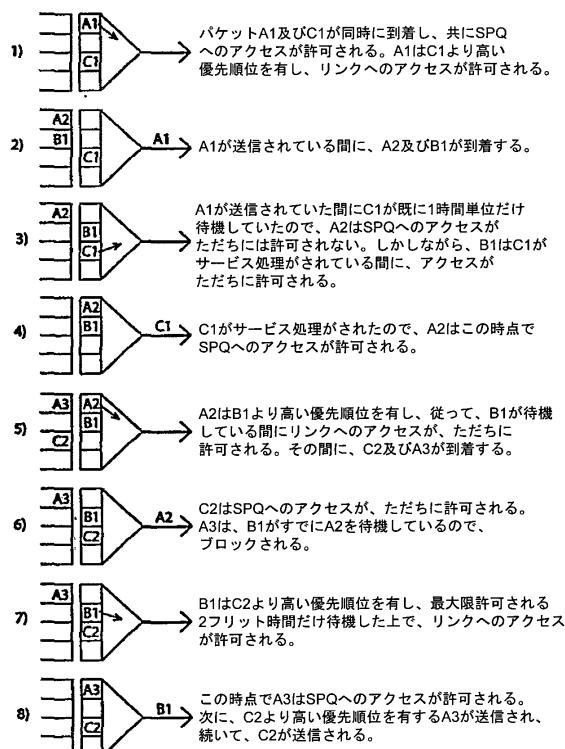
【図3】



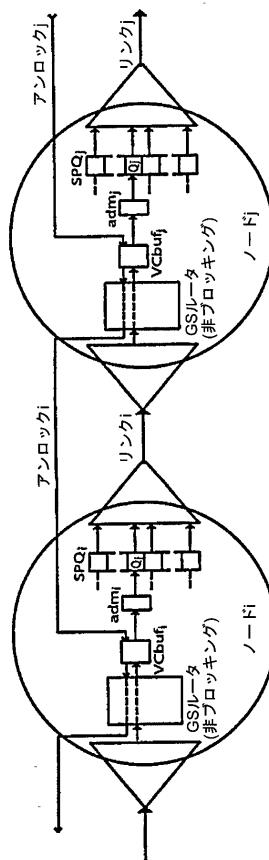
【図2】



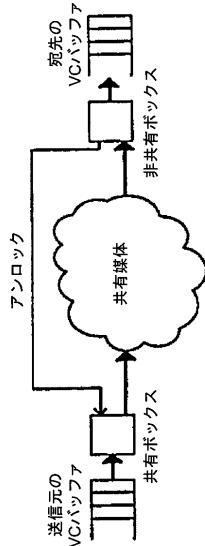
【図4】



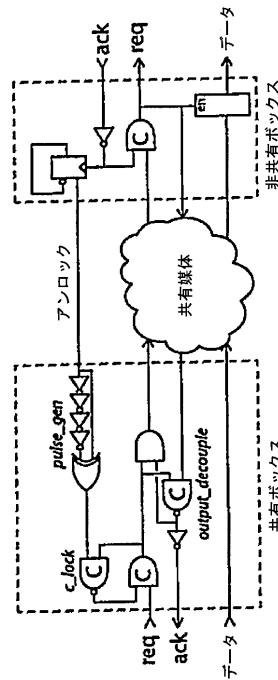
【図5】



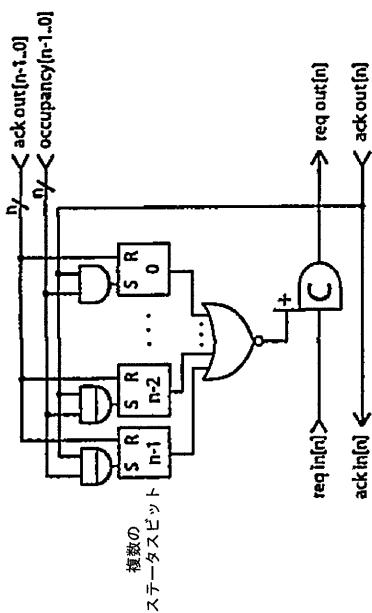
【図6】



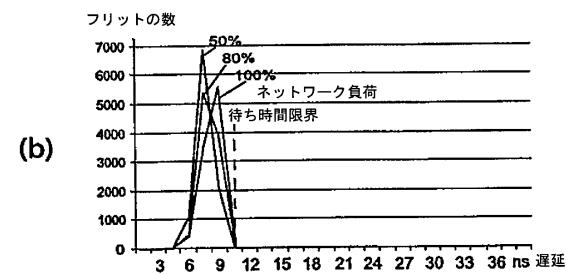
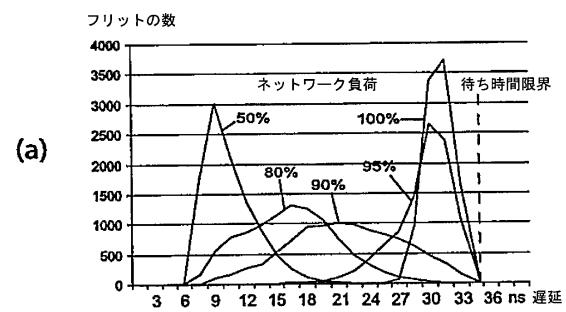
【 四 7 】



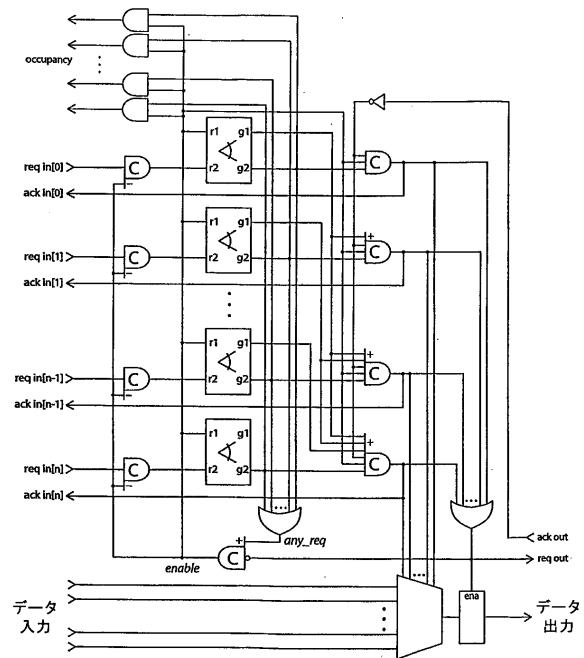
【 四 8 】



【図9】



【図10】



フロントページの続き

審査官 衣鳩 文彦

(56)参考文献 特開平10-200550(JP, A)
特開平11-046197(JP, A)

(58)調査した分野(Int.Cl., DB名)
H04L 12/56