

(19) 日本国特許庁 (JP)

(12) 公 開 特 許 公 報 (A)

(11) 特許出願公開番号

特開2020-9417

(P2020-9417A)

(43) 公開日 令和2年1月16日 (2020.1.16)

(51) Int.Cl.	F I	テーマコード (参考)
G 0 6 F 13/38 (2006.01)	G 0 6 F 13/38 3 1 0 C	5 B 0 4 2
G 0 6 F 15/173 (2006.01)	G 0 6 F 15/173 6 8 5 M	5 B 0 4 5
G 0 6 F 11/34 (2006.01)	G 0 6 F 13/38 3 4 0 C	5 B 0 7 7
	G 0 6 F 11/34 1 4 2	

審査請求 未請求 請求項の数 10 O L 外国語出願 (全 13 頁)

(21) 出願番号	特願2019-90545 (P2019-90545)	(71) 出願人	512033279 ブル・エス・アー・エス フランス国、エフー78340・レ・クレ イ・スー・ボア、リュ・ジャン・ジヨレ
(22) 出願日	令和1年5月13日 (2019.5.13)	(74) 代理人	100126572 弁理士 村越 智史
(31) 優先権主張番号	1854232	(72) 発明者	ピエール アクセル、ラガデック フランス共和国 ヴィユヌーヴ ルーベ 06270 アヴニユー ドゥラ バット ゥリール パロネ 1001
(32) 優先日	平成30年5月22日 (2018.5.22)	(72) 発明者	サイド、ドゥラッジ フランス共和国 イシー レ ムリーノ 92130 エスプラナード ラウル フ ォルロー 21
(33) 優先権主張国・地域又は機関	フランス (FR)		

最終頁に続く

(54) 【発明の名称】 データを送信する方法、並びにそれに対応するコンピュータプログラム及びシステム

(57) 【要約】 (修正有)

【課題】大容量の入力バッファメモリを必要とせずにデータを送信する方法並びに対応するコンピュータプログラム及びシステムを提供する。

【解決手段】システムにおいて、複数のスイッチのうちの2つである上流装置202から下流装置204にデータを送信する方法は、下流装置の同一の入力バッファメモリ208を共有する複数の仮想チャネルのうちの1つの仮想チャネルが、入力バッファメモリのメモリロケーションのうち、厳密に現在の上限值よりも小さい数のメモリロケーションを使用している場合に、前記1つの仮想チャネルから1つのデータを送信する工程を含む。方法はさらに、上流装置と下流装置との間の通信レイテンシを測定する工程と、レイテンシから現在の上限值を計算する工程と、を備える。

【選択図】 図2

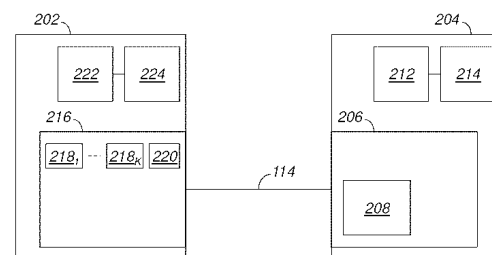


Figure 2

【特許請求の範囲】**【請求項 1】**

上流装置から下流装置にデータを送信する方法であって、

前記下流装置の同一の入力バッファメモリを共有する複数の仮想チャネルのうちの 1 つの仮想チャネルが、前記入力バッファメモリのメモリロケーションのうち、厳密に現在の上限値よりも小さい数のメモリロケーションを使用している場合に、前記 1 つの仮想チャネルから 1 つのデータを送信する工程を備える方法において、

前記上流装置と前記下流装置との間の通信レイテンシを測定する工程と、

前記レイテンシから前記現在の上限値を計算する工程と、

をさらに備えることを特徴とする方法。

10

【請求項 2】

前記レイテンシを測定する工程は、前記上流装置と前記下流装置との間でメッセージを交換するのに要した時間を計測する工程を含む、請求項 1 に記載の方法。

【請求項 3】

前記入力バッファメモリの稼働率を決定する工程をさらに備え、

前記現在の上限値が前記レイテンシ及び前記稼働率から計算される、請求項 1 又は 2 に記載の方法。

【請求項 4】

前記現在の上限値を計算する工程は、

前記レイテンシから複数の選択可能な上限値を計算する工程と、

前記稼働率に基づいて前記複数の選択可能な上限値のうちの 1 つを前記現在の上限値として選択する工程と、

を含む、請求項 3 に記載の方法。

20

【請求項 5】

前記複数の選択可能な上限値はそれぞれ稼働率の複数の隣接するインターバルに関連付けられ、前記選択された選択可能な上限値はその範囲内に前記決定された稼働率を含む、請求項 4 に記載の方法。

【請求項 6】

互いに隣接する下方インターバルと上方インターバルとの境界となる閾値が、前記選択された選択可能な上限値が前記下方インターバルに関連付けられた上限値であるときに高い値を取り、前記選択された選択可能な上限値が前記上方インターバルに関連付けられた上限値であるときに低い値を取り、当該低い値は前記高い値より低い値である、請求項 5 に記載の方法。

30

【請求項 7】

前記レイテンシから各仮想チャネルのために確保されるメモリロケーションの数を計算する工程をさらに備え、

前記 1 つの仮想チャネルが厳密に前記上限値よりも小さい数のメモリロケーションを使用しており、かつ前記入力バッファメモリが、確保されているが未使用であるメモリロケーションの数のすべての仮想チャネルについての合計よりも多くの未使用のメモリロケーションを有している場合に、前記 1 つのデータが送信される、請求項 1 から 6 のいずれか 1 項に記載の方法。

40

【請求項 8】

前記上流装置及び前記下流装置は、

ノードと、

前記ノードの相互接続ネットワークであって、スイッチとコミュニケーションリンクとを含み、当該コミュニケーションリンクは前記スイッチ間の接続及び前記スイッチと前記ノードとの接続を行う、相互接続ネットワークと、
を含むシステムに属し、

前記上流装置及び前記下流装置の各々は、前記ノードの一つ又は前記スイッチの一つのいずれかである、請求項 1 から 7 のいずれか 1 項に記載の方法。

50

【請求項 9】

通信ネットワークからダウンロード可能、及び／又はコンピュータ可読媒体に記録可能、及び／又はプロセッサによって実行可能なコンピュータプログラムであって、コンピュータ上で実行されたときに、請求項 1 から 8 のいずれか 1 項に記載の方法の各工程を実行するための命令を含む、コンピュータプログラム。

【請求項 10】

データを送信するためのシステムであって、
入力バッファメモリを含む下流装置と、
前記入力バッファメモリを共有する複数の仮想チャネルのうちの 1 つの仮想チャネルが、前記入力バッファメモリのメモリロケーションのうち、厳密に現在の上限値よりも小さい数のメモリロケーションを使用している場合に、前記 1 つの仮想チャネルから 1 つのデータを送信するように構成された上流装置と、
前記上流装置と前記下流装置との間の通信レイテンシを測定する手段と、
前記レイテンシから前記現在の上限値を計算する手段と、
を備えるシステム。

10

【発明の詳細な説明】**【技術分野】****【0001】**

本発明はデータを送信する方法、並びにそれに対応するコンピュータプログラム及びシステムに関する。

20

【0002】

本発明は特に、上流装置から下流装置にデータを送信する方法であって、前記下流装置の同一の入力バッファメモリを共有する複数の仮想チャネルのうちの 1 つの仮想チャネルが、前記入力バッファメモリのメモリロケーションのうち、厳密に現在の上限値よりも小さい数のメモリロケーションを使用している場合に、前記 1 つの仮想チャネルから 1 つのデータを送信する工程を備える方法に適用される。

【背景技術】**【0003】**

例えば、前記上流装置及び前記下流装置は、ノードと、前記ノードの相互接続ネットワークであって、スイッチとコミュニケーションリンクとを含み、当該コミュニケーションリンクは前記スイッチ間の接続及び前記スイッチと前記ノードとの接続を行う、相互接続ネットワークと、を含むシステムに属する。前記上流装置及び前記下流装置の各々は、前記ノードの一つ又は前記スイッチの一つのいずれかである。前記スイッチ及び前記ノードは通信リンクが接続された接続ポートを有する。

30

【0004】

このフレームワークにおいて、システム内に存在するすべての入力バッファメモリについて、前記上限値を同一の値に設定することが知られている。この値は大きな値となるよう選択され、よって、各仮想チャネルがその傾向において十分なメモリロケーションを有し、各仮想チャネルのデータを円滑に送信できるようになっている。

【発明の概要】

40

【発明が解決しようとする課題】**【0005】**

この解決法の問題点は大容量の入力バッファメモリを必要とすることである。

【0006】

そこで、この問題点及び制約の少なくとも一部を解決できる、データを送信する方法が求められている。

【課題を解決するための手段】**【0007】**

よって、本発明の目的は、上流装置から下流装置にデータを送信する方法であって、前記下流装置の同一の入力バッファメモリを共有する複数の仮想チャネルのうちの 1 つの仮

50

想チャンネルが、前記入力バッファメモリのメモリロケーションのうち、厳密に現在の上限値よりも小さい数のメモリロケーションを使用している場合に、前記1つの仮想チャンネルから1つのデータを送信する工程を備える方法において、前記上流装置と前記下流装置との間の通信レイテンシを測定する工程と、前記レイテンシから前記現在の上限値を計算する工程と、をさらに備えることを特徴とする方法を提供することである。

【0008】

実際、通信リンクの長さはシステムのノード間で一定ではない。本発明により、上流装置と下流装置の対ごとに異なる上限値を計算することが可能になる。この上限値は、通信リンクの実際の長さに依存する通信レイテンシにしたがって可能な限り最善の値として計算され、それによって、高性能を維持しつつ仮想チャンネルごとに必要な量のメモリのみを使用することが可能になる。

10

【0009】

また、前記レイテンシを測定する工程は、前記上流装置と前記下流装置との間でメッセージを交換するのに要した時間を計測することを含んでもよい。

【0010】

また、前記方法はさらに、前記入力バッファメモリの稼働率を決定する工程を備えてもよく、前記現在の上限値が前記レイテンシ及び前記稼働率から計算されてもよい、

【0011】

また、前記現在の上限値を計算する工程は、前記レイテンシから複数の選択可能な上限値を計算する工程と、前記稼働率に基づいて前記複数の選択可能な上限値のうちの1つを前記現在の上限値として選択する工程と、を含んでもよい。

20

【0012】

また、前記複数の選択可能な上限値はそれぞれ稼働率の複数の隣接するインターバルに関連付けられ、前記選択された選択可能な上限値はその範囲内に前記決定された稼働率を含んでもよい。

【0013】

また、互いに隣接する下方インターバルと上方インターバルとの境界となる閾値が、前記選択された選択可能な上限値が前記下方インターバルに関連付けられた上限値であるときに高い値を取り、前記選択された選択可能な上限値が前記上方インターバルに関連付けられた上限値であるときに低い値を取り、当該低い値は前記高い値より低い値であってもよい。

30

【0014】

また、前記方法はさらに、前記レイテンシから各仮想チャンネルのために確保されるメモリロケーションの数を計算する工程を備えてもよく、前記1つの仮想チャンネルが厳密に前記上限値よりも小さい数のメモリロケーションを使用しており、かつ前記入力バッファメモリが、確保されているが未使用であるメモリロケーションの数のすべての仮想チャンネルについての合計よりも多くの未使用のメモリロケーションを有している場合に、前記1つのデータが送信されてもよい。

【0015】

また、前記上流装置と前記下流装置は、ノードと、前記ノードの相互接続ネットワークであって、スイッチとコミュニケーションリンクとを含み、当該コミュニケーションリンクは前記スイッチ間の接続及び前記スイッチと前記ノードとの接続を行う、相互接続ネットワークと、を含むシステムに属し、前記上流装置及び前記下流装置の各々は、前記ノードの一つ又は前記スイッチの一つのいずれかであってもよい。

40

【0016】

本発明の別の目的は、通信ネットワークからダウンロード可能、及び／又はコンピュータ可読媒体に記録可能、及び／又はプロセッサによって実行可能なコンピュータプログラムであって、コンピュータ上で実行されたときに、本発明に記載の方法の各工程を実行するための命令を含む、コンピュータプログラムを提供することである。

【0017】

50

本発明の別の目的は、データを送信するためのシステムであって、入力バッファメモリを含む下流装置と、前記入力バッファメモリを共有する複数の仮想チャネルのうちの1つの仮想チャネルが、前記入力バッファメモリのメモリロケーションのうち、厳密に現在の上限値よりも小さい数のメモリロケーションを使用している場合に、前記1つの仮想チャネルから1つのデータを送信するように構成された上流装置と、前記上流装置と前記下流装置との間の通信レイテンシを測定する手段と、前記レイテンシから前記現在の上限値を計算する手段と、を備えるシステムを提供することである。

【0018】

本発明は、単に例として提示される以下の説明によってよりよく理解されるであろう。

【図面の簡単な説明】

10

【0019】

【図1】本発明の一実施形態によるシステムの全体構造を示す概略図である。

【図2】図1のシステムの上流装置及び下流装置の全体構造を示す概略図である。

【図3】本発明の一実施形態による、図2の上流装置から下流装置にデータを送信する方法の連続的な工程を示す図である。

【発明を実施するための形態】

【0020】

図1を参照して、本発明を実装するシステム100を説明する。

【0021】

システム100は複数の独立したコンピュータ装置102をグループ化したものであり、外部からは非常に高い計算能力を有する単一のコンピュータ、いわゆる高性能処理コンピュータもしくはスーパーコンピュータ、又はHPC(High Performance Computing)コンピュータのように見える。

20

【0022】

各コンピュータ装置102は少なくとも1つのネットワークインターフェースを有する。各コンピュータ装置102はさらに、中央演算処理装置、当該中央演算処理装置が処理する命令が記録されるメインメモリ（これらを有するコンピュータ装置102を以下で「計算用コンピュータ装置」という）、及び／又はデータ記憶装置（これを有するコンピュータ装置102を以下で「入力／出力（IO）用コンピュータ装置」という）を備えることができる。データ記憶装置は、例えば、計算用コンピュータ装置によって使用及び／又は生成されるデータの記録及び／又は復旧のための1つ以上のハードドライブなどの大容量メモリである。

30

【0023】

各ネットワークインターフェースはそれぞれシステム100のノード106を形成する。複数のネットワークインターフェース（すなわち複数のノード106）は同一のコンピュータ装置102に属してもよく、あるいは、説明する例におけるように、各コンピュータ装置102はノード106の1つを形成する単一のネットワークインターフェースのみを含んでもよい。計算用コンピュータ装置102のネットワークインターフェースによって形成されるノード106は「計算ノード」と呼ばれ、入力／出力用コンピュータ装置102のネットワークインターフェースによって形成されるノード106は「入力／出力ノード」と呼ばれる。

40

【0024】

システム100はさらに、管理サーバ108を含む。管理サーバ108は特にシステム100の使用を望むクライアントとの対話を行うように設計されており、システム100はクライアントによって単一のコンピュータと見なされる。管理サーバ108はさらに、クライアントから受けたタスクを計算用コンピュータ装置間で順番に配分するように設計されている。

【0025】

このようなシステム100では、計算用コンピュータ装置の少なくとも一部にわたって、複雑な処理の配分及び／又は並行計算処理の配分が可能である。

50

【0026】

システム100はさらに、ノード106の相互接続ネットワーク110を含む。

【0027】

ネットワーク110は、スイッチ112及びコミュニケーションリンク114を含み、コミュニケーションリンク114はスイッチ112間の接続及びスイッチ112とノード106との接続を行う。各スイッチ112及び各ノード106はいくつかの接続ポート（図1においてスイッチ112については小さい四角で示す）を有し、各コミュニケーションリンク114はスイッチ112の接続ポートと他のスイッチ112の接続ポートとの間、又はノード106の接続ポートとの間を接続する。

【0028】

システム100は、ノード106、スイッチ112、及び通信リンク114の相対配置を定義する特定のトポロジを有する。説明する例において、このトポロジは並列一般化ファットツリー（PGFT）のトポロジである。

【0029】

図2を参照して、上流装置202、下流装置204、及びそれらを接続する通信リンク114を説明する。説明する例において、上流装置202及び下流装置204は図1のスイッチ112のうちの2つである。あるいは、上流装置202及び下流装置204の一方はノード106のうちの1つであってもよい。

【0030】

説明する例において、ポートと通信リンク114は双方向に機能する。よって、各スイッチ112及び各ノード106はいずれも、下流装置でありかつ上流装置でもある。これらの各ポートは入力ポートであるとともに出力ポートでもあり、以下に説明する上流装置202の出力ポート及び下流装置204の入力ポートのそれぞれの構成要素のすべてを備える。

【0031】

下流装置204の入力ポート206には通信リンク114が接続され、データの受け取りを行う。入力ポート206は所定数Kの仮想チャネル及び例えば毎秒25GBのスループットDに対応する。入力ポート206は入力バッファメモリ208を備え、この入力バッファメモリ208が有する特定の数のメモリロケーションの各々が単位量のデータを受け取るようになっている。この特定の数を入力バッファメモリ208のキャパシティCという。説明する例において、単位量のデータは32バイトで構成されるフロー制御単位（FLIT）である。入力バッファメモリ208のキャパシティCは、例えば、4096FLITである。

【0032】

下流装置204はさらに、マイクロプロセッサなどの処理装置212と、処理装置212に作動的に結合されたメモリ214とを備え、メモリ214は処理装置212が実行するための命令からなるコンピュータプログラムを含む。

【0033】

上流装置202の出力ポート216には通信リンク114が接続される。出力ポート216は、入力ポート206と同じ所定数Kの仮想チャネルと、入力ポート206と同じスループットDに対応する。出力ポート216は、K個の仮想チャネルの各々に使用するクレジットカウンタ $218_1 \sim 218_K$ を備え、仮想チャネルのすべてによって使用されているメモリロケーションをカウントする入力バッファメモリ208の使用カウンタ220を備える。よって、入力バッファメモリ208は、使用カウンタ220の値と入力バッファメモリ208のキャパシティCとの間の比率と等しい稼働率TUを有する。

【0034】

上流装置202はさらに、マイクロプロセッサなどの処理装置222と、処理装置222に作動的に結合されたメモリ224とを備え、メモリ224は処理装置222が実行するための命令からなるコンピュータプログラムを含む。

【0035】

10

20

30

40

50

このように、説明する例において、上流装置 202 及び下流装置 204 はコンピュータプログラムの形態でソフトウェア手段を備え、このコンピュータプログラムは通信ネットワークからダウンロード可能、及び／又はコンピュータ可読媒体に記録可能、及び／又はプロセッサによって実行可能であり、上流装置 202 及び下流装置 204 によって実行されたときに、以下に説明する方法の各工程を実行するための命令を含む。あるいは、これらの手段の全部または一部は、専用集積回路内のマイクロプログラム式 (micro-programmed) またはマイクロワイヤ式 (micro-wired) ハードウェア手段であり得る。したがって、別の形態として、上流装置 202 及び下流装置 204 の各々は、同様の動作を実行するためのデジタル回路からのみ構成される (コンピュータプログラムなしの) 電子装置であり得る。

10

【0036】

さらに、入力バッファメモリ 208 のキャパシティ C 及びスループット D はメモリ 224 内に記録され、後述の計算に使用できるようになっている。

図 3 を参照して、上流装置 202 から下流装置 204 にデータを送信する方法 300 を説明する。

【0037】

工程 302 において、上流装置 202 は上流装置 202 と下流装置 204 との間の通信レイテンシ (以下では記号 L で表す) を測定する。

【0038】

説明する例において、システム 100 は起動処理中に初期化フェーズを実行するが、このときスイッチ 112 はスイッチ相互の間及びノード 106 との間で初期化メッセージを交換する。これらの交換メッセージを用いて、初期化フェーズ中に上流装置 202 と下流装置 204 との間でメッセージを交換するのに要した時間を計測することによって、レイテンシ L が決定される。より正確には、上流装置 202 は通信リンク 114 を介して第 1 の初期化メッセージを下流装置 204 に送信する。下流装置 204 は第 1 の初期化メッセージを受信し、それに応答して、やはり通信リンク 114 を介して第 2 の初期化メッセージを上流装置 202 に送信する。上流装置 202 はこの第 2 の初期化メッセージを受信する。この受信は第 1 の初期化メッセージの送信からある一定の時間だけ間隔があり、この時間を上流装置 202 が計測する。説明する例において、レイテンシ L はこの計測された時間に対応する。

20

30

【0039】

工程 304 において、上流装置 202 は入力バッファメモリ 208 の稼働率 TU の複数の隣接するインターバルの各々について、選択可能な上限値をレイテンシ L から計算する。後述するように、これらの選択可能な上限値の各々は、各仮想チャネルによって使用可能なメモリロケーションの最大数を表す現在の上限値 P を形成するために選択される。

【0040】

例えば、これらの選択可能な上限値の各々は、レイテンシ L 及び通信リンク 114 が接続された出力ポートのスループット D から計算される。

【0041】

より好ましくは、選択可能な上限値の各々は、入力バッファメモリ 208 のキャパシティ C を仮想チャネルの数 K で除算した値と等しい最小値を有する。例えば、キャパシティ C が 4096 メモリロケーションであり、仮想チャネルの数 K が 20 である場合、選択可能な上限値の最小値は 205 メモリロケーションである。

40

【0042】

より好ましくは、稼働率が上昇する方向において、1つのインターバルから次のインターバルに移るにしたがって、選択可能な上限値は低下する。

【0043】

説明する例において、2つのインターバルが用いられる。第 1 のインターバルは 75 % 以下の稼働率 TU に対応し、第 2 のインターバルは厳密に 75 % より大きい稼働率 TU に対応する。第 1 のインターバルは以下の式で表される例における第 1 の選択可能な上限値

50

P 1に関連付けられる。

$$P 1 = \max (L \times D ; C / K)$$

ここで、Lはレイテンシ、Dはスループット（F L I T数で表す）、Cは入力バッファメモリ208のキャパシティ、Kは仮想チャネルの数である。例えば、測定されたレイテンシLが400nsであり、スループット（バイト数で表す）が25GB/s（すなわち 0.78125×10^9 F L I T/s）である場合、第1の選択可能な上限値P 1は313である。第2のインターバルは以下の式で表される例における第2の選択可能な上限値P 2に関連付けられる。

$$P 2 = \max (50 \% \times L \times D ; C / K)$$

ここで、Lはレイテンシ、Dはスループット（F L I T数で表す）、Cは入力バッファメモリ208のキャパシティ、Kは仮想チャネルの数である。例えば、測定されたレイテンシLが400nsであり、スループット（バイト数で表す）が25GB/s（すなわち 0.78125×10^9 F L I T/s）である場合、 $50 \% \times L \times D$ は157であり、第2の選択可能な上限値P 2は最小値の205である。

【0044】

稼働率TUが上昇する方向において、2つの隣接するインターバルのうちの一方（下方インターバルという）が他方（上方インターバルという）の下側に位置しているが、閾値がこれら2つのインターバルの境界となることが理解されよう。説明する例において、2つのインターバルを分ける閾値は75%であり、これに固定されている。しかしながら、他の実施形態において、この閾値はヒステリシスの現象にしたがって変化し得る。より正確には、選択された選択可能な上限値が下方インターバルに関連付けられたものであるとき、閾値は高い値を取り、選択された選択可能な上限値が上方インターバルに関連付けられたものであるとき、閾値は低い値を取る。この低い値は上記高い値より低い値である。例えば、選択可能な上限値P 1が選択されたとき、第2のインターバルに移るための閾値は80%であり、選択可能な上限値P 2が選択されたとき、下方インターバルに移るための閾値は70%である。このヒステリシスの現象によって、稼働率TUが閾値周辺を変動するときに現在の上限値Pが頻繁に変化することを回避できる。

【0045】

工程306において、上流装置202は仮想チャネルごとに確保されるメモリロケーションの数（この数は N_R と表す）をレイテンシLから計算する。説明する例において、以下の式を用いる。

$$N_R = 10 \% \times L \times D$$

ここで、Lはレイテンシ、Dはスループット（F L I T数で表す）である。

【0046】

工程308において、上流装置202は第1の選択可能な上限値P 1を現在の上限値Pとして選択する。説明する例において、この工程は各クレジットカウンタ218₁~218_Kを第1の数P 1に初期化することによって実行される。

【0047】

工程310において、上流装置202は使用カウンタ220をゼロに初期化する。

【0048】

その後、上流装置202が通信リンク114を介して下流装置204にF L I Tを送信しようとする度に、以下の第1の工程ループが実行される。ここで、送信するF L I Tの仮想チャネルをk 1で表す。

【0049】

工程312において、上流装置202は、例えば入力バッファメモリ208のキャパシティC及び使用カウンタ220を用いて、入力バッファメモリ208の稼働率TUを決定する。

【0050】

工程314において、上流装置202は決定された稼働率TUが含まれるインターバルに関連付けられた数を現在の上限値Pとして選択する。

【0051】

工程316において、上限値Pが工程314で変化した場合、上流装置202は変化前の上限値Pと変化後の上限値Pの差を計算し、この差を各クレジットカウンタ2181～218Kに適用する。

【0052】

例えば、現在の上限値Pが工程314の前には選択可能な上限値P2であり、工程314の終了時には選択可能な上限値P1である場合、再び上述の例によれば、その差は $P1 - P2$ 、すなわち $313 - 205 = 108$ となる。よって、各クレジットカウンタ2181～218Kに108のクレジットが加算される。

【0053】

あるいは、現在の上限値Pが工程314の前には選択可能な上限値P1であり、工程314の終了時には選択可能な上限値P2である場合、やはり上述の例によれば、その差は $P2 - P1$ 、すなわち -108 となる。よって、各クレジットカウンタ2181～218Kから108のクレジットが減算される。これによってクレジットカウンタが負の値を持つこともあり得る。その場合、対応する仮想チャネルが現在の上限値Pを超過しており、よって入力バッファメモリ208のメモリロケーションを多く使い過ぎていることを意味する。

【0054】

工程318において、上流装置202は処理対象のFLITの仮想チャネルが厳密に現在の上限値Pよりも小さい数のメモリロケーションを使用しているかを確認する。

【0055】

説明する例において、この工程は対応するクレジットカウンタ218kが厳密にゼロより大きいことを確認することによって実行される。

【0056】

工程320において、上流装置202は、確保されているが未使用であるメモリロケーションの数のすべての仮想チャネルについての合計よりも多くの未使用のメモリロケーションを入力バッファメモリ208が有しているかを確認する。

【0057】

説明する例において、この工程は各仮想チャネルについてこの仮想チャネルが使用しているメモリロケーションの数（この数を $N_U(k)$ と表す。kは仮想チャネルを表す）を、例えば以下の式を用いて計算することによって実行される。

$$N_U(k) = P - C(k)$$

ここで、 $C(k)$ はクレジットカウンタ218kの値である。その後、この仮想チャネルのために確保されているが未使用であるメモリロケーションの数（ $N_{RL}(k)$ と表す）を、確保されているメモリロケーションの数 N_R からこの仮想チャネルが使用しているメモリロケーションの数を減算し、この結果をゼロ以上に限定することによって計算する。例えば、以下の式を用いる。

$$N_{RL}(k) = \max(N_R - N_U(k); 0)$$

その後、確保されているが未使用であるメモリロケーションの数の合計（すなわち、すべての仮想チャネルについての合計であり、 N_{RL} で表す）を、各仮想チャネルの確保されているが未使用であるメモリロケーションの数をすべての仮想チャネルについて加算することによって計算する。例えば、以下の式を用いる。

$$N_{RL} = \sum_k N_{RL}(k)$$

入力バッファメモリ208の未使用のメモリロケーションの数は、例えば以下の式で与えられる。

$$N_L = C - E_U$$

ここで、 E_U は使用されているメモリロケーションの数であり、使用カウンタ220によって与えられる。その後、確保されているが未使用であるメモリロケーションの合計数を、入力バッファメモリ208内の未使用のメモリロケーションの数と比較する。

【0058】

10

20

30

40

50

工程 3 2 2 において、仮想チャネル k_1 が厳密に現在の上限値 P より小さい数のメモリロケーションを使用しており、かつ入力バッファメモリ 2 0 8 が、確保されているが未使用であるメモリロケーションの数のすべての仮想チャネルについての合計よりも多くの未使用のメモリロケーションを有している場合に、上流装置 2 0 2 は F L I T F 1 を送信し、クレジットカウンタ 218_{k_1} をデクリメントし、使用カウンタ 2 2 0 をインクリメントする。それ以外の場合は、上流装置 2 0 2 は F L I T F 1 を送信せず、工程 3 1 2 に戻って他の F L I T の処理を行う。

【0059】

工程 3 2 4 において、下流装置 2 0 4 は F L I T F 1 を受信して入力バッファメモリ 2 0 8 内に記録する。

【0060】

並行して、例えば他の装置に送信するために下流装置 2 0 4 がその入力バッファメモリ 2 0 8 から F L I T F 2 を削除する度に、以下の第 2 の工程ループが実行される (F L I T F 2 の仮想チャネルを k_2 で表す)。

【0061】

工程 3 2 6 において、下流装置 2 0 4 はその入力バッファメモリ 2 0 8 から F L I T F 2 を削除する。

【0062】

工程 3 2 8 において、下流装置 2 0 4 は仮想チャネル k_2 のクレジットを通信リンク 1 1 4 を介して上流装置 2 0 2 に送信する。

【0063】

工程 3 3 0 において、上流装置 2 0 2 は仮想チャネル k_2 のクレジットを受信し、仮想チャネル k_2 に対応するクレジットカウンタ 218_{k_2} をインクリメントし、使用カウンタ 2 2 0 をデクリメントする。

【0064】

なお、他の実施形態においては、現在の上限値が仮想チャネルごとに異なることもあり得る。例えば、各仮想チャネルの現在の上限値はさらに仮想チャネルの使用履歴から計算することもできる。これによって、過去の使用頻度が低い仮想チャネルの上限値を低くし、逆に使用頻度が高い仮想チャネルの上限値を高くすることができる。

【0065】

このような方法により、高性能を維持しつつ入力バッファメモリの量を低減することが可能になる。

【0066】

さらに、本発明は上述の実施形態に限定されないことに留意されたい。本明細書の記載に鑑み、上述の実施形態には種々の変更が可能であることが当業者には明らかであろう。上記の本発明の詳細な説明において、使用された用語は本発明をこの説明で開示された実施形態に限定するものとして解釈されるべきではなく、当業者がその一般的な知識を本明細書の内容の実施に適用する範囲内で予見されるすべての均等物を含むものとして解釈されるべきである。

10

20

30

【図 1】

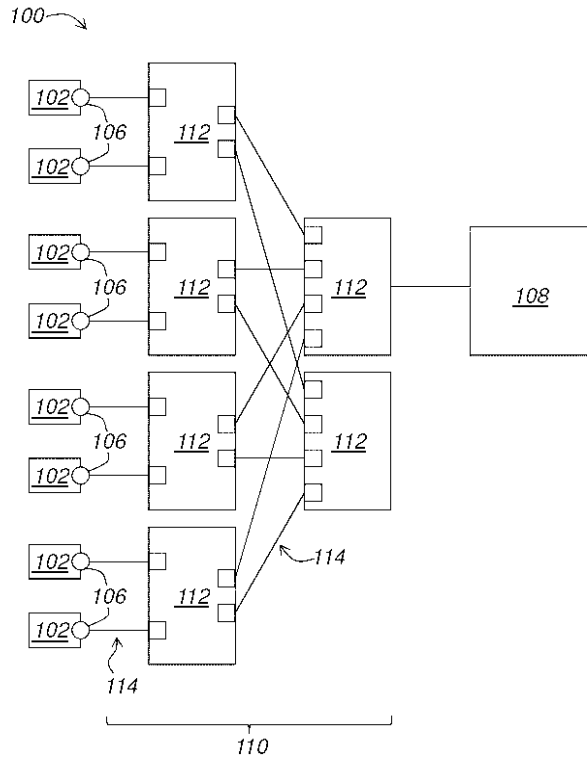


Figure 1

【図 2】

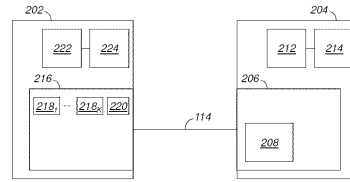


Figure 2

【図 3】

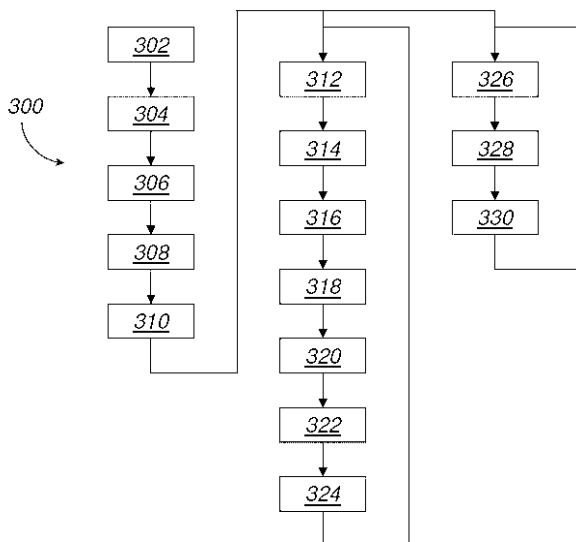


Figure 3

フロントページの続き

(72)発明者 ドミニック、リガル

フランス共和国 ニース 0 6 2 0 0 アヴニュー サント マルゲリット レジダンス レ ヴ
ォワリエ 1 5 0

(72)発明者 ローラン、マーリアック

フランス共和国 サン ローラン デュ ヴァー 0 6 7 0 0 アブニュー デ プランティエ
3 5

F ターム(参考) 5B042 JJ23 MA08 MA14 MC29 MC33

5B045 BB15 BB38 JJ08

5B077 DD22

【外国語明細書】
2020009417000001.pdf