(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5404947号
(P5404947)

(45)発行日　平成26年2月5日(2014.2.5)　(24)登録日　平成25年11月8日(2013.11.8)

(51)Int.Cl.　　F I
G06F 3/06 (2006.01)　　G06F 3/06 301J
G06F 13/10 (2006.01)　　G06F 13/10 340B
　　G06F 3/06 301X

請求項の数 16　(全 59 頁)

(21)出願番号　特願2013-64768(P2013-64768)
(22)出願日　平成25年3月26日(2013.3.26)
(62)分割の表示　特願2010-252336(P2010-252336)の分割
　原出願日　平成22年11月10日(2010.11.10)
(65)公開番号　特開2013-145592(P2013-145592A)
(43)公開日　平成25年7月25日(2013.7.25)
　審査請求日　平成25年3月26日(2013.3.26)

早期審査対象出願

(73)特許権者　000003078
株式会社東芝
東京都港区芝浦一丁目１番１号
(74)代理人　100108855
弁理士　蔵田　昌俊
(74)代理人　100109830
弁理士　福原　淑弘
(74)代理人　100088683
弁理士　中村　誠
(74)代理人　100103034
弁理士　野河　信久
(74)代理人　100095441
弁理士　白根　俊郎
(74)代理人　100075672
弁理士　峰　隆司

最終頁に続く

(54)【発明の名称】ストレージ装置

(57)【特許請求の範囲】
【請求項１】
第１の入力ポートと、第１の不揮発性メモリと、前記第１の不揮発性メモリを制御する第１の制御部と、を含む第１のメモリノードと、
前記第１のメモリノードと接続される第２の入力ポートと、第２の不揮発性メモリと、前記第２の不揮発性メモリを制御する第２の制御部と、いずれか１つは前記第１の入力ポートと接続される複数の第１の出力ポートと、を含む第２のメモリノードと、
前記複数の第１の出力ポートのうちの１つに接続される第３の入力ポートを含み、前記第１乃至第２のメモリノードとは異なる複数の第３のメモリノードと、
を具備するストレージ装置であって、
前記第２のメモリノードは、前記ストレージ装置内における前記第２のメモリノードの位置情報である第１の位置情報を記憶し、
前記第１の制御部は、送信先アドレスと送信元アドレスを少なくとも含むパケットを前記第２のメモリノードへ出力し、
前記第２の入力ポートは、前記第１のメモリノードから出力された前記パケットを受信し、
前記第２の制御部は、前記パケット内の前記送信先アドレスと前記送信元アドレス、前記第１の位置情報を少なくとも含む情報に基づいて、少なくとも前記複数の第３のメモリノード及び前記第１のメモリノードの中から前記パケットの出力先を決定するように構成されているストレージ装置。

【請求項２】
第１の不揮発性メモリと、前記第１の不揮発性メモリを制御する第１の制御部と、を含む第１のメモリノードと、
前記第１のメモリノードと接続される第１の入力ポートと、第２の不揮発性メモリと、前記第２の不揮発性メモリを制御する第２の制御部と、複数の第１の出力ポートと、を含む第２のメモリノードと、
前記複数の第１の出力ポートのうちの１つに接続される第２の入力ポートを含み、前記第１乃至第２のメモリノードとは異なる複数の第３のメモリノードと、
前記第１乃至第３のメモリノードとは異なる１以上の第４のメモリノードと、
を具備するストレージ装置であって、
前記第１の制御部は、送信先アドレスと送信元アドレスを少なくとも含むパケットを前記第２のメモリノードへ出力し、
前記第１の入力ポートは、前記第１のメモリノードから出力された前記パケットを受信し、
前記第２の制御部は、前記第１のメモリノードと前記複数の第３のメモリノードと前記１以上の第４のメモリノードのうち、前記ストレージ装置内における位置情報が前記送信元アドレスと一致する第５のメモリノードと、前記ストレージ装置内における位置情報が前記送信先アドレスと一致する第６のメモリノードとを結ぶ第１の直線で区切られる２つの領域のうち第１の領域に前記第２のメモリノードが属する場合に、前記複数の第３のメモリノードのうち第１の方向に存在する第７のメモリノードに前記パケットを出力し、前記第１の領域とは異なる第２の領域に前記第２のメモリノードが属する場合に、前記複数の第３のメモリノードのうち第２の方向に存在する第８のメモリノードに前記パケットを出力するストレージ装置。
【請求項３】
前記第２の制御部は、前記ストレージ装置内における前記第２のメモリノードの位置情報が前記送信先アドレスと一致する場合に、前記パケットに含まれる情報を、前記第２の不揮発性メモリに記録する請求項２に記載のストレージ装置。
【請求項４】
前記第２の制御部は、前記パケット内の前記送信先アドレス、前記送信元アドレス、前記第１の位置情報、及び前記複数の第１の出力ポートにおける前記パケットとは別のパケットの出力に使用されているかを示す占有情報を少なくとも含む情報に基づいて、少なくとも前記複数の第３のメモリノード及び前記第１のメモリノードの中から前記パケットの出力先を決定する請求項１に記載のストレージ装置。
【請求項５】
前記第２の制御部は、前記複数の第１の出力ポートにおける前記パケットとは別のパケットの出力に使用されているかを示す占有情報を少なくとも含む情報に基づいて、前記パケットを、前記複数の第３のメモリノードのいずれかに出力する請求項２に記載のストレージ装置。
【請求項６】
前記第２の制御部は、前記第１の領域に前記第２のメモリノードが属する場合に、前記複数の第１の出力ポートのうち前記第７のメモリノードと接続される出力ポートが前記パケットとは別のパケットの出力用に使用されている場合、前記第８のメモリノードに前記パケットを出力し、前記第２の領域に前記第２のメモリノードが属する場合に、前記複数の第１の出力ポートのうち前記第８のメモリノードと接続される出力ポートが前記パケットとは別のパケットの出力用に使用されている場合、前記第７のメモリノードに前記パケットを出力する請求項２に記載のストレージ装置。
【請求項７】
前記第２の制御部は、前記第１の領域において、前記第１の直線と直行し、かつ前記第６のメモリノードを通過する第２の直線で区切られる領域であって、前記第５のメモリノードに近い領域を第３の領域、他方を第４の領域とし、前記第２の領域において、前記第

２の直線で区切られる領域のうち、前記第５のメモリノードに近い領域を第５の領域、他方を第６の領域とした場合に、前記第３の領域に前記第２のメモリノードが属する場合、前記第７のメモリノードに前記パケットを出力し、前記第４の領域に前記第２のメモリノードが属する場合、前記複数の第３のメモリノードのうち前記第２の方向と逆の方向に存在するメモリノードに前記パケットを出力し、

前記第５の領域に前記第２のメモリノードが属する場合、前記第８のメモリノードに前記パケットを出力し、前記第６の領域に前記第２のメモリノードが属する場合、前記複数の第３のメモリノードのうち前記第１の方向と逆の方向に存在するメモリノードに前記パケットを出力する請求項６に記載のストレージ装置。

【請求項８】

前記第１乃至第３のメモリノードとは異なる１以上の第４のメモリノードをさらに具備し、

前記第２の制御部は、前記第１のメモリノードと前記複数の第３のメモリノードと前記１以上の第４のメモリノードのうち、前記ストレージ装置内における位置情報が前記送信先アドレスと一致するメモリノードへの転送経路が最も短いメモリノードに前記パケットを出力する請求項１に記載のストレージ装置。

【請求項９】

前記第２のメモリノードは前記第１の入力ポートの他にコンピュータと接続可能な１以上の入力ポートを含む請求項１に記載のストレージ装置。

【請求項１０】

前記第１乃至第３のメモリノードとは異なる１以上の第４のメモリノードをさらに具備し、

前記複数の第３のメモリノードのうち前記第２のメモリノードに隣接しない第１の非隣接メモリノードと、前記１以上の第４のメモリノードのうち前記第１の非隣接メモリノードと接続され前記第１の非隣接メモリノードに隣接しない１以上の第２の非隣接メモリノードと、前記第２のメモリノードは夫々メモリノード同士の相対的な物理位置によって決定する付加アドレスを持ち、

前記第２の制御部は、前記第２のメモリノードの付加アドレス、前記第１の非隣接メモリノードの付加アドレス、前記第２の非隣接メモリノードの付加アドレスの少なくともいずれか一つを含む情報に基づいて、少なくとも前記複数の第３のメモリノード及び前記第１のメモリノードの中から前記パケットの出力先を決定する請求項１に記載のストレージ装置。

【請求項１１】

前記第１乃至第３のメモリノードとは異なる１以上の第４のメモリノードと、

前記第１のメモリノードと前記第２のメモリノードと前記複数の第３のメモリノードと前記１以上の第４のメモリノードのうちいずれか１つ以上と接続されるリレーと、
をさらに具備し、

前記リレーは、前記リレーと接続されるメモリノードの中で、前記第１のメモリノードと前記第２のメモリノードと前記複数の第３のメモリノードと前記１以上の第４のメモリノードのうち前記ストレージ装置内における位置情報が前記送信先アドレスと一致するメモリノードへの転送経路が最も短いメモリノードに前記パケット送信することを特徴とする請求項１に記載のストレージ装置。

【請求項１２】

前記第１のメモリノードと前記第２のメモリノードと前記複数の第３のメモリノードと前記１以上の第４のメモリノードのうちいずれか１つ以上と接続されるリレーと、
をさらに具備し、

前記リレーは、前記リレーと接続されるメモリノードの中で、前記第６のメモリノードへの転送経路が最も短いメモリノードに前記パケット送信することを特徴とする請求項２に記載のストレージ装置。

【請求項１３】

前記第１乃至第３のメモリノードとは異なる１以上の第４のメモリノードをさらに具備し、
前記第１のメモリノードと前記第２のメモリノードと前記複数の第３のメモリノードと前記１以上の第４のメモリノードのうち、他のメモリノードと接続されない入力ポートを少なくとも１以上有するメモリノードに接続可能なコンピュータの総数をＮｃとし、前記第１のメモリノードと前記第２のメモリノードと前記複数の第３のメモリノードと前記１以上の第４のメモリノードの総数をＮｎｏｄｅとした場合、Ｎｃ＜√Ｎｎｏｄｅを満たすことを特徴とする請求項１に記載のストレージ装置。
【請求項１４】
前記第１のメモリノードと前記第２のメモリノードと前記複数の第３のメモリノードと前記１以上の第４のメモリノードのうち、他のメモリノードと接続されない入力ポートを少なくとも１以上有するメモリノードに接続可能なコンピュータの総数をＮｃとし、前記第１のメモリノードと前記第２のメモリノードと前記複数の第３のメモリノードと前記１以上の第４のメモリノードの総数をＮｎｏｄｅとした場合、Ｎｃ＜√Ｎｎｏｄｅを満たすことを特徴とする請求項２に記載のストレージ装置。
【請求項１５】
Ｎｃ＜０．１×√Ｎnodeの関係式を満たすことを特徴とする請求項１３または１４に記載のストレージ装置。
【請求項１６】
前記第２のメモリノードは前記パケットに含まれるkeyをアドレスに変換するアドレス変換器をさらに具備し、
前記第２の制御部は、key-value型データの各レコードを保有し、前記送信先アドレスを前記アドレス変換器により変換された前記アドレスとし、前記パケットにvalueを含めて出力する請求項１または２に記載のストレージ装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明の実施形態は、転送機能を有するメモリノードを相互に接続したストレージ装置及びデータ処理方法に関わり、例えばストレージ装置におけるデータパケットの転送制御方式に関わる。
【背景技術】
【０００２】
容量を容易に拡張できるストレージ装置として、転送機能を有するメモリノードを相互に接続したストレージ装置が考えられる。各メモリノードは、自身のメモリノード宛のデータパケットを受信した場合は読み出しもしくは書き込みなどの所定の処理を行う。一方、各メモリノードは、自身のメモリノード宛でないパケットを受信した場合は、受信パケットを適切な他のメモリノードに転送する。各メモリノードによって適切な転送が繰り返されることにより、データパケットは目的のメモリノードに到達できる。
【０００３】
各メモリノードは、メモリと、転送機能を有するコントローラ、複数のポートを有している。各メモリノードは、パケットの転送先を示した経路指定表（ルーティングテーブル）を維持・管理し、それに従ってパケットの転送を行う。経路指定表を管理すれば、物理的な位置に関わらず任意の論理的なパケット転送ネットワークを構築することが可能である。
【０００４】
しかし、容量拡張のため新たなメモリノードを追加する場合、もしくは故障等の理由により既存のメモリノードを除去する場合には、各メモリノードの経路指定表を更新する必要があり、その手続きは煩雑である。経路指定表の維持・管理コストは、特にメモリノード数が大規模になった場合に膨大になり、それが容量の拡張性に制限を課すことになる。
【０００５】

また、転送機能を有するメモリノードを相互に接続したストレージ装置おいて、複数のデータを複数のメモリノードに書き込み・読み出しを行う場合において、複数のデータを同一配線で同時に通信させることは一般に困難であるため、データの転送待機が生じやすい。このようなデータの転送待機は、データの書き込み・読み出しに必要な時間の増加を招く。

【先行技術文献】

【特許文献】

【０００６】

【特許文献１】米国特許出願公開第２００９／０２１６９２４号明細書

【発明の概要】

【発明が解決しようとする課題】

【０００７】

メモリノードが経路指定表を管理する必要がなく、効率的にパケットを転送することができるストレージ装置、及びデータ処理方法を提供する。

【課題を解決するための手段】

【０００８】

一実施態様のストレージ装置は、第１の入力ポートと、第１の不揮発性メモリと、前記第１の不揮発性メモリを制御する第１の制御部と、を含む第１のメモリノードと、前記第１のメモリノードと接続される第2の入力ポートと、第２の不揮発性メモリと、前記第２の不揮発性メモリを制御する第２の制御部と、いずれか１つは前記第１の入力ポートと接続される複数の第１の出力ポートと、を含む第２のメモリノードと、前記複数の第１の出力ポートのうちの１つに接続される第3の入力ポートを含み、前記第１乃至第２のメモリノードとは異なる複数の第３のメモリノードとを具備するストレージ装置であって、前記第２のメモリノードは、前記ストレージ装置内における前記第２のメモリノードの位置情報である第１の位置情報を記憶し、前記第１の制御部は、送信先アドレスと送信元アドレスを少なくとも含むパケットを前記第２のメモリノードへ出力し、前記第2の入力ポートは、前記第１のメモリノードから出力された前記パケットを受信し、前記第２の制御部は、前記パケット内の前記送信先アドレスと前記送信元アドレス、前記第１の位置情報を少なくとも含む情報に基づいて、少なくとも前記複数の第３のメモリノード及び前記第１のメモリノードの中から前記パケットの出力先を決定するように構成されている。

【図面の簡単な説明】

【０００９】

【図１】第１実施形態のストレージ装置の構成を示す図である。

【図２】第１実施形態におけるメモリノードの構成を示す図である。

【図３】第１実施形態におけるメモリノードの配置例を示す図である。

【図４】第１実施形態のストレージ装置における転送アルゴリズム１を示す図である。

【図５Ａ】第１実施形態における転送アルゴリズム１によるパケットの転送過程を示す図である。

【図５Ｂ】第１実施形態における転送アルゴリズム１によるパケットの転送過程を示す図である。

【図５Ｃ】第１実施形態における転送アルゴリズム１によるパケットの転送過程のフロー図である。

【図５Ｄ】第１実施形態における転送アルゴリズム１によるパケットの転送過程の具体例を示す図である。

【図６】第１実施形態のストレージ装置を含むストレージシステムの構成を示す図である。

【図７】第１実施形態のストレージシステムにおける書き込み動作を示す図である。

【図８】第１実施形態のストレージシステムにおける読み出し動作を示す図である。

【図９】第１実施形態のストレージ装置におけるメモリノードの自動アドレス取得方式を示す図である。

【図１０】第２実施形態のストレージ装置における転送アルゴリズム２を示す図である。
【図１１Ａ】第２実施形態における転送アルゴリズム２によるパケットの転送過程を示す図である。
【図１１Ｂ】第２実施形態における転送アルゴリズム２によるパケットの転送過程を示す図である。
【図１１Ｃ】第２実施形態における転送アルゴリズム２によるパケットの転送過程のフロー図である。
【図１１Ｄ】第２実施形態における転送アルゴリズム２によるパケットの転送過程の具体例を示す図である。
【図１２Ａ】転送アルゴリズム１によるパケットの転送過程の具体例を示す図である。
【図１２Ｂ】転送アルゴリズム２によるパケットの転送過程の具体例を示す図である。
【図１３】第３実施形態のストレージ装置における転送アルゴリズム３を示す図である。
【図１４Ａ】第３実施形態における転送アルゴリズム３によるパケットの転送過程を示す図である。
【図１４Ｂ】第３実施形態における転送アルゴリズム３によるパケットの転送過程を示す図である。
【図１４Ｃ】第３実施形態における転送アルゴリズム３によるパケットの転送過程のフロー図である。
【図１４Ｄ】第３実施形態における転送アルゴリズム３によるパケットの転送過程の具体例を示す図である。
【図１５】第３実施形態のストレージ装置における転送アルゴリズム４を示す図である。
【図１６】第３実施形態のストレージ装置における転送アルゴリズム５を示す図である。
【図１７】第４実施形態のストレージシステムの構成を示す図である。
【図１８】第４実施形態のストレージシステムにおけるバイパス転送発生率と渋滞発生率を示す図である。
【図１９】第５実施形態のストレージ装置の構成を示す図である。
【図２０】第６実施形態のストレージシステムの構成を示す図である。
【図２１】第６実施形態のストレージシステムの他の構成例を示す図である。
【図２２】第７実施形態のストレージシステムの構成を示す図である。
【図２３】第７実施形態のストレージシステムに対する比較例を示す図である。
【図２４】第７実施形態におけるパケットのヘッダー部に記録されたアドレス情報を示す図である。
【図２５】第７実施形態のストレージシステムにおける書き込み動作を示す図である。
【図２６】第７実施形態のストレージシステムの他の構成例を示す図である。
【図２７】第８実施形態のストレージシステムの構成を示す図である。
【図２８】第８実施形態のストレージシステムに対する比較例を示す図である。
【図２９】第８実施形態におけるパケットのヘッダー部に記録されたアドレス情報を示す図である。
【図３０】第８実施形態のストレージシステムにおける書き込み動作を示す図である。
【図３１】第８実施形態のストレージシステムの他の構成例を示す図である。
【図３２】第８実施形態のストレージシステムの他の構成例を示す図である。
【図３３Ａ】第９実施形態のストレージシステムの構成を示す図である。
【図３３Ｂ】ストレージシステムにおいて転送待機が発生する読み出し動作を示す図である。
【図３３Ｃ】ストレージシステムにおいて転送待機が発生する読み出し動作を示す図である。
【図３３Ｄ】ストレージシステムにおいて転送待機が発生する読み出し動作を示す図である。
【図３４Ａ】第９実施形態のストレージシステムにおいて転送待機の発生を回避する読み出し動作を示す図である。

【図３４Ｂ】第９実施形態のストレージシステムにおいて転送待機の発生を回避する読み出し動作を示す図である。
【図３４Ｃ】第９実施形態のストレージシステムにおいて転送待機の発生を回避する読み出し動作を示す図である。
【図３４Ｄ】第９実施形態のストレージシステムにおいて転送待機の発生を回避する読み出し動作を示す図である。
【図３４Ｅ】第９実施形態のストレージシステムにおいて転送待機の発生を回避する読み出し動作を示す図である。
【図３５Ａ】第９実施形態のストレージシステムにおいて転送待機の発生を回避する読み出し動作の他の例を示す図である。
【図３５Ｂ】第９実施形態のストレージシステムにおいて転送待機の発生を回避する読み出し動作の他の例を示す図である。
【図３６Ａ】第９実施形態のストレージシステムにおいて転送待機の発生を回避する読み出し動作の他の例を示す図である。
【図３６Ｂ】第９実施形態のストレージシステムにおいて転送待機の発生を回避する読み出し動作の他の例を示す図である。
【図３６Ｃ】第９実施形態のストレージシステムにおいて転送待機の発生を回避する読み出し動作の他の例を示す図である。
【図３６Ｄ】第９実施形態のストレージシステムにおいて転送待機の発生を回避する読み出し動作の他の例を示す図である。
【図３６Ｅ】第９実施形態のストレージシステムにおいて転送待機の発生を回避する読み出し動作の他の例を示す図である。
【図３６Ｆ】第９実施形態のストレージシステムにおいて転送待機の発生を回避する読み出し動作の他の例を示す図である。
【図３７Ａ】第１０実施形態のストレージシステムの構成を示す図である。
【図３７Ｂ】ストレージシステムにおいて転送待機が発生する書き込み動作を示す図である。
【図３７Ｃ】ストレージシステムにおいて転送待機が発生する書き込み動作を示す図である。
【図３７Ｄ】ストレージシステムにおいて転送待機が発生する書き込み動作を示す図である。
【図３７Ｅ】ストレージシステムにおいて転送待機が発生する書き込み動作を示す図である。
【図３７Ｆ】ストレージシステムにおいて転送待機が発生する書き込み動作を示す図である。
【図３７Ｇ】ストレージシステムにおいて転送待機が発生する書き込み動作を示す図である。
【図３８Ａ】第１０実施形態のストレージシステムにおいて転送待機の発生を回避する書き込み動作を示す図である。
【図３８Ｂ】第１０実施形態のストレージシステムにおいて転送待機の発生を回避する書き込み動作を示す図である。
【図３８Ｃ】第１０実施形態のストレージシステムにおいて転送待機の発生を回避する書き込み動作を示す図である。
【図３８Ｄ】第１０実施形態のストレージシステムにおいて転送待機の発生を回避する書き込み動作を示す図である。
【図３８Ｅ】第１０実施形態のストレージシステムにおいて転送待機の発生を回避する書き込み動作を示す図である。
【発明を実施するための形態】
【００１０】
以下、図面を参照して実施形態について説明する。なお、以下の説明において、同一の

機能及び構成を有する構成要素については、同一符号を付し、重複説明は必要な場合にのみ行う。

【００１１】

［第１実施形態］

第１実施形態は、転送機能を有する複数のメモリノードを相互に接続したストレージ装置を備え、各メモリノードが効率的にデータパケットを転送する転送方式を有する。

【００１２】

［１］ストレージ装置の構成

図１は、第１実施形態のストレージ装置の構成を示す図であり、複数のメモリノードの物理的な配置法と、それに対するアドレスの割り当て法の一例を示す。

【００１３】

図示するように、ストレージ装置１０はデータ転送機能を有するメモリノード１１を複数備える。各々のメモリノード１１は、正方格子の格子点に置かれる。格子点の座標である格子座標（ｘ，ｙ）に位置するメモリノードの論理アドレスは、位置座標と一致して（ｘ，ｙ）とする。すなわち、メモリノード１１の論理アドレスは、物理アドレス（格子座標（ｘ，ｙ））に一致する。

【００１４】

各メモリノード１１は、４つの入力ポート１２と４つの出力ポート１３を持つ。各メモリノードは、隣接する４つのメモリノードと入力ポート１２及び出力ポート１３を介して相互に接続される。具体的には、対向する２つの隣接したメモリノードは、互いの入力ポート１２と出力ポート１３とが接続される。

【００１５】

図２に、各メモリノード１１の構成を示す。メモリノード１１は、入力ポート１２、入力ポートバッファ１２Ａ、出力ポート１３、出力ポートバッファ１３Ａ、セレクタ１４、パケットコントローラ１５、メモリ１６、メモリコントローラ１７、ＭＰＵ１８、及びローカルバス１９を有する。

【００１６】

入力ポート１２に入力されたパケットは、入力ポートバッファ１２Ａに一時的に記憶される。セレクタ１４には入力ポートバッファ１２Ａからパケットが入力され、またパケットコントローラ１５から制御信号が入力される。セレクタ１４は、制御信号に従い、入力されたパケットからいずれかのパケットを選択して出力ポートバッファ１３Ａに出力する。出力ポートバッファ１３Ａは、セレクタ１４から出力されたパケットを一時的に記憶すると共に、出力ポート１３に出力する。パケットコントローラ１５は、セレクタ１４の出力を制御する。パケットとは、送信先アドレス及び送信元アドレスを少なくとも含むヘッダー部と、データ部とからなる転送データの単位である。

【００１７】

メモリ１６は、データを記憶する複数のメモリセルを有する。メモリ１６は、例えばＮＡＮＤ型フラッシュメモリ等からなる。メモリコントローラ１７は、メモリ１６への書き込み、読み出し、及び消去の動作を制御する。ＭＰＵ１８は、メモリノード内で必要な演算処理を行う。ローカルバス１９は、入力ポートバッファ１２Ａ、パケットコントローラ１５、メモリコントローラ１７、及びＭＰＵ１８間を相互に接続し、これらの間の信号伝送を行う。

【００１８】

メモリノード１１が受信したパケットは、入力ポート１２を介して入力ポートバッファ１２Ａに格納される。パケットコントローラ１５は、パケットが含む送信先（宛先）アドレスとそのメモリノード自身（以下、自ノード）のアドレスの２つの情報に基づいて、受信したパケットが自ノード宛であるか否かを判断する。

【００１９】

もし、自ノード宛であれば、パケットコントローラ１５は、自ノードのメモリ１６に対する書き込み、読み出し、もしくはその他所定の処理を行う。もし、自ノード宛でなけれ

ば、パケットの送信先アドレスと自ノードのアドレスの２つの情報に基づいて、パケットを転送する隣接するメモリノードを決定し、セレクタ１４により対応する出力ポートバッファ１３Ａにパケットを出力する。

【００２０】

図１に示したストレージ装置では、正方格子の格子点にメモリノードが配置されたが、本実施形態はこの例に限定されるものではない。これを図３を参照して、例を挙げて説明する。

【００２１】

図３（ａ）に示す構成が図１に示した例に相当する。より一般的には、本実施形態において、各メモリノードは格子点に配置される。格子とは、幾何学においては平面上の点のうち、ｘ座標、ｙ座標がともに整数である点を言う。ここでは、ｘ方向とｙ方向の単位ベクトル、e_xとe_yの長さが異なる場合、即ちｘ方向とｙ方向の繰り返し周期が異なる場合も含む。この例を図３（ｂ）に示す。

【００２２】

また、ｘ方向とｙ方向の単位ベクトルが直交しない場合、即ちｘ軸とｙ軸が直交しない場合も含む。この例を図３（ｃ）に示す。また、メモリノードの相互接続数は４とは限らない。相互接続数が６の例を図３（ｄ）に示す。

【００２３】

本実施形態においては、格子の定義に関わらず、いずれの場合においても、格子座標（ｘ，ｙ）に位置するメモリノードの論理アドレスは、位置座標と一致して（ｘ，ｙ）とする。すなわち、メモリノードの論理アドレスは、物理アドレス（格子座標（ｘ，ｙ））に一致する。

【００２４】

また、より一般的には、本実施形態においては、各格子点に２以上のメモリノードの組が配置される場合を含む。各格子点に２つのメモリノードが配置される例を、図３（ｅ）に示す。この例では、格子点（１，０）などに２つのメモリノードが配置される。これを、（１，０，０）と（１，０，１）とする。即ち、１つのメモリノードの座標は（ｘ，ｙ，ｚ）の３つの整数の組によって表される。この場合においても、ノード座標（ｘ，ｙ，ｚ）に位置するメモリノードの論理アドレスは、位置座標と一致して（ｘ，ｙ，ｚ）とする。すなわち、メモリノードの論理アドレスは、物理アドレス（格子座標（ｘ，ｙ，ｚ））に一致する。さらに、格子点が配置される平面が三次元的に曲げられたり、もしくは折りたたまれた場合も含む。

【００２５】

隣接したメモリノードとは、図３（ａ）～図３（ｅ）において以下のような位置関係にあるメモリノードを指す。図３（ａ）～図３（ｃ）において、例えば自ノードが座標（１，１）に存在する場合、自ノードに隣接したメモリノードは座標（０，１）、（１，２）、（２，１）、（１，０）に存在する４つのメモリノードを指す。また図３（ｄ）において、例えば自ノードが座標（１，１）に存在する場合、自ノードに隣接したメモリノードは座標（０，１）、（０，２）、（１，２）、（２，１）、（２，０）、（１，０）に存在する６つのメモリノードを指す。さらに、図３（ｅ）において、例えば自ノードが座標（１，０，１）に存在する場合、自ノードに隣接したメモリノードとは座標（０，１，０）、（１，１，０）、（１，０，０）に存在する３つのメモリノードを指す。

【００２６】

［２］ストレージ装置の転送アルゴリズム１

転送アルゴリズム１では、パケットが含む送信先アドレス、及び自ノードのアドレスの２つの情報に基づいて、転送先のメモリノードを決定する。

【００２７】

図４は、第１実施形態のストレージ装置における転送アルゴリズム１を示す図である。図４を参照して、パケットの送信先アドレスと自ノードのアドレスの２つの情報に基づいて、転送先の隣接ノードを決定する方法の一例を示す。これを転送アルゴリズム１とする

。

【００２８】

パケットを受け取ったメモリノードは、パケットの送信先ノード(to)と自ノード(PP: Present position)との距離が最も小さくなる隣接するメモリノードにパケットを転送する。

【００２９】

図５Ａ～図５Ｄを参照して、転送アルゴリズム１に従うパケットの転送過程の具体例を示す。

【００３０】

図５Ａに示すように、送信先ノード(to)のアドレスを(x_{to}, y_{to})、自ノード(PP)のアドレスを(x_{pp}, y_{pp})とし、さらに$dx=x_{to}-x_{pp}$、$dy=y_{to}-y_{pp}$とする。転送先を示す方向として、図５Ｂに示すように、ｙが増加する方向をＮ(North)、ｘが増加する方向をＥ(East)、ｙが減少する方向をＳ(South)、ｘが減少する方向をＷ(West)とする。

【００３１】

ａに対してその絶対値をあらわす記号を｜ａ｜として、｜ｄｘ｜＞｜ｄｙ｜ならばｘ方向へ、｜ｄｘ｜＜｜ｄｙ｜ならｙ方向へ進む。ｘ方向へ進む場合は、ｄｘ＞０ならＥへ、ｄｘ＜０ならＷに転送する。同様に、ｙ方向へ進む場合は、ｄｙ＞０ならＮへ、ｄｙ＜０ならＳに転送する。

【００３２】

図５Ｃに、ストレージ装置における転送アルゴリズム１のフローを示す。転送アルゴリズム１はパケットコントローラ１５に記憶されており、パケットコントローラ１５により実行される。

【００３３】

まず、パケットコントローラ１５は、$dx=x_{to}-x_{pp}$、$dy=y_{to}-y_{pp}$を算出する（ステップＳ１）。続いて、パケットコントローラ１５は、ｄｘが０であるか否かを判定する（ステップＳ２）。ｄｘが０であるとき、$y_{to}>y_{pp}$が成り立つか否かを判定する（ステップＳ３）。$y_{to}>y_{pp}$が成り立つとき、パケットをＮへ転送する（ステップＳ４）。一方、$y_{to}>y_{pp}$が成り立たないとき、パケットをＳへ転送する（ステップＳ５）。

【００３４】

次に、ステップＳ２においてｄｘが０でないとき、パケットコントローラ１５は、ｄｙが０であるか否かを判定する（ステップＳ６）。ｄｙが０であるとき、$x_{to}>x_{pp}$が成り立つか否かを判定する（ステップＳ７）。$x_{to}>x_{pp}$が成り立つとき、パケットをＥへ転送する（ステップＳ８）。一方、$x_{to}>x_{pp}$が成り立たないとき、パケットをＷへ転送する（ステップＳ９）。

【００３５】

次に、ステップＳ６においてｄｙが０でないとき、すなわちｄｘとｄｙが０でないとき、パケットコントローラ１５は、ｄｘ＞０かつｄｙ＞０が成り立つか否かを判定する（ステップＳ１０）。ｄｘ＞０かつｄｙ＞０が成り立つとき、ｄｘ＞ｄｙが成り立つか否かを判定する（ステップＳ１１）。ｄｘ＞ｄｙが成り立つとき、パケットをＥへ転送する（ステップＳ１２）。一方、ｄｘ＞ｄｙが成り立たないとき、パケットをＮへ転送する（ステップＳ１３）。

【００３６】

次に、ステップＳ１０においてｄｘ＞０かつｄｙ＞０が成り立たないとき、パケットコントローラ１５は、ｄｘ＜０かつｄｙ＞０が成り立つか否かを判定する（ステップＳ１４）。ｄｘ＜０かつｄｙ＞０が成り立つとき、（－１）・ｄｘ＞ｄｙが成り立つか否かを判定する（ステップＳ１５）。（－１）・ｄｘ＞ｄｙが成り立つとき、パケットをＷへ転送する（ステップＳ１６）。一方、（－１）・ｄｘ＞ｄｙが成り立たないとき、パケットをＮへ転送する（ステップＳ１７）。

【００３７】

次に、ステップＳ１４においてｄｘ＜０かつｄｙ＞０が成り立たないとき、パケットコ

ントローラ１５は、ｄｘ＜０かつｄｙ＜０が成り立つか否かを判定する（ステップＳ１８）。ｄｘ＜０かつｄｙ＜０が成り立つとき、ｄｘ＞ｄｙが成り立つか否かを判定する（ステップＳ１９）。ｄｘ＞ｄｙが成り立つとき、パケットをＳへ転送する（ステップＳ２０）。一方、ｄｘ＞ｄｙが成り立たないとき、パケットをＷへ転送する（ステップＳ２１）。

【００３８】

次に、ステップＳ１８においてｄｘ＜０かつｄｙ＜０が成り立たないとき、パケットコントローラ１５は、ｄｘ＞（－１）・ｄｙが成り立つか否かを判定する（ステップＳ２２）。ｄｘ＞（－１）・ｄｙが成り立つとき、パケットをＥへ転送する（ステップＳ２３）。一方、ｄｘ＞（－１）・ｄｙが成り立たないとき、パケットをＳへ転送する（ステップＳ２４）。

【００３９】

以上の処理により、メモリノードに入力されたパケットは、送信先ノード(to)と自ノード(PP)との距離が最も小さくなるような隣接するメモリノードに転送される。

【００４０】

図５Ｄに、パケットの転送過程の具体例を示す。転送アルゴリズム１においては｜ｄｘ｜＝｜ｄｙ｜となるまでは｜ｄｘ｜＞｜ｄｙ｜ならばｘ方向へ、｜ｄｘ｜＜｜ｄｙ｜ならｙ方向へ連続して進み、｜ｄｘ｜＝｜ｄｙ｜となった後はｘ方向とｙ方向へ交互に進む。

【００４１】

例えば、ケース１では、ｄｘ＝０かつｄｙ＞０であるため、送信先ノード(to)に到達するまでＮへ進む。ケース４では、ｄｘ＞０かつｄｙ＞０であり、ｄｘ＜ｄｙであるため、ｄｘ＝ｄｙとなるまでＮへ連続して進み、ｄｘ＝ｄｙとなった後はＥとＮへ交互に進む。

【００４２】

［３］ストレージ装置を含むストレージシステム

図６は、第１実施形態のストレージ装置を含むストレージシステムの構成を示す図である。

【００４３】

ストレージシステム２０は、ストレージ装置をクライアントが使用するためのシステムであり、以下のような構成を備える。

【００４４】

ストレージ装置１０はゲートウェイサーバを介してクライアントと接続される。ストレージ装置１０内部の通信規格とゲートウェイサーバ２１Ａ，２１Ｂの通信規格が異なる場合は、両者の間にアダプタ２２Ａ，２２Ｂをそれぞれ設置しても良い。

【００４５】

詳述すると、ストレージ装置１０の外周部に配置されたメモリノード（１，４）は、アダプタ２２Ａ、ゲートウェイサーバ２１Ａを介してクライアント３１Ａと接続される。同様に、メモリノード（１，１）は、アダプタ２２Ｂ、ゲートウェイサーバ２１Ｂを介してクライアント３１Ｂ１，３１Ｂ２と接続される。なお、“メモリノード（ｘ，ｙ）”は、アドレス（ｘ，ｙ）のメモリノードであることを表す。以降も同様である。

【００４６】

ゲートウェイサーバ２１Ａ，２１Ｂは、例えばコンピュータから構成され、ストレージ装置１０内部のメモリノード１１と同じ規約に基づいてアドレスを持つ。図６では、ゲートウェイサーバ２１Ａはアドレス（０，４）を持ち、ゲートウェイサーバ２１Ｂはアドレス（０，１）を持つ。

【００４７】

次に、ストレージシステムにおける書き込み動作を説明する。

【００４８】

図７は、ストレージ装置１０にクライアントがファイルを書き込む手続きを示す。ここでは、クライアント３１Ａがストレージ装置１０にファイルを書き込む場合を述べる。

【００４９】

クライアント３１Ａは、ファイルとファイルＩＤをゲートウェイサーバ２１Ａに送信する（（１）参照）。ファイルＩＤは、ファイルを一意に特定できる識別子である。ファイルＩＤとしては、\\strage_system \home \cliantA \file1.txtといった所定のファイルシステムにおけるフルパスのファイル名を使うことができる。

【００５０】

次に、ゲートウェイサーバ２１Ａは、ファイルを規定サイズのデータパケットに分割し、各パケットにパケットＩＤを割り振る。続いて、ファイルＩＤと、分割されたパケットのパケットＩＤをファイルテーブルに書き込む（（２）参照）。パケットＩＤは、パケットを一意に特定できる識別子である。パケットＩＤとしては、・・・\file1.txt~1、・・・\file1.txt~2というように、「ファイルＩＤ＋連続する番号」を割り当てることができる。

【００５１】

次に、ゲートウェイサーバ２１Ａは、パケットＩＤの情報に基づき、そのパケットを書き込むメモリノードのアドレス（以下、書き込みノードアドレス）を決定する（（３）参照）。この際、大規模分散ファイルシステムで使用されるコンシステントハッシング（文献１参照）と呼ばれるノード決定手法を使っても良い。コンシステントハッシングは、ノードアドレスのハッシュ値、およびパケットＩＤのハッシュ値の両方を使って、書き込みアドレスを決定する点に特徴がある。

【００５２】

[文献１]：丸山不二夫/首藤一幸編、「雲の世界の向こうをつかむクラウドの技術」、株式会社アスキー・メディアワークス、2009年11月6日、p.88、ISBN978-4-04-868064-6

次に、ゲートウェイサーバ２１Ａは、書き込みアドレスを送信先アドレス、ゲートウェイサーバ２１Ａのアドレスを送信元アドレスとして、書き込みパケットをストレージ装置１０に送信する（（４）、（５）参照）。

【００５３】

メモリノード（１，４）に送信されたパケットは、ストレージ装置１０内で転送アルゴリズム１に従って適切な転送が繰り返されることにより、送信先アドレスのメモリノードに到達する（（６）参照）。目的のメモリノードでは、受け取ったパケットを自ノードのメモリ１６に書き込む（（７）参照）。その後、書き込み完了パケットをゲートウェイサーバ２１Ａへ返信する（（８）参照）。

【００５４】

次に、ストレージシステムにおける読み出し動作を説明する。

【００５５】

図８は、ストレージ装置１０からクライアントがファイルを読み出す手続きを示す。ここでは、クライアント３１Ａがストレージ装置１０からファイルを読み出す場合を述べる。

【００５６】

クライアント３１Ａは、読み出し要求（ファイルＩＤ）をゲートウェイサーバ２１Ａに送信する（（１）参照）。

【００５７】

ゲートウェイサーバ２１Ａは、ファイルテーブルからファイルＩＤに対応するパケットＩＤを取得する（（２）参照）。続いて、ゲートウェイサーバ２１Ａは、パケットＩＤの情報に基づき、読み出しを行うメモリノードのアドレス（以下、読み出しノードアドレス）を決定する（（３）参照）。この際、大規模分散ファイルシステムで使用されるコンシステントハッシングと呼ばれるノード決定手法を使っても良い。

【００５８】

次に、ゲートウェイサーバ２１Ａは、読み出しノードアドレスを送信先アドレス、ゲートウェイサーバ２１Ａのアドレスを送信元アドレスとして、読み出しパケットをストレージ装置１０に送信する（（４）、（５）参照）。

【００５９】

メモリノード（１，４）に送信されたパケットは、ストレージ装置１０内で転送アルゴリズム１に従って適切な転送が繰り返されることにより、送信先アドレスのメモリノードに到達する（（６）参照）。目的のメモリノードでは、読み出しパケットに従い、自ノードのメモリ１６から目的のデータを読み出す（（７）参照）。その後、読み出したデータをデータパケットとしてゲートウェイサーバ２１Ａに返信する（（８）参照）。

【００６０】

図６に示したシステムにおいては、書き込みデータのランダマイズ処理（Rand処理）、書き込みデータへの誤り訂正符号（ＥＣＣ：error-correcting code）の付与、及びＥＣＣを用いた読み出しデータの誤り検出および訂正の機能（ＥＣＣ機能）をゲートウェイサーバが実施してよい。この場合、各メモリノードのメモリコントローラはランダマイズ処理とＥＣＣ機能を持つ必要がない。さらに、メモリノード間の通信の際にＥＣＣ処理を実施する必要がない。このため、メモリノード一つ当たりのコストを低減することが可能となる。

【００６１】

［４］ストレージ装置の拡張性

ストレージ装置に新たなメモリノードを追加する際の方式について説明する。

【００６２】

図９は、ストレージ装置に新たなメモリノードを追加した際の自動アドレス取得方式を示す。

【００６３】

本実施形態のストレージ装置１０においては、格子座標（ｘ，ｙ）に位置するメモリノードの論理アドレスは位置座標と一致して（ｘ，ｙ）とするため、追加するメモリノードは隣接ノードにアドレスを問い合わせることにより、自ノードのアドレスを簡易に割り出すことができる。

【００６４】

例えば、追加するメモリノードがW方向にある隣接ノードにアドレスを問い合わせる場合、隣接ノードのアドレスを（ｘ，ｙ）とすると、追加ノードのアドレスは（ｘ＋１，ｙ）となる。同様に、Ｎ，Ｅ，Ｓにある隣接ノード（ｘ，ｙ）に問い合わせた場合、自ノードのアドレスはそれぞれ（ｘ，ｙ－１），（ｘ－１，ｙ），（ｘ，ｙ＋１）となる。

【００６５】

図９に示した場合においては、メモリノード（４，４）のＥ方向に新しいメモリノードを追加した状況を例示している。追加するメモリノードは、toとfromがnoneであるaddress_requestパケットをW方向にある隣接ノード（４，４）に送信する。address_requestパケットを受信したメモリノード（４，４）は、fromを自ノードアドレスとし、toを（５，４）としたaddress_answerパケットをＥ方向にあるメモリノードの出力ポートバッファにセットする。追加ノードは、address_answerパケットを受信することにより、自ノードのアドレスが、（５，４）であることを割り出すことができる。

【００６６】

また、別の方式として、手動アドレス設定方式も考えられる。手動アドレス設定方式では、ストレージ装置１０の運用者が予め追加するメモリノードのアドレスを割り出し、それをプリセットしてから、新規のメモリノードを追加する。

【００６７】

前述した自動アドレス取得方式は、各メモリノードのパケットコントローラがAddress_requestパケットに対して回答する機能を有することが必要である。このため、パケットコントローラ１つ当たりのコストが高価になる傾向がある。一方、追加するメモリノードのアドレスをプリセットする必要がないため、メモリノードを追加する時の手続きが簡易化され、運用コストを低下させることができる。

【００６８】

以上説明したように第１実施形態によれば、メモリノードの論理的なアドレスと物理的な位置（物理アドレス）が一致するため、各メモリノードが経路指定表を管理する必要の

ない効率的なデータ転送方式を実現することができる。このため、ストレージ装置の拡張性に優れる、すなわちストレージ装置にメモリノードを容易に追加することが可能である。

【００６９】

［第２実施形態］

第２実施形態のストレージ装置について説明する。第２実施形態は、第１実施形態が有する転送アルゴリズム１と異なる転送アルゴリズム２を有する。転送アルゴリズム２では、パケットが含む送信先アドレス及び送信元アドレス、自ノードのアドレスの３つの情報に基づいて、転送先のメモリノードを決定する。なお、このストレージ装置は、図１及び図２と同様な構成を有するため、記載を省略する。

【００７０】

［１］ストレージ装置の転送アルゴリズム２

図１０は、第２実施形態のストレージ装置における転送アルゴリズム２を示す図である。図１０を参照して、パケットの送信先アドレス及び送信元アドレス、自ノードのアドレスの３つの情報に基づいて、転送先の隣接ノードを決定する方法の一例を示す。これを転送アルゴリズム２とする。

【００７１】

図１０に示すように、送信先ノード(to)と送信元ノード(from)を結ぶ直線で区切られる２つの領域のどちらに自ノード(PP)が存在するかを判断し、その領域に割り当てられた方向の隣接するメモリノードにパケットを転送する。

【００７２】

図１１Ａ～図１１Ｄを参照して、転送アルゴリズム２に従うパケットの転送過程の具体例を示す。

【００７３】

図１１Ａに示すように、送信先ノード(to)のアドレスを(x_{to}, y_{to})、送信元ノード(from)のアドレスを(x_{from}, y_{from})、自ノード(PP)のアドレスを(x_{pp}, y_{pp})とする。さらに、$dx=x_{to}-x_{from}$、$dy=y_{to}-y_{from}$、$Dx=x_{pp}-x_{from}$、$Dy=y_{pp}-y_{from}$とする。転送先を示す方向として、図１１Ｂに示すように、ｙが増加する方向をＮ、ｘが増加する方向をＥ、ｙが減少する方向をＳ、ｘが減少する方向をＷとする。

【００７４】

ｙ＝（ｄｘ／ｄｙ）・ｘという送信先ノード(to)と送信元ノード(from)を結ぶ直線の式に基づき、Ｄｙと（ｄｙ／ｄｘ）・Ｄｘのどちらが大きいかを比較することで、その直線で区切られる２つの領域のどちらに自ノード(PP)が存在しているかを判断する。

【００７５】

ｄｘ＞０かつｄｙ＞０の場合を例に転送方向の割り当て例を説明する。Ｄｙと（ｄｙ／ｄｘ）・Ｄｘを比較した場合に、Ｄｙが（ｄｙ／ｄｘ）・Ｄｘよりも大きくなる領域をＡ領域、もう一方の領域をＢ領域とする。転送において、Ａ領域の第一優先方向にはＥを割り当て、Ｂ領域の第一優先方向にはＮを割り当てる。転送アルゴリズム２においては、送信先ノード(to)と送信元ノード(from)を結ぶ直線に沿ってパケットが進むように転送が行われる。

【００７６】

図１１Ｃに、ストレージ装置における転送アルゴリズム２のフローを示す。転送アルゴリズム２はパケットコントローラ１５に記憶されており、パケットコントローラ１５により実行される。

【００７７】

まず、パケットコントローラ１５は、$dx=x_{to}-x_{from}$、$dy=y_{to}-y_{from}$、$Dx=x_{pp}-x_{from}$、$Dy=y_{pp}-y_{from}$を算出する（ステップＳ３１）。続いて、パケットコントローラ１５は、ｄｘが０であるか否かを判定する（ステップＳ３２）。ｄｘが０であるとき、$y_{to}>y_{pp}$が成り立つか否かを判定する（ステップＳ３３）。$y_{to}>y_{pp}$が成り立つとき、パケットをＮへ転送する（ステップＳ３４）。一方、$y_{to}>y_{pp}$が成り立たないとき、パケッ

トをＳへ転送する（ステップＳ３５）。

【００７８】

次に、ステップＳ３２においてｄｘが０でないとき、パケットコントローラ１５は、ｄｙが０であるか否かを判定する（ステップＳ３６）。ｄｙが０であるとき、$x_{to}>x_{pp}$が成り立つか否かを判定する（ステップＳ３７）。$x_{to}>x_{pp}$が成り立つとき、パケットをＥへ転送する（ステップＳ３８）。一方、$x_{to}>x_{pp}$が成り立たないとき、パケットをＷへ転送する（ステップＳ３９）。

【００７９】

次に、ステップＳ３６においてｄｙが０でないとき、すなわちｄｘとｄｙが０でないとき、パケットコントローラ１５は、Ｄｙ・ｄｘ＞ｄｙ・Ｄｘが成り立つか否かを判定する（ステップＳ４０）。Ｄｙ・ｄｘ＞ｄｙ・Ｄｘが成り立つとき、ｄｘ＞０かつｄｙ＞０が成り立つか否かを判定する（ステップＳ４１）。ｄｘ＞０かつｄｙ＞０が成り立つとき、パケットをＥへ転送する（ステップＳ４２）。一方、ｄｘ＞０かつｄｙ＞０が成り立たないとき、ｄｘ＜０かつｄｙ＞０が成り立つか否かを判定する（ステップＳ４３）。ｄｘ＜０かつｄｙ＞０が成り立つとき、パケットをＮへ転送する（ステップＳ４４）。

【００８０】

次に、ステップＳ４３においてｄｘ＜０かつｄｙ＞０が成り立たないとき、ｄｘ＜０かつｄｙ＜０が成り立つか否かを判定する（ステップＳ４５）。ｄｘ＜０かつｄｙ＜０が成り立つとき、パケットをＷへ転送する（ステップＳ４６）。ｄｘ＜０かつｄｙ＜０が成り立たないとき、パケットをＳへ転送する（ステップＳ４７）。

【００８１】

次に、ステップＳ４０においてＤｙ・ｄｘ＞ｄｙ・Ｄｘが成り立たないとき、ｄｘ＞０かつｄｙ＞０が成り立つか否かを判定する（ステップＳ４８）。ｄｘ＞０かつｄｙ＞０が成り立つとき、パケットをＮへ転送する（ステップＳ４９）。ｄｘ＞０かつｄｙ＞０が成り立たないとき、ｄｘ＜０かつｄｙ＞０が成り立つか否かを判定する（ステップＳ５０）。ｄｘ＜０かつｄｙ＞０が成り立つとき、パケットをＷへ転送する（ステップＳ５１）。ｄｘ＜０かつｄｙ＞０が成り立たないとき、ｄｘ＜０かつｄｙ＜０が成り立つか否かを判定する（ステップＳ５２）。ｄｘ＜０かつｄｙ＜０が成り立つとき、パケットをＳへ転送する（ステップＳ５３）。ｄｘ＜０かつｄｙ＜０が成り立たないとき、パケットをＥへ転送する（ステップＳ５４）。

【００８２】

以上の処理により、メモリノードに入力されたパケットは、送信先ノード(to)と送信元ノード(from)を結ぶ直線に沿ってパケットが進むように、隣接するメモリノードに転送される。

【００８３】

図１１Ｄに、パケットの転送過程の具体例を示す。転送アルゴリズム２においては、送信先ノード(to)と送信元ノード(from)を結ぶ直線に沿ってパケットが進むように転送が行われる。例えば、ケース３では、ｄｘ＜０かつｄｙ＞０であり、送信先ノード(to)と送信元ノード(from)を結ぶ直線からできるだけ離れないように、ＷまたはＮへ進む。ケース４では、ｄｘ＞０かつｄｙ＞０であり、送信先ノード(to)と送信元ノード(from)を結ぶ直線からできるだけ離れないように、ＮまたはＥへ進む。

【００８４】

図１２Ａ及び図１２Ｂを参照して、転送アルゴリズム１に対する転送アルゴリズム２の優位性を示す。

【００８５】

図１２Ａ及び図１２Ｂは、２つの送信元ノード(from1)と(from2)からそれぞれ送信先ノード(to)宛てのパケットを送信した場合のパケットの転送過程を例示している。図１２Ａは転送アルゴリズム１による転送過程を示し、図１２Ｂは転送アルゴリズム２による転送過程を示す。

【００８６】

図１２Ａに示す転送アルゴリズム１では、２つのパケットが合流する地点で、合流待ちの渋滞（jamで示す）が発生している。一方、図１２Ｂに示す転送アルゴリズム２では渋滞は発生していない。したがって、転送アルゴリズム２では、転送アルゴリズム１よりも渋滞発生確率を減少させることができる。

【００８７】

以上説明したように第２実施形態によれば、第１実施形態に比べて渋滞発生確率を減少させることができ、多くのクライアントが同時に接続した場合でも、応答速度を維持することができるストレージ装置を提供可能である。

【００８８】

また、第１実施形態と同様に、メモリノードの論理アドレスと物理アドレスが一致するため、各メモリノードが経路指定表を管理する必要のない効率的なデータ転送方式を実現することができる。このため、ストレージ装置の拡張性に優れる、すなわちストレージ装置にメモリノードを容易に追加することが可能である。その他の構成及び効果は、前述した第１実施形態と同様である。

【００８９】

［第３実施形態］

第３実施形態のストレージ装置について説明する。第３実施形態は、第１，第２実施形態が有する転送アルゴリズム１，２と異なる転送アルゴリズム３を有する。転送アルゴリズム３では、パケットが含む送信先アドレス及び送信元アドレス、自ノードのアドレス、及び自ノードの出力ポート占有情報の４つの情報に基づいて、転送先のメモリノードを決定する。なお、このストレージ装置は、図１及び図２と同様な構成を有するため、記載を省略する。

【００９０】

［１］ストレージ装置の転送アルゴリズム３

図１３は、第３実施形態のストレージ装置における転送アルゴリズム３を示す図である。図１３を参照して、パケットの送信先アドレス及び送信元アドレス、自ノードのアドレス、自ノードの出力ポート占有情報の４つの情報に基づいて、転送先の隣接ノードを決定する方法の一例を示す。これを転送アルゴリズム３とする。

【００９１】

図１３に示すように、送信先ノード(to)と送信元ノード(from)を結ぶ直線で区切られる２つの領域のどちらに自ノード(PP)が存在するかを判断し、その領域に割り当てられた２つの方向のうち、自ノードの出力ポートの占有情報によって決まる方向の隣接ノードにパケットを転送する。各領域には、第一優先方向と、第二優先方向を割り当てる。第一優先方向の出力ポートバッファが別のパケットにより占有されていた場合に、第二優先方向が選択される。

【００９２】

図１３、図１４Ａ～図１４Ｄを参照して、転送アルゴリズム３に従うパケットの転送過程の具体例を示す。

【００９３】

図１４Ａに示すように、送信先ノード(to)のアドレスを(x_{to}, y_{to})、送信元ノード(from)のアドレスを(x_{from}, y_{from})、自ノードのアドレス(PP)を(x_{pp}, y_{pp})とする。さらに、$dx = x_{to} - x_{from}$、$dy = y_{to} - y_{from}$、$Dx = x_{pp} - x_{from}$、$Dy = y_{pp} - y_{from}$とする。

【００９４】

図１３では、$dx > 0$かつ$dy > 0$の場合を例示している。Ｄｙと(ｄｙ／ｄｘ)・Ｄｘを比較した場合に、Ｄｙが(ｄｙ／ｄｘ)・Ｄｘよりも大きくなる領域をＡ領域、もう一方の領域をＢ領域とする。転送において、Ａ領域の第一優先方向にはＥを割り当て、Ｂ領域の第一優先方向にはＮを割り当てる。また、Ａ領域の第二優先方向(bypass方向)にはＮを割り当て、Ｂ領域の第二優先方向にはＥを割り当てる。例えば、図示されているようにパケットを転送するメモリノードがＢ領域に属す場合において、第一優先方向であるＮ方向の出力ポートバッファが別のパケットに占有されていれば、そのメモリノードは第二優先

方向であるE方向にパケットを転送する。

【００９５】

図１４Ｃに、ストレージ装置における転送アルゴリズム３のフローを示す。転送アルゴリズム３はパケットコントローラ１５に記憶されており、パケットコントローラ１５により実行される。なお、図１４Ｂに示すように、Ｎへパケットを出力するための出力ポートバッファをＯＰＢＮとし、Ｅへパケットを出力するための出力ポートバッファをＯＰＢＥ、Ｗへパケットを出力するための出力ポートバッファをＯＰＢＷ、Ｓへパケットを出力するための出力ポートバッファをＯＰＢＳとする。

【００９６】

これら出力ポートバッファＯＰＢＮ、ＯＰＢＥ、ＯＰＢＷ、ＯＰＢＳが空いているか（パケットが記憶できるか）、あるいはパケットに占有されているかの情報（出力ポート占有情報）は、以下のようにパケットコントローラ１５に記憶される。パケットコントローラ１５は、出力ポートバッファと入力ポートバッファの総数分だけバッファ占有フラグビットを有する。パケットコントローラ１５は、出力ポートバッファにパケットが書き込まれた場合、そのバッファに対応するバッファ占有フラグビットを“１”にし、出力ポートバッファからパケットが出力された場合、そのバッファに対応するバッファ占有フラグビットを“０”にする。パケットコントローラ１５は、バッファ占有フラグビットを評価することで、対応する出力ポートバッファが空いているか否か（パケットに占有されているか否か）を判断できる。

【００９７】

まず、パケットコントローラ１５は、$dx=x_{to}-x_{from}$、$dy=y_{to}-y_{from}$、$Dx=x_{pp}-x_{from}$、$Dy=y_{pp}-y_{from}$を算出する（ステップＳ６１）。続いて、パケットコントローラ１５は、ｄｘが０であるか否かを判定する（ステップＳ６２）。ｄｘが０であるとき、$y_{to}>y_{pp}$が成り立つか否かを判定する（ステップＳ６３）。$y_{to}>y_{pp}$が成り立つとき、パケットをＮへ転送する（ステップＳ６４）。一方、$y_{to}>y_{pp}$が成り立たないとき、パケットをＳへ転送する（ステップＳ６５）。

【００９８】

次に、ステップＳ６２においてｄｘが０でないとき、パケットコントローラ１５は、ｄｙが０であるか否かを判定する（ステップＳ６６）。ｄｙが０であるとき、$x_{to}>x_{pp}$が成り立つか否かを判定する（ステップＳ６７）。$x_{to}>x_{pp}$が成り立つとき、パケットをＥへ転送する（ステップＳ６８）。一方、$x_{to}>x_{pp}$が成り立たないとき、パケットをＷへ転送する（ステップＳ６９）。

【００９９】

次に、ステップＳ６６においてｄｙが０でないとき、すなわちｄｘとｄｙが０でないとき、パケットコントローラ１５は、Ｄｙ・ｄｘ＞ｄｙ・Ｄｘが成り立つか否かを判定する（ステップＳ７０）。Ｄｙ・ｄｘ＞ｄｙ・Ｄｘが成り立つとき、ｄｘ＞０かつｄｙ＞０が成り立つか否かを判定する（ステップＳ７１）。ｄｘ＞０かつｄｙ＞０が成り立つとき、Ｅへ出力する出力ポートバッファＯＰＢＥが空いているか否か、すなわち別のパケットにより占有されていないかを判定する（ステップＳ７２）。出力ポートバッファＯＰＢＥが空いているとき、パケットをＥへ転送する（ステップＳ７３）。一方、出力ポートバッファＯＰＢＥが空いていないとき、パケットをＮへ転送する（ステップＳ７４）。

【０１００】

次に、ステップＳ７１においてｄｘ＞０かつｄｙ＞０が成り立たないとき、ｄｘ＜０かつｄｙ＞０が成り立つか否かを判定する（ステップＳ７５）。ｄｘ＜０かつｄｙ＞０が成り立つとき、Ｎへ出力する出力ポートバッファＯＰＢＮが空いているか否かを判定する（ステップＳ７６）。出力ポートバッファＯＰＢＮが空いているとき、パケットをＮへ転送する（ステップＳ７７）。一方、出力ポートバッファＯＰＢＮが空いていないとき、パケットをＷへ転送する（ステップＳ７８）。

【０１０１】

次に、ステップＳ７５においてｄｘ＜０かつｄｙ＞０が成り立たないとき、ｄｘ＜０か

つｄｙ＜０が成り立つか否かを判定する（ステップＳ７９）。ｄｘ＜０かつｄｙ＜０が成り立つとき、Ｗへ出力する出力ポートバッファＯＰＢＷが空いているか否かを判定する（ステップＳ８０）。出力ポートバッファＯＰＢＷが空いているとき、パケットをＷへ転送する（ステップＳ８１）。一方、出力ポートバッファＯＰＢＷが空いていないとき、パケットをＳへ転送する（ステップＳ８２）。

【０１０２】

次に、ステップＳ７９においてｄｘ＜０かつｄｙ＜０が成り立たないとき、Ｓへ出力する出力ポートバッファＯＰＢＳが空いているか否かを判定する（ステップＳ８３）。出力ポートバッファＯＰＢＳが空いているとき、パケットをＳへ転送する（ステップＳ８４）。一方、出力ポートバッファＯＰＢＳが空いていないとき、パケットをＥへ転送する（ステップＳ８５）。

【０１０３】

次に、ステップＳ７０においてＤｙ・ｄｘ＞ｄｙ・Ｄｘが成り立たないとき、ｄｘ＞０かつｄｙ＞０が成り立つか否かを判定する（ステップＳ８６）。ｄｘ＞０かつｄｙ＞０が成り立つとき、Ｎへ出力する出力ポートバッファＯＰＢＮが空いているか否かを判定する（ステップＳ８７）。出力ポートバッファＯＰＢＮが空いているとき、パケットをＮへ転送する（ステップＳ８８）。一方、出力ポートバッファＯＰＢＮが空いていないとき、パケットをＥへ転送する（ステップＳ８９）。

【０１０４】

次に、ステップＳ８６においてｄｘ＞０かつｄｙ＞０が成り立たないとき、ｄｘ＜０かつｄｙ＞０が成り立つか否かを判定する（ステップＳ９０）。ｄｘ＜０かつｄｙ＞０が成り立つとき、Ｗへ出力する出力ポートバッファＯＰＢＷが空いているか否かを判定する（ステップＳ９１）。出力ポートバッファＯＰＢＷが空いているとき、パケットをＷへ転送する（ステップＳ９２）。一方、出力ポートバッファＯＰＢＷが空いていないとき、パケットをＮへ転送する（ステップＳ９３）。

【０１０５】

次に、ステップＳ９０においてｄｘ＜０かつｄｙ＞０が成り立たないとき、ｄｘ＜０かつｄｙ＜０が成り立つか否かを判定する（ステップＳ９４）。ｄｘ＜０かつｄｙ＜０が成り立つとき、Ｓへ出力する出力ポートバッファＯＰＢＳが空いているか否かを判定する（ステップＳ９５）。出力ポートバッファＯＰＢＳが空いているとき、パケットをＳへ転送する（ステップＳ９６）。一方、出力ポートバッファＯＰＢＳが空いていないとき、パケットをＷへ転送する（ステップＳ９７）。

【０１０６】

次に、ステップＳ９４においてｄｘ＜０かつｄｙ＜０が成り立たないとき、Ｅへ出力する出力ポートバッファＯＰＢＥが空いているか否かを判定する（ステップＳ９８）。出力ポートバッファＯＰＢＥが空いているとき、パケットをＥへ転送する（ステップＳ９９）。一方、出力ポートバッファＯＰＢＥが空いていないとき、パケットをＳへ転送する（ステップＳ１００）。

【０１０７】

以上の処理により、第一優先方向において渋滞が発生している場合には、第二優先方向へパケットを転送することにより、渋滞を回避しながら送信先ノード(to)と送信元ノード(from)を結ぶ直線に沿って、隣接するメモリノードにパケットが転送される。

【０１０８】

図１４Ｄに、パケットの転送過程の具体例を示す。ケース１は渋滞が発生しない例であり、送信先ノード(to)と送信元ノード(from)を結ぶ直線に沿ってパケットが転送される。ケース２，３は、渋滞を回避する例を示す。ケース２では、パケットをＮへ転送しようとしているとき、Ｎ方向に渋滞が発生している。このため、ケース３に示すように、パケットをＥへ転送して渋滞を回避している。ケース４，５は、渋滞を回避する他の例を示す。ケース４では、パケットをＥへ転送しようとしているとき、Ｅ方向に渋滞が発生している。このため、ケース５に示すように、パケットをＮへ転送して渋滞を回避している。

【０１０９】
転送アルゴリズム３においては、第一優先方向で渋滞が発生している場合には、第二優先方向へパケットを転送することにより、渋滞を回避することができる。この際、渋滞がそもそも発生しなかった場合と比べて、送信先ノードにパケットが到達するまでの転送回数は変わらない。
【０１１０】
［２］ストレージ装置の転送アルゴリズム４
図１５は、ストレージ装置における転送アルゴリズム４を示す図である。図１５を参照して、パケットの送信先アドレス及び送信元アドレス、自ノードのアドレス、自ノードの出力ポート占有情報の４つの情報に基づいて、転送先の隣接ノードを決定する方法の別の第１例を示す。これを転送アルゴリズム４とする。
【０１１１】
図１５に示すように、送信先ノード(to)と送信元ノード(from)を結ぶ直線、及びその直線と送信先ノード(to)の位置で直交する直線で区切られる４つの領域を定義する。それら４つの領域のうち、どの領域に自ノード(PP)が存在するかを判断し、自ノード(PP)が存在する領域に割り当てられた第一優先と第二優先の２つの方向のうち、自ノード(PP)の出力ポートの占有情報によって決まる方向の隣接ノードにパケットを転送する。
【０１１２】
アルゴリズム３と比較したアルゴリズム４の優位性は次の通りである。アルゴリズム３に対しては、送信先ノード(to)と送信元ノード(from)を結ぶ直線を対角線とする長方形により規定される領域の外にパケットを転送することを禁止する制限を設ける必要がある。一方、アルゴリズム4に対しては、その制限を設ける必要がない。
【０１１３】
［３］ストレージ装置の転送アルゴリズム５
図１６は、ストレージ装置における転送アルゴリズム５を示す図である。図１６を参照して、パケットの送信先アドレス及び送信元アドレス、自ノードのアドレス、自ノードの出力ポート占有情報の４つの情報に基づいて、転送先の隣接ノードを決定する方法の別の第２例を示す。これを転送アルゴリズム５とする。
【０１１４】
図１６に示すように、送信先ノード(to)と送信元ノード(from)を結ぶ直線、その直線と送信先ノード(to)の位置で直交する直線、送信先ノード(to)の位置でｘ方向とy方向に伸びる２つの直線の４つの直線で区切られる８つの領域を定義する。それら８つの領域のうち、どの領域に自ノード(PP)が存在するかを判断し、自ノード(PP)が存在する領域に割り当てられた第一優先と第二優先の２つの方向のうち、自ノード(PP)の出力ポートの占有情報によって決まる方向の隣接ノードにパケットを転送する。前記送信先ノード(to)の位置でｘ方向とy方向に伸びる２つの直線は、例えば送信先ノード(to)を通りメモリノードの配列方向に沿った２つの直線を含む。
【０１１５】
アルゴリズム４と比較したアルゴリズム５の優位性は次の通りである。アルゴリズム４では、送信先ノード(to)と送信元ノード(from)を結ぶ直線を対角線とする長方形により規定される領域の外でパケットを第二優先方向(bypass方向)に転送させた場合、送信先ノード(to)にパケットが到達するまでの転送回数が少なくとも１つ増加する。一方、アルゴリズム５の場合では転送回数の増加がない。
【０１１６】
以上説明したように第３実施形態によれば、第１，第２実施形態に比べて渋滞発生確率を減少させることができ、多くのクライアントが同時に接続した場合でも、応答速度を維持することができるストレージ装置を提供可能である。
【０１１７】
また、第１実施形態と同様に、メモリノードの論理アドレスと物理アドレスが一致するため、各メモリノードが経路指定表を管理する必要のない効率的なデータ転送方式を実現

することができる。このため、ストレージ装置の拡張性に優れる、すなわちストレージ装置にメモリノードを容易に追加することが可能である。その他の構成及び効果は、前述した第１実施形態と同様である。

【０１１８】

［第４実施形態］

第４実施形態では、転送機能を有するメモリノードを複数備えたストレージ装置と、ストレージ装置に接続された制御コンピュータからなるストレージシステムにおいて、パケット転送時の渋滞が発生しない、つまり渋滞フリーとなるシステムの運用条件について説明する。

【０１１９】

［１］ストレージシステムの構成

図１７は、第４実施形態のストレージシステムの構成を示す図である。

【０１２０】

図示するように、ストレージシステムは、複数のメモリノード１１を含むストレージ装置１０、及びストレージ装置１０に接続された複数の制御コンピュータ４１を備える。ストレージ装置１０は、図１に示したように、データ転送機能を有するメモリノードが相互に複数接続された構成を有する。ストレージ装置１０の外周部に配置されたメモリノード１１には、制御コンピュータ４１が接続されている。

【０１２１】

［２］ストレージシステムの渋滞発生フリーの運用条件

図１７を参照して、ストレージシステムにおけるパケット転送シミュレーションの枠組みを説明する。ストレージ装置１０に接続された制御コンピュータ４１から送信されたリクエストパケットが送信先アドレスのメモリノードに到達するまでの過程、および目的のメモリノードから返送されたデータパケットが元の制御コンピュータ４１に到着するまでの過程をシミュレートする。パケットの転送アルゴリズムとしては、図１３を用いて説明したバイパス機能付ルーティングアルゴリズム（転送アルゴリズム３）を使用する。

【０１２２】

ストレージ装置１０のメモリノード数をN_{node}＝Nx×Nyとし、ストレージ装置１０に接続されている制御コンピュータ数をNcとする。メモリノード間で１回のパケット転送が行われる時間を単位ステップ時間とする。制御コンピュータ４１が、単位ステップ時間当たりに、リクエストパケットを送信する確率をRrとする。Rrは最大で１である。バイパス転送発生率をR_{bypass}、渋滞発生率をR_{jam}とする。

【０１２３】

バイパス転送発生率R_{bypass}は、メモリノードが１つのパケットの転送を試みた際に、バイパス転送により渋滞が回避された確率を示す。一方、渋滞発生率R_{jam}は、メモリノードが１つのパケットの転送を試みた際に、第一優先方向の出力ポートバッファのみならず、第二優先方向の出力ポートバッファも別のパケットにより占有されていたために、パケット転送を実施できず渋滞が発生し、パケットが滞留した確率を示す。また、負荷係数をR_{load}とする。負荷係数R_{load}は、１つのメモリノードに存在しているパケット数の平均である。

【０１２４】

図１８は、バイパス転送発生率R_{bypass}と負荷係数R_{load}の関係、ならびに渋滞発生率R_{jam}と負荷係数R_{load}の関係を示す。多数のプロットは、メモリノード数N_{node}、制御コンピュータ数をNc、確率Rrを様々に変化させた場合の結果を示している。

【０１２５】

渋滞発生率R_{jam}は負荷係数R_{load}が０．２より小さい場合ほぼゼロであり、渋滞フリーのシステム運用条件は負荷係数R_{load}＜０．２であることが分かる。渋滞発生率R_{jam}は負荷係数R_{load}が０．２以上になると増加し始める。しかし、負荷係数R_{load}＝２であっても渋滞発生率R_{jam}は０．０５程度であり、実用上ほとんど問題にならない。一方、渋滞発生率R_{jam}は負荷係数R_{load}が２より大きくなると急増する。これは、負荷係数R_{l}

$_{oad}$が２より大きくなると連鎖的に渋滞が発生するハングアップ現象が起こるためである。

【０１２６】

バイパス転送発生率R_{bypass}は負荷係数R_{load}の増加に対して、渋滞発生率R_{jam}よりも一桁早く立ち上がっている。つまり、バイパス転送アルゴリズムは渋滞発生要件を一桁程度改善する効果がある。

【０１２７】

次に、このシミュレーション結果に基づき、渋滞フリーもしくは渋滞発生が実用上問題とならない、制御コンピュータ数とメモリノード数の関係を求める。平均パケット滞在ステップ時間をS_{avg}とする。平均パケット滞在ステップ時間S_{avg}は、制御コンピュータ４１がリクエストパケットを送信してからそのリクエストパケットが送信先アドレスのメモリノードに到達するまでのステップ時間$S_{request}$と、目的のメモリノードから返送されたデータパケットが元の制御コンピュータに到着するまでのステップ時間S_{dat}の和である。ステップ時間$S_{request}$とS_{dat}は、渋滞が発生しない場合（Ｎｘ／２＋Ｎｙ／２）である。従って、平均パケット滞在ステップ時間S_{avg}は（Ｎｘ＋Ｎｙ）であり、これはおよそ$2\sqrt{N_{node}}$である。

【０１２８】

一方、制御コンピュータ４１が、１ステップ時間当たりに送信するリクエストパケットの総数は、Ｎｃ×Ｒｒである。従って、ストレージシステム内に存在するパケット総数は平均で$Nc\times Rr\times 2\sqrt{N_{node}}$である。このため、$R_{load}\sim Nc\times Rr\times 2\sqrt{N_{node}}/N_{node}=Nc\times Rr\times 2/\sqrt{N_{node}}$である。ここで、渋滞発生要件を$R_{load}{}^{limit}$とすると、好ましいシステム運用条件は、$R_{load}<R_{load}{}^{limit}$となる。従って、渋滞フリーもしくは渋滞発生が実用上問題とならない最大の制御コンピュータ数Nc^{max}は、$Nc^{max}<R_{load}{}^{limit}\times\sqrt{N_{node}}/(Rr\times 2)$となる。Ｒｒは最大１であるから、より厳しい条件は、$Nc^{max}<R_{load}{}^{limit}\times\sqrt{N_{node}}/2$である。

【０１２９】

前述したように、パケット転送時に渋滞発生フリーの運用条件は$R_{load}{}^{limit}=0.2$であり、渋滞が実用上問題とならない運用条件は$R_{load}{}^{limit}=2$である。従って、好ましい最大の制御コンピュータ数Nc^{max}は$Nc^{max}<\sqrt{N_{node}}$であり、より好ましくは$Nc^{max}<0.1\times\sqrt{N_{node}}$である。よって、$Nc<\sqrt{N_{node}}$という関係が成立している。その他の構成及び効果は、前述した第１実施形態と同様である。

【０１３０】

［第５実施形態］

第５実施形態は、図１に示した転送機能を有する複数のメモリノードを相互に接続したストレージ装置を備え、各メモリノードが分散処理機能を有する。

【０１３１】

［１］ストレージ装置の構成

図１９は、第５実施形態のストレージ装置の構成を示す図である。

【０１３２】

図示するように、ストレージ装置は、図１に示した複数のメモリノード１１を含むストレージ装置１０を備えると共に、各メモリノード１１はアドレス変換器４２を備え、分散処理機能を有している。

【０１３３】

［２］ストレージ装置の分散処理機能

ここでは、図１９を参照して、appleとorangeというキーワードを共に含むＵＲＬを見つけるＡＮＤ検索処理を例に挙げて、分散処理機能を説明する。

【０１３４】

ここで、メモリノード（１，３）と（４，３）がそれぞれappleとorangeの転置ファイルを保持しているものとする。転置ファイルとは、キーワードの一つ一つについて作成される索引ファイルで、例えばappleというキーワードに対応する転置ファイルにはappleを

含む全ＵＲＬのリストが格納される。

【０１３５】

ゲートウェイサーバは、ＡＮＤ検索命令をappleとorangeの転置ファイルを管理するメモリノードに送信する。転置ファイルを管理するメモリノードは、転置ファイルの内容をkey-value型データにマッピングする。ここで、keyとvalueはそれぞれＵＲＬとその転置ファイルのキーワードとする。

【０１３６】

各メモリノードは、自らが保有するkey-value型データの各レコードについて、keyをアドレス変換器４２によりアドレスに変換し、そのアドレスにvalueを含むパケットを送信する。アドレス変換器４２は、あるルールに従い、自らkeyからアドレスを算出する場合もあるし、keyをアドレスに変換する機能を有するサーバに問い合わせて、keyに対応するアドレスを得ても良い。

【０１３７】

例えば、アドレス変換器４２は、次のアドレッシングルールに従い、keyからアドレスを算出する。

【０１３８】

アドレス＝hash(key) mod N

ここで、hash()は暗号学的なハッシュ関数、Ｎはストレージ装置１０内のメモリノード数、mod ＮはＮを法とする剰余演算を表す。

【０１３９】

また例えば、アドレス変換器４２は、keyに対応するアドレスを別のサーバに問い合わせ、そして別のサーバがコンシステントハッシングによりkeyをアドレスに変換して元のアドレス変換器４２に回答する。

【０１４０】

図１９に示す例では、appleとorangeの転置ファイルは共にＵＲＬ１を含むとし、ＵＲＬ１はアドレス（２，１）に変換されたものとする。この場合、メモリノード（２，１）が、ＵＲＬ１がＡＮＤ検索式を満たすかどうかを判定する役割を担うことになる。メモリノード（２，１）には、valueがappleとorangeである２つのパケットが届く。このため、メモリノード（２，１）はＵＲＬ１がＡＮＤ検索条件を満たしていることが分かる。

【０１４１】

メモリノード（２，１）はその旨をクエリーを出したゲートウェイサーバに通知する。同様の判定を転置ファイルに記載されたＵＲＬすべてについて多数のメモリノードが行うことにより、クエリーを出したゲートウェイサーバはＡＮＤ検索条件を満たしたＵＲＬのリストを得ることができる。

【０１４２】

以上のＡＮＤ検索演算を、単一メモリノードで行う場合は、次のコードで示す演算を行う必要がある。

【数１】

$$\begin{aligned} &\mathrm{for}(i=0;i<N_{apple};i++)\\ &\quad \mathrm{for}(j=i;j<N_{orange};j++) \qquad \cdots(1)\\ &\quad\quad \mathrm{if}(URLlist_{apple}[i]==URLlist_{orange}[j]) \end{aligned}$$

【０１４３】

ここでは、appleとorangeの転置ファイルがそれぞれ格納するＵＲＬの数を、それぞれN_{apple}とN_{orange}とし、==は一致演算を示す。メモリノードは一致演算を、$N_{apple}\times N_{orange}/2$回だけ繰り返す必要がある。本実施形態のストレージ装置では、この一致演算を数多くのメモリノードで分散して実施することができる。

【０１４４】

以上説明したように第５実施形態によれば、第１実施形態に記載したストレージング機

能だけでなく、分散処理（分散コンピューティング）機能も併せ持つことができる。その他の構成及び効果は、前述した第１実施形態と同様である。

【０１４５】

［第６実施形態］

第６実施形態は、転送機能を有するメモリノードを複数備えたストレージ装置と、ストレージ装置の一部のメモリノードが、隣接するメモリノード（以下、隣接ノード）と接続された入出力ポート（以下、隣接ポート）以外の入出力ポート（以下、非隣接ポート）を有し、前記一部のメモリノードが、前記非隣接ポートにより、制御コンピュータもしくは隣接しないメモリノード（以下、非隣接ノード）と接続された構成を有する。

【０１４６】

［１］ストレージシステムの構成

図２０は、第６実施形態のストレージシステムの構成を示す図である。

【０１４７】

図示するように、ストレージシステムは、複数のメモリノード１１を含むストレージ装置１０、及びストレージ装置１０に接続された複数のゲートウェイサーバ２１Ａ，２１Ｂを備える。ストレージ装置１０は、図１に示したように、データ転送機能を有するメモリノードが相互に複数接続された構成を有する。ストレージ装置１０の外周部に配置されたメモリノード（１，４）には、アダプタ２２Ａを介してゲートウェイサーバ２１Ａが接続されている。ストレージ装置１０の中央部に配置されたメモリノード（３，５）は、隣接ノードと接続された入出力ポート以外の入出力ポート（非隣接ポート）を有し、この非隣接ポートにはアダプタ２２Ｂを介してゲートウェイサーバ２１Ｂが接続されている。

【０１４８】

［２］ストレージシステムのパケット転送

図２０を参照して、隣接ノードと接続された入出力ポート以外の入出力ポート（非隣接ポート）を有するメモリノードを導入した効果について述べる。ここで、ストレージ装置１０のｘ方向のメモリノード数をＮｘ、ｙ方向のメモリノード数をＮｙとする。

【０１４９】

ゲートウェイサーバの接続先がストレージ装置１０の外周部に配置されたメモリノードに限定される場合、ゲートウェイサーバから目的のメモリノードに到達するまでのパケットの平均転送回数は、Ｎｘ／２＋Ｎｙ／２である。

【０１５０】

一方、図２０に示すように、ストレージ装置１０の中央部に位置するメモリノード（３，５）の入出力ポート数を増やし、追加した入出力ポート（前記非隣接ポート）によって、メモリノード（３，５）をゲートウェイサーバ２１Ｂと接続する。すると、ゲートウェイサーバ２１Ｂから目的のメモリノードまでのパケットの平均転送回数は、Ｎｘ／４＋Ｎｙ／４となる。

【０１５１】

このように、ストレージ装置１０内の一部のメモリノードの入出力ポート数を増やし、追加した入出力ポートでメモリノードとゲートウェイサーバを接続すると、パケットが目的のメモリノードに到達するまでの平均転送回数を低減させることができる。

【０１５２】

また、図２１に示すように、追加した入出力ポートによってゲートウェイサーバに接続されたメモリノードをストレージ装置１０に複数配置しても良い。すなわち、メモリノード（４，７）は隣接ノードと接続された入出力ポート以外の入出力ポート（非隣接ポート）を有し、この非隣接ポートはアダプタ２２Ｃを介してゲートウェイサーバ２１Ｃに接続されている。

【０１５３】

さらに、ストレージ装置１０内のメモリノードに接続された複数のゲートウェイサーバ２１Ａ，２１Ｂ，２１Ｃ，２１Ｄ同士を、サーバとメモリノードとを接続する配線とは別の配線２３で接続しても良い。また、ゲートウェイサーバ間、例えばゲートウェイサーバ

２１Ａ，２１Ｂ間にアダプタ２２Ｅを接続しても良い。これにより、配線２３を使用して、ゲートウェイサーバ間、及びゲートウェイサーバとメモリノード間でパケット転送を行うことができる。例えば、ゲートウェイサーバ２１Ａ→ゲートウェイサーバ２１Ｂ→ゲートウェイサーバ２１Ｃ→メモリノード（４，７）の順でパケットを転送してもよい。また、ゲートウェイサーバ２１Ａ→ゲートウェイサーバ２１Ｄ→メモリノード（５，４））の順でパケットを転送してもよい。

【０１５４】

以上説明したように第６実施形態によれば、ゲートウェイサーバとストレージ装置内のメモリノードとの間でパケット転送を行う際に、転送時間を短縮させることができる。その他の構成及び効果は、前述した第１実施形態と同様である。

【０１５５】

［第７実施形態］

第７実施形態は、転送機能を有するメモリノードを複数備えたストレージ装置と、ストレージ装置の一部のメモリノードが、隣接ノードと接続された入出力ポート以外の入出力ポート（非隣接ポート）を有し、前記一部のメモリノードが、前記非隣接ポートにより、制御コンピュータもしくは隣接しないメモリノード（非隣接ノード）と接続された構成を有する。

【０１５６】

［１］ストレージシステムの構成

図２２は、第７実施形態のストレージシステムの構成を示す図である。

【０１５７】

図示するように、ストレージシステムは、複数のメモリノード１１を含むストレージ装置１０、及びストレージ装置１０に接続されたゲートウェイサーバ２１Ａを備える。ストレージ装置１０は、図１に示したように、データ転送機能を有するメモリノードが相互に複数接続された構成を有する。ストレージ装置１０の外周部に配置されたメモリノード（７，１）には、アダプタ２２Ａを介してゲートウェイサーバ２１Ａが接続されている。

【０１５８】

ストレージ装置１０の内部に配置されたメモリノード（３，３）、（３，８）、（８，３）、（８，８）は、隣接ノードと接続された入出力ポート以外の入出力ポート（非隣接ポート）を有する。非隣接ポート間は配線２４により接続されている。例えば、メモリノード（３，３）と（３，８）間、メモリノード（３，８）と（８，８）間、メモリノード（８，８）と（８，３）間、及びメモリノード（８，３）と（３，３）間が、配線２４により相互に接続されている。

【０１５９】

非隣接ノードに接続されたメモリノードは、物理的な位置によって決まるアドレスに加えて、非隣接ノードと接続されたメモリノード同士の相対的な物理位置によって決まる付加アドレスを持つ。非隣接ノードと接続されたメモリノードは、自ノード以外のメモリノードを宛先とするパケットを受信した場合、自ノードのパケットコントローラが、自ノードの付加アドレス、もしくは自ノード以外の、非隣接ノードに接続されたメモリノードの付加アドレスのいずれか一つを少なくとも含む情報に基づいてパケットの出力ポートを決定する。

【０１６０】

［２］ストレージシステムのパケット転送

まず、隣接しないメモリノード（非隣接ノード）同士の接続が、ストレージ装置１０内のパケット転送回数に与える影響について説明する。前述したように、一部の非隣接ノードが非隣接ポートによって相互接続されたストレージ装置を図２２に、全メモリノードの入出力ポート数が同じで、非隣接ノード間の接続が無いストレージ装置を図２３に示す。

【０１６１】

これらストレージ装置において、アドレス（７，０）のゲートウェイサーバ２１Ａから、メモリノード（７，８）にデータパケットを送付するケースを考える。

【０１６２】
非隣接ノード間の接続が無い場合（図２３）、最短のデータ転送経路は、ゲートウェイサーバ２１Ａ（７，０）→メモリノード（７，１）→（７，２）→（７，３）→（７，４）→（７，５）→（７，６）→（７，７）→（７，８）である。よって、ゲートウェイサーバ２１Ａからメモリノード（７，８）にパケット転送を行う際の（ゲートウェイサーバとメモリノード間、及びメモリノードとメモリノード間の）総パケット転送回数は８回である。
【０１６３】
他方、非隣接ノード間が相互接続されている場合（図２２）、ゲートウェイサーバ２１Ａ→メモリノード（７，１）→（７，２）→（７，３）→（８，３）→（８，８）→（７，８）が最短経路になり、６回の転送でパケットをゲートウェイサーバ２１Ａから送信先メモリノード（７，８）に送ることができる。
【０１６４】
このように、一部の隣接しないメモリノード同士を増設した入出力ポートによって接続し、それら入出力ポート間の接続配線を用いてパケット転送を行うことにより、送信先ノードに到達するまでのパケット転送回数を低減でき、転送時間を短縮することができる。
【０１６５】
但し、図２２に示すように、非隣接ノード間が接続されたストレージ装置では、メモリノードとメモリノード間、或いはメモリノードとゲートウェイサーバ２１Ａ間の物理的な距離と、それらのデータ転送上の距離が必ずしも一致しない点に注意が必要である。
【０１６６】
前述したように、アドレス（７，０）のゲートウェイサーバ２１Ａから、メモリノード（７，８）のにデータパケットを送付する場合において、非隣接ノード間の接続が存在しないストレージ装置（図２３）では、パケットの最低総転送回数（８回）は、パケット送信元と送信先の物理的な位置で決まるアドレス（ゲートウェイサーバ２１Ａ（７，０）、メモリノード（７，８））の差に等しい。
【０１６７】
他方、前述したように、非隣接ノード間が接続された図２２のストレージ装置では、同じ物理アドレスを有するゲートウェイサーバ２１Ａ（７，０）～送信先メモリノード（７，８）間のパケット通信を、６回の転送で行うことができる。このように、非隣接ノード間が接続されたストレージ装置では、物理的な最短距離と、データ転送上の最短距離が異なるため、物理的な位置によって決まるアドレスのみでパケットの最短転送経路を決定することはできない。
【０１６８】
そこで、非隣接ノード間が接続されたストレージ装置では、これら非隣接ノードに接続されたメモリノード同士の相対的な物理位置を反映したアドレス(以下、サブアドレスと記す)を非隣接ノードに接続されたメモリノードに追加付与し、全メモリノードに与えられた、物理位置を反映したアドレス(以下、メインアドレスと記す)とサブアドレスの両者からデータ転送経路・パケット転送先を決定すれば、前述したアルゴリズムに従って効率良くパケット転送を行うことができる。なお、図２２では、メインアドレスを丸括弧で、サブアドレスを四角括弧で表示している。また、前記非隣接ノードに接続されたメモリノードは、例えば図２２に示したメモリノード（３，３）、（３，８）、（８，３）、（８，８）を指す。
【０１６９】
但し、前述したように効率良くパケット転送を行うためには、パケットのヘッダー部に必要なアドレス情報が全て書かれていること、転送先決定に用いられる、パケットの送信元アドレス及び送信先アドレスを適宜書き換えること等が必要である。これらについて、次に説明する。
【０１７０】
非隣接ノードに接続されたメモリノードを有するストレージシステムでパケット転送を

行う場合、望ましいパケットのヘッダー部のアドレス情報の一例を図２４に示す。

【０１７１】

ここで、最終送信先ノードと送信元ノードは、それぞれパケットを最終的に届ける宛先のノードと、パケットを最初に作成・発信するノードを意味する。他方、一時送信先ノードアドレスと一時送信元ノードアドレスは、各メモリノードでパケット転送先決定に用いられるアドレスであり、パケットの通信過程で書き換えられる。このときの書き換えルールについては後述する。

【０１７２】

なお、一時送信先／送信元アドレスのタイプは、そのアドレスが、メインアドレスとサブアドレスのどちらであるか区別するためのものである。また、第１中継ノード、第２中継ノードは、パケット送信時に経由すべき非隣接ノードに接続されたメモリノードであり、送信元ノードに最も近い非隣接ノードに接続されたメモリノードが第１中継ノード、最終送信先ノードに最も近い非隣接ノードに接続されたメモリノードが第２中継ノードである。

【０１７３】

図２２に示したように、隣接しないメモリノードと一部のメモリノードが接続されたストレージ装置を含むシステムで、非隣接ノード間の接続を経由したパケット転送を行う場合、パケットを受信したメモリノードにおけるパケット転送先決定やパケットのヘッダー部の修正等は、例えば次のようなルールで行うことが好ましい。

【０１７４】

（ａ）追加の入出力ポート（非隣接ポート）を有しないメモリノードがパケットを受信した場合、パケットコントローラがパケットのヘッダー部に記録されたアドレス情報を調べ、

(i) 最終送信先ノードのアドレスが自ノードのアドレスと一致する場合には、パケット転送を行わない。

【０１７５】

(ii) 最終送信先ノードのアドレスが自ノードのアドレスと異なる場合には、一時送信元ノード、一時送信先ノード、自ノードのメインアドレスを参照して転送先を決定し、隣接したメモリノードにパケットを送信する。

【０１７６】

（ｂ）非隣接ノードに接続されたメモリノードがパケットを受信した場合、パケットコントローラはパケットのヘッダー部に記録されたアドレス情報を調べ、

(i) 自ノードのアドレスが最終送信先ノードのアドレスと一致する場合には、パケット転送を行わない。

【０１７７】

(ii) 自ノードのアドレスが、最終送信先ノードのアドレスと異なり、

(1) 第１中継ノードのアドレスと一致する場合は、パケットのヘッダー部の一時送信先、一時送信元のアドレスを、それぞれ第２中継ノードのサブアドレス、自ノードのサブアドレスに書き換える。さらに、一時送信元ノード、一時送信先ノード、自ノードのサブアドレスを参照して転送先を決定し、他の非隣接ノードに接続されたメモリノードにパケットを送信する。

【０１７８】

(2) 第２中継ノードのアドレスと一致する場合は、一時送信先アドレスを最終送信先ノードのメインアドレス、一時送信元アドレスを自ノードのメインアドレスに変更する。さらに、一時送信元ノード、一時送信先ノード、自ノードのメインアドレスを参照して転送先を決定し、隣接した他のメモリノードにパケットを転送する。

【０１７９】

(3) 第１中継ノード、第２中継ノードのいずれのアドレスとも異なり、さらにパケットのヘッダー部に書かれた一時送信先アドレスと一時送信元アドレスがサブアドレスである場合には、一時送信元ノード、一時送信先ノード、自ノードのサブアドレスを参照し

て転送先を決定し、他の非隣接ノードに接続されたメモリノードにパケットを転送する。
【０１８０】
(4) 第１中継ノード、第２中継ノードのいずれのアドレスとも異なり、さらに一時送信先アドレスと一時送信元アドレスがメインアドレスである場合には、一時送信元ノード、一時送信先ノード、自ノードのメインアドレスを参照して転送先を決定し、他の隣接したメモリノードにパケットを転送する。
【０１８１】
次に、非隣接ノード間の接続を併用してパケット送信を行う例として、図２２に示したシステムで、クライアントがゲートウェイサーバ２１Ａを介してストレージ装置にファイルを書き込む手続きを説明する。
【０１８２】
図２５は、図２２に示したストレージシステムにおける書き込み動作を示す図である。
【０１８３】
まず、クライアントは、ファイルとファイルＩＤとをゲートウェイサーバ２１Ａに送信する（（１）参照）。ファイルＩＤは、ファイルを一意に特定できる識別子である。
【０１８４】
次に、ゲートウェイサーバ２１Ａは、ファイルを規定サイズのデータに分割し、分割した各データに分割データＩＤを割り振る。さらに、ファイルＩＤと分割データＩＤをファイルテーブルに書き込む。分割データＩＤは、分割されたデータを一意に特定できる識別子である（（２）参照）。
【０１８５】
次に、ゲートウェイサーバ２１Ａは、分割データＩＤの情報に基づき、分割データを書き込む書き込み先のメモリノード（以下、書き込みノード）のアドレスを決定する（図２２ではメインアドレス（７，８）のメモリノード)。
【０１８６】
次に、下記手順で、ゲートウェイサーバ２１Ａから書き込みノードにパケットを転送する際に、通信時間(パケット転送回数)が最小となる経路を求める：
1．ゲートウェイサーバ２１Ａに最も近い、非隣接ノードに接続されたメモリノード(第１中継ノード)と、書き込みノードに最も近接した非隣接ノードに接続されたメモリノード(第２中継ノード)のそれぞれのアドレスを調べる。図２２で（７，８）に書き込む場合、ゲートウェイサーバ２１Ａと書き込みノードのそれぞれの最近接の非隣接ノードに接続されたメモリノードは、メインアドレス（８，３）、（８，８）のメモリノードである。
【０１８７】
2．非隣接ノード間の接続を含む最短経路と、含まない最短経路を求める。図２２でメモリノード（７，８）に書込む場合、前者は、ゲートウェイサーバ２１Ａ→メモリノード（７，１）→（７，２）→（７，３）→（８，３）（［２，１］）→（８，８）（［２，２］）→（７，８）であり、後者は、ゲートウェイサーバ２１Ａ→メモリノード（７，１）→（７，２）→（７，３）→（７，４）→（７，５）→（７，６）→（７，７）→（７，８）である。
【０１８８】
3．隣接ノードとのみ接続されたメモリノード間の経路についてはメインアドレスで、非隣接ノードに接続されたメモリノード間の経路は、サブアドレスを元にして、それぞれの経路でパケット送信を行う際に生じる転送回数を算出する。図２２でメモリノード（７，８）に書き込む場合、非隣接ノードに接続されたメモリノードを介する最短経路の転送回数は６回であり、非隣接ノードに接続されたメモリノードを介さない最短経路では、転送回数は８回である。
【０１８９】
4．パケット転送回数が小さい方の経路をデフォルト経路として決定する（（３）参照）。以下では、非隣接ノード間の接続を含む経路の方が、含まない経路よりもパケット転送回数が少ないケースについて説明する。

【０１９０】
・アドレスデータ、書き込み命令等からなるヘッダー部を、書き込みデータに付加した書き込みパケットをゲートウェイサーバ２１Ａで作成する。なお、パケットのヘッダー部に記録されたアドレス情報の一時送信先は第１中継ノードのメインアドレス（図２２では（８，３））とし、一時送信元はパケット送信元のメインアドレス（図２２では（７，０））とする。
【０１９１】
・作成した書き込みパケットを、ゲートウェイサーバ２１Ａからゲートウェイサーバ２１Ａに接続されたメモリノードに転送する（（４）参照）。
【０１９２】
・前述したアルゴリズムに従って、一時送信先メモリノード（図２２ではメインアドレス（８，３）の第１中継ノード）に到達するまで、メモリノード間で書き込みパケットの転送を行う(各メモリノードは一時送信先、一時送信元、自ノードのメインアドレスを参照して転送先を決定する)（（５）参照）。
【０１９３】
・書き込みパケットを受け取った第１中継ノードは、パケットのヘッダー部を読み、パケットの最終送信先ノードが他のメモリノードであり、第１中継ノードが自ノードであることから、一時送信先を第２中継ノードのサブアドレス（図２２では［２，２］）、一時送信元を自ノードのサブアドレス（図２２では［２，１］））に変更する（（６）参照）。さらに、隣接した非隣接ノードに接続されたメモリノードにパケットを転送する（（７）参照）。
【０１９４】
・前述したアルゴリズムに従い、一時送信先メモリノードである第２中継ノードに到達するまで、非隣接ノードに接続されたメモリノード間で書き込みパケットの転送を行う(各メモリノードは一時送信先、一時送信元、自ノードのサブアドレスを参照して転送先を決定する)。
【０１９５】
・書き込みパケットを受信した第２中継ノードは、パケットのヘッダー部を読み、最終送信先が他のメモリノードで、第２中継ノードが自ノードであることから、一時送信先を最終送信先ノードのメインアドレス（図２２では書き込みノード（７，８））、一時送信元を自ノードのメインアドレス（図２２では（８，８））に変更する（（８）参照）。さらに、隣接したメモリノードに書き込みパケットを転送する。
【０１９６】
・一時送信元ノード、一時送信先ノード、自ノードのメインアドレスを参照して、前述したアルゴリズムによりメモリノード間で転送が繰り返され、最終送信先である書き込みノード（図８ではメインアドレス（７，８））に書き込みパケットが到達する（（９）参照）。
【０１９７】
・パケットを受信した書き込みノードは、最終送信先が自ノードであることから、パケットを転送せず、ヘッダー部のアドレス情報や書き込みデータ等を自ノードのメモリ１６に書き込む（（１０）参照）。その後、宛先や中継先を逆にした書き込み完了報告のパケットを作成する。このパケットでは、発信元ノードを書き込みノードの（７，８）とし、最終送信先ノードをゲートウェイサーバ２１Ａの（７，０）、第１中継ノードをメインアドレス（８，８）（サブアドレス［２，２］）のノード、第２中継ノードをメインアドレス（８，３）（サブアドレス［２，１］）のノードとする。その後、書き込み完了報告のパケットを前述した手順と同様の手続きで、ゲートウェイサーバ２１Ａへ返信する（（１１）参照）。全分割データの書き込みが終わった後、ゲートウェイサーバ２１Ａはクライアントにファイル書き込みの終了報告を行う。
【０１９８】
前述した例では、隣接ノードと接続された入出力ポート以外の入出力ポート（非隣接ポ

ート）を有し、隣接しないメモリノード同士が非隣接ポートによって相互接続されたメモリノードを有するストレージ装置にクライアントがファイルを書き込む手順を説明したが、メモリノードに書き込まれたデータの読み出し・消去やメモリの空き容量確認などのため、読み出し命令・消去命令や空き容量の返信命令等をゲートウェイサーバからストレージ装置のメモリノードに送る場合も、非隣接ノード間の接続を介して通信を行うことでパケット転送時間を短縮することができる。さらに、読み出したデータや問合せ結果などをメモリノードからゲートウェイサーバに送る場合などにおいても、非隣接ノード間の接続を介して通信を行うことでパケット転送時間を短縮することができる。

【０１９９】

前述したように、隣接しないノード間の接続を介してデータ送信を行うことでパケットの平均転送回数を減らすことが可能である。なお、追加入出力ポート（非隣接ポート）によって非隣接ノードと接続されるメモリノードは、ストレージ装置内で均一に分散して配置されていることが好ましい。

【０２００】

例えば、１０×１０のメモリノードから構成されている図２２に示したストレージ装置では、５×５のメモリノードを含む４つの領域に分割（境界を点線で示す)し、各領域の中心に位置するメモリノード（３，３）、（８，３）、（３，８）、（８，８）同士を増設ポート（非隣接ポート）で接続している。この場合には、どの５×５のメモリノードブロックにも、１つの追加入出力ポート（非隣接ポート）を有するメモリノードが存在する。

【０２０１】

このように、含まれるメモリノード数が同じであるような、複数の領域にストレージ装置を分割し、その中心のメモリノードを追加入出力ポート付きのメモリノードとして相互接続することで、追加入出力ポート付きのメモリノードをストレージ装置内で均一に分散配置させることができる。

【０２０２】

なお、例えば、ａｌ×ｂｍのノードから構成されたストレージ装置をｌ×ｍずつの領域ａｂ個に分割し、各領域の中心に位置するメモリノード同士を追加入出力ポートにより接続する場合、（ｃｌ＋ｄ、ｅｍ＋ｆ）（ａ≧ｃ≧０、ｂ≧ｅ≧０、ｌ≧ｄ≧０、ｍ≧ｆ≧０）で表わされるメモリノードに最も近く、各領域の中心に位置する追加入出力ポート付きのメモリノードアドレスは、（ｃｌ＋round（ｌ／２，０）、ｅｍ＋round（ｍ／２，０））で表わされる。（round（Ａ，０）は、Ａの小数点以下を四捨五入する関数)。

【０２０３】

さて、ストレージ装置において、隣接しない一部のメモリノード間を接続することで生じる課題は、その接続を介して通信を行うパケットの数が増えると、パケット通信の渋滞が生じやすくなることである。

【０２０４】

これに対して、隣接したメモリノード間よりも、隣接しないメモリノード間のパケット通信速度を高くすることで、このパケット転送の渋滞を緩和させることができる。但し、この場合には、ゲートウェイサーバにおいて、パケット転送のデフォルト経路決定時に行う、非隣接ノード間の接続を介する最短経路と、非隣接ノード間の接続を経由しない最短経路の総パケット通信時間の算出に注意が必要である。総パケット通信時間は、メモリノード間の通信時間とパケット転送回数の積に等しい。従って、隣接したメモリノード間と隣接しないメモリノード間のデータ通信速度が等しい場合には、前述したように、各経路の転送回数でデータ通信時間を比較することができる。しかし、パケット送信速度が異なる場合には、転送回数だけで総パケット通信時間を比べることができない。

【０２０５】

通信速度の逆数（速度＝距離/時間であるから、これは通信時間に比例）と転送回数の積を参照することで、非隣接ノード同士の接続を経由する経路と、非隣接ノード同士の接続を経由しない経路の総パケット通信時間を正確に比較することができる。しかし、前述

した例のように、隣接したノード間の接続と非隣接ノード間の接続がパケット通信経路で混在している場合、この計算は煩雑なものとなる。

【０２０６】

隣接ノード間と非隣接ノード間でパケット通信速度が異なるストレージ装置では、付与するアドレスのステップをパケット通信速度に逆比例させ、そのアドレス差を転送回数と見なすのも一つの方法である。

【０２０７】

隣接ノード間よりも非隣接ノード間の方が、パケット通信速度が１０倍高い場合のアドレス付与例を図２６に示す。図２６では、隣接しないノードと接続されたメモリノードには、それら物理位置を反映させながら、あるメモリノードと、そのメモリノードと接続された非隣接ノードとの間で１ずつ異なるようにサブアドレスを付与している(四角括弧で表示)。さらに、図２６に示す全メモリノードには、互いの物理的な位置関係を反映させながら、隣り合うメモリノード間で、１０ずつ異なるメインアドレスを与えている(丸括弧で表示)。

【０２０８】

このようなストレージ装置で、非隣接ノードに接続されたメモリノード間でパケット送信を行う場合は各メモリノードのサブアドレスの差を計算し、隣接メモリノード間でパケット送信を行う場合はメインアドレスの差を計算し、それらを転送回数と見なす。転送回数の算出ルールをこのように定めれば、隣接ノード間と非隣接ノード間で接続速度が異なる場合においても、転送回数だけでパケット通信時間を見積もり、比較することができる。

【０２０９】

以上説明したように第７実施形態によれば、ゲートウェイサーバとストレージ装置内のメモリノードとの間でパケット転送を行う際に、転送時間を短縮させることができる。その他の構成及び効果は、前述した第１実施形態と同様である。

【０２１０】

［第８実施形態］

第８実施形態では、パケットの平均転送回数を減らすため、ゲートウェイサーバとストレージ装置との間にスイッチングリレーを追加したストレージシステムについて説明する。

【０２１１】

［１］ストレージシステムの構成

図２７は、第８実施形態のストレージシステムの構成を示す図である。

【０２１２】

図示するように、ストレージシステムは、複数のメモリノード１１を含むストレージ装置１０、ストレージ装置１０に接続されたスイッチングリレー８１、及びスイッチングリレー８１に接続されたゲートウェイサーバ２１Ａを備える。

【０２１３】

ストレージ装置１０は、図１に示したように、データ転送機能を有するメモリノードが相互に複数接続された構成を有する。スイッチングリレー８１は、ストレージ装置１０の一端側（図２７において左端）に配置されたメモリノード（１，１）、（１，２）、（１，３）、（１，４）、（１，５）、（１，６）、（１，７）、（１，８）、（１，９）の全てに接続されている。スイッチングリレー８１には、アダプタ２２Ａを介してゲートウェイサーバ２１Ａが接続されている。

【０２１４】

スイッチングリレー８１は、パケットのヘッダー部に記録されたアドレス情報に従い、受信したパケットを指定された送信先ノードに転送する。スイッチングリレー８１には、ストレージ装置内のメモリノードとは異なるアドレスが付与されている(図２７では四角括弧で表示)。

【０２１５】

［２］ストレージシステムのパケット転送
第８実施形態のストレージシステムにおいてパケットを転送する手順について説明する。
【０２１６】
ゲートウェイサーバ２１Ａから送り出されたパケットは、アダプタ２２Ａを通ってスイッチングリレー８１に入る。スイッチングリレー８１に入ったパケットは、スイッチングリレー８１と接続された、メモリノード（１，１）、（１，２）、（１，３）、（１，４）、（１，５）、（１，６）、（１，７）、（１，８）、（１，９）のいずれか一つに送られ、その後、送信先アドレスのメモリノードまで転送される。
【０２１７】
逆に、ストレージ装置１０内のメモリノードから発信されたパケットは、メモリノード（１，１）～（１，９）のいずれか一つのメモリノードに送られ、スイッチングリレー８１とアダプタ２２Ａを介してゲートウェイサーバ２１Ａに転送される。
【０２１８】
図２７に示したシステムにおいて、ゲートウェイサーバ２１Ａからスイッチングリレー８１とメモリノード（１，９）を介して、メモリノード（５，９）にパケットを送る場合は以下となる。最短経路はスイッチングリレー８１→メモリノード（１，９）→（２，９）→（３，９）→（４，９）→（５，９）で、ストレージ装置１０内の転送回数は４回である。
【０２１９】
他方、図２８に示すように、スイッチングリレー８１を介さずに、ゲートウェイサーバ２１Ａがストレージ装置１０のメモリノード（１，４）に接続されているシステムでは、ゲートウェイサーバ２１Ａからメモリノード（５，９）にパケットを送る場合は以下となる。ゲートウェイサーバ２１Ａ→メモリノード（１，４）→（１，５）→（１，６）→（１，７）→（１，８）→（１，９）→（２，９）→（３，９）→（４，９）→（５，９）が最短経路の一つであり、ストレージ装置１０内でパケットの転送は最低９回必要である。
【０２２０】
このように、ゲートウェイサーバ２１Ａとストレージ装置１０との間にスイッチングリレー８１を導入すると、パケットの転送回数を減らし、転送時間を短縮することができる。
【０２２１】
なお、ゲートウェイサーバ２１Ａからストレージ装置１０にパケットを送信するケースについて説明したが、ストレージ装置１０内のメモリノードに保存されたデータをゲートウェイサーバ２１Ａに送信する場合においても、ゲートウェイサーバ２１Ａとストレージ装置１０との間にスイッチングリレー８１が存在するシステム(図２７)の方が、アダプタを介してゲートウェイサーバ２１Ａとストレージ装置１０が直接接続されたシステム(図２８)よりも、一般に、パケットの平均転送回数が少なく、転送時間が短い。
【０２２２】
但し、スイッチングリレーを導入したシステムにおけるパケット転送回数は、スイッチングリレーとストレージ装置間でパケットの送受信を行う際に中継するメモリノードに依存する。例えば、前述したように、図２７に示したシステムでゲートウェイサーバ２１Ａがメモリノード（５，９）との間でパケット転送を行う場合、中継するメモリノードがメモリノード（１，９）であれば、ストレージ装置１０内の転送回数は最低４回である。しかし、他のメモリノードを中継した場合には、５回以上の転送が必要である。
【０２２３】
このように、パケットの転送回数を最小限にするためには、スイッチングリレーに直接接続されたメモリノードのうち、送信先ノードに最も近いメモリノードを中継メモリノードとして選択することが重要である。パケット転送時間が最短となる、中継メモリノードのアドレス算出手順については後述する。

【０２２４】
なお、スイッチングリレーを介さずにゲートウェイサーバとストレージ装置が接続されたシステムにおいて、ゲートウェイサーバとストレージ装置内のメモリノードとの間でパケット転送を行う場合、送信元と送信先は、ゲートウェイサーバ、メモリノードのいずれかであり、これはパケット転送が行われている間変更されない。
【０２２５】
他方、スイッチングリレーを介してゲートウェイサーバとストレージ装置が接続されたシステムにおいて、パケット転送を行う場合には、まず中継メモリノードを宛先としてパケット転送が行われる。パケットが中継メモリノードに到達した後は、メモリノード(ゲートウェイサーバからメモリノードにパケットを送信する場合)或いはスイッチングリレー(メモリノードからゲートウェイサーバにパケット送信する場合)を送信先としてパケット転送が行われる。すなわち、パケットが中継メモリノードに到達する前後で、パケットの送信先の変更が必要である。
【０２２６】
スイッチングリレーを使用したシステムにおいて、パケットのヘッダー部に記録されたアドレス情報の例を図２９に示す。ここで、最終送信先ノードと送信元ノードは、それぞれパケットを最終的に届ける宛先のメモリノード、パケットを最初に作成・発信するメモリノードを意味する。他方、一時送信先ノードアドレスと一時送信元ノードアドレスは、各メモリノードでパケット転送先の決定に用いられるアドレスであり、スイッチングリレーに接続された中継メモリノードで書き換えられる(書き換えルールについては後述する)。
【０２２７】
なお、アドレスタイプとは、そのアドレスが、ストレージ装置１０内のメモリノードのアドレスであるか、スイッチングリレー８１のアドレスであるかを、区別するためのものである。また、中継ノードアドレスは、パケット転送時に経由すべき、スイッチングリレー８１と接続されたメモリノードのアドレスである。
【０２２８】
前述したように、スイッチングリレーに接続されたメモリノードでは、パケットのヘッダー部に記録されたアドレス情報の書き換えが行われる。パケットのアドレス情報が図２９に示したものである場合、書き換えルールは例えば次のようなものである。
【０２２９】
1．最終送信先ノードが自ノード以外のメモリノードで、送信元がスイッチングリレーである場合、一時送信先を最終送信先ノードに、一時送信元を自ノードに変更する。
【０２３０】
2．最終送信先ノードがスイッチングリレーである場合、一時送信先をスイッチングリレーに、一時送信元を自ノードに変更する。
【０２３１】
3．最終送信先が自ノード以外のメモリノードで、送信元もメモリノードである場合、一時送信先、一時送信元共に変更しない。
【０２３２】
例えば、図２７に示したシステムで、クライアントがストレージ装置１０にファイルを書き込む手続きは次のようになる。
【０２３３】
図３０は、図２７に示したストレージシステムにおける書き込み動作を示す図である。
【０２３４】
まず、クライアントは、ファイルとファイルＩＤとをゲートウェイサーバ２１Ａに送信する（（１）参照）。ファイルＩＤは、ファイルを一意に特定できる識別子である。
【０２３５】
次に、ゲートウェイサーバ２１Ａは、ファイルを規定サイズのデータに分割し、分割した各データに分割データＩＤを割り振る。さらに、ファイルＩＤと分割データＩＤをファ

イルテーブルに書き込む。分割データＩＤは、分割されたデータを一意に特定できる識別子である（（２）参照）。

【０２３６】

次に、ゲートウェイサーバ２１Ａは、分割データＩＤの情報に基づき、分割データを書き込む書き込み先のメモリノード（書き込みノード）のアドレスを決定する（図２７ではアドレス（５，９））。さらに、スイッチングリレー８１に接続されたメモリノードのうち、書き込みノードに最も近いメモリノードを中継メモリノードとして、そのアドレスを算出する（図２７ではアドレス（１，９））（（３）参照）。

【０２３７】

次に、ゲートウェイサーバ２１Ａは、書き込みデータに、前述したアドレス情報を含むヘッダー部等を付け加えた書き込みパケットを作成する。ここで、一時送信先アドレスは中継ノードアドレス（図２７ではアドレス（１，９））、一時送信元ノードアドレスは、スイッチングリレー８１のアドレス（図２７では［１］）とする。その後、パケットをスイッチングリレー８１に送信する（（４）参照）。

【０２３８】

書き込みパケットを受け取ったスイッチングリレー８１は、指定された一時送信先メモリノード（中継メモリノード（１，９））にこの書き込みパケットを送信する（（５）参照）。

【０２３９】

スイッチングリレー８１から書き込みパケットを受信した中継メモリノードは、パケットのヘッダー部を読み、最終送信先ノードがストレージ装置１０内の他のメモリノードであることから、以下のようにヘッダー部を書き換えた書き込みパケットを作成する。ヘッダー部において、一時送信元アドレスを自ノードアドレスに、一時送信先ノードアドレスを、最終送信先である書き込みノードのアドレス（図２７では（５，９））に変更する（（６）参照）。

【０２４０】

その後、書き込みパケットを隣接したメモリノードに転送する。ストレージ装置１０内で適切な転送を繰り返すことにより、書き込みパケットを書き込みノード（図２７ではメモリノード（５，９））に到達させる（（７）参照）。

【０２４１】

書き込み先のメモリノードでは、受け取ったパケットの書き込みデータ、パケット送信元や中継ノードのアドレス等を自ノードのメモリ１６に書き込む（（８）参照）。その後、書き込み完了報告のパケットを作成し、上記と逆の経路でゲートウェイサーバ２１Ａに返信する（（９）参照）。書き込み完了報告のパケットでは、ヘッダー部のアドレス情報の最終送信先ノードはスイッチングリレーとし、送信元は書き込みノード、中継ノードはゲートウェイサーバ２１Ａから書き込みノードにパケット転送を行った際と同じメモリノード、一時送信先ノードは中継ノード、一時送信元ノードは書き込みノードとする。前記パケットは、中継メモリノードにおいて、一時送信先をスイッチングリレーに、一時送信元を中継ノードに、ヘッダー部のアドレス情報が書き換えられる（（１０）参照）。

【０２４２】

全分割データの書き込みが終わった後、ゲートウェイサーバ２１Ａは、クライアントにファイル書き込みの終了報告を行う。

【０２４３】

なお、クライアントがストレージ装置１０にファイルを書き込む場合だけでなく、書き込まれたデータの読み出し・消去やメモリの空き容量の確認などのため、読み出し・消去命令や空き容量の返信命令等をストレージ装置１０のメモリノードに送る場合も、前述した手続きに従ってスイッチングリレー８１に接続されたメモリノードを介してパケットを転送することができる。さらに、書き込みや消去等の命令実施完了報告、メモリノードから読み出したデータなどをメモリノードからゲートウェイサーバ２１Ａに送信する際などにも、同様に、前述した手続きに従ってスイッチングリレー８１に接続されたメモリノー

ドを介してパケットを転送することができる。これにより、転送時間を短縮することができる。

【０２４４】

なお、ここまでは、図２７に示したように、左端のメモリノードのみがスイッチングリレー８１と接続されたストレージシステムについて説明したが、図３１に示すように、全ての外周部のメモリノードがスイッチングリレー８１，８２，８３，８４と接続されたストレージシステムにおいても、スイッチングリレーを導入しないシステムより、パケットの平均転送回数が少なく、転送時間を短くすることができる。図３１に示すストレージシステムでは、ストレージ装置の左端に配置されたメモリノードにスイッチングリレー８１が接続され、上端に配置されたメモリノードにスイッチングリレー８２、右端に配置されたメモリノードにスイッチングリレー８３、下端に配置されたメモリノードにスイッチングリレー８４がそれぞれ接続されている。また、全てのスイッチングリレーは、アダプタ２２Ａを介してゲートウェイサーバ２１Ａと接続されている。

【０２４５】

さらに、図３２に示すように、隣接ノードと接続された入出力ポート以外の入出力ポート（非隣接ポート）を有する複数のメモリノードが、非隣接ポートによってスイッチングリレー８１と接続されたストレージ装置を含むストレージシステムにおいても、スイッチングリレーを導入しないシステムより、パケットの平均転送回数が少なく、転送時間を短くすることができる。図３２に示すストレージシステムでは、ストレージ装置１０内のメモリノード（３，２）、（３，７）、（８，２）、（８，７）にスイッチングリレー８１が接続されている。

【０２４６】

さて、前述したように、スイッチングリレーに接続されているメモリノードを介してパケット転送を行う場合、パケット転送時間(転送回数)が最小となるメモリノードを中継メモリノードとして選択することが重要である。以下では、その中継メモリノードのアドレス算出について説明する。

【０２４７】

まず、図２７に示したように、左端のメモリノード全てがスイッチングリレー８１に接続されたストレージ装置１０とゲートウェイサーバ２１Ａとの間でデータ転送を行う場合、ストレージ装置１０内のアドレス（ｘ，ｙ）のパケット送信先／送信元メモリノードにとって、パケット転送距離（転送回数）が最小となる、スイッチングリレー８１に接続されたメモリノードは、アドレス（１，ｙ）のメモリノードである。

【０２４８】

角に配置されたメモリノードのアドレスが（ａ０，ｂ０）で、横ｃ０個、縦ｄ０個のメモリノードから成るストレージ装置において、外周部の全メモリノードがスイッチングリレーに接続されている場合、ストレージ装置１０内のアドレス（ｘ，ｙ）のメモリノードがゲートウェイサーバ２１Ａとの間でパケット転送を行うとき、転送時間（転送回数）が最小となる中継メモリノードのアドレスは、以下の通りである。ここで、図３１に示すように、スイッチングリレーと接続されたメモリノードのアドレスは、（ａ０，ｙ）、（ａ０＋ｃ０，ｙ）、（ｘ，ｂ０）、（ｘ，ｂ０＋ｄ０）で表わされる（ｘはａ０～ａ０＋ｃ０の任意の整数値、ｙはｂ０～ｂ０＋ｄ０の任意の整数値）。

【０２４９】

min(ｘ－ａ０，ａ０＋ｃ０－ｘ)≦min(ｙ－ｂ０，ｂ０＋ｄ０－ｙ)かつmin(ｘ－ａ０，ａ０＋ｃ０－ｘ)＝ｘ－ａ０の場合は(ｘ－ａ０，ｙ)；

min(ｘ－ａ０，ａ０＋ｃ０－ｘ)≦min(ｙ－ｂ０，ｂ０＋ｄ０－ｙ)かつmin(ｘ－ａ０，ａ０＋ｃ０－ｘ)＝ａ０＋ｃ０－ｘの場合は(ａ０＋ｃ０－ｘ，ｙ)；

min(ｘ－ａ０，ａ０＋ｃ０－ｘ)≧min(ｙ－ｂ０，ｂ０＋ｄ０－ｙ)かつmin(ｙ－ｂ０，ｂ０＋ｄ０－ｙ)＝ｙ－ｂ０の場合は(ｘ，ｙ－ｂ０)；

min(ｘ－ａ０，ａ０＋ｃ０－ｘ)≧min(ｙ－ｂ０，ｂ０＋ｄ０－ｙ)かつmin(ｙ－ｂ０，ｂ０＋ｄ０－ｙ)＝ｂ０＋ｄ０－ｙの場合は(ｘ，ｂ０＋ｄ０－ｙ)。

【０２５０】
なお、min（ｘ，ｙ）は、二つの引数ｘ，ｙのうち、小さい方を与える関数である。
【０２５１】
他方、図３２に示すように、スイッチングリレー８１と接続されたメモリノードのアドレスが（ａｍ＋ｂ，ｃｎ＋ｄ）（ｍ，ｎは整数）で与えられる場合（図３２では、ａ＝５、ｂ＝３、ｃ＝５、ｄ＝２、ｍ＝０または１、ｎ＝０または１）、任意のメモリノード（アドレス（ｘ，ｙ））に対して、パケット転送時間（転送回数）が最小となる、スイッチングリレー８１に接続されるメモリノード（中継メモリノード）は、(ａ(round(ｘ／ａ，０)＋ｂ、ｃ(round(ｙ／ｃ，０)＋ｄ)で与えられる。ここで、round(ｕ，０)は、数値ｕを小数点で四捨五入する関数である。
【０２５２】
以上説明したように第８実施形態によれば、ストレージ装置内でメモリノード間のパケット転送回数を低減させることができる。その他の構成及び効果は、前述した第１実施形態と同様である。
【０２５３】
［第９実施形態］
第９実施形態は、複数のデータを複数のメモリノードに保存するデータ処理において、ゲートウェイサーバとのデータ転送時間が互いに異なるメモリノードを、複数データの保存先として選択するデータ処理手順を備える。
【０２５４】
［１］ストレージシステムの構成
図３３Ａは、第９実施形態のストレージシステムの構成を示す図である。
【０２５５】
図示するように、ストレージシステムは、複数のメモリノード１１を含むストレージ装置１０、及びストレージ装置１０に接続された複数のゲートウェイサーバ２１Ａを備える。ストレージ装置１０は、図１に示したように、データ転送機能を有するメモリノードが相互に複数接続された構成を有する。ストレージ装置１０の外周部に配置されたメモリノード（１，４）には、アダプタ２２Ａを介してゲートウェイサーバ２１Ａが接続されている。
【０２５６】
［２］ストレージシステムのデータ処理方法
第９実施形態のストレージシステムにおけるデータ処理手順について説明する。
【０２５７】
図３３Ａ～図３３Ｄ及び図３４Ａ～図３４Ｅは、ストレージ装置１０がアダプタ２２Ａを介してゲートウェイサーバ２１Ａと接続されたストレージシステムである。ストレージ装置は、隣接したメモリノード１１が相互接続された複数のメモリノードから構成される。
【０２５８】
各メモリノード１１は、送信されたパケットが自ノード宛であれば受け取り、他のメモリノードが宛先の場合は、隣接するメモリノードにパケットを転送する。このデータ転送機能により、ゲートウェイサーバ２１Ａと指定メモリノードとの間でパケット通信を行うことができる。
【０２５９】
但し、パケット転送に必要な転送回数は、メモリノードにより異なる。例えば、ゲートウェイサーバ２１Ａとのパケット通信に必要な最低転送回数は、アドレス（１，４）のメモリノードは０回、アドレス（１，５）、（２，４）、（１，３）のメモリノードは１回、アドレス（１，６）、（２，５）、（３，４）、（２，３）、（１，２）のメモリノードは２回である。図３３Ａ～図３３Ｄ及び図３４Ａ～図３４Ｅでは、同一の最低転送回数を持つメモリノードを同一のハッチングで表している。
【０２６０】

またここでは、ストレージ装置１０のメモリノード内のパケット転送時間とメモリノード間のパケット送信時間が、メモリノードに依らず一定とする。この場合、総パケット転送時間はどのメモリノード間でも同じである。なお、パケット転送時間は、入力ポートにパケットを受信した後、パケットのヘッダー部に記録されたアドレスから自ノード宛かどうかを判断し、出力ポートからパケットを出力するまでの時間である。総パケット転送時間は、メモリノードがパケットを受け取ってから隣接ノードにパケットを送信し、そのパケットが隣接ノードに到達するまでの時間である。

【０２６１】

さて、一つのファイルを３つに分割し、分割後のデータの順番に合わせてＩＤ＝１、ＩＤ＝２、ＩＤ＝３とＩＤを付与した３つのデータを、３つのメモリノードにそれぞれ保存する。その後、メモリノードから３つのデータを読み出して、ゲートウェイサーバに送信するデータ処理について考える。

【０２６２】

データを保存する３つのメモリノードが、ゲートウェイサーバとのパケット通信に必要な最低転送回数が同じになるメモリノードである場合と、最低転送回数が全て異なるメモリノードの場合のそれぞれについて、読み出されたデータのストレージ装置内の転送過程を比較する。なお、読み出し命令は一斉に行われ、読み出されたデータの送信は同時に開始されること、従って３つのデータの転送も全く同じタイミングで進行するものとする。

【０２６３】

ゲートウェイサーバ２１Ａとのパケット通信に必要な最低転送回数が３回であるメモリノード（１，１）、（２，２）、（３，５）に保存された３つのデータが、各メモリノードから読み出された後、ゲートウェイサーバ２１Ａに向けて転送されていく過程の一例を図３３Ａ～図３３Ｄに示す。

【０２６４】

図３３Ａは、各メモリノード内のメモリからデータが読み出された直後（転送前）のストレージ装置１０の状態を示し、図３３Ｂは、図３３Ａから一定時間経過し、３つのデータ全てが、保存されていたメモリノードの隣接ノードに転送された状態を示す。図３３Ｂに示すように、図３３Ａにおいてメモリノード（１，１）、（２，２）、（３，５）に保存されていた３つのデータは、メモリノード（１，２）、（２，３）、（２，５）にそれぞれ転送されている。

【０２６５】

図３３Ｃ、図３３Ｄは、それぞれデータの転送が２回、３回行われた状態を示す。図３３Ｃに示すように、図３３Ｂにおいてメモリノード（１，２）、（２，３）、（２，５）に保存されていた３つのデータは、メモリノード（１，３）、（２，４）、（１，５）にそれぞれ転送されている。さらに、図３３Ｄに示すように、図３３Ｃにおいてメモリノード（１，３）、（２，４）、（１，５）に保存されていた３つのデータは、メモリノード（１，４）にそれぞれ転送されている。なお図３３Ａ～図３３Ｄは、保存されていたメモリノードからゲートウェイサーバ２１Ａへ向けてデータ転送を行う過程であるため、時間の経過と共に、ゲートウェイサーバ２１Ａとのパケット通信に必要な、最低転送回数が少ないメモリノードにデータが転送されている。

【０２６６】

さて、図３３Ａ～図３３Ｃに示したストレージ装置１０では、３つのデータはそれぞれ別のメモリノードに存在しているが、図３３Ｄに示したストレージ装置１０では、全てのデータがアドレス（１，４）のメモリノード上に位置している。これは、３つのデータが同時にアドレス（１，４）のメモリノードに到達した状態を表している。

【０２６７】

但し、メモリノードの一時保存用メモリ（入力ポートバッファ）の記憶容量サイズが、複数のデータを保存できるほど大きくない場合には、メモリノード（１，４）では一度に一つのデータしか受け入れることができない。一時保存用メモリは、自ノード宛でないデータを受信した場合に、そのデータを転送するまでの間、保管するためのメモリである。

【０２６８】
この場合には、受信した一つのデータのアダプタ２２Ａへの転送が終了するまで、他のデータはアドレス（１，４）のメモリノードに転送できず、隣接ノードで待機する必要がある。
【０２６９】
メモリノードの一時保存用メモリのサイズが十分大きく、３つのデータを同時に一時保存できる場合においても、アドレス（１，４）のメモリノードからゲートウェイサーバ２１Ａに３つのデータを一度に送信することは通常できないため、一つのデータがゲートウェイサーバ２１Ａに送信されるまでの間、他のデータはアドレス（１，４）のメモリノードで待機する必要がある。このようなデータ転送の待機が生じると、全データの転送に要する時間が増加する。
【０２７０】
また、図３３Ａ～図３３Ｄからわかるように、ゲートウェイサーバ２１Ａとのパケット通信に必要な最低転送回数が同じであるメモリノードの数は、転送回数が少なくなるにつれて減少する。従って、図３３Ａに示すように、ゲートウェイサーバ２１Ａとのパケット通信に必要な最低転送回数が同じで異なるメモリノードに複数データを保存し、同時に読み出して転送を開始すると、時間が経過してデータがゲートウェイサーバ２１Ａに近づくほど、転送先のメモリノードの数が減少し、複数データが同一のメモリノードに転送される確率が高まる。このように、ゲートウェイサーバ２１Ａとのパケット通信に必要な最低転送回数が同じであるメモリノードに複数データを保存すると、読み出し後の転送時に無駄な待機時間が発生しやすくなる。
【０２７１】
図３４Ａ～図３４Ｅ、図３５Ａ、及び図３５Ｂを参照して、このような転送時の待機時間を低減するためのデータ処理手順を説明する。
【０２７２】
図３４Ａ～図３４Ｅは、メモリノードから読み出されたデータがゲートウェイサーバ２１Ａに向けて転送されていく過程の一例を示す図である。図３４Ａ～図３４Ｅは、ゲートウェイサーバ２１Ａとのパケット通信に必要な最低転送回数が１回であるメモリノード（１，５）にＩＤ＝２のデータを、最低転送回数が２回であるメモリノード（１，２）にＩＤ＝１のデータを、３回であるメモリノード（４，４）にＩＤ＝３のデータを保存させた後、これらデータを同時に読み出してゲートウェイサーバ２１Ａに転送する過程を示している。
【０２７３】
この場合も、ゲートウェイサーバ２１Ａとのパケット通信に必要な最低転送回数が少ないメモリノードへとデータは転送されている。しかし、図３４Ａ～図３４Ｅの場合、全ての過程で、図３３Ｄに示したような同一メモリノードへの複数データの転送が生じていない。これは、ゲートウェイサーバ２１Ａとのパケット通信に必要な最低転送回数が異なるメモリノードにデータを保存したためである。この場合にも、最低転送回数が少ないメモリノードへとデータ転送が行われるが、最低転送回数が異なるメモリノードから全データの転送が開始されるため、どの時間においても各データは最低転送回数が異なるメモリノードに位置する。
【０２７４】
このように、ゲートウェイサーバ２１Ａとのパケット通信に必要な最低転送回数が異なるメモリノードにデータを保存することにより、読み出しデータの転送時における複数データの同一メモリノードへの同時転送が回避され、データ転送時間を短縮することができる。
【０２７５】
なお、当然ではあるが、ゲートウェイサーバとのパケット通信に必要な最低転送回数が異なるメモリノードに保存したデータを読み出してゲートウェイサーバに転送する場合、データを保存したメモリノードがゲートウェイサーバに近い順番でゲートウェイサーバに

データが到達する。図３４Ａでは、ゲートウェイサーバ２１Ａに最も近いメモリノードにＩＤ＝２のデータを、次にサーバ２１Ａに近いメモリノードにＩＤ＝１のデータを、最もサーバ２１Ａから遠いメモリノードにＩＤ＝３のデータを保存しているため、これらメモリノードから読み出し・転送されたデータは、図３４Ｃ～図３４Ｅに示したように、ＩＤ＝２のデータ、ＩＤ＝１のデータ、ＩＤ＝３のデータの順番でゲートウェイサーバ２１Ａに到達する。

【０２７６】

前述したように、これらデータは一つのファイルを３分割したものであり、分割後のデータの順番に合わせてＩＤ＝１、ＩＤ＝２、ＩＤ＝３とＩＤを付与されている。このため、分割データからファイルを再構成するためには、ゲートウェイサーバ２１Ａに到達したデータをＩＤ順に入れ替える必要がある。

【０２７７】

他方、図３５Ａでは、ＩＤ＝１のデータをゲートウェイサーバ２１Ａからの最低転送回数１回のメモリノードに、ＩＤ＝２のデータを最低転送回数２回のメモリノードに、ＩＤ＝３のデータを最低転送回数３回のメモリノードに保存している。このため、これらのデータをメモリノードから読み出し、ゲートウェイサーバ２１Ａに転送すると、図３５Ｂに示すように、ＩＤ＝１のデータ、ＩＤ＝２のデータ、ＩＤ＝３のデータの順番でゲートウェイサーバ２１Ａに到達する。従って、ファイルの再構成を行う際に、データの並べ替えを行う必要が無い。

【０２７８】

このように、データの読み出し順番が意味を持つ複数のデータを、ゲートウェイサーバとのパケット通信に必要な最低転送回数が異なるメモリノードに保存する場合、その順番に合わせて、ゲートウェイサーバに近いメモリノードから遠いメモリノードへ順に保存する。これにより、ゲートウェイサーバ２１Ａに転送されたデータの並び替え作業を省略することができる。

【０２７９】

さて、全て等価なメモリノードから構成されたストレージ装置の単一メモリノードとゲートウェイサーバが、アダプタを介して接続されているストレージシステムで、ストレージ装置に複数データを保存する手順について説明したが、図３６Ａに示すように、隣接ノードと接続された入出力ポート以外の入出力ポート（非隣接ポート）を有し、非隣接ポートによって相互接続された複数のメモリノードを有するストレージ装置においても、ゲートウェイサーバ２１Ｂとのパケット通信に必要な最低転送回数が異なるメモリノードにデータを保存することで、保存データの転送時に無駄な待機時間が発生するのを回避することができる。

【０２８０】

さらに、図３６Ｂ～図３６Ｄに示すように、スイッチングリレーを介してストレージ装置１０とゲートウェイサーバ２１Ａが接続されている場合においても、ゲートウェイサーバ２１Ａとのパケット通信に必要な最低転送回数が異なるメモリノードにデータを保存することで、保存データの転送時に無駄な待機時間が発生するのを回避することができる。なお、図３６Ａ～図３６Ｄでは、ＩＤ＝１、ＩＤ＝２、ＩＤ＝３のデータを、パケット通信に必要な最低転送回数が１回、２回、３回のメモリノードにそれぞれ保存した例を示している。

【０２８１】

なお前述したように、メモリノードのアドレスのハッシュ値とパケットＩＤのハッシュ値から、パケットデータを保存するメモリノードのアドレスを決める方法（コンシステントハッシング）がある。これは例えば、０～$2^{160}-1$の整数値を取るＩＤ空間を考え、メモリノードとパケットのそれぞれに対して、前者はアドレス、後者はパケットＩＤで暗号学的ハッシュ関数ＳＨＡ－１を計算（計算結果は０～$2^{160}-1$のいずれかの整数値になる）することで、それらをＩＤ空間のいずれかのＩＤに割り当てる。次に、各パケットに対して、ハッシュ値のＩＤからＩＤ空間を時計回りにたどり、最初に行き当たった

メモリノードを、そのパケットを保存するメモリノードとして決定するものである。

【０２８２】

本方法は、高い均一性でパケットをメモリノードに分散保存できるだけでなく、メモリノードを増減させても、変更が必要なパケットが少ない（すなわち、メモリノードのスケーラビリティが高い）特長を有している。しかし、この方法でパケットを保存するメモリノードを決めると、ゲートウェイサーバとの通信時に必要な転送回数が同じメモリノードにパケットが保存されうる(前述したように、これは、データ読み出し時にパケット衝突を引き起こす)。

【０２８３】

ゲートウェイサーバとの通信でパケット転送回数が異なるメモリノードに、Ｎ個のパケットを均一に分散させて保存するためには、例えばパケットＩＤのハッシュ値mod Ｎを計算して（modは割り算の余りを返す関数）、そのパケットを保存するメモリノードの転送回数（０～（Ｎ－１）のいずれか）を決める（計算の結果、転送回数が一致するパケットがあれば、異なる値になるように調整する）。さらに、各パケットに対して、決められた転送回数を有するメモリノードの中から、そのパケットを保存するメモリノードを前述したコンシステントハッシングにより決めれば良い。

【０２８４】

なお、このようなパケット保存先のメモリノードを決定する手順では、決められた転送回数を有するメモリノードのアドレス把握が必要である。ある転送回数を有するメモリノードのアドレスは次のように表わされる。

【０２８５】

図３３Ａ～図３５Ｂのように、ゲートウェイサーバ２１Ａが、アダプタ２２Ａを介してストレージ装置（ここでは、メモリノードのアドレスは全て正の整数値とする）の端に位置するメモリノード（１，ａ０）と接続されているシステムにおいて、メモリノード（１，ａ０）との間でパケット通信を行う場合に、最低転送回数がｎ回であるメモリノードのアドレスは、(１＋ｂ，ａ０＋(ｎ－ｂ))(ｎ≧ｂ≧０)、及び(１＋ｃ，ａ０－(ｎ－ｃ))(ａ０－１≧ｎ－ｃ≧０)で表わされる。

【０２８６】

図３６Ａ、図３６Ｄ示したように、増加ポート（非隣接ポート）によって、隣接しないメモリノード同士やメモリノードとスイッチングリレーが接続されたストレージ装置（メモリノードのアドレスは正の整数値とする）において、増加ポートを有するノード（ａ０，ｂ０）との間でパケット通信を行う場合に、最低転送回数がｎ回であるメモリノードのアドレスは、(ａ０＋ｃ，ｂ０＋(ｎ－ｃ))(ｎ≧ｃ≧０)、(ａ０＋ｄ，ｂ０－(ｎ－ｄ))(ｎ≧ｄ≧０，ｂ０－１≧ｎ－ｄ)、(ａ０－ｅ，ｂ０＋(ｎ－ｅ))(min(ａ０－１，ｎ)≧ｅ≧０)、(ａ０－ｆ，ｂ０－(ｎ－ｆ))(ａ０－１≧ｆ≧０，ｂ０－１≧ｎ－ｆ≧０)で表わされる。

【０２８７】

図３６Ｂに示すように、ストレージ装置（全メモリノードのアドレスは正の整数値とする）の左端（アドレス（１，ｙ））のメモリノードがスイッチングリレーと接続されているシステムにおいて、アドレス（１，ａ０）の中継メモリノードとの間でパケット通信を行う場合に、最低転送回数がｎ回であるメモリノードのアドレスは、（１，ａ０＋ｎ）で表わされる。

【０２８８】

なお、データ転送時間が同じメモリノードが相互接続されたストレージ装置を含むシステムについて説明してきたが、図３６Ｅに示すようなデータ転送時間が同じメモリノードがツリー状に接続されたストレージシステムや、図３６Ｆに示すようなデータ転送時間の異なるメモリノードから成るストレージシステムなど、サーバとのデータ通信時間が異なるメモリノードが存在するストレージ装置を含むシステムであれば、本実施形態を適用することができる。

【０２８９】

図３６Ｅに示すシステムは、データ転送時間が同じメモリノードがツリー状に接続されたストレージシステムであり、ゲートウェイサーバ２１Ａとの通信時間が同じメモリノードを同一のハッチングで示している。図３６Ｆに示すシステムでは、メモリノード間とメモリノード～スイッチングリレー間のデータ通信速度は同じであり、さらにアドレス(a，l)、(a，m)、(a，n)、(b，l)、(b，m)、(b，n)のメモリノード間、及びこれらメモリノード～スイッチングリレー８１間のデータ通信速度に対して、アドレス(α，δ)、(β，γ)、(β，ε)のメモリノード間、及びこれらメモリノード～スイッチングリレー８１間のデータ通信速度は２倍、アドレス(Ａ，Ｂ)とスイッチングリレー８１間の通信速度は４倍の場合を示している。なお、スイッチングリレー８１とのデータ転送時間が同じメモリノードを同一のハッチングで示している。但し、メモリノード内でのデータ転送時間は、メモリノード間及びスイッチングリレー～メモリノード間のデータ転送時間より十分に小さいと仮定する。

【０２９０】

以上説明したように第９実施形態によれば、複数のメモリノードに保存したデータを各メモリノードから読み出した後、ゲートウェイサーバに送信する際の転送時間を短縮することができる。これにより、高速なデータ読み出しが可能となる。その他の構成及び効果は、前述した第１実施形態と同様である。

【０２９１】

［第１０実施形態］

第１０実施形態は、ゲートウェイサーバから複数のデータを複数のメモリノードに送信するデータ処理において、データ通信時間が長いメモリノードを宛先とするデータから、データ通信時間が短いメモリノードを宛先とするデータの順に、データ転送を行うデータ処理手順を備える。

【０２９２】

［１］ストレージシステムの構成

図３７Ａは、第１０実施形態のストレージシステムの構成を示す図である。

【０２９３】

図示するように、ストレージシステムは、複数のメモリノード１１を含むストレージ装置１０、及びストレージ装置１０に接続された複数のゲートウェイサーバ２１Ａを備える。ストレージ装置１０は、図１に示したように、データ転送機能を有するメモリノードが相互に複数接続された構成を有する。ストレージ装置１０の外周部に配置されたメモリノード（１，４）には、アダプタ２２Ａを介してゲートウェイサーバ２１Ａが接続されている。

【０２９４】

［２］ストレージシステムのデータ処理方法

第１０実施形態のストレージシステムにおけるデータ処理手順について説明する。

【０２９５】

図３７Ａ～図３７Ｇ及び図３８Ａ～図３８Ｅは、第９実施形態と同様に、ストレージ装置がアダプタ２２Ａを介してゲートウェイサーバ２１Ａと接続されたストレージシステムである。ストレージ装置は、隣接したメモリノード１１が相互接続された複数のメモリノードから構成される。

【０２９６】

ここでは、ゲートウェイサーバ２１Ａから、ＩＤ＝１、ＩＤ＝２、ＩＤ＝３の３つのデータを、それぞれアドレス（１，５）、（１，２）、（４，４）のメモリノードに送信する場合の、ゲートウェイサーバ２１Ａからのデータ送信順番と、全データ送信に必要な時間の関係について考える。なお、メモリノードの一時保存用メモリ（入力ポートバッファ）にパケットは１つしか保存できず、また一度に１つのパケットしか、ゲートウェイサーバ２１Ａからアドレス（１，４）のメモリノードに送信できないものとする。

【０２９７】

図３７Ａ～図３７Ｇでは、ゲートウェイサーバ２１Ａに近いメモリノードを宛先とする

データから遠いメモリノードを宛先とするデータの順番（すなわち、ＩＤ＝１のデータ、ＩＤ＝２のデータ、ＩＤ＝３のデータの順）で、ゲートウェイサーバ２１Ａからストレージ装置にパケットを送信した場合のパケット転送過程を示している。

【０２９８】

この場合当然のことながら、図３７Ｂ～図３７Ｃに示すように、ゲートウェイサーバ２１Ａに最も近い、アドレス（１，５）のメモリノード宛のデータ（ＩＤ＝１）送信は最も早く終了する。

【０２９９】

しかし、ゲートウェイサーバ２１Ａから最も遠い、アドレス（４，４）のメモリノードを宛先としたデータ（ＩＤ＝３）は、図３７Ｄに示すように、２つのデータ（ＩＤ＝１、ＩＤ＝２）が、アドレス（１，４）のメモリノードから隣接メモリノードに転送されてしまうまで、アドレス（１，４）に送信することができず、それまでの間、ゲートウェイサーバ２１Ａで待機する必要がある。その待機が終わり、アドレス（１，４）のメモリノードに送信されても、図３７Ｄ～図３７Ｇに示すように、アドレス（４，４）までデータが到達するためには最低３回の転送が必要である。

【０３００】

他方、図３８Ａ～図３８Ｅでは、ゲートウェイサーバ２１Ａから最も遠いメモリノードを宛先とするデータから、近いメモリノードを宛先とするデータの順番（すなわちＩＤ＝３のデータ、ＩＤ＝２のデータ、ＩＤ＝１のデータの順）で、ゲートウェイサーバ２１Ａからストレージ装置にパケットを送信した場合のパケット転送過程を示している。

【０３０１】

この場合においても、図３８Ｂ～図３８Ｄに示すように、ＩＤ＝３のデータとＩＤ＝２のデータがアドレス（１，４）のメモリノードから隣接ノードに転送されてしまうまで、ＩＤ＝１のデータはアドレス（１，４）のメモリノードに送信できず、ゲートウェイサーバ２１Ａで待機する必要がある。

【０３０２】

しかし、後から発信されるデータ（ＩＤ＝１）の宛先であるアドレス（１，５）のメモリノードは、ゲートウェイサーバ２１Ａに近く、少ない転送回数で宛先のメモリノードに到達する。他方、先にゲートウェイサーバ２１Ａから送信されたデータ（ＩＤ＝３）は、宛先がゲートウェイサーバ２１Ａから離れているが、他のデータがゲートウェイサーバ２１Ａで待機している間に転送が開始されているため、その分早く宛先に到達する。これらの結果、図３８Ｅに示すように、３つのデータは、同時に宛先のメモリノードに到達する。

【０３０３】

図３７Ａ～図３７Ｇと図３８Ａ～図３８Ｅの比較からわかるように、宛先のメモリノードがゲートウェイサーバ２１Ａから離れたデータから、ゲートウェイサーバ２１Ａから近いデータの順番で、ゲートウェイサーバ２１Ａからデータ送信を行うことで、全てのデータ送信に必要な時間を最小にすることができる。

【０３０４】

なお、データ転送時間が同じメモリノードが相互接続されたストレージ装置を含むシステムについて説明してきたが、データ転送時間が同じメモリノードがツリー状に接続されたストレージシステム(図３６Ｅ)や、データ転送時間の異なるメモリノードから成るストレージシステム(図３６Ｆ)など、サーバとのデータ通信時間が異なるメモリノードが存在するストレージ装置に複数のデータを送信する場合には、本実施形態を同様に適用することができる

以上説明したように第１０実施形態によれば、ゲートウェイサーバとの通信時間が異なる複数のメモリノードに複数のデータを送る際、必要となる通信時間を最小にすることができる。その他の構成及び効果は、前述した第１実施形態と同様である。

【０３０５】

以上述べたように第１～第１０実施形態によれば、メモリノードが経路指定表を管理す

る必要がなく、効率的にパケットを転送することができるストレージ装置、及びデータ処理方法を提供することができる。

【０３０６】

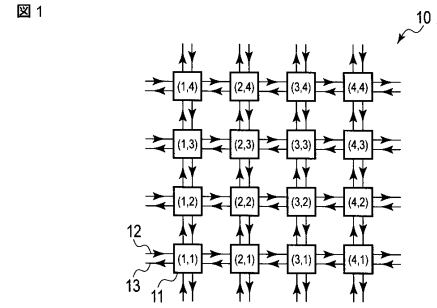
本発明のいくつかの実施形態を説明したが、これらの実施形態は、例として提示したものであり、発明の範囲を限定することは意図していない。これら新規な実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。これら実施形態やその変形は、発明の範囲や要旨に含まれるとともに、特許請求の範囲に記載された発明とその均等の範囲に含まれる。

【符号の説明】

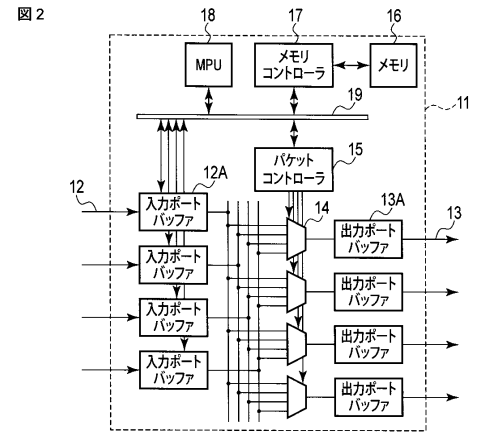
【０３０７】

１０…ストレージ装置、１１…メモリノード、１２…入力ポート、１２Ａ…入力ポートバッファ、１３…出力ポート、１３Ａ…出力ポートバッファ、１４…セレクタ、１５…パケットコントローラ、１６…メモリ、１７…メモリコントローラ、１８…ＭＰＵ、１９…ローカルバス、２０…ストレージシステム、２１Ａ，２１Ｂ，２１Ｃ，２１Ｄ…ゲートウェイサーバ、２２Ａ，２２Ｂ，２２Ｃ，２２Ｄ，２２Ｅ…アダプタ、３１Ａ，３１Ｂ１，３１Ｂ２…クライアント、４１…制御コンピュータ、４２…アドレス変換器、２３，２４…配線、８１，８２，８３，８４…スイッチングリレー。

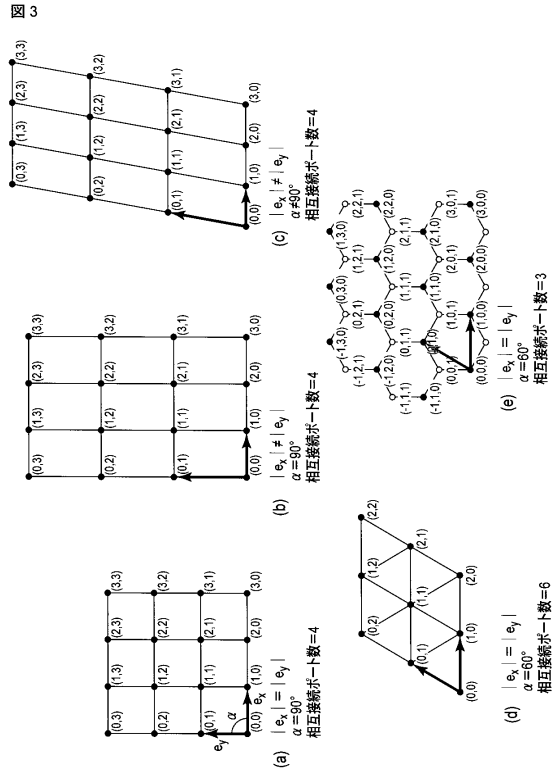
【図１】



【図２】

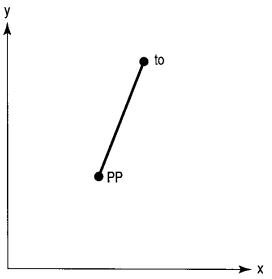


【図３】



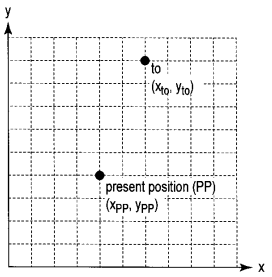
【図４】

図4



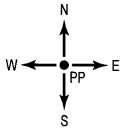
【図５Ａ】

図5A



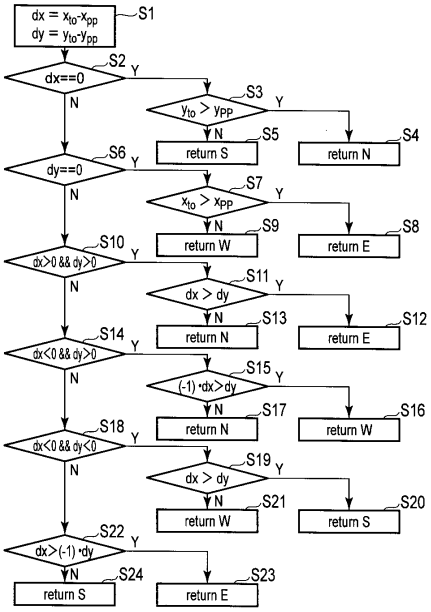
【図５Ｂ】

図5B



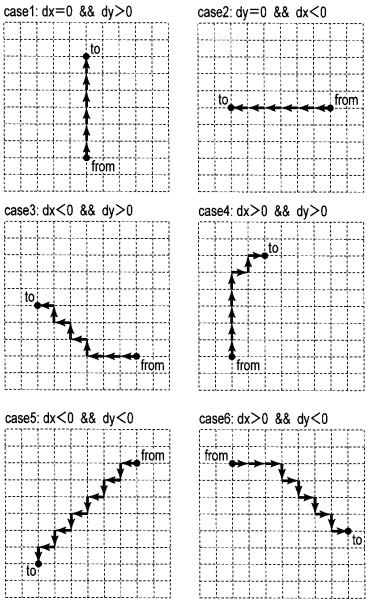
【図５Ｃ】

図5C



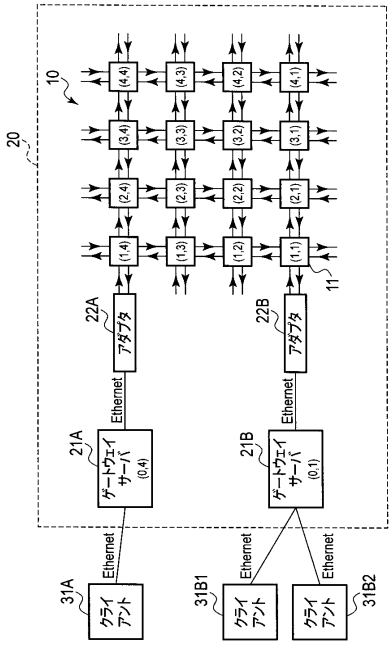
【図５Ｄ】

図5D



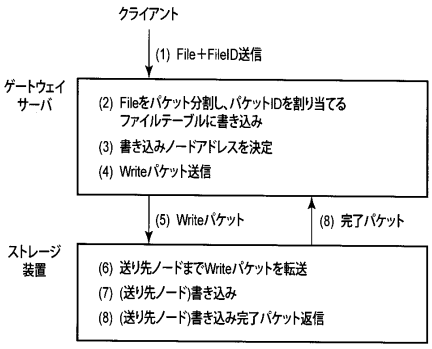
【図６】

図6



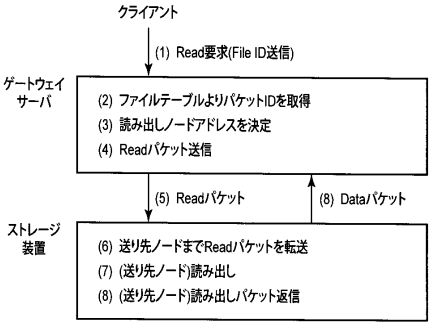
【図７】

図 7



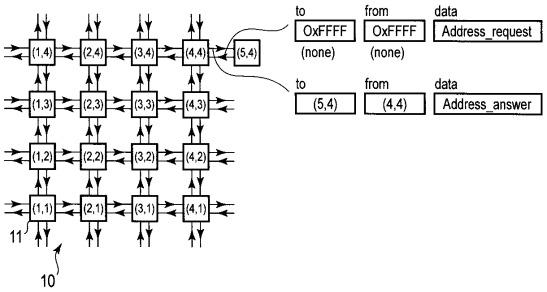
【図８】

図 8



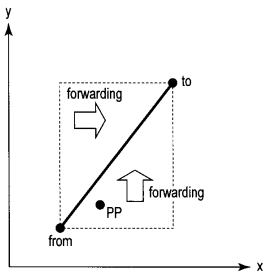
【図９】

図 9



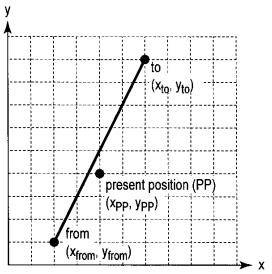
【図１０】

図 10



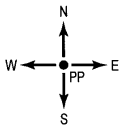
【図１１Ａ】

図 11A



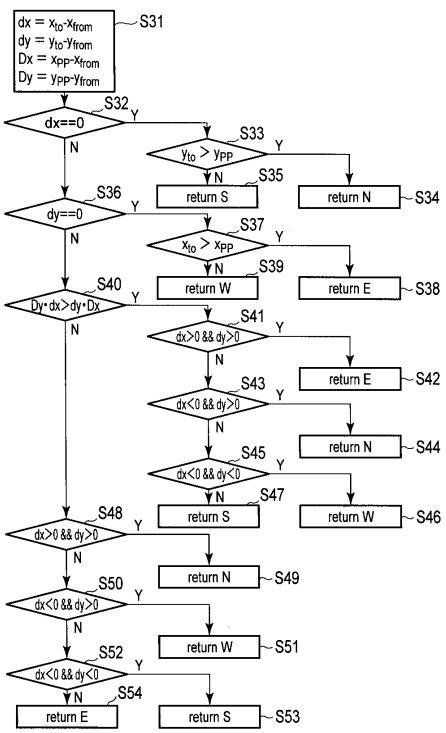
【図１１Ｂ】

図 11B



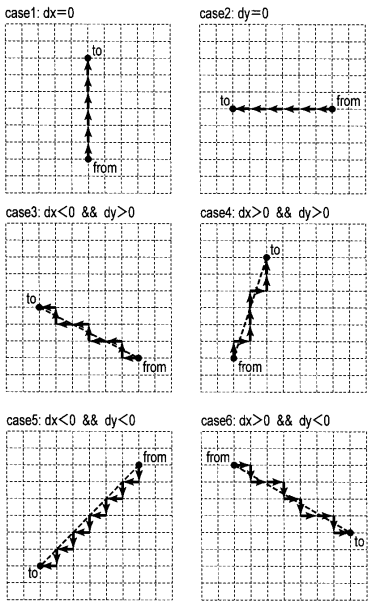
【図１１Ｃ】

図 11C



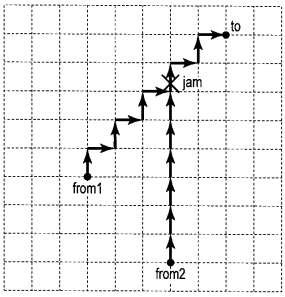
【図１１Ｄ】

図 11D



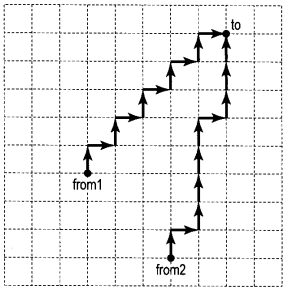
【図１２Ａ】

図 12A



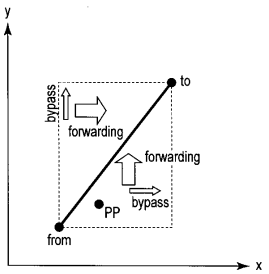
【図１２Ｂ】

図 12B



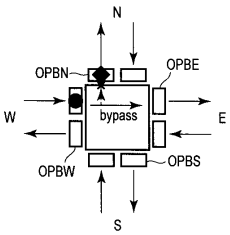
【図１３】

図 13



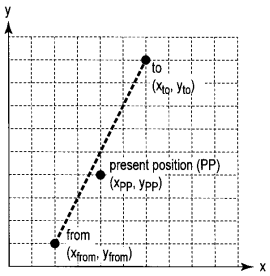
【図１４Ｂ】

図 14B



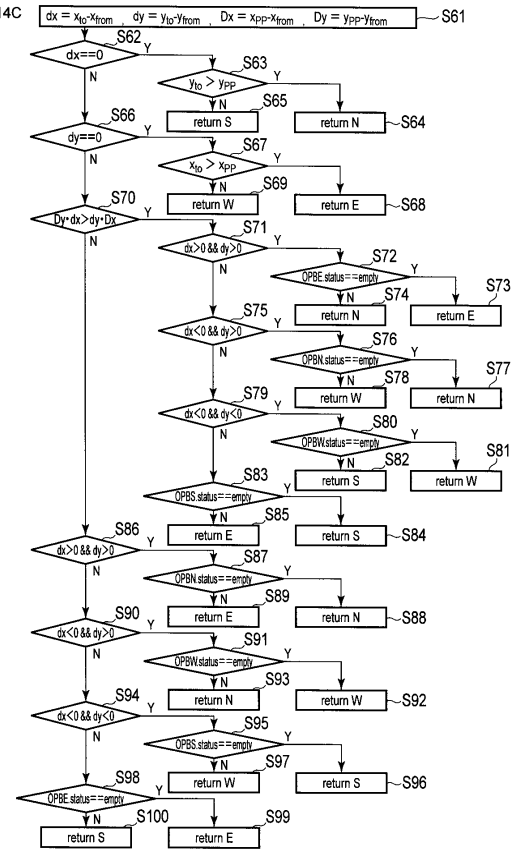
【図１４Ａ】

図 14A



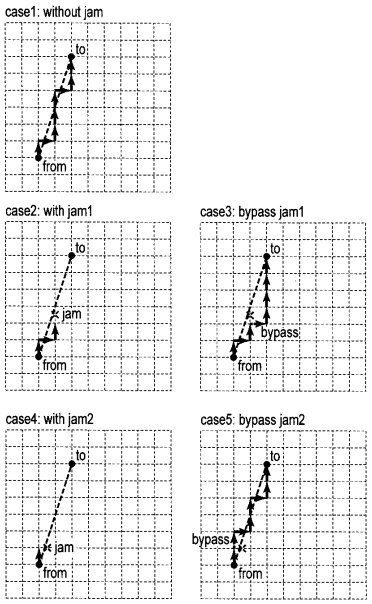
【図１４Ｃ】

図 14C



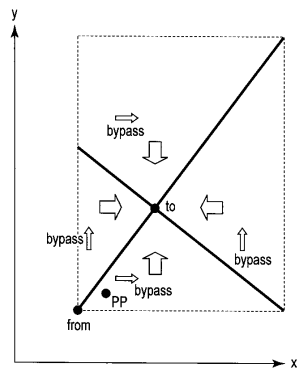
【図１４Ｄ】

図 14D



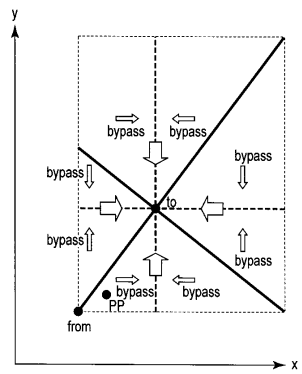
【図１５】

図 15



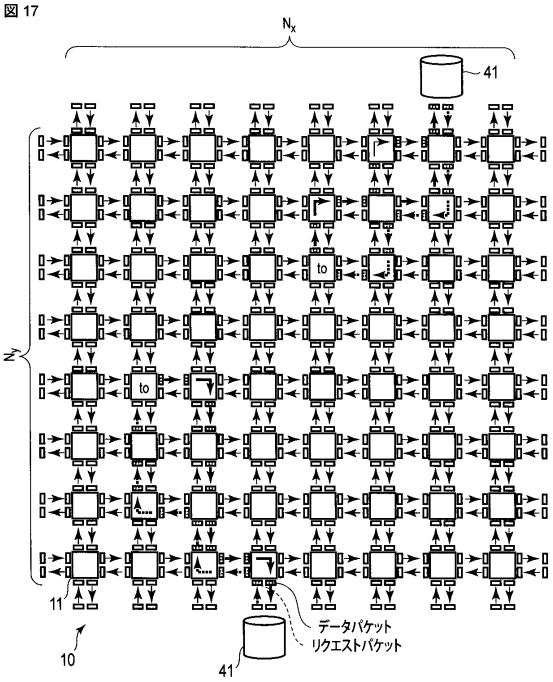
【図１６】

図 16



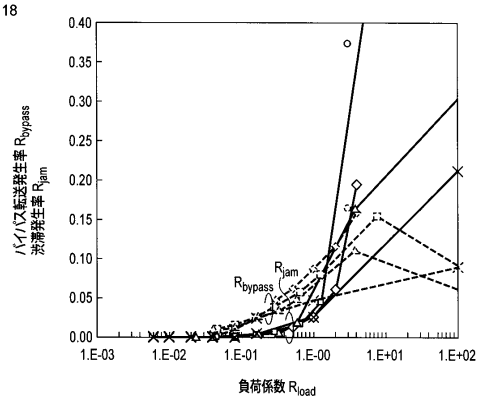
【図１７】

図 17



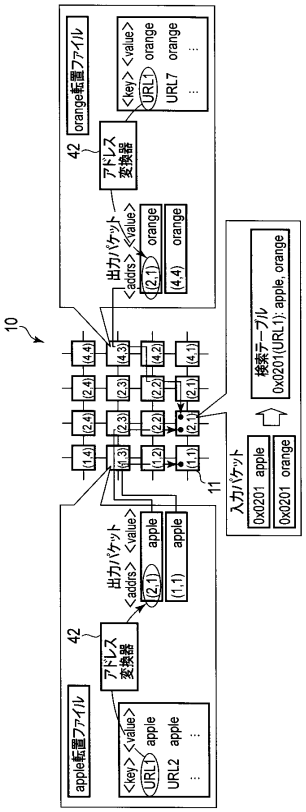
【図１８】

図 18



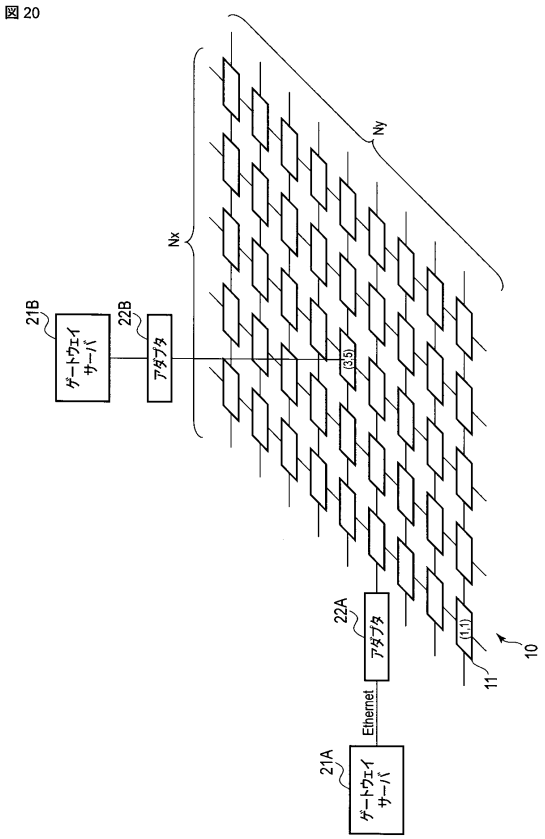
【図１９】

図 19



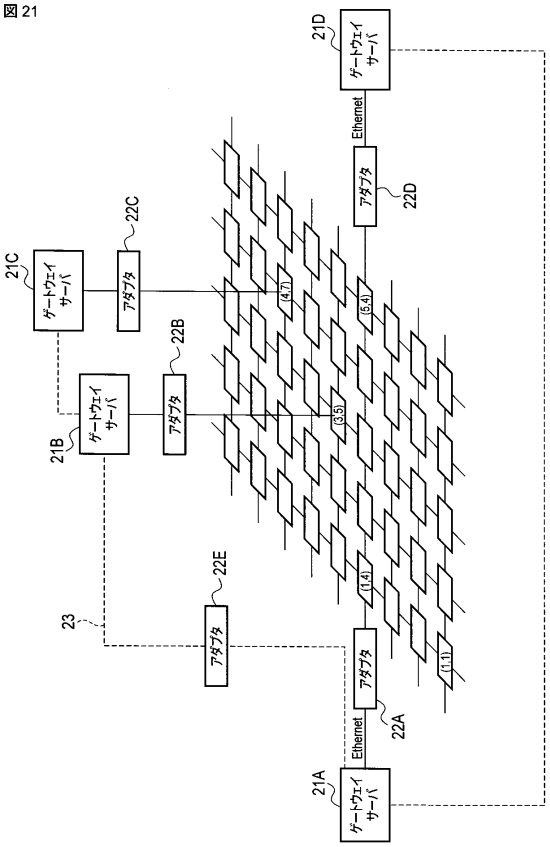
【図２０】

図 20



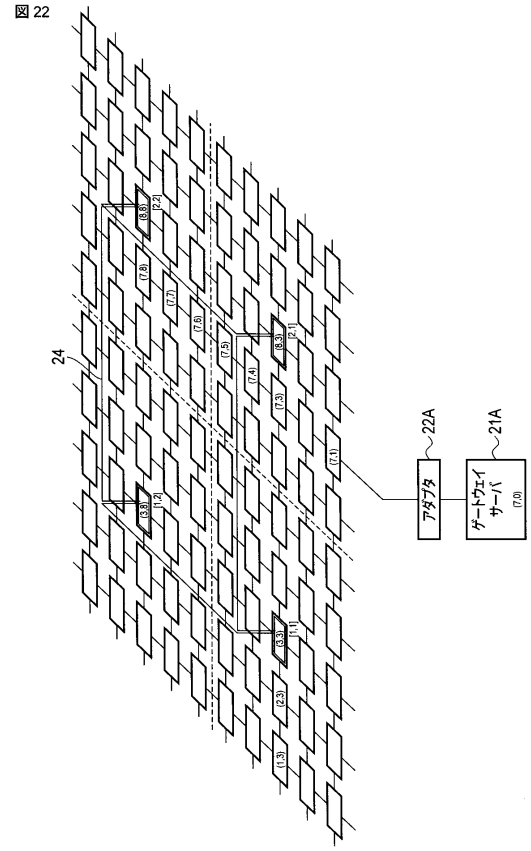
【図２１】

図 21



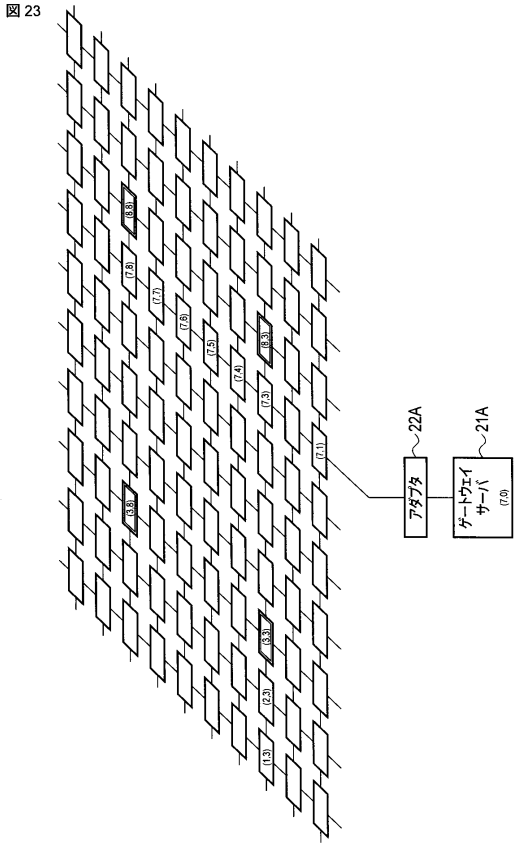
【図２２】

図22



【図２３】

図23



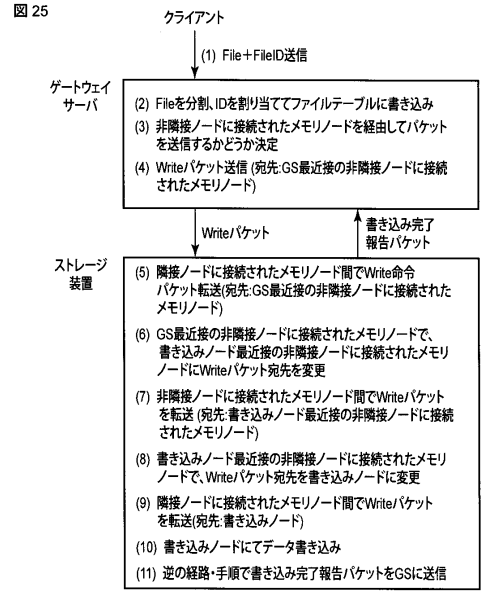
【図２４】

図24

一時送信先ノードアドレス及びタイプ	一時送信元ノードアドレス及びタイプ	第1中継ノードサブアドレス	第2中継ノードサブアドレス	最終送信先ノードメインアドレス	送信元ノードメインアドレス

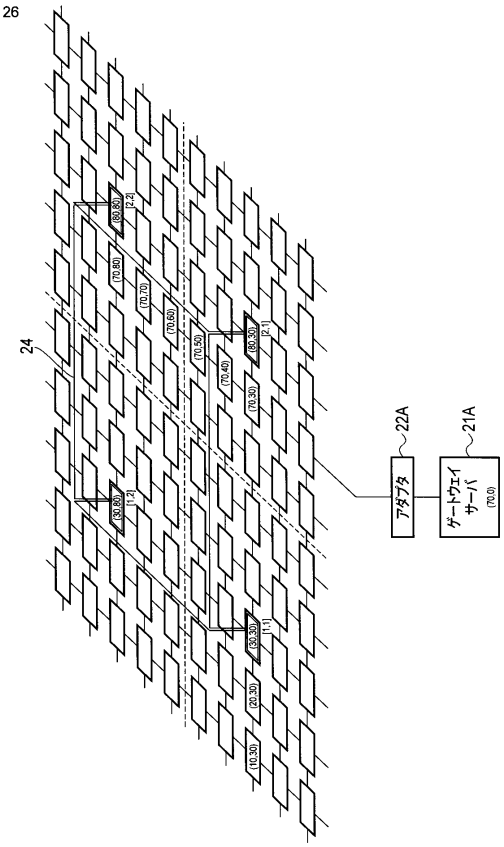
【図２５】

図25



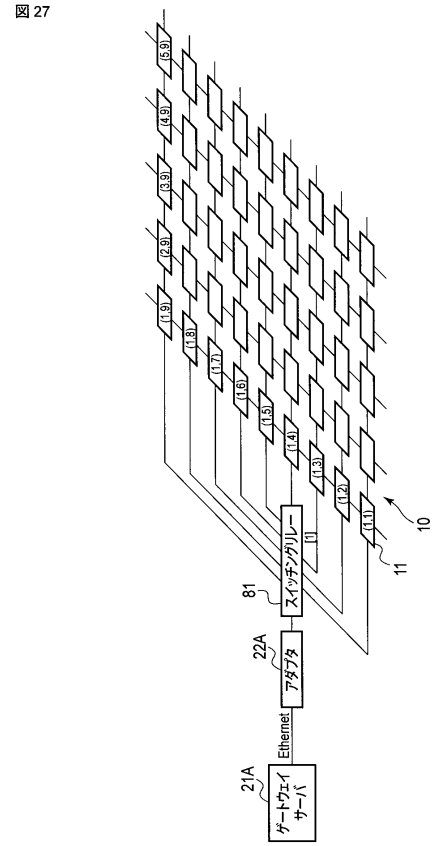
【図２６】

図26



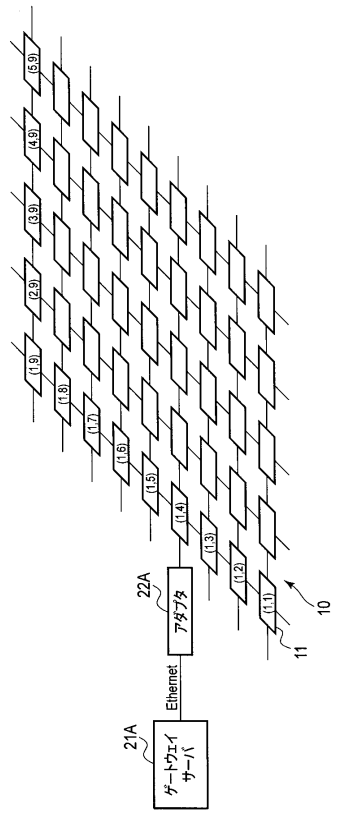
【図２７】

図 27



【図２８】

図 28



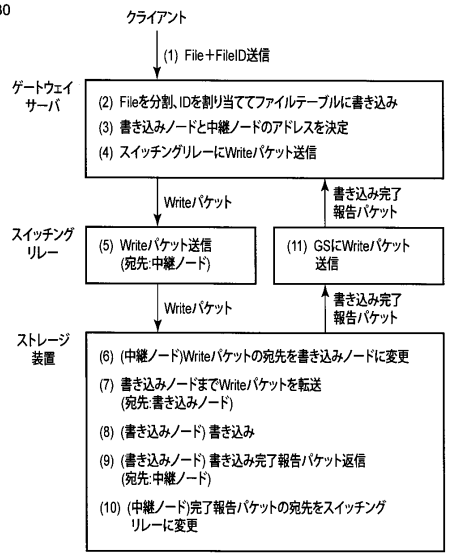
【図２９】

図 29

一時送信先ノードアドレス及びタイプ	一時送信元ノードアドレス及びタイプ	中継ノードアドレス	最終送信先ノードアドレス及びタイプ	送信元ノードアドレス及びタイプ

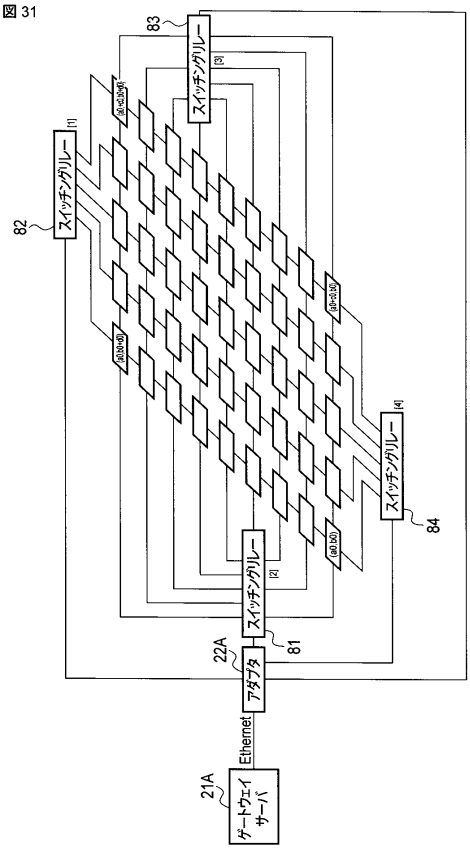
【図３０】

図 30



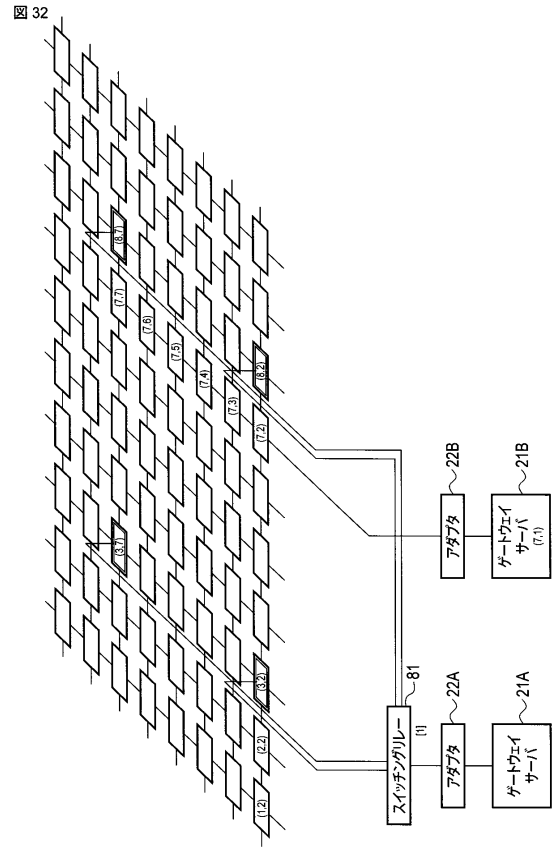
【図３１】

図 31



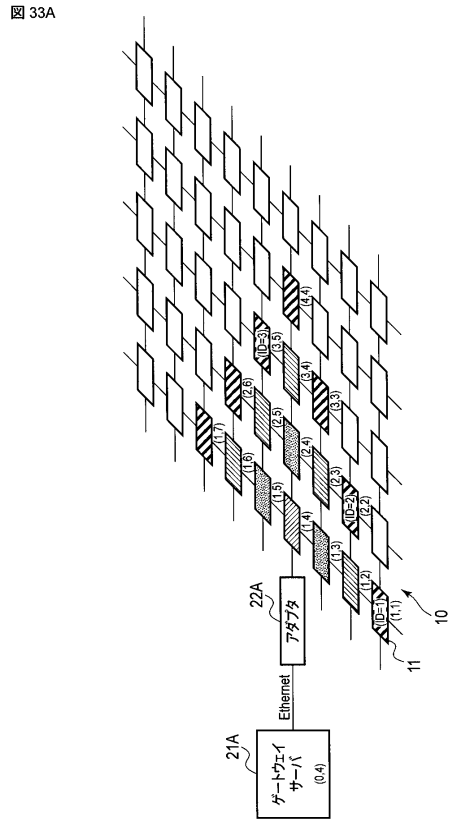
【図３２】

図 32



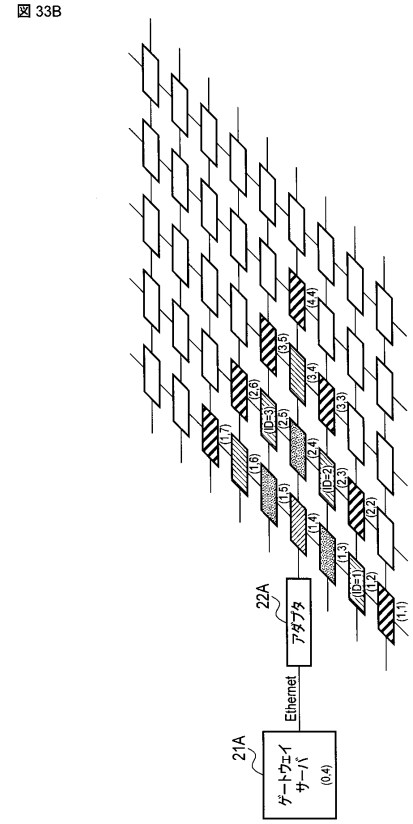
【図３３Ａ】

図 33A



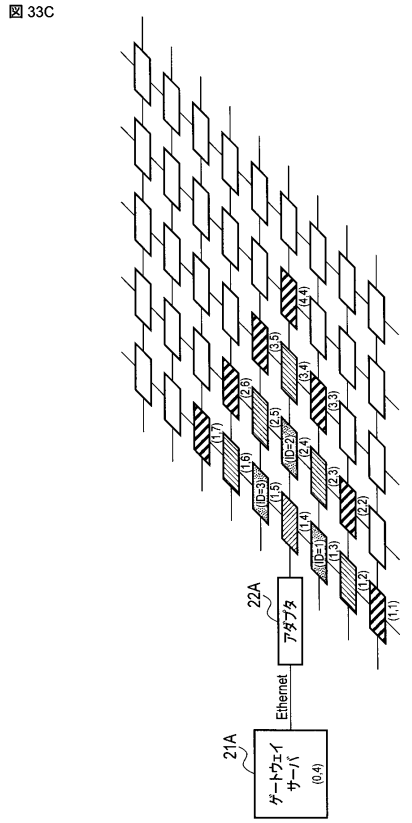
【図３３Ｂ】

図 33B



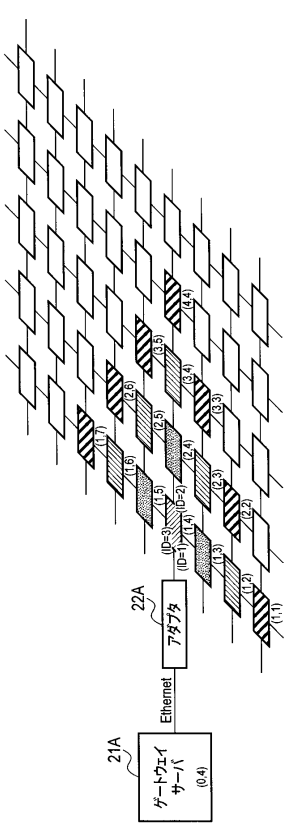
【図３３Ｃ】

図 33C



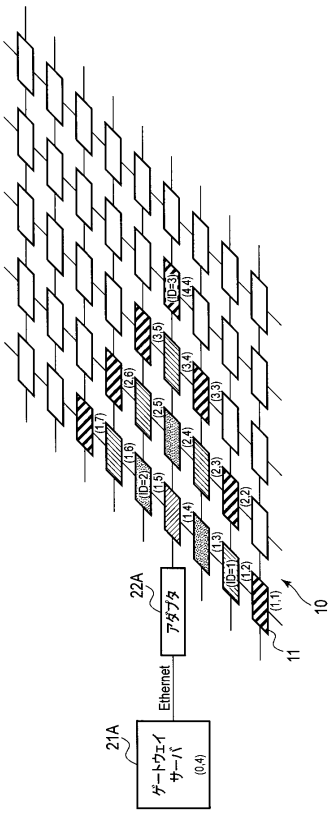
【図３３Ｄ】

図 33D



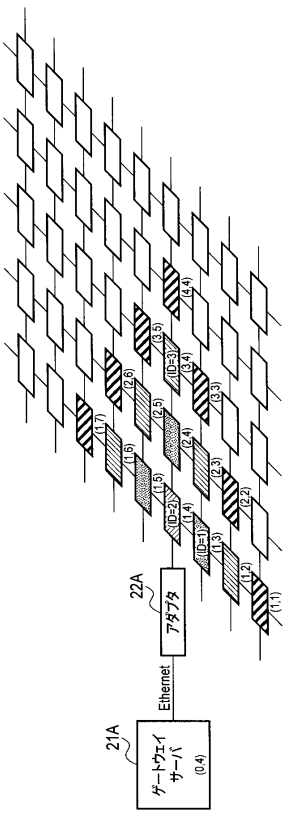
【図３４Ａ】

図 34A



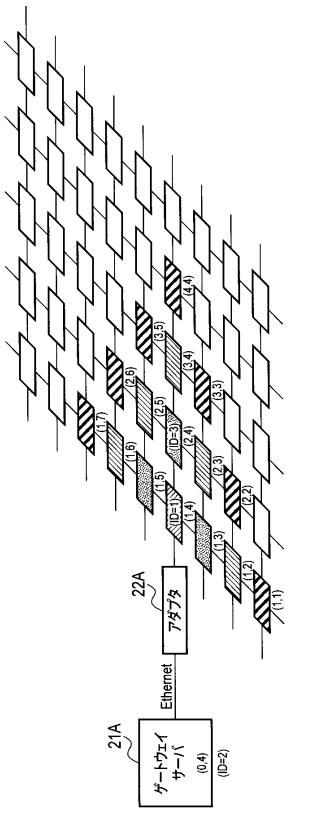
【図３４Ｂ】

図 34B



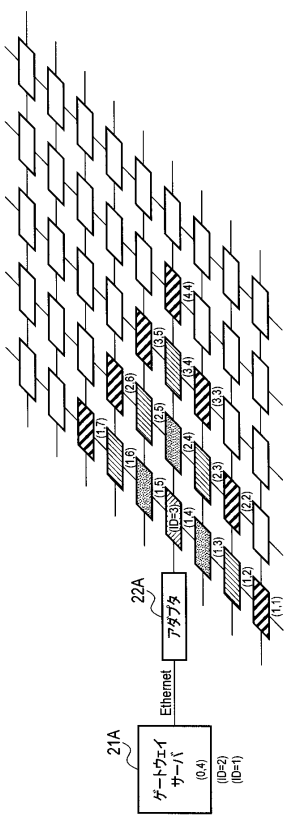
【図３４Ｃ】

図 34C



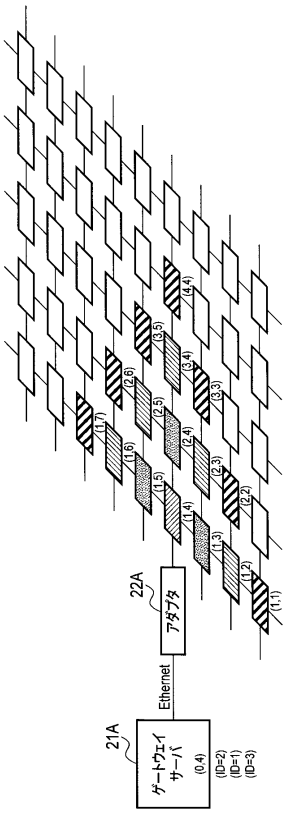
【図３４Ｄ】

図 34D



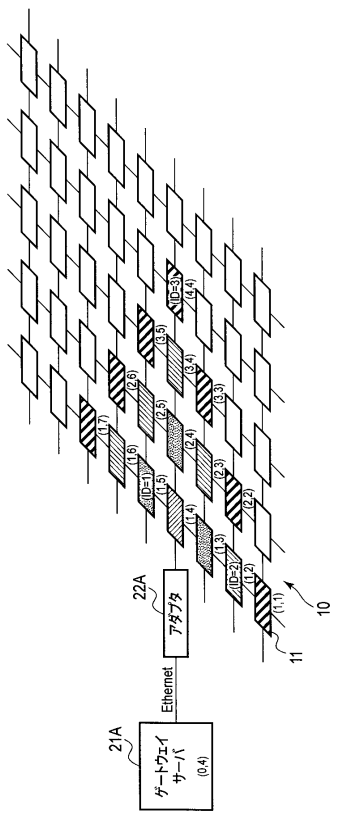
【図３４Ｅ】

図 34E



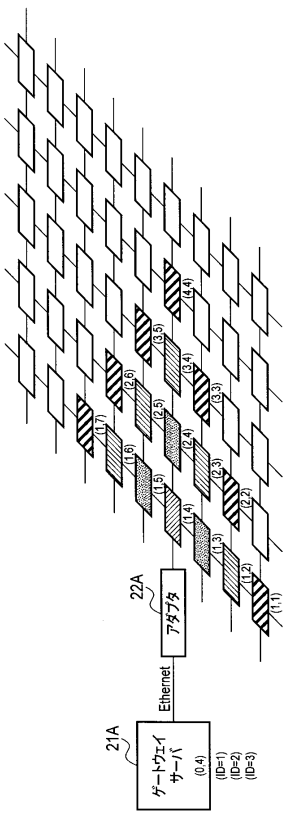
【図３５Ａ】

図 35A

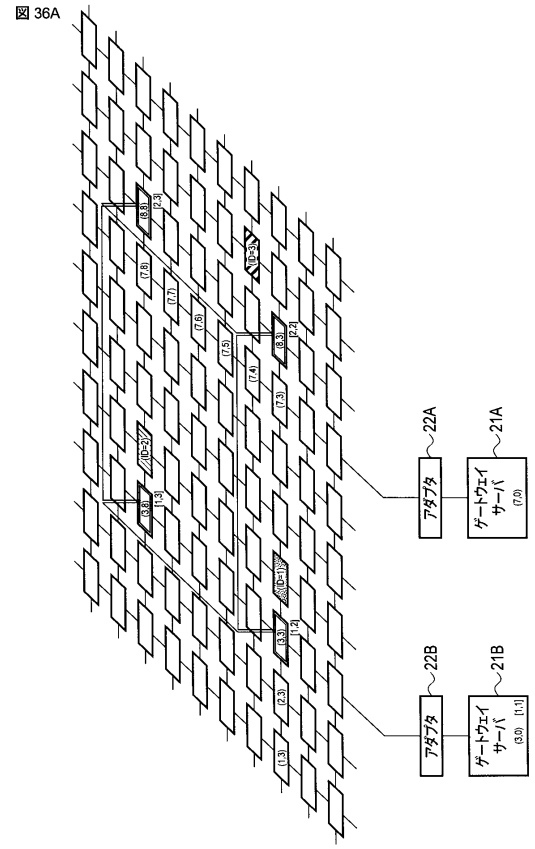


【図３５Ｂ】

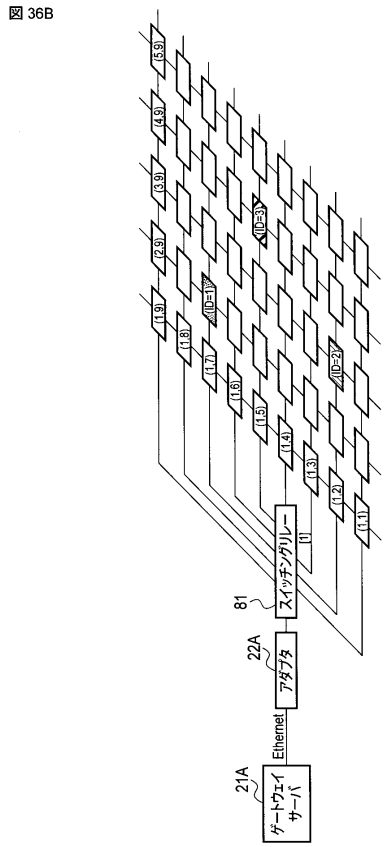
図 35B



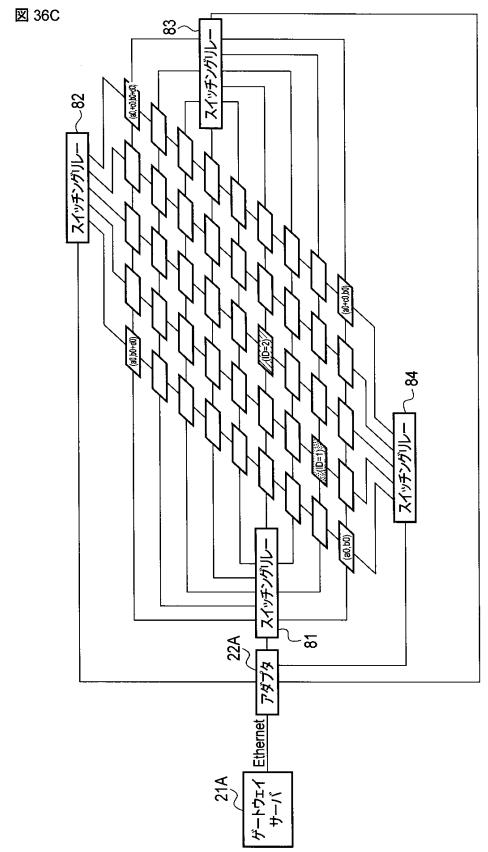
【図３６Ａ】



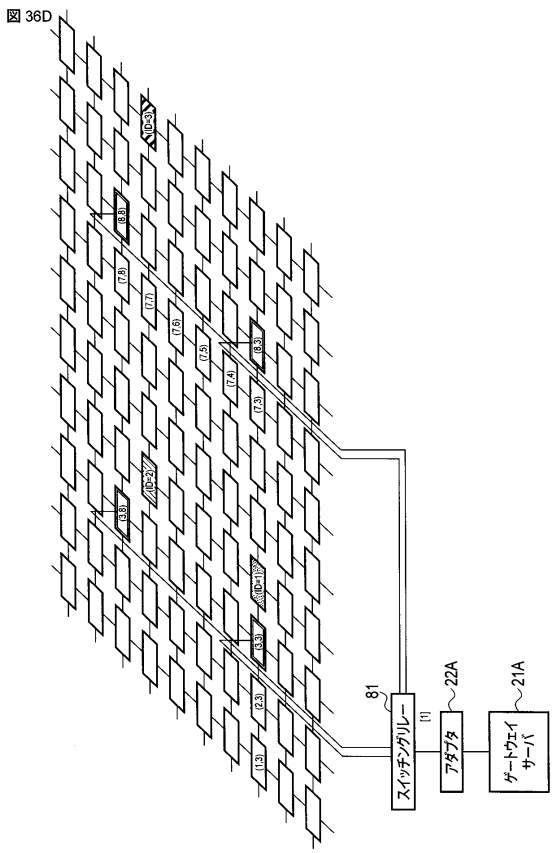
【図３６Ｂ】



【図３６Ｃ】

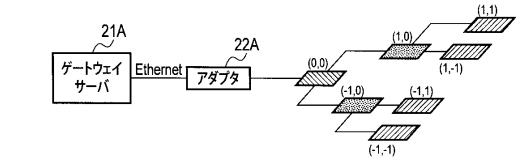


【図３６Ｄ】



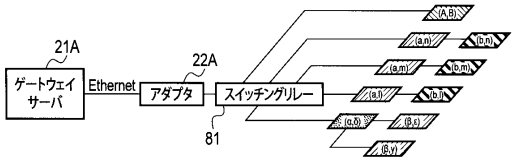
【図３６Ｅ】

図 36E



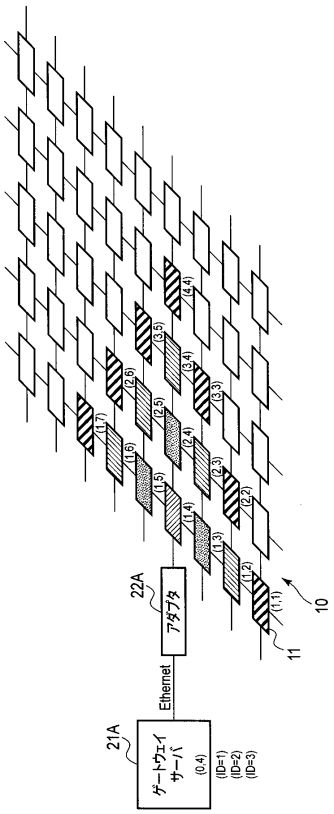
【図３６Ｆ】

図 36F



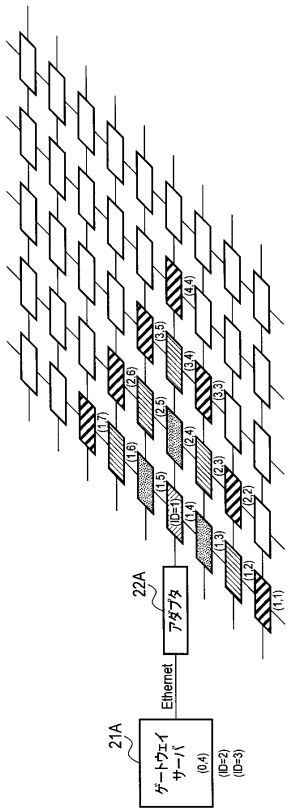
【図３７Ａ】

図 37A



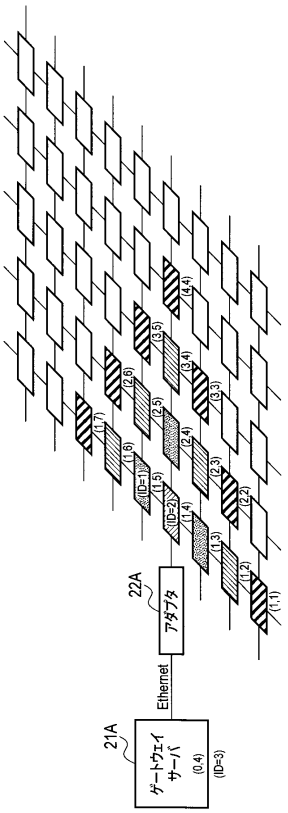
【図３７Ｂ】

図 37B



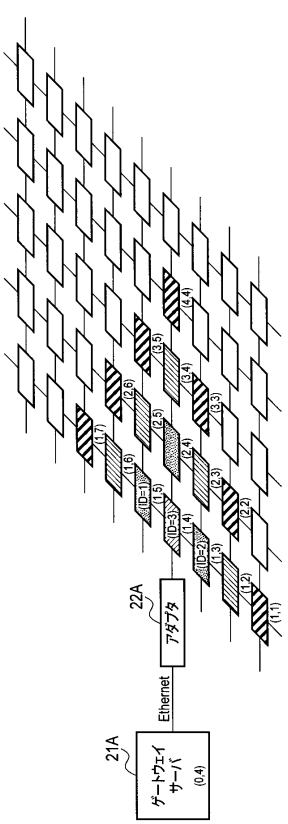
【図３７Ｃ】

図 37C



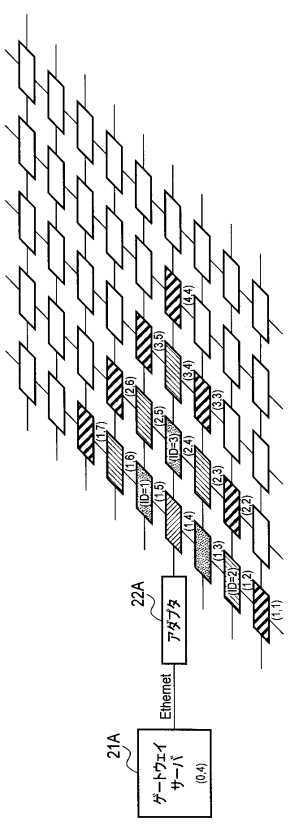
【図３７Ｄ】

図 37D



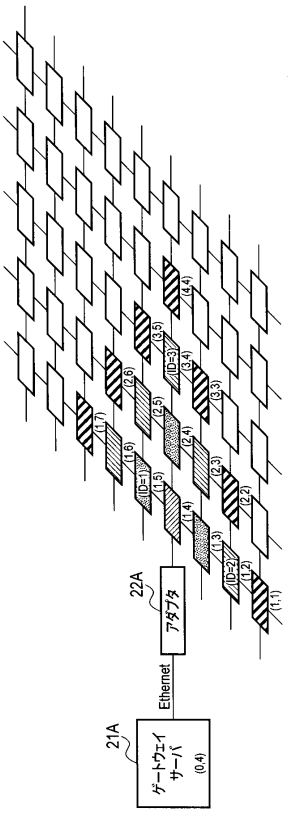
【図３７Ｅ】

図 37E



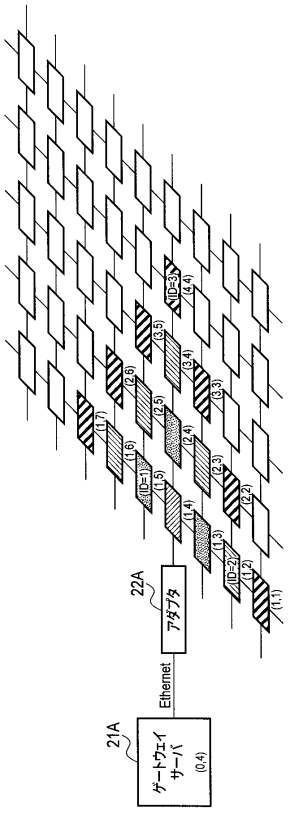
【図３７Ｆ】

図 37F



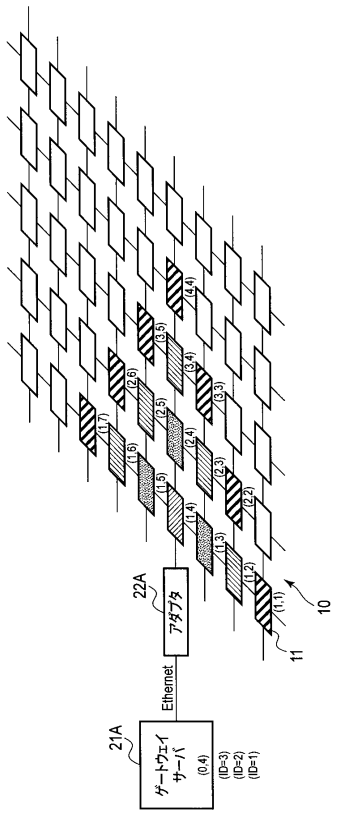
【図３７Ｇ】

図 37G



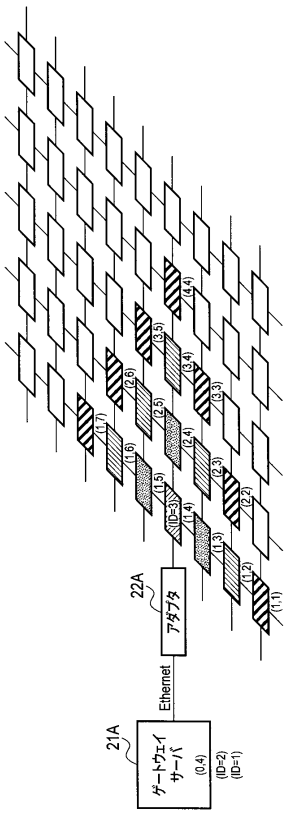
【図３８Ａ】

図 38A



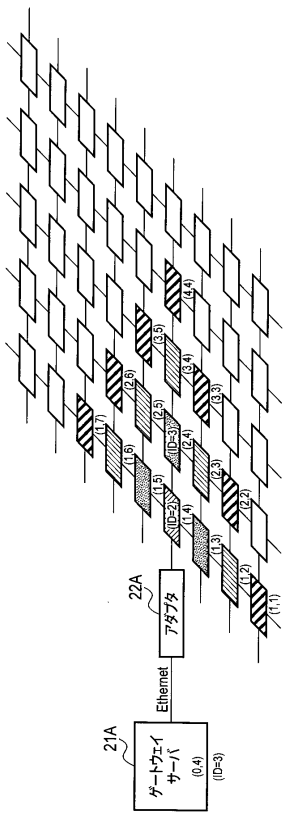
【図３８Ｂ】

図 38B



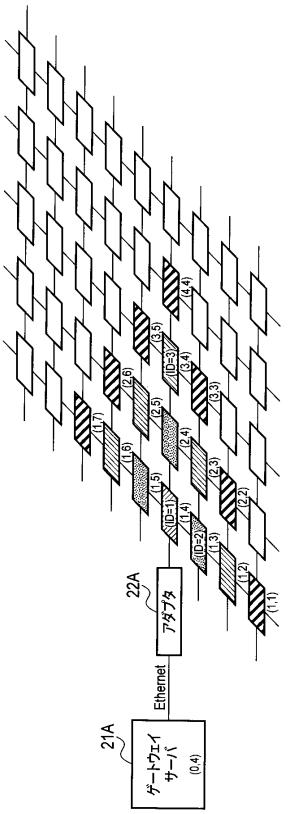
【図３８Ｃ】

図 38C



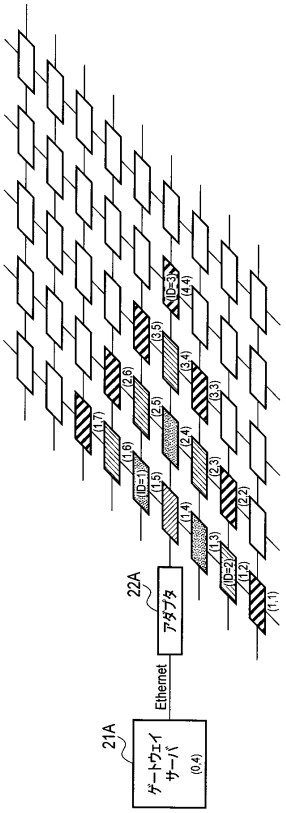
【図３８Ｄ】

図 38D



【図３８Ｅ】

図 38E


(4,4)
ID=3
(3,5)
(3,4)
(3,3)
(2,6)
(2,5)
(2,4)
(2,3)
(2,2)
(1,7)
(1,6)
ID=1
(1,5)
(1,4)
(1,3)
ID=2
(1,2)
(1,1)
22A
アダプタ
Ethernet
21A
ゲートウェイ
サーバ
(0,4)

フロントページの続き

(74)代理人 100119976
弁理士 幸長 保次郎
(74)代理人 100153051
弁理士 河野 直樹
(74)代理人 100140176
弁理士 砂川 克
(74)代理人 100158805
弁理士 井関 守三
(74)代理人 100172580
弁理士 赤穂 隆雄
(74)代理人 100179062
弁理士 井上 正
(74)代理人 100124394
弁理士 佐藤 立志
(74)代理人 100112807
弁理士 岡田 貴志
(74)代理人 100111073
弁理士 堀内 美保子
(74)代理人 100134290
弁理士 竹内 将訓
(72)発明者 辰村 光介
東京都港区芝浦一丁目１番１号 株式会社東芝内
(72)発明者 木下 敦寛
東京都港区芝浦一丁目１番１号 株式会社東芝内
(72)発明者 西野 弘剛
東京都港区芝浦一丁目１番１号 株式会社東芝内
(72)発明者 鈴木 正道
東京都港区芝浦一丁目１番１号 株式会社東芝内
(72)発明者 西 義史
東京都港区芝浦一丁目１番１号 株式会社東芝内
(72)発明者 丸亀 孝生
東京都港区芝浦一丁目１番１号 株式会社東芝内
(72)発明者 栗田 貴宏
東京都港区芝浦一丁目１番１号 株式会社東芝内

審査官 稲葉 崇

(56)参考文献 再公表特許第２００８／１４９７８４（ＪＰ，Ａ１）
特開２００３－１７４４６０（ＪＰ，Ａ）
特表２００８－５３７２６５（ＪＰ，Ａ）
特開２００９－０３７２７３（ＪＰ，Ａ）
特許第４３８５３８７（ＪＰ，Ｂ２）
特開２０１０－２１８３６４（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ ３／０６－３／０８
Ｇ０６Ｆ １３／１０－１３／１４
Ｇ０６Ｆ １２／００

G06F 15/16-15/177
G06F 15/80