(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2016-501474
(P2016-501474A)
(43)公表日　平成28年1月18日(2016.1.18)

| (51)Int.Cl. | F I | | テーマコード（参考） |
|---|---|---|---|
| H04L 12/721 (2013.01) | H04L 12/721 | Z | 5K030 |

審査請求 未請求　予備審査請求 未請求　（全 27 頁）

(21)出願番号　特願2015-542376 (P2015-542376)
(86)(22)出願日　平成25年11月13日 (2013.11.13)
(85)翻訳文提出日　平成27年7月13日 (2015.7.13)
(86)国際出願番号　PCT/IB2013/003051
(87)国際公開番号　WO2014/076573
(87)国際公開日　平成26年5月22日 (2014.5.22)
(31)優先権主張番号　13/678,382
(32)優先日　平成24年11月15日 (2012.11.15)
(33)優先権主張国　米国 (US)

(71)出願人　512039835
コンパス・エレクトロ－オプティカル・システムズ・リミテッド
イスラエル国　４２５０４　ネタニヤ，ギボリー・イスラエル・ストリート　７，ポレグ・インダストリアル・エリア，ピーオーボックス　８７６３
(74)代理人　100101454
弁理士　山田　卓二
(74)代理人　100081422
弁理士　田中　光雄
(74)代理人　100125874
弁理士　川端　純市

最終頁に続く

(54)【発明の名称】分散型スイッチレス相互接続

(57)【要約】
フルメッシュ分散型スイッチレスシステムは、発信元ノード及び宛先ノードの間の直接通信及び間接通信を可能にする。直接通信において、データは、発信元ノード及び宛先ノードを接続するリンクを介して伝搬する。間接通信において、データはまず、発信元ノード及び中間ノードを接続するリンクを介して中間ノードに送られる。中間ノードは、中間ノード及び宛先ノードを接続するリンクを介して宛先ノードにデータを送る。トラフィックは、発信元ノード及び宛先ノードを接続するリンクだけではなく、複数のノードにわたる利用可能なすべてのリンクに分割することができる。直接通信よりも間接通信のほうがより多くのリンクを使用するので、各リンク中のトラフィックはより小さくなる。従って、スイッチレス分散型相互接続システムは、任意の２つのノード間のリンク数を削減し、より小さな帯域幅のリンクを用いて、動作可能である。

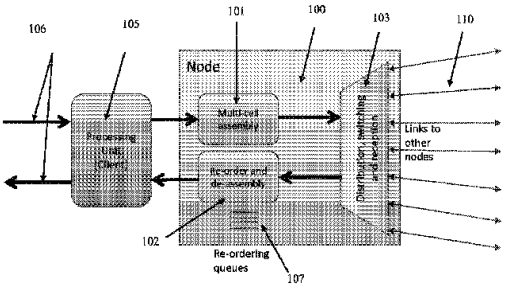



FIG. 1

【特許請求の範囲】
【請求項１】
フルメッシュ分散型スイッチレス相互接続システムにおける第１の処理エンジンによって、宛先処理エンジンを示すデータを受信することと、
上記第１の処理エンジンによって、上記第１の処理エンジンが上記フルメッシュ分散型スイッチレス相互接続システムにおける上記宛先処理エンジンであるか否かを決定することと、
上記第１の処理エンジンが上記宛先処理エンジンであるとき、上記第１の処理エンジンによって上記データを処理することと、
上記第１の処理エンジンが上記宛先処理エンジンでないとき、上記第１の処理エンジンによって、上記データを上記宛先処理エンジンに送るか、又は上記データを上記フルメッシュ分散型スイッチレス相互接続システムにおける中間処理エンジンに送ると決定することと、
上記第１の処理エンジンが上記宛先処理エンジンでないとき、上記第１の処理エンジンによる決定に基づいて、上記データを上記宛先処理エンジン及び上記中間処理エンジンのいずれかに送信することとを含む方法。
【請求項２】
上記データは、少なくとも１つのパケットを含む少なくとも１つのマスターセルを含む請求項１記載の方法。
【請求項３】
上記少なくとも１つのマスターセルは、キューシーケンス番号を有するヘッダを含む請求項２記載の方法。
【請求項４】
上記データを処理することは、
上記キューシーケンス番号に基づいて上記少なくとも１つのマスターセルを順序づけることと、
上記少なくとも１つのパケットを抽出することと、
上記少なくとも１つのパケットを、上記第１の処理エンジンに接続された処理装置に送ることとを含む請求項３記載の方法。
【請求項５】
上記第１の処理エンジンによって、少なくとも１つパケットを含む少なくとも１つのマスターセルを形成することをさらに含む請求項４記載の方法。
【請求項６】
リンク容量、処理エンジン能力、リンクの個数、及び中間処理エンジンの個数の少なくとも１つに基づいて、上記マスターセルのサイズを決定することをさらに含む請求項５記載の方法。
【請求項７】
形成することは、タイムアウトパラメータが終了するまで、上記少なくとも１つのパケットを上記少なくとも１つのマスターセルに追加することを含む請求項５記載の方法。
【請求項８】
形成することは、利用可能なパケットの量及び配送の緊急度に基づいて、上記少なくとも１つのパケットを上記少なくとも１つのマスターセルに追加することを含む請求項５記載の方法。
【請求項９】
上記方法は、上記第１の処理エンジンによって、上記データから複数のマスターセルを形成することを含み、
上記第１の処理エンジンは、上記第１の処理エンジンに接続された処理装置から上記データを受信し、
送信することは、上記複数のマスターセルを複数の中間処理エンジンに送信することを含み、

上記複数のマスターセルは複数の異なるパケットを含む請求項１記載の方法。
【請求項１０】
上記第１の処理エンジンが第２の処理エンジンからのキープアライブメッセージの受信に失敗したとき、上記第１の処理エンジンによって、上記第１の処理エンジンを上記第２の処理エンジンに接続する第１のリンクが不活性であることをブロードキャストすることをさらに含む請求項１記載の方法。
【請求項１１】
上記第１の処理エンジンによって、第１のキープアライブメッセージを第２の処理エンジンに送ることと、
上記第１の処理エンジンが、上記第１のキープアライブメッセージに応答した上記第２の処理エンジンからのメッセージの受信に失敗したとき、上記第１の処理エンジンによって、上記第１の処理エンジンを上記第２の処理エンジンに接続する第１のリンクが不活性であることをブロードキャストすることとをさらに含む請求項１記載の方法。
【請求項１２】
上記第１の処理エンジンによって、第１のキープアライブメッセージを第２の処理エンジンに送ることと、
上記第１の処理エンジンが、上記第１のキープアライブメッセージの受信に失敗したことを示すメッセージを上記第２の処理エンジンから受信したとき、上記第１の処理エンジンによって、上記第１の処理エンジンを上記第２の処理エンジンに接続する第１のリンクが不活性であることをブロードキャストすることとをさらに含む請求項１記載の方法。
【請求項１３】
フルメッシュ分散型スイッチレス相互接続システムにおける装置であって、上記装置は、
処理装置と、
上記処理装置に接続され、上記フルメッシュ分散型スイッチレス相互接続システムを介して、宛先処理エンジンを示すデータを受信するように構成された第１の処理エンジンとを備え、
上記第１の処理エンジンは、上記第１の処理エンジンが上記フルメッシュ分散型スイッチレス相互接続システムにおける上記宛先処理エンジンであるとき、データを処理し、上記処理されたデータを上記処理装置へ配送するように構成され、
上記第１の処理エンジンは、上記第１の処理エンジンが上記宛先処理エンジンでないとき、上記宛先処理エンジンに、又は上記フルメッシュ分散型スイッチレス相互接続システムにおける中間処理エンジンに、上記データを送信するように構成される装置。
【請求項１４】
上記第１の処理エンジンは、上記受信されたデータが上記処理装置から受信されたとき、上記受信されたデータに基づく少なくとも１つのパケットを含む少なくとも１つのマスターセルを形成するように構成されたアセンブリ装置を備える請求項１３記載の装置。
【請求項１５】
上記第１の処理エンジンは、上記少なくとも１つのマスターセルを複数の中間処理エンジンに送信するように構成される請求項１４記載の装置。
【請求項１６】
上記第１の処理エンジンは、第２の処理エンジンから受信された少なくとも１つのマスターセルを並べ替え、上記受信されたマスターセルを少なくとも１つのパケットに分解するように構成された並べ替え装置を備える請求項１３記載の装置。
【請求項１７】
上記第１の処理エンジンは、上記第１の処理エンジンが第２の処理エンジンからのキープアライブメッセージの受信に失敗したとき、上記第２の処理エンジンを上記第１の処理エンジンに接続する第１のリンクが不活性であることをブロードキャストするように構成される請求項１３記載の装置。
【請求項１８】

上記第１の処理エンジンは、第１のキープアライブメッセージを第２の処理エンジンに送信するように構成され、
上記第１の処理エンジンは、上記第１のキープアライブメッセージに応答した上記第２の処理エンジンからのメッセージの受信に失敗したとき、上記第１の処理エンジンを上記第２の処理エンジンに接続する第１のリンクが不活性であることをブロードキャストするように構成される請求項１３記載の装置。
【請求項１９】
上記第１の処理エンジンは、第１のキープアライブメッセージを第２の処理エンジンに送信するように構成され、
上記第１の処理エンジンは、上記第１のキープアライブメッセージの受信に失敗したことを示すメッセージを上記第２の処理エンジンから受信したとき、上記第１の処理エンジンを上記第２の処理エンジンに接続する第１のリンクが不活性であることをブロードキャストするように構成される請求項１３記載の装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本開示は、フルメッシュネットワークにおいてデータを送信するための方法に関する。具体的には、本開示は、複数のリンク及び複数の中間ノードを介してデータの分配及び送信を行うことに関する。
【背景技術】
【０００２】
高速なシステムは、複数の処理エンジン間の完全な接続性を必要とする。１つの処理エンジン（「ノード」）は所定の処理能力を有する。処理エンジンは、典型的には、物理的位置に関して制限されている特定のハードウェア資源の集合、例えば、特定の回線カード、棚、又はラックに関連付けられる。
【０００３】
伝統的に、通信及びコンピュータの産業では、ノード間接続性、すなわち、２つの方法である（ｉ）スイッチ接続性及び（ｉｉ）フルメッシュ接続性のうちの１が使用される。スイッチ接続性の一例は、ノード間のスイッチングステージを使用するクロス（Clos）スイッチングである。他のノードにデータを送ろうとする各ノードは、スイッチにデータを送る。スイッチは、単一の宛先ノード又は複数の宛先ノード（ブロードキャスト又はマルチキャスト）のいずれかにデータを送る。スイッチ接続性の１つの欠点は、ノードの個数が増大するとき、スイッチのサイズ及び複雑さが増大するということにある。より大きな個数のノードを有するスイッチは、例えば、複数のノードから同じ宛先ノードへのトラフィックをスケジューリングする際にスイッチのオーバヘッド及び非能率性に適応するために、より多くの処理パワーを必要とする。
【０００４】
フルメッシュ接続性において、各ノードは他のすべてのノードと接続されている（ポイントツーポイントの接続性）。発信元ノードが宛先ノードにデータを送る場合、それは、宛先ノードに直接的に接続されたリンクを介してデータを送る。マルチキャストトラフィックの場合には、発信元ノードは、データをローカルに増加させて、宛先ノードに直接的に接続されたリンクを介して各宛先ノードにコピーを送る。フルメッシュ接続性は所定の欠点を有する。例えば、システムに追加のノードが追加される場合、すべての既存のノードからの少なくとも１つのリンクは互いに接続されていない状態にあり、新規なノードに差し込まれる。再度差し込まれるべきリンクの個数は、既存のノードの個数以上であるか、又は、新規なリンクが追加されなければならない。
【０００５】
典型的なフルメッシュ接続性において、各ノードが容量$C_N$及びＭ個のリンクを有するＮ個のノードのシステムは、各ノードが少なくとも（Ｎ－１）×$C_N$の総容量を提供可能であることを必要とする。この場合、$C_L$が単一のリンクの容量に対応するとき、リンク

の個数は少なくとも（Ｎ－１）×$C_N$／$C_L$であってもよい。さらに、いったんＮ個のノードがフルメッシュで接続されると、１つのノード当たりのリンクの合計個数Ｍが増大しない限り追加のノードをシステムに追加することは性能を低下させる。

【発明の概要】

【発明が解決しようとする課題】

【０００６】

本開示は、発信元ノードが複数の中間ノードにデータを送信し、複数の中間ノードが宛先ノードにデータを送るフルメッシュネットワークに関する。本開示はまた、フルメッシュネットワークにおいて複数の中間ノードを用いてデータを送信する方法に関する。

【課題を解決するための手段】

【０００７】

本開示の１つの態様において、本方法は、フルメッシュ分散型スイッチレス相互接続システムにおける第１のノードによって、宛先ノードを示すデータを受信することを含む。本方法はまた、上記第１のノードによって、上記第１のノードが上記宛先ノードであるか否かを決定することを含む。本方法はまた、上記第１のノードが上記フルメッシュ分散型スイッチレス相互接続システムにおける上記宛先ノードであるとき、上記第１のノードによって上記データを処理することを含む。上記方法はまた、上記第１のノードが上記宛先ノードでないとき、上記第１のノードによって、上記データを上記宛先ノードに送るか、又は上記データを上記フルメッシュ分散型スイッチレス相互接続システムにおける中間ノードに送ると決定することと、上記宛先ノードが上記第１のノードでないとき、上記第１のノードによる決定に基づいて、上記データを上記１のノード及び上記中間ノードのいずれかに送信することとを含む。

【０００８】

本開示のもう１つの態様では、上記データは、少なくとも１つのパケットを含む少なくとも１つのマスターセルを含んでもよい。上記少なくとも１つマスターセルは、キューシーケンス番号を有するヘッダを含んでもよい。データを処理することは、キューシーケンス番号に基づいて少なくとも１つのマスターセルを順序づけることと、少なくとも１つのパケットを抽出することと、上記少なくとも１つのパケットを、上記第１の処理エンジンに接続された処理装置に送ることとを含んでもよい。

【０００９】

本開示の態様では追加の特徴が現れてもよい。例えば、本方法は、上記第１の処理エンジンによって、少なくとも１つパケットを含む少なくとも１つのマスターセルを形成することをさらに含んでもよい。本方法は、リンク容量、処理エンジン能力、リンクの個数、及び中間処理エンジンの個数の少なくとも１つに基づいて、上記マスターセルのサイズを決定することをさらに含んでもよい。形成することは、タイムアウトパラメータが終了するまで、上記少なくとも１つのパケットを上記少なくとも１つのマスターセルに追加することを含んでもよい。形成することは、また、利用可能なパケットの量及び配送の緊急度に基づいて、上記少なくとも１つのパケットを上記少なくとも１つのマスターセルに追加することを含んでもよい。

【００１０】

本開示の態様では、本方法は、上記第１の処理エンジンによって、上記データから複数のマスターセルを形成することをさらに含んでもよい。上記第１の処理エンジンは、また、上記第１の処理エンジンに接続された処理装置から上記データを受信してもよい。送信することは、複数の異なるパケットを含む上記複数のマスターセルを、複数の中間処理エンジンに送信することを含んでもよい。

【００１１】

本開示のもう１つの態様では、本方法は、上記第１の処理エンジンが第２の処理エンジンからのキープアライブメッセージの受信に失敗したとき、上記第１の処理エンジンによって、上記第１の処理エンジンを上記第２の処理エンジンに接続する第１のリンクが不活性であることをブロードキャストすることを含んでもよい。本方法はまた、上記第１の処

理エンジンによって、第１のキープアライブメッセージを第２の処理エンジンに送ることをさらに含んでもよい。本方法は、上記第１の処理エンジンが、上記第１のキープアライブメッセージに応答した上記第２の処理エンジンからのメッセージの受信に失敗したとき、上記第１の処理エンジンによって、上記第１の処理エンジンを上記第２の処理エンジンに接続する第１のリンクが不活性であることをブロードキャストすることをさらに含んでもよい。代替として、本方法は、上記第１の処理エンジンによって、第１のキープアライブメッセージを第２の処理エンジンに送ることを含んでもよい。本方法は、上記第１の処理エンジンが、上記第１のキープアライブメッセージの受信に失敗したことを示すメッセージを上記第２の処理エンジンから受信したとき、上記第１の処理エンジンによって、上記第１の処理エンジンを上記第２の処理エンジンに接続する第１のリンクが不活性であることをブロードキャストすることをさらに含んでもよい。

【００１２】

本開示の態様では、フルメッシュ分散型スイッチレス相互接続システムにおける装置であって、上記装置は、処理装置と、上記処理装置に接続され、上記フルメッシュ分散型スイッチレス相互接続システムを介して、宛先処理エンジンを示すデータを受信するように構成された第１の処理エンジンとを備える。上記第１の処理エンジンは、上記第１の処理エンジンが上記フルメッシュ分散型スイッチレス相互接続システムにおける上記宛先処理エンジンであるとき、データを処理し、上記処理されたデータを上記処理装置へ配送するように構成される。一方、上記第１の処理エンジンは、上記第１の処理エンジンが上記宛先処理エンジンでないとき、上記宛先処理エンジンに、又は上記フルメッシュ分散型スイッチレス相互接続システムにおける中間処理エンジンに、上記データを送信するように構成される。

【００１３】

本開示のもう１つの態様では、上記第１の処理エンジンは、上記受信されたデータが上記処理装置から受信されたとき、上記受信されたデータに基づく少なくとも１つのパケットを含む少なくとも１つのマスターセルを形成するように構成されたアセンブリ装置を備えてもよい。

【図面の簡単な説明】

【００１４】

【図１】本開示の態様に係る、処理装置をネットワークに接続するノードを示す。

【図２】本開示の態様に係る、マスターセルを組み立てるフローチャートを示す。

【図３】本開示の態様に係る、受信されたマスターセルを処理するフローチャートを示す。

【図４】本開示の態様に係る、論理回路の並べ替えのフローチャートを示す。

【図５】本開示の態様に係る、マスターセルに対応するビットを有するシーケンス番号及びパケット終了部のデータベースを示す。

【図６】本開示の態様に係る、分散型スイッチレス相互接続システムにおける直接リンク及び間接リンクを介して発信元ノードから宛先ノードに伝搬するデータを示す。

【図７】本開示の態様に係る、分散型スイッチレス相互接続システムにおけるデータを受信して当該データを他のノードに送信するノードを示す。

【図８】本開示の態様に係る、異なる容量のリンクを有する分散型スイッチレス相互接続システムを示す。

【図９】本開示の態様に係る、ノード間に異なる個数のリンクを有する分散型スイッチレス相互接続システムを示す。

【図１０】本開示の態様に係る、切断されたリンクを有する分散型スイッチレス相互接続システムを示す。

【図１１】フルメッシュで接続されたシステムを示す。

【図１２】本開示の態様に係る、複数のノードを接続するＰａｓｓＣＯＭを示す。

【図１３】本開示の態様に係る、複数のノードを接続するＰａｓｓＣＯＭの内部リンク構成を示す。

【図１４Ａ】本開示の態様に係る、４つのノードを接続するように設計された、複数のプラグで２つのノードを接続するＰａｓｓＣＯＭの内部リンク構成を示す。
【図１４Ｂ】本開示の態様に係る、２つのノードを接続するように設計された、複数のプラグで２つのノードを接続するＰａｓｓＣＯＭの内部リンク構成を示す。
【図１５】本開示の態様に係る、２つのプラグで４つのノードを接続するＰａｓｓＣＯＭの内部リンク構成を示す。
【図１６】本開示の態様に係る、２つのフロントエンドコネクタ及び４つのバックエンドコネクタで４つのノードを接続するＰａｓｓＣＯＭの内部リンク構成を示す。
【図１７】本開示の態様に係る、複数のノードを接続する２つのＰａｓｓＣＯＭを示す。
【図１８】本開示の態様に係る、複数のノードを接続する２つのＰａｓｓＣＯＭの内部リンク構成を示す。
【発明を実施するための形態】
【００１５】
本開示の態様によれば、受動相互接続及び分散型スイッチレススイッチングを行うマルチシャーシルータが提供される。本システムは、複数のノードをフルメッシュで接続し、データの直接転送及び間接転送を可能にする。マルチシャーシルータは、マルチクラスタ計算環境における複数の計算プロセッサを接続することにも使用可能である。さらに、マルチシャーシルータは、計算プロセッサ及びストレージプロセッサを接続することができる。ルータは、フルメッシュネットワークのセットアップ又は更新の処理を簡単化するために受動接続性光学モジュール（Passive Connectivity Optical Module：「ＰａｓｓＣＯＭ」）を使用する。ＰａｓｓＣＯＭは、電子部品を含まない受動装置である。
【００１６】
物理的なノード構造及びノード機能
【００１７】
図１は、処理装置１０５をネットワークに接続するノード１００を示す。ノード１００は、処理装置１０５に接続され、リンク１１０を用いて他のノードに接続されている。リンク１１０を使用して、ノード１００は、処理装置１０５と、回線カード、棚、ラック、又は他の物理的位置にわたって分布した他の処理装置との間の接続性を提供する。
【００１８】
処理装置１０５はノード１００のクライアントである。通信システムにおいて、処理装置１０５は、例えば、外部インターフェース１０６に接続されたネットワークプロセッサーであってもよい。処理装置１０５は、インターフェース１０６から受信されたパケットを検査し、ルーティング及び／又はスイッチング動作に基づいて、受信されたパケットの宛先を決定する。宛先を決定するために、他のパケット情報、例えば、サービス品質（ＱｏＳ）、キューイング、及び変更が使用されてもよい。宛先は、ユニキャスト通信の場合は単一の処理装置になり、又はマルチキャスト通信の場合は複数の処理装置になる可能性がある。
【００１９】
処理装置１０５からのデータのパケットが到着したとき、パケットはスイッチング装置１０３に送られてもよく、次いで、それは他のノードにパケットを送る。この場合、パケットは他のノードに直接的に送られる。アセンブリ装置１０１は、他のノードに送られるパケットのヘッダに、パケットシーケンス番号を書き込んでもよい。代替として、パケットが到着したとき、アセンブリ装置１０１は、複数のパケットをマスターセルへ組み立てて、スイッチング装置１０３を介してマスターセルを他のノードに送ってもよい。アセンブリ装置１０１は、マスターセルのシーケンス番号をマスターセルのヘッダに書き込んでもよい。
【００２０】
図２は、マスターセルを組み立てるフローチャートを示す。アセンブリ装置１０１は、パケットを受信し（ステップ２１０）、それらの宛先に基づいてパケットを分割する（ステップ２２０）。アセンブリ装置１０１は、パケットのそれぞれを、同じ宛先のパケット

を含むマスターセルに追加する（ステップ２３０）。完全には満たされていない既存のマスターセルがある場合、アセンブリ装置１０１は、パケットをその既存のマスターセルに追加する。そうでなければ、アセンブリ装置１０１は、新規なマスターセルを形成し、満たされたマスターセルを解放して仮想的な出力キューイング（ＶＯＱ）に送る（ステップ２４０）。パケットがマスターセルの長さを超過する場合、アセンブリ装置は、マスターセルを満たし、満たされたマスターセルを送る。また、パケットの残りの部分は、次のマスターセルに配置される。次のマスターセルは次のパケットを待機する。

【００２１】

解放されたマスターセルは、外部メモリにおけるＶＯＱに書き込まれる（ステップ２５０）。ＶＯＱ論理回路は、１つの宛先ノード当たりに少なくとも１つのキューを保持する。特定の宛先へのバックプレッシャ（特定のリンクにおけるデータの強化）がある場合には、特定の宛先に関連する１つ又は複数のキューにおけるマスターセルの送信だけが停止される。他の宛先のためのキューにおける他のマスターセルを送り続けることができる。マスターセルを送る準備ができている場合、マスターセルは、メモリから読み出され（ステップ２６０）、他のノードに送られる（ステップ２７０）。ＶＯＱは、マスターセル生成処理によって独立して管理されてもよい。マスターセル生成処理では、アセンブリ装置が、各宛先のための現在のマスターセルを単に保持し、ＶＯＱ外部バッファが外部から管理される。その後、アセンブリ装置は、マスターセルのリリースレートに基づいてＶＯＱパケットからスイッチング装置１０３へ引く。

【００２２】

マスターセルは、パケットの部分的なペイロードを含んでもよい。この場合、アセンブリ装置１０１は、パケットの残りを次のマスターセルに追加する。アセンブリ装置１０１は、異なるサービスクラス（ＣＯＳ）をそれぞれ表す複数のキュー（各キュー）を宛先ノード毎に保持してもよい。

【００２３】

１つのサイズで複数のマスターセルを作成することができるので、マスターセルの使用は、効率的なスイッチ間のメモリ管理及びキューイング管理を可能にする。固定サイズのセルを使用することは、いくつかの性能上の利点を提供する。例えば、固定サイズのセルは効率的な処理及び外部メモリ管理を可能にするが、これは、ＤＲＡＭバンクの競合を回避しながら、比較的大きなブロックサイズでダイナミックランダムアクセスメモリ（ＤＲＡＭ）に書き込み、また読み出すことが、より効率的になるからである。典型的には、平均パケットサイズは、マスターセルのサイズに合わせられる最適なブロックサイズより小さい。さらに、リンクが小さな帯域幅を有する場合、大きなパケットは複数のマスターセルに分割されることが可能である。複数のマスターセルを送るために複数のリンクを使用することは、単一のレーンの使用に比較して削減された遅延及びジッタで、宛先ノードがパケット全体を並列に受信することを可能にする。さらに、多数の小さなパケットの場合に比較してディスクリプタの個数を少なくすることができるので、本来の処理装置のサイズが限定されている場合、キュー及びバッファの管理はより簡単になる。マスターセルの使用は、リンク負荷の平衡を保つために必要なアカウント処理を簡単化するが、これは、アカウント処理を可変パケットサイズではなく固定サイズセルに基づいて行うことができるからである。

【００２４】

マスターセルがフルになったとき、マスターセルはスイッチング装置１０３に配送される。スイッチング装置１０３は、マスターセルをその宛先に送るためにどのリンクを使用するのかを決定する。特定の宛先／ＣＯＳの組み合わせについてスイッチング装置１０３からのバックプレッシャがない場合、マスターセルは、その宛先に直ちに送られてもよい。

【００２５】

マスターセルのサイズは変動する可能性がある。例えば、マスターセルのサイズは、マスターセルを追加のパケットで満たすことから発生する過度の遅延を回避するために、優

先度が高いＣＯＳの場合には、より小さくすることができる。マスターセルのサイズは、ネットワークにおける利用可能な帯域幅に依存して変動することも可能である。利用可能な帯域幅は、ノードの個数及びリンクの個数によって決定可能である。アセンブリ装置１０１はまた、過度の遅延を回避するために、タイムアウトを使用し、部分的に満たされたマスターセルを送ってもよい。さらに、アセンブリ装置１０１は、相互接続リンクにおける不必要な帯域幅の消費を回避するために、部分的に満たされたマスターセルをリリースしてもよい。

【００２６】

スイッチング装置１０３は、他のノードからデータを受信することもできる。スイッチング装置１０３は、データの宛先がローカルな処理装置１０５であるか、それとも他のノードの処理装置であるかを決定する。最終的な宛先がノードの処理装置である場合、スイッチング装置１０３は、宛先ノード又は他の中間ノードにデータを送る。

【００２７】

スイッチング装置１０３がローカルな処理装置のためのデータを受信したとき、スイッチング装置１０３は、並べ替え及びデアセンブリ装置１０２へデータをわたす。並べ替え及びデアセンブリ装置１０２は、メモリにおける並べ替えキュー１０７にデータを格納する。データがマスターセルの形式を有している可能性がある場合、並べ替え及びデアセンブリ装置１０２は、複数のマスターセルを並べ替えて分解する。並べ替え及びデアセンブリ処理は、２つの別個の装置上で別個に動作しても、又は１つの装置上で動作してもよい。

【００２８】

図３は、並べ替え及びデアセンブリ装置１０２において、受信されたマスターセルを処理するフローチャートを示す。並べ替え及びデアセンブリ装置１０２がマスターセルを受信したとき（ステップ３１０）、並べ替え及びデアセンブリ装置１０２は、メモリにおける並べ替えキュー１０７にマスターセルを格納してもよい（ステップ３２０）。並べ替え及びデアセンブリ装置１０２は、マスターセルが最初に送信されたノードにおいて定義されたマスターセルシーケンス番号に基づいて、マスターセルをそれらの発信元によって分割する（ステップ３３０）。並べ替え及びデアセンブリ装置１０２は、受信されたマスターセルのマスターセルシーケンス番号を比較し、並べ替え論理回路を用いてそれらをキューに正しく配置し、並べ替えられたマスターセルをリリースする（ステップ３４０）。

【００２９】

図４は、ステップ３４０の並べ替え論理回路のフローチャートを示す。各マスターセルは、マスターセルシーケンス番号に基づいて決定された、その対応するビットを有する。並べ替え及びデアセンブリ装置１０２は、受信されたマスターセルに対応するビットを設定する（ステップ４１０）。ビットの設定は、ビットを０から１に変化させることで行われてもよい。これは、マスターセルが受信されていることを示す。マスターセルがパケット終了部（ＥＯＰ）を含む場合、並べ替え及びデアセンブリ装置１０２は、ＥＯＰを示すように追加のビットを設定する（ステップ４１０）。

【００３０】

図５は、マスターセルに対応するビットを有するシーケンス番号及びＥＯＰデータベースを示す。シーケンス番号データベース５００において、２つのタイプのビット、すなわち受信ビット及びＥＯＰビットがある。１つの実施形態において、第１の列におけるビットは、対応するマスターセルが到着したか否かを示す受信ビットである。第２の列におけるビットは、対応するマスターセルがＥＯＰを含むか否かを示すＥＯＰビットである。実施形態において、値０を有する受信ビットは、その対応するマスターセルが到着していないことを示し、値１を有する受信ビットは、その対応するマスターセルが到着したことを示す。値１を有するＥＯＰビットは、その対応するマスターセルがパケット終了部を含むことを示し、値０を有するＥＯＰビットは、そうでないことを示す。

【００３１】

受信ビット５０１はキューの先頭におけるビットであり、受信ビット５０４は、先行す

るビットが失われていない場合におけるキューの最後のビットである。ＥＯＰが受信ビット５０１及び受信ビット５０４の間に存在する場合、並べ替え及びデアセンブリ装置１０２は、ＥＯＰで終わるパケット全体が受信されたことを認識している。例えば、ＥＯＰビット５１３は、ＥＯＰが受信ビット５０１及び５０４の間にあることを示す。対照的に、ＥＯＰの前に０を有する受信ビットを有することは、パケットを含むすべてのマスターセルが到着していないことを示す。ＥＯＰビット５１７の前に、例えば、０を有する受信ビット５０５及び５０６がある。従って、ＥＯＰビット５１７の対応するマスターセルで終わるパケット全体は到着していない。

【００３２】

並べ替え及びデアセンブリ装置１０２は、受信ビットをチェックすることにより、パケットを含むすべてのマスターセルが到着しているか否かをチェックする（ステップ４２０）。パケットを含むすべてのマスターセルが到着していない場合、並べ替え及びデアセンブリ装置１０２は次のマスターセルを待機する（ステップ４３０）。パケットを含むすべてのマスターセルが到着している場合、並べ替え及びデアセンブリ装置１０２はマスターセルをリリースする（ステップ４４０）。並べ替え及びデアセンブリ装置１０２は、タイムアウトが終了するとき、マスターセルをリリースしてもよい。

【００３３】

リリースされたマスターセルは、調停論理回路を用いてデキュー（de-queue）される（ステップ３６０）。デキュー調停論理回路は、処理装置１０５に送られる準備ができているすべてのマスターセルのうちの最も優先度が高いマスターセルを選択する。並べ替え及びデアセンブリ装置１０２は、メモリからマスターセルを読み出し、マスターセルを元のパケットへ分解し（ステップ３７０）、その後、元のパケットは処理装置１０５に送られる（ステップ３８０）。並べ替え処理は、発信元／ＣＯＳの各組み合わせについて内部キュー１０７を保持してもよい。内部キューは、例えばノードに位置したバッファであってもよい。

【００３４】

システムは、パケットが他のパケットとともにマスターセルへ組み立てられることがない簡単化された実施形態を使用してもよい。パケットは複数のマスターセルに分割されなくてもよい。さらに、１つのマスターセルは１つのパケットを含んでいてもよい。この場合、並べ替え及びデアセンブリ装置１０２は、上述したものと同じ論理回路を使用するが、ＥＯＰビットをマークする必要はない。同じ論理回路が並べ替えのみを目的として使用される。

【００３５】

各マスターセルの到着時間は、並べ替え及びデアセンブリ装置１０２によって記録される。マスターセルが処理装置１０５に送られた後、並べ替え及びデアセンブリ装置１０２は、キューの先頭のシーケンス番号を、処理装置１０５に送られた最後のマスターセルのマスターセルシーケンス番号に設定する。マスターセルがキューの先頭よりも小さいマスターセルシーケンス番号で受信された場合、マスターセルは、そのデアセンブリの直後に処理装置１０５に送られる。

【００３６】

処理装置１０５は、トラフィック管理及びキューイングのようなＱｏＳ機能をさらに配備してもよい。処理装置１０５は、マルチキャストトラフィックのために、必要であれば、ローカルコピーを作成することができる。処理装置１０５はまた、受信されたパケットのフォーマットを変更し、それらをインターフェース１０６に送ってもよい。典型的なネットワークシステムにおいて、入力インターフェースにおけるパケットフォーマットは、ルーティングヘッダ及びＭＰＬＳラベルのような追加された情報に起因して、出力インターフェースにおけるパケットフォーマットとは異なる可能性がある。この場合、パケットフォーマット変更のうちの一部は、入力回線カードにおける処理装置によって扱われ、ヘッダ操作のうちの一部は、出力回線カードにおける処理装置において行われる。

【００３７】

ノード１００が分散型スイッチレス相互接続システムでの一時的輻輳に起因して処理装置１０５へのバックプレッシャ状態を示す場合、ノード１００は複数のＱｏＳ方法を実行してもよい。バックプレッシャは、すべてのトラフィックにおいて、又はトラフィックの一部において、特定宛先について、又は優先度毎に生じる可能性がある。

【００３８】

１つの例示的方法は、バックプレッシャが宛先毎に存在する否かにかかわらず、バックプレッシャが存在する場合には処理装置１０５からノード１００へのパケットの送信を停止することである。バックプレッシャ期間に輻輳するノードに送信側ノードがデータを送らないので、行頭ブロッキング（head-of-line blocking）が生じる可能性がある。行頭ブロッキングは、一部のノードへのトラフィックにおける輻輳に起因して、すべてのノードへのトラフィックが停止される状況である。これが生じるのは、キューの先頭にあるパケットが、バックプレッシャを経験し、パケットを受信できる他のノードへの次のパケットを阻止するからである。行頭ブロッキングは効率的でなく、多くの場合、はなはだ不適当である。

【００３９】

優先度毎にバックプレッシャが存在する場合、ノードは、優先度が高いトラフィックだけを許可し、優先度が低いトラフィックを停止してもよい。このアプローチは、それが処理装置１０５及び／又はアセンブリ装置１０１における少数のキューのみを必要とするので、有益である可能性がある。

【００４０】

もう１つの方法は、宛先ノード毎に複数のキューを使用することである。特定のノードへのルートにおける輻輳を示すバックプレッシャが受信されるとき、処理装置１０５及びアセンブリ装置１０１は、特定のノードにデータを送信することを停止することができる。処理装置１０５及びアセンブリ装置１０１は、ノードへの特定の優先度での送信を許可してもよい。例えば、優先度が高いトラフィックだけが許可されてもよい。この方法は、行頭ブロッキングを回避するが、より複雑なトラフィック管理を必要とする。

【００４１】

ＱｏＳ管理に関するノード１００及び処理装置１０５の間の論理的な分離は、例示のみを目的としている。代替の実施形態によれば、アセンブリ装置１０１及び処理装置１０５は、マスターセルの組み立て及びキューイングを組み合わせる単一の物理的装置であってもよい。単一の物理的な装置は、処理装置１０５における負荷を軽減することができる。

【００４２】

さらに、アセンブリ装置１０１、並べ替え及びデアセンブリ装置１０２、及びスイッチング装置１０３は、さまざまな種類のインターフェースを有する複数の物理的な装置から構成されていてもよい。各機能は、例えば、ＦＰＧＡとして、ＡＳＩＣとして、又は本開示に述べられた複数の論理的機能を利用する組み合わせとして、実装可能である。

【００４３】

複数ノードの接続

【００４４】

分散型スイッチレス相互接続システムは、物理的なフルメッシュ接続性を介してスイッチレスなスケーラブルノード接続性を可能にする。いくつかの実施形態において、システムは内部ノード間スイッチングを使用する。具体的には、システムは、１）発信元ノード及び宛先ノードを直接的に接続するリンクと、２）発信元ノードからデータを受信し、それを宛先ノードに向けて再ルーティングする中間ノードとを用いて、発信元ノードから宛先ノードへのデータ伝送を可能にする。

【００４５】

Ｎ個のノードをフルメッシュで接続するために、Ｍ個のリンクを有する各ノードは、それ自体のＭ個のリンクを他の（Ｎ－１）個のノードに分割してもよい。従って、システムにおいて接続可能であるノードの個数は、リンクの個数と１との和以下、すなわち、Ｎ≦Ｍ＋１である。リンクは、ノード間で均等に分割されていても、いなくてもよい。すべて

のノードが同じ容量を有する場合、各ノードは、等しい個数のリンクを用いて、他のノードに接続することができる。分散型スイッチレス相互接続システムは、ノード間において、光リンク、電子リンク、又はこれら２つの組み合わせを使用してもよい。

【００４６】

ノードは所定のノード容量、例えば$C_N$を有する。ノードに接続されたＭ個のリンクのそれぞれは、所定のリンク容量、例えば$C_L$を有する。$C_N$は、ノードから、通信又は他の処理を制御するその対応する処理装置への接続によって決定される。例えば、１０Ｇｂｐｓの１０個のインターフェースと、すべてのインターフェースを処理できる処理装置とを有するノードは、１００Ｇｂｐｓに等しい容量$C_N$を有する。

【００４７】

任意の２つのノード間に１つのリンクを有する典型的なフルメッシュシステムにおいて、ネットワークが完全な処理能力を扱うことを可能にするために、各リンクに係るノード容量に対するリンク容量の比率$C_L／C_N$は、１より大きい必要がある。そうでなければ、あるノードによる処理のレートは、あるリンクを用いた転送のレートを超過し、従ってトラフィックをブロックする。本開示の態様によれば、トラフィックは複数のリンクに分割されることが可能である。従って、ノードがＭ個のリンクを有する場合の通信システムは、$M\times C_L／C_N＞１$を満たすように設計可能である。比率$M\times C_L／C_N$は、ローカル過剰速度と呼ばれる。ローカル過剰速度が１より大きい場合、ネットワークは、処理装置の完全な処理能力を扱うことができる。

【００４８】

本開示の態様によれば、分散型スイッチレス相互接続システムは、$N\times C_N$個の可能な同時システム入力を有する、$N\times M\times C_L$の実効スイッチング容量を有する。

【００４９】

各ノードは、他のすべての発信元から、最大で$M\times C_L$の帯域幅までのピークトラヒックを受信することができる。$M\times C_L／C_N$は、ノードの物理インターフェース容量に対する、ピーク接続性の間の一時的に利用可能な過剰速度を表す。各発信リンクにおいて、ノードは、ローカルに発信されるデータを、他のノードから受信されたデータと集める。その後、ノードは、リンクの後方の他のノードにデータを送る。

【００５０】

図６は、本開示の態様に係る分散型スイッチレス相互接続システムの例を示す。図における各直線６１２は、２つのノードを接続する１つ以上物理リンクを表す。この実施形態において、発信元ノード６２０は宛先ノード６２４にデータを送る。発信元ノード６２０は、完全なシステム負荷が均等に分配されるように、すべてのアクティブリンク６１２間でトラフィックを均等に分配してもよい。すべてのノードがトラフィックを均等に分配する場合、グローバルなシステム負荷平衡を達成することができる。分散型スイッチレス相互接続システムは、システム内の部分的なバックプレッシャが存在する場合には、トラフィックを分配するために不足ラウンドロビン（Deficit Round Robin：ＤＲＲ）又は重みづけられたＤＲＲを使用することができる。

【００５１】

少なくとも１つのリンク、リンク６１１は、発信元ノード６２０及び宛先ノード６２４を直接的に接続する。他のリンクは、発信元ノード６２０及び宛先ノード６２４のいずれかを中間ノード、すなわちノード６２１～６２３及び６２５～６２７に接続する。中間ノードは、発信元ノード６２０からデータを受信し、ノード６２４を宛先ノードであると識別し、ノード６２４にデータを送る。中間ノードは、ノード６２４への直接のリンクを用いてデータを送ってもよい。この場合、データは２ホップを介して送られる。ホップは、あるノードから他のノードへの直接の転送を表し、従って、２ホップは１つの中間ノードを有することを表す。発信元ノードは中間ノードにパケットを送り、それは宛先ノードにパケットを送る。代替として、データが２つよりも多くのホップを用いて送られるように、中間ノードは他の中間ノードにデータを送ることができる。

【００５２】

図７は、分散型スイッチレス相互接続システムのノード７２６を示す。リンク７１２は、ノード７２６に他のノードを接続する。リンク７１２を介して、ノード７２６は、他のノードからデータを受信することができ、他のノードにデータを送ることができる。ノード７２６が他のノードからデータを受信する場合、ノード２０６は、発信元ノードの処理装置によって挿入されたパケットヘッダに示されたデータの宛先を決定する。パケットヘッダは宛先ノード番号を含む。

【００５３】

宛先が他のノードである場合、ノード７２６は、宛先ノードにデータを送る中間ノードとして機能する。ノード７２６は、発信元ノード７２０からデータを受信し、宛先ノード７２４にデータを送る。ノード７２６は、宛先ノード番号を識別することによって明示的に、又は、パケットが１ホップであることを識別することによって暗黙的に、パケットの宛先がローカルノードであると決定してもよい。それは、パケットヘッダに表示されてもよい。

【００５４】

本開示の実施形態によれば、分散型スイッチレス相互接続システムは選択的な負荷バランス化を実装する。例えば、負荷バランス化は、中間ノードを介してのみ宛先にデータを送ることで達成可能である。選択的な負荷バランス化は、それが異なる個数のリンクを通って流れたトラフィックから生じた到着時間の差を減少させるので、有益である可能性がある。選択的な負荷バランス化の例示的な方法において、いくつかのリンクは特定のトラフィック優先度に専用であってもよい。

【００５５】

本開示のもう１つの実施形態によれば、分散型スイッチレス相互接続システムは、まず、宛先ノードに直接的に送信することを決定し、直接のリンクに過負荷がかかっている場合のみ、中間ノードを使用してもよい。そのような選択的な負荷バランス化は、それが直接のリンク上で進むトラフィックの遅延を最小化するので、有益である可能性がある。それは、特定のトラフィック優先度に属するトラフィックにのみ使用されてもよい。

【００５６】

マルチレベル優先度は、各負荷バランス化エンティティにおいて、複数のノードにわたるバックプレッシャ及び輻輳を回避するために使用可能である。バックプレッシャは、グローバルなバックプレッシャ、宛先ノード毎のバックプレッシャ、及び宛先インターフェース毎のバックプレッシャのように、複数のレベルで発生する可能性がある。バックプレッシャは、優先度毎である可能性がある。グローバルなバックプレッシャは、転送ノードからスイッチングノードへのすべてのトラフィックを制御する。宛先ノードにおけるバックプレッシャは、特定のノードを宛先とするトラフィックを制御する。宛先インターフェース毎のバックプレッシャは、特定のノードの特定のインターフェースを宛先とするトラフィックを制御する。それは、優先度毎であってもよく、又は同様に、より高い粒度であってもよい。より高い粒度のＱｏＳの一例は、宛先ポート及びＣＯＳの組み合わせについてキューを使用することであり、又は、サービス毎のキューなど、より高い粒度であってもよい。

【００５７】

変動するリンク数及びリンク容量を有する分散型スイッチレス相互接続システム

【００５８】

本開示のいくつかの実施形態によれば、本システムは、異なる個数のリンク及び異なるリンク容量を有する複数のノードをサポートすることができる。図８は、異なる容量のリンクを有する分散型スイッチレス相互接続システムを示す。図８において、ノード８２０、８２５、８２６、及び８２７間の破線８１３は、より大きな容量を有するリンクを表す。より大きな容量のリンクを有するノードは、より大きな容量のリンクを用いた高速接続性を作成し、より小さな容量のリンクを用いたより低速の接続性を作成することができる。

【００５９】

図９は分散型スイッチレス相互接続システムを示す。図９において、ノード９２０、９２５、９２６、及び９２７間の二重の破線９１４は、より多数のリンクを表す。より多数のリンクを有するノードは、より少数のリンクを有する他のノードへの接続性を保持しながら、より大きな総容量を有する任意ノード間の接続性を作成することができる。

【００６０】

不活性リンクを回避するためのトラフィックの再ルーティング

【００６１】

発信元ノード及び宛先ノードの間のトラフィックが特定のリンクに限定されないので、所定のリンクが不活性であるか輻輳している場合であっても、本システムは動作可能である。トラフィックは、残りの機能するリンクを使用するように再ルーティングされることが可能である。本開示の実施形態によれば、本システムにおけるすべてのノードは、固定の間隔毎に、他のすべてのノードにキープアライブメッセージを送る。代替として、リンク障害は、他の手段、例えば、光リンクにおける光の喪失によって、同様に検出されることが可能である。

【００６２】

リンクに障害が発生したとき、又はリンクが輻輳したとき、ノードは、キープアライブメッセージの受信に失敗するであろう。受信側ノードは、受信側ノードがキープアライブメッセージの受信に失敗したとき、キープアライブメッセージを送信するノードにメッセージを送ってもよい。受信側ノードからメッセージは、リンクが不活性であると報告する。代替として、受信側ノードは、キープアライブメッセージの受信に肯定応答する返信メッセージを送ってもよい。送信側ノードが、最初のキープアライブメッセージに応答した返信メッセージを受信しないとき、送信側ノードはまた、リンクが不活性であることを検出する。

【００６３】

不活性なリンクが検出されたとき、２つのメッセージをブロードキャストすることができる。不活性なリンクの受信側のノードは、不活性なリンクを用いてデータを送るべきでないことを示すメッセージを、他のすべてのノードへブロードキャストしてもよい。以下で説明するように、不活性なリンクの送信側ノードはまた、不活性なリンクを用いてデータをわたすべきでないことを他のノードに示す同じメッセージを、他のすべてのノードにブロードキャストしてもよい。

【００６４】

不活性なリンクがアクティブになったとき、リンクの受信側ノード及び送信側ノードの両方は、リンクがアクティブであることを示す２つの別個のメッセージをブロードキャストしてもよい。リンク状態変化を示すメッセージを２つの異なる発信元から送信させることによって、本システムは、メッセージの配送の信頼性を増大する。

【００６５】

図１０は、不活性なリンクを有する分散型スイッチレス相互接続システムの例を示す。この例において、ノード１０２１からノード１０２２へのリンクは不活性である。まず、ノード１０２２は、ノード１０２１からキープアライブメッセージが受信されていないことを識別する。その後、ノード１０２２は、ノード３０１に、それらの間のリンクが不活性であることを示すメッセージを送る。ノード１０２２は、ノード１０２１を介してノード１０２２にデータを送るべきでないことを示すブロードキャストメッセージを、すべてのノードに送る。ノード１０２１はまた、同じブロードキャストメッセージを他のすべてのノードに送る。他のすべてのノードは、ノード１０２１を介してノード１０２２にデータを送ることを停止するように、それらの動的トラフィック分配論理回路を更新する。

【００６６】

動的トラフィック分配論理回路は、バックプレッシャメカニズムによって利用可能にされ、システムにわたるリンクの状態及びバッファの状態を用いて、システムにわたるトラフィックをできるだけ一様に分配する。典型的には、各発信元ノードが利用可能なリンクに基づいて各宛先にトラフィックをできるだけ等しく分配するとき、グローバルな負荷バ

ランス化が達成される。利用可能なリンクはと、直接のリンクと、中間ノードを介した間接のリンクとの両方を含んでもよい。しかしながら、動的なトラフィック分配論理回路は、より少数の間接のリンクが利用可能である場合、トラフィックの少ないほうの部分を間接のリンクを介して送る。動的トラフィック分配論理回路は、複数のルートを介して宛先に配送されたトラフィックの部分を、重み付けラウンドロビンを用いて制御する。ネットワークにおけるノードが他のノードにデータを送る場合、発信元ノードは、ノード１０２１～１０２２からリンクを使用することを回避し、他の機能するリンクを用いてデータを再分配する。

【００６７】

トラフィック分配論理回路は、すべてのリンクの利用可能性をリストするテーブルを保持してもよい。そのようなテーブルの一例は、１つの宛先ノード当たりに１行を含む。各行は、宛先のための中間ノードとして利用可能なノードのリストを、そのような各中間ノードのためのリンクのリストとともに含む。テーブルはまた、各宛先のための直接的に接続されたリンクのリストを含んでもよい。テーブルにおけるすべてのリンクは、その発信元ノード及び宛先ノードでマーキングされてもよい。このテーブルは、宛先への直接のリンク及び間接のリンクの両方を含んでもよい。

【００６８】

テーブルにおける各リンクの状態は、上述のように自動的に、又はコントロールプレーンによって手動で更新可能である。コントロールプレーンは、ノードのＣＰＵ上で実行される処理であり、システムの動作を制御する。コントロールプレーンのための入力は、システムの状態についてのオペレータ構成コマンド及びハードウェア表示であってもよい。自動更新は、システムにおけるエラーを検出して修復するのに有用である可能性があり、手動更新はメンテナンスのために有用である可能性がある。

【００６９】

特定のリンク上でメンテナンスが予想される場合、コントロールプレーンはリンクをディセーブルにするかイネーブルにしてもよい。メンテナンス又はアップグレードを行うとき、オペレータは、テーブルを手動で更新することで、特定のリンクを手動でディセーブルにしてもよい。メンテナンスを完了した後、オペレータは、テーブルを更新し、リンクをイネーブルにすることができる。複数のリンクをアップグレードする必要がある場合、リンクは、上述のシーケンスに従って１つずつ置き換えられてもよい。この機能は、追加のノードを導入する処理を簡単化する可能性がある。

【００７０】

ＰａｓｓＣＯＭ

【００７１】

本開示の態様によれば、受動接続性光学モジュールが提供される。ＰａｓｓＣＯＭは、複数のノードをフルメッシュで接続すること、及び、ノードを既存のフルメッシュネットワークに追加することを簡単にする。ＰａｓｓＣＯＭの物理的形状又は接続インターフェースは従来のスイッチに類似しているかもしれない。しかしながら、従来のスイッチと異なり、ＰａｓｓＣＯＭは、電子部品を含まない受動装置である。

【００７２】

図１１において、６つのノード１１２１～１１２６がフルメッシュで接続されている。フルメッシュの接続性は、Ｎがノードの個数であるとき、少なくともＮ（Ｎ－１）／２個のリンク１１５０を必要とする。フルメッシュにおけるノードをＮ（Ｎ－１）／２個のリンクに接続させることは、システムを物理的にセットアップし、ノードを既存システムに追加することを困難にする可能性があった。

【００７３】

図１２は、ＰａｓｓＣＯＭを用いて複数のノードを接続するシステムの例を示す。ＰａｓｓＣＯＭ１２０１はノード１２２１～１２２６を接続する。すべてのリンク群１２６０は、単一のノードをＰａｓｓＣＯＭ１２０１に接続する。ループバックループ、又はそれ自体へのループを含めて、この構成は、ノード及びＰａｓｓＣＯＭ１２０１の間のリンク

群１２６０毎に、Ｎ個のリンクを必要とする。例えば、図１２において、各リンク群１２６０は５つのリンクを含む。要するに、システムには少なくともＮ$^2$個のリンクがある。全体的な帯域幅及び弾力性を改善するために、１つのリンク群当たりのＮ個よりも多くのリンクを使用することが可能である。しかしながら、すべてのリンクがＰａｓｓＣＯＭ１２０１の形式を有する中央のハブに接続されているので、リンクの物理的配列はより簡単である。

【００７４】

ＰａｓｓＣＯＭは、Ｎ個のフロントエンドコネクタと、Ｋ個のプラグと、Ｋ個のバックエンドコネクタと、フロントエンドコネクタ及びバックエンドコネクタを接続する内部光ファイバーとを含むことができる。図１３は、４つのノード１３２１～１３２４を接続するＰａｓｓＣＯＭ１３０１を示す。ＰａｓｓＣＯＭ　１３０１は４つのフロントエンドコネクタ１３３１～１３３４を含んでいる。また、外部リンクグループ１３６０は、それらの対応するフロントエンドコネクタにノードを接続する。内部リンク１３７０は、フロントエンドコネクタ１３３１～１３３４のそれぞれを、バックエンドコネクタ１３８１～１３８４のそれぞれに接続する。バックエンドコネクタ１３８１～１３８４は、置き換え可能なプラグ１３４１～１３４４を受ける。置き換え可能なプラグ１３４１～１３４４から構成されたプラグセットは、識別されたフロントエンドコネクタへの接続を提供する。

【００７５】

プラグ１３４１はループバック接続を提供する。ある場合には、あらゆる可能な経路にわたって同じ遅延を維持することが重要である。ループバック接続性は、すべての経路を正確に同じにすることができる。ノードは、それ自体にデータを内部的に送るか、又はＰａｓｓＣＯＭを介してデータを送ることができる。ノードがそれ自体にデータを内部的に送るとき、ノードは、それ自体にデータを送ることのための別個の論理回路を実装しなければならない。さらに、データを内部的に受信するための待ち時間は、ＰａｓｓＣＯＭを介してデータを受信するための待ち時間とは異なる。さらに、内部ループは、潜在的な輻輳を扱うための、追加の電線、マルチプレクサ、デマルチプレクサー、及びメモリを必要としてもよい。従って、すべてのデータをＰａｓｓＣＯＭに送り、ＰａｓｓＣＯＭによりループバック接続を通じて返送するようにルーティングさせるほうが簡単である可能性がある。プラグ１３４２は、２つのフロントエンドコネクタ１３３１及び１３３２を接続し、２つのフロントエンドコネクタ１３３３及び１３３４をさらに接続する。プラグ１３４３は、２つのフロントエンドコネクタ１３３１及び１３３３を接続し、２つのフロントエンドコネクタ１３３２及び１３３４をさらに接続する。同様に、プラグ１３４４は、２つのフロントエンドコネクタ１３３１及び１３３４を接続し、２つのフロントエンドコネクタ１３３２及び１３３３をさらに接続する。

【００７６】

本開示の態様によれば、各フロントエンドコネクタは、バックエンドコネクタのそれぞれに接続されている。このように、すべてのノードは、最もバランスのとれた方法で、フルメッシュで接続されることが可能である。例えば、１つのプラグが抜かれるとき、各ノードは、同じ個数のリンク接続を失う。ＰａｓｓＣＯＭ１３０１について、与えられたどのプラグが除去される場合でも、各ノードは１つの接続を失う。従って、各ノードについてバランスのとれた帯域幅低下が生じる。

【００７７】

プラグの切断に起因して物理的な接続性は部分的に失われるが、論理的な接続性（すなわち、ノード間でパケットを送る能力）は、２ホップスイッチング方法に起因して、なお存在する。

【００７８】

本開示の態様によれば、Ｋ個のプラグ１３４１～１３４４のプラグセットは、複数のノードをフルメッシュトポロジーで物理的に接続する。ノードからノードへのすべての接続は、等しい帯域幅を有してもよく、又は異なる帯域幅を有してもよい。さらに、各接続は、一対のノードを接続するために同じ個数のリンク１３７０又は異なる個数のリンクを使

用してもよい。

【００７９】

ＰａｓｓＣＯＭ１３０１は、同じ個数のフロントエンドコネクタ、バックエンドコネクタ、及びプラグを有する。しかしながら、ＰａｓｓＣＯＭはそのような構成に限定されない。ＰａｓｓＣＯＭは、任意個数のフロントエンドコネクタ、バックエンドコネクタ、及びプラグを有してもよい。

【００８０】

実施形態において、本システムにおけるすべてのノードからの受信する側リンク及び送信側リンクの束は、フロントエンドＰａｓｓＣＯＭコネクタに接続されている。プラグセットは、１つのノードからの１つ以上の送信側リンクを、もう１つのノードにおける同じ個数の受信側リンクに接続する。受信側リンクと及び送信側リンクは光ファイバーであってもよい。内部リンクもまた光ファイバーであってもよい。受信側リンク及び送信側リンクはＫ個のグループに分割される。また、内部リンクは、受信側リンク及び送信側リンクの各グループを、その対応するＫ個のプラグへ接続する。

【００８１】

ＰａｓｓＣＯＭは、各ノードが他のノードにデータを送ることができる、内部ノード間スイッチングシステムとともに使用可能である。発信元ノードは特定の送信側リンクを介してデータを送ってもよく、それは他のノードの受信側のリンクに接続されている。

【００８２】

ＰａｓｓＣＯＭの１つの利点は、その簡単なアップグレード処理である。ノードの個数を増大する場合、オペレータは、新規なノードからＰａｓｓＣＯＭのフロントエンドコネクタにリンクを接続することで、既存のＰａｓｓＣＯＭに新規なノードを接続することができる。すべてのリンクが個々のノードにではなくＰａｓｓＣＯＭに接続されているので、配線の処理は簡単である。プラグセットを変化させることがＰａｓｓＣＯＭの接続性を変化させるので、内部リンクは静的に接続される。言いかえれば、内部リンクは同じままである。

【００８３】

アップグレード処理は、異なる個数のノードを接続するために異なるプラグセットをＰａｓｓＣＯＭが使用してもよいので、プラグを置き換えることを必要としてもよい。任意の２つのノードを接続する、より多くのリンクを有することは、システムの帯域幅を増大させることができる。フルメッシュ接続性において、比率$M \times C_L / C_N$はローカルの過剰速度である。ここで、Ｍはシステムにおける利用可能なリンクの個数であり、$C_L$はリンクの容量であり、$C_N$は、ノードの容量（すなわち、処理能力及び平均通過容量）である。ローカルの過剰速度が１より大きい場合、システムは、そのノードのフル処理能力を扱うことができる。ローカルの過剰速度を、１より大きいしきい値数より高く維持することが望ましい。

【００８４】

ローカルの過剰速度が１より大きくなるように、リンク容量が大きいか、ノード容量が小さい場合、部分的なトラフィック損失のみで、又は無損失で、システム規模（すなわちノードの個数）のアップグレード又はダウングレードを行うことができる。このトラフィック損失は、Ｍの値に依存する。

【００８５】

図１４Ａは、図１３で用いたプラグ１３４１～１３４４を有するプラグセットと同じものであるが、２つのノードを接続するプラグセット、すなわちプラグ１４４１～１４４４を有するＰａｓｓＣＯＭを示す。２つのノード１４２１及び１４２２は、２つのフロントエンドコネクタ１４３１及び１４３２にそれぞれ接続されている。プラグ１４４１～１４４４の中で、プラグ１４４２だけが、ノード１４２１及びノード１４２２の間の接続を提供する。さらに、プラグ１４４１はループバック接続を提供する唯一のプラグである。

【００８６】

対照的に、図１４Ｂは、異なるプラグセット、すなわちプラグ１４４６～１４４９を有

するＰａｓｓＣＯＭを示す。この例において、２つのノード１４２１及び１４２２は２つのリンクによって互いに接続され、各ノードは２つのループバックリンクを有するように、プラグセットは定義される。２つのノード１４２１及び１４２２は、２つのフロントエンドコネクタ１４３１及び１４３２にそれぞれ接続されている。プラグ１４４６～１４４９の中で、２つのプラグ１４４８及び１４４９は、帯域幅を２倍にして、ノード１４２１及びノード１４２２を接続する。この設定内容において、２つのプラグ１４４６及び１４４７はループバック接続を可能にする。

【００８７】

提示した例は、２つのフロントエンドコネクタＮ＝２及び４つのバックエンドコネクタＫ＝４を有するＰａｓｓＣＯＭを含む。しかしながら、ＮがＫ未満である場合、Ｎ及びＫの任意の組み合わせが可能である。ＮがＫより小さな場合、又はノードの個数がバックエンドコネクタの個数より小さい場合であっても、図１４Ｂに示すように、１つより多くのリンクによって複数ペアのノードを接続することで、同じ全体相互接続帯域幅を達成することができる。

【００８８】

ノード間で必要とされる最小帯域幅に依存して、より少ない個数の接続で複数のノードを接続することが可能である。例えば、１つのリンクの帯域幅が１４２１及び１４２１の接続に十分である場合、１４４１～１４４４を使用することは十分な帯域幅を提供するだろう。従って、ＰａｓｓＣＯＭは、プラグ１４４６～１４４９を有するプラグセットを使用する必要はない。

【００８９】

既存のノード１４２１及び１４２２に加えて、プラグ１４４６～１４４９を有するＰａｓｓＣＯＭ１４０１にノードを追加するとき、オペレータはプラグを変更するであろう。プラグ１４４６～１４４９を有するプラグセットは２つのノードを接続するように設計されているので、フロントエンドコネクタ１４３３及び１４３４は、フロントエンドコネクタ１４３１及び１４３２に接続されない。その結果、新たに導入されるノードは、既存のノード１４２１及び１４２２に接続されない。代わりに、ノードを追加する場合、プラグ１４４１～１４４４を有するプラグセットが使用可能である。

【００９０】

ＰａｓｓＣＯＭ及び分散型スイッチレススイッチングを備えたルータ

【００９１】

分散型スイッチレス相互接続システムにおいてＰａｓｓＣＯＭを使用する利益は、間接のデータ転送が可能にされている場合、トラフィックの大きな中断なしでシステムをアップグレードすることができる、ということにある。まず、オペレータは、置き換えられるプラグに接続されたリンクを使用することを停止するためのコマンドを、すべてのノードに送ってもよい。オペレータがプラグを置き換えた後、オペレータは、リンクを使用することを開始するためのコマンドを、すべてのノードに送ってもよい。

【００９２】

従来のフルメッシュネットワークにおいて、アップグレードされる接続を使用するトラフィックは、アップグレード処理の間に停止される。しかしながら、マルチホップ転送を可能にする、提案された分散型スイッチレスシステムにおいて、トラフィックは流れ続けることができる。まず、新規なノードがＰａｓｓＣＯＭのフロントエンドコネクタに接続される。ＰａｓｓＣＯＭにおけるプラグが置き換えられるとき、置き換えられているプラグを使用するトラフィックは、中間ノードとして新規なノードを介して再ルーティングされる。

【００９３】

ＰａｓｓＣＯＭは、より少数の大きなプラグ（すなわち、より多数の接続を有するプラグ）、より多数の小さなプラグ（すなわち、より少数の接続を有するプラグ）、又は小さなプラグ及び大きなプラグの組み合わせとともに、動作することができる。プラグのサイズにかかわらず、同じ接続性を達成することができる。しかしながら、システムアップグ

レード（すなわち、ノードの追加又は除去）の間に、異なるサイズを有するプラグを使用することのトレードオフがある。プラグが除去される場合、システムは、帯域幅の１／Ｋ$^{th}$部分を失う。プラグが利用不可能である間、システムは、フル帯域幅の（Ｋ－１）／Ｋ$^{th}$部分を使用する。大きなプラグが使用される場合、又は、プラグの個数Ｋが小さい場合、帯域幅の削減は大きくなる。しかしながら、より多くのリンクを接続するために大きなプラグを使用することは、アップグレードを完了するために必要なステップ数を少なくする。小さなプラグを使用する場合、反対のことがあてはまる。小さなプラグを交換する場合、帯域幅の削減は小さいが、システムをアップグレードするためのステップ数は、より多数の動作が必要とされることに起因してより大きくなる。

【００９４】

図１５は、大きなプラグ１５４１及び１５４２を有するＰａｓｓＣＯＭ１５０１を示す。接続は、図１３のプラグ１３４１～１３４４を有するＰａｓｓＣＯＭ１３０１と同一であるが、プラグ１５４１及び１５４２の個々は、個別のプラグ１３４１～１３４４の２倍のリンクを接続する。図１５に示すように、大きなプラグは、より多数の接続のための差し込みとなる可能性がある。プラグ１５４１はバックエンドコネクタ１５８１及び１５８２に差し込まれ、プラグ１５４２はバックエンドコネクタ１５８３及び１５８４に差し込まれる。この例示において、２つのバックエンドコネクタは１つのプラグを受ける。しかしながら、２つよりも多くのバックエンドコネクタが１つのプラグを受けてもよい。

【００９５】

図１６は、４つのノード１６２１、１６２２、１６２３、及び１６２４を接続する２つのフロントエンドコネクタ１６３１及び１６３２を有するＰａｓｓＣＯＭ１６０１を示す。この場合、各コネクタは２つのノードを接続する。しかしながら、本開示は図に示す構成に限定されない。フロントエンドコネクタ及びノードの任意の組み合わせが可能である。提示しているように、ノードの個数は、フロントエンドコネクタの個数より小さくされてもよい。代替として、ノードの個数は、フロントエンドコネクタの個数より大きくされてもよい。

【００９６】

ＰａｓｓＣＯＭの故障確率は、それが受動素子を有するので低い。しかしながら、ＰａｓｓＣＯＭの障害からシステムをさらに保護するために、接続性を複数のＰａｓｓＣＯＭへ分割し、最大でプラグの個数まで分割することができる。実施形態において、すべてのノードは２つ以上のＰａｓｓＣＯＭに接続されている。図１７は、ＰａｓｓＣＯＭ１７０１及び１７０２に接続されたノード１７２１～１７２６を示す。

【００９７】

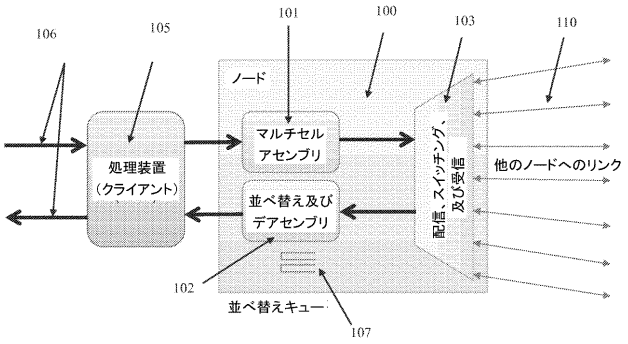
図１８は、ＰａｓｓＣＯＭ１８０１ａ及び１８０１ｂにおけるフロントエンドコネクタ及びプラグの間でリンクがどのように接続されているかについての論理図を示す。各フロントエンドコネクタは、２つのＰａｓｓＣＯＭからの２つのフロントエンドコネクタに接続されている。例えば、ノード１８２１、１８２２、１８２３、及び１８２４は、２つのフロントエンドコネクタ１８３１ａ及び１８３１ｂ、１８３２ａ及び１８３２ｂ、１８３３ａ及び１８３３ｂ、及び１８３４ａ及び１８３４ｂにそれぞれ接続されている。ＰａｓｓＣＯＭ１８０１ａにおける内部リンクは、フロントエンドコネクタ１８３１ａ～１８３４ａをバックエンドコネクタ１８８１ａ～１８８４ａに接続する。同様に、ＰａｓｓＣＯＭ１８０１ｂにおける内部リンクは、フロントエンドコネクタ１８３１ｂ～１８３４ｂをバックエンドコネクタ１８８１ｂ～１８８４ｂに接続する。

【００９８】

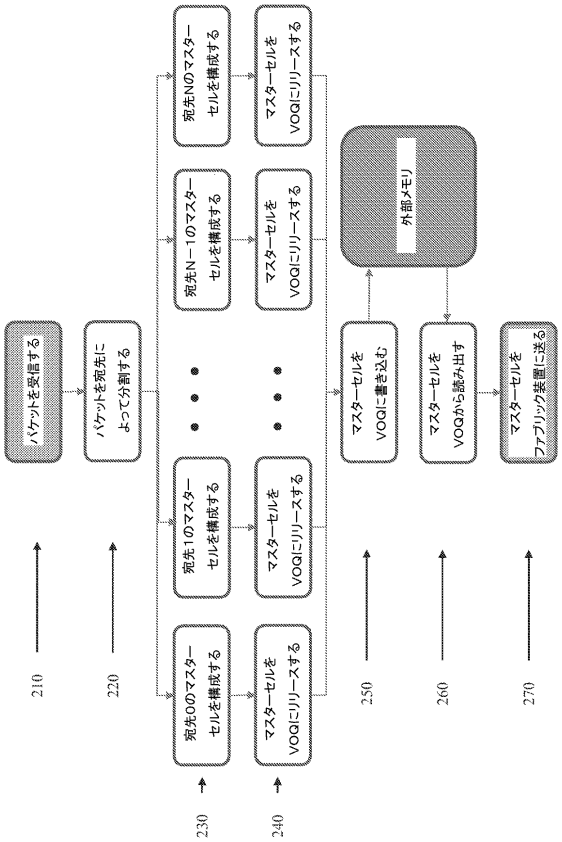
本開示の読み取りから当業者に明らかになるように、本開示は、上記で具体的に開示されたものとは異なる形式で実施することができる。従って、上述した特定の実施形態は、例示であって、限定的でないものとして考慮されるべきである。当業者は、通常の実験を越えるものを使用することなく、本明細書で説明した特定の実施形態への多数の等価物を認識するか又は確認することができるであろう。本発明の範囲は、先の説明に含まれていた例示に限定されるものではなく、添付された特許請求の範囲及びその等価物に示される

。

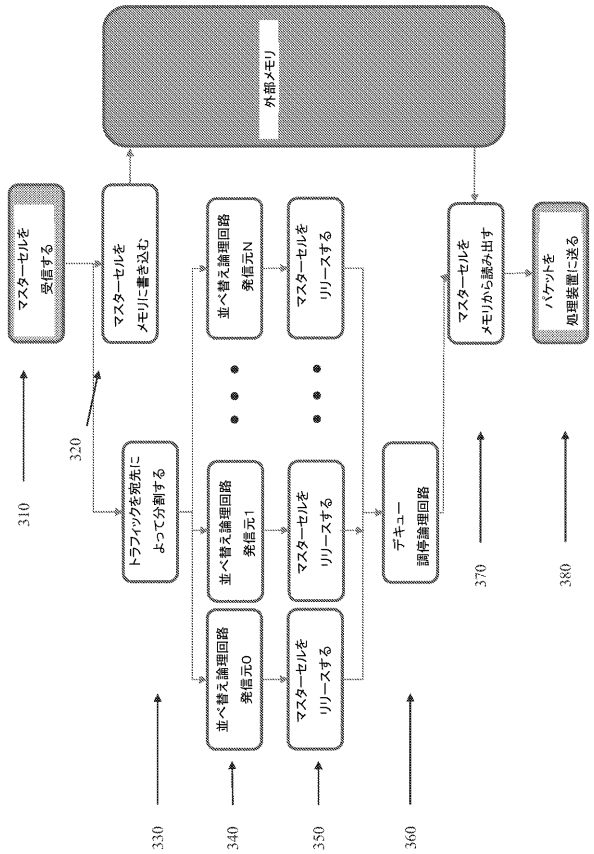
【図１】


106
105
101
100
103
110
ノード
処理装置
（クライアント）
マルチセル
アセンブリ
並べ替え及び
デアセンブリ
配信、スイッチング、
及び受信
他のノードへのリンク
102
並べ替えキュー
107

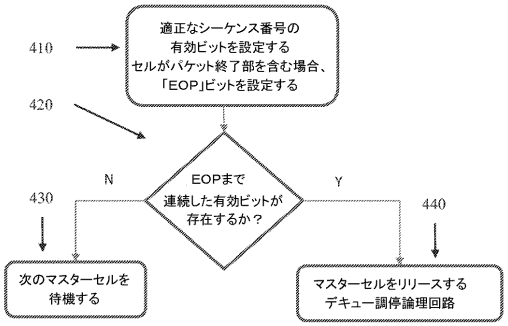
【図２】


パケットを受信する
パケットを宛先に
よって分割する
宛先0のマスター
セルを構成する
宛先1のマスター
セルを構成する
宛先N－1のマスター
セルを構成する
宛先Nのマスター
セルを構成する
マスターセルを
VOQにリリースする
マスターセルを
VOQにリリースする
マスターセルを
VOQにリリースする
マスターセルを
VOQにリリースする
外部メモリ
マスターセルを
VOQに書き込む
マスターセルを
VOQから読み出す
マスターセルを
ファブリック装置に送る
210
220
230
240
250
260
270

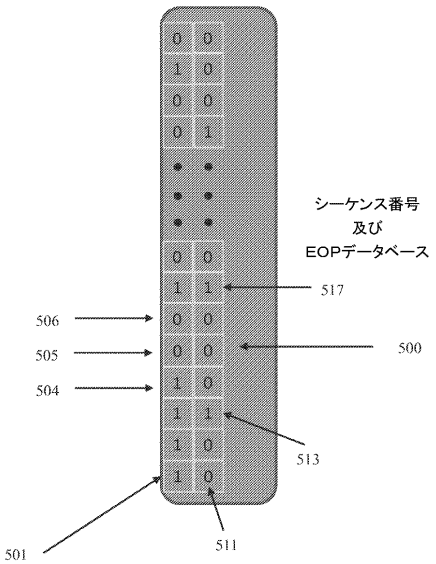
【図３】


外部メモリ
マスターセルを受信する
マスターセルをメモリに書き込む
並べ替え論理回路 発信元N
マスターセルをリリースする
マスターセルをメモリから読み出す
パケットを処理装置に送る
トラフィックを宛先によって分割する
並べ替え論理回路 発信元1
マスターセルをリリースする
デキュー調停論理回路
並べ替え論理回路 発信元0
マスターセルをリリースする
310
320
330
340
350
360
370
380

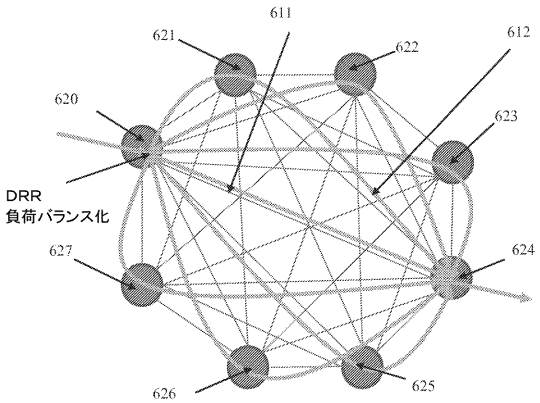
【図４】


410
適正なシーケンス番号の有効ビットを設定する
セルがパケット終了部を含む場合、「EOP」ビットを設定する
420
EOPまで連続した有効ビットが存在するか？
N
Y
430
440
次のマスターセルを待機する
マスターセルをリリースするデキュー調停論理回路

【図５】


シーケンス番号及びEOPデータベース
517
500
513
511
501
506
505
504

【図６】


611
621
622
612
620
623
DRR
負荷バランス化
627
624
626
625

【図７】

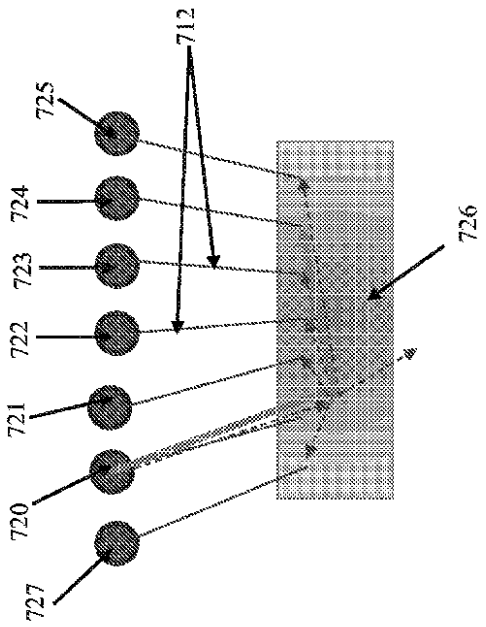



FIG. 7

【図８】

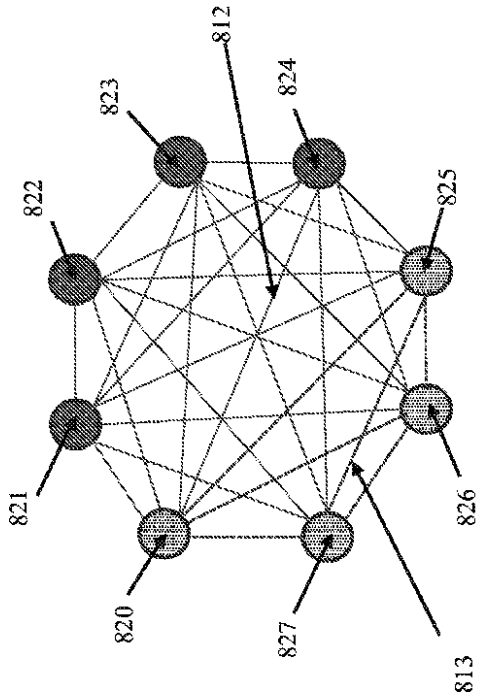



FIG. 8

【図９】

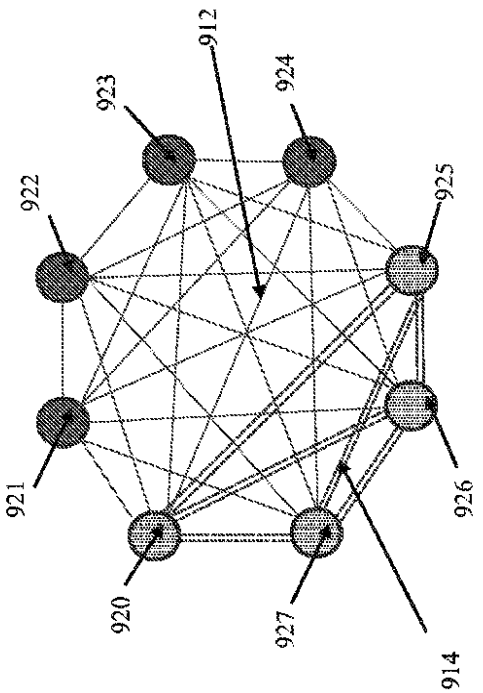



FIG. 9

【図１０】

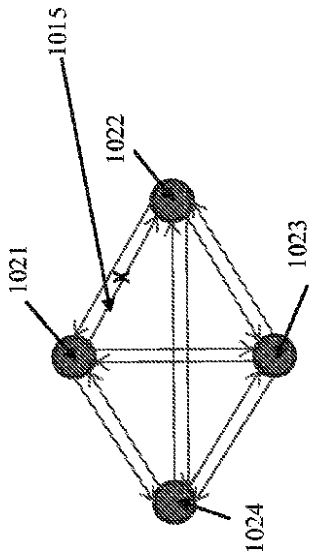



FIG. 10

【図１１】

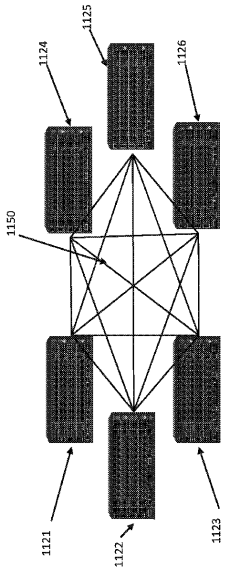



FIG. 11

【図１２】

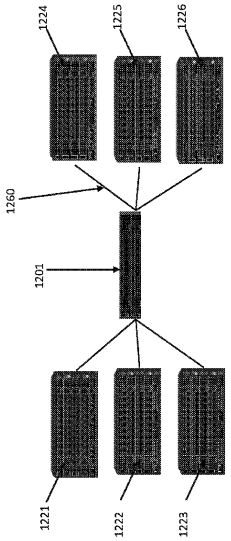
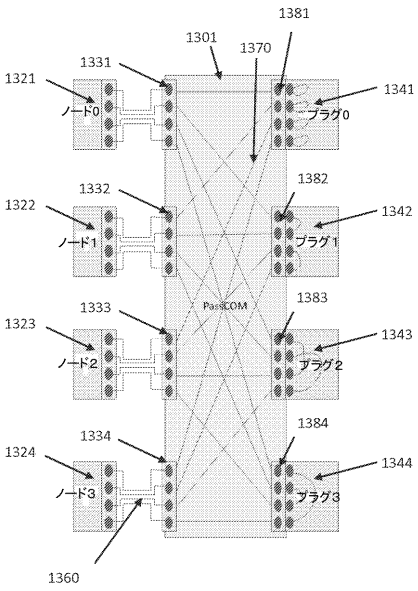
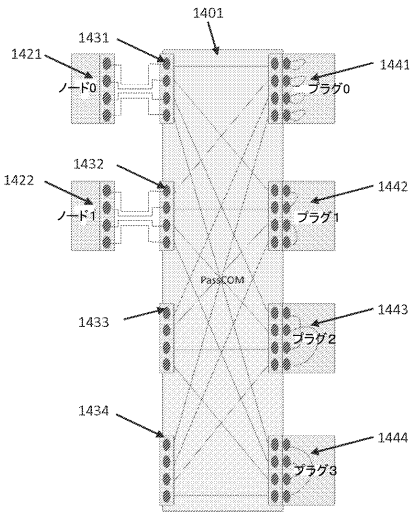



FIG. 12

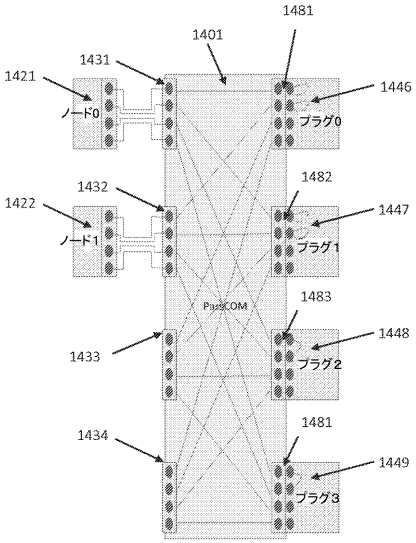
【図１３】



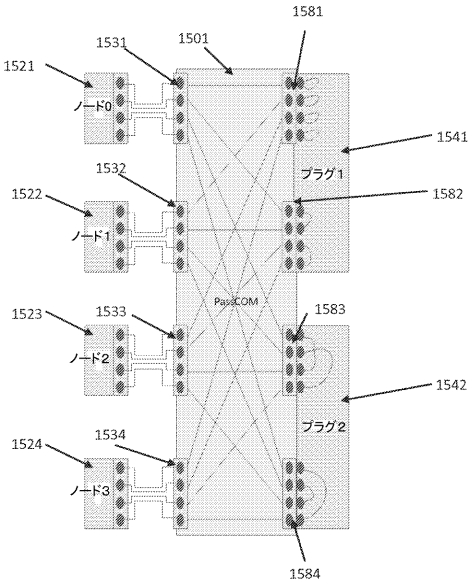

【図１４Ａ】



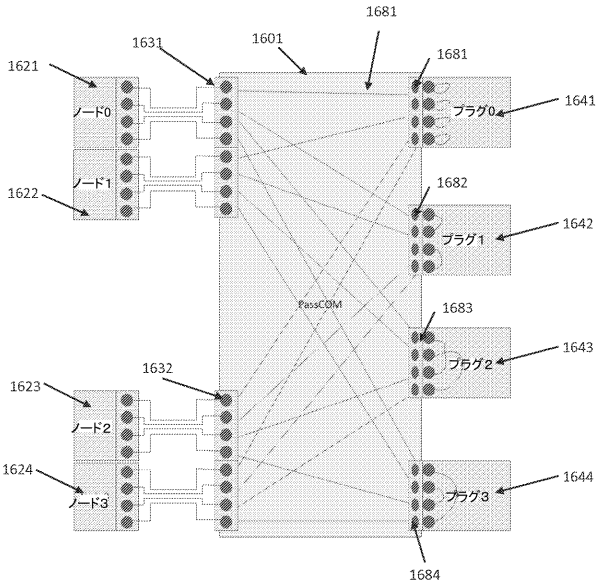
【図１４Ｂ】




【図１５】




【図１６】




【図１７】

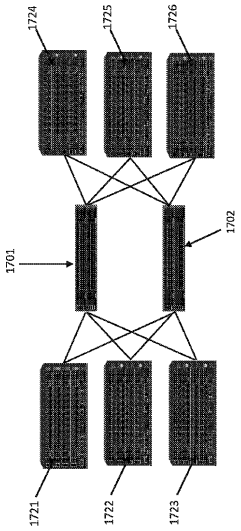



FIG. 17

【図１８】

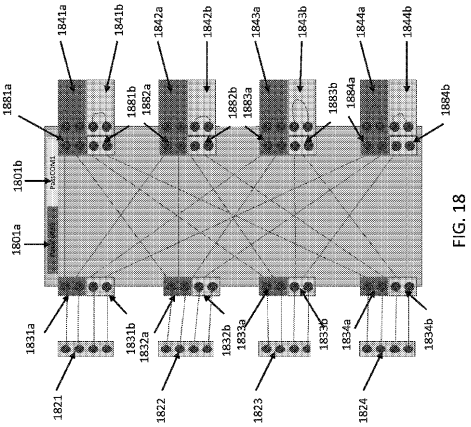

1841a
1841b
1842a
1842b
1843a
1843b
1844a
1844b
1881a
1881b
1882a
1882b
1883a
1883b
1884a
1884b
1801b
1801a
1831a
1831b
1832a
1832b
1833a
1833b
1834a
1834b
1821
1822
1823
1824

FIG. 18

【国際調査報告】

**INTERNATIONAL SEARCH REPORT**

International application No.
PCT/IB 13/03051

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - H04L 12/56 (2014.01)
USPC - 370/406
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
USPC: 370/406

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 370/371, 357, 400, 406

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Patbase; Google Patents; Google Scholar; Google Web
Search Terms Used: Mesh, full, switchless, distributed, processing, engine, node, destination, packet, master, cell, size, extract, keepalive, keep, alive, interconnect

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | US 2010/0061272 A1 (VEILLETTE) 11 March 2010 (11.03.2010), para [0161]-[0167], [0180]-[0181], [0187]-[0190], [0194]-[0195], [0226], [0266] | 1-19 |
| Y | US 2008/0068989 A1 (WYK et al.) 20 March 2008 (20.03.2008), para [0128]-[0129], [0142], [0282], [0586]-[0588], [0641]-[0642], [0732]-[0735], [0876]-[0878], [0982]-[0983] | 1-19 |
| A | US 6154449 A (RHODES et al.) 28 March 2000 (28.03.2000), see entire document | 1-19 |

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 17 June 2014 (17.06.2014) | 30 JUL 2014 |
| Name and mailing address of the ISA/US<br>Mail Stop PCT, Attn: ISA/US, Commissioner for Patents<br>P.O. Box 1450, Alexandria, Virginia 22313-1450<br>Facsimile No. 571-273-3201 | Authorized officer:<br>Lee W. Young<br>PCT Helpdesk: 571-272-4300<br>PCT OSP: 571-272-7774 |

Form PCT/ISA/210 (second sheet) (July 2009)

フロントページの続き

(81)指定国 AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CZ,DE,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,JP,KE,KG,KN,KP,KR,KZ,LA,LC,LK,LR,LS,LT,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,SM,ST,SV,SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US

(72)発明者 ウラディミール・ミラフスキー
イスラエル、ペター・ティクバ、グロート・アレクサンダー３番
(72)発明者 デイビッド・チェアマン
イスラエル４２８２３ゾラン、ハサヴィオニム１２番
(72)発明者 ニヴ・マルガリット
イスラエル４７２６７ラマト・ハシャロン、シヴティ・イスラエル２４番
(72)発明者 イフター・メイロン
イスラエル、キルヤット・オノ、ハパルデス８／２２番
(72)発明者 ダヴィド・ゼリグ
イスラエル３０９００ジフロン・ヤアコヴ、ハトメル・ストリート１９番
(72)発明者 アレクサンダー・ゼルツァー
イスラエル４２４９０ネタニヤ、カルトシット・ストリート３／１４番

Ｆターム(参考) 5K030 LB05 LB17