(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2005-217815
(P2005-217815A)
(43)公開日　平成17年8月11日(2005.8.11)

(51)Int.Cl.[7]	F I		テーマコード（参考）
H04L 12/56	H04L 12/56	100A	5B014
G06F 3/06	G06F 3/06	301A	5B065
G06F 13/10	G06F 13/10	340A	5K030
H04L 12/46	H04L 12/46	D	5K033
H04L 12/66	H04L 12/66	A	

審査請求 未請求　請求項の数 19　OL　（全 31 頁）

(21)出願番号　特願2004-22463(P2004-22463)
(22)出願日　平成16年1月30日(2004.1.30)

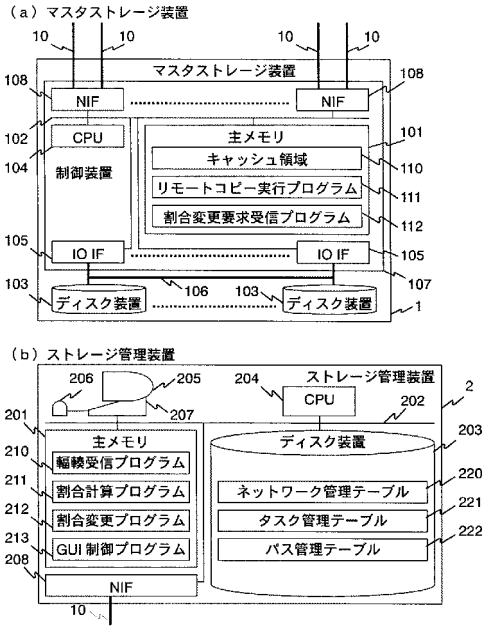
(71)出願人　000005108
株式会社日立製作所
東京都千代田区丸の内一丁目６番６号
(74)代理人　100075096
弁理士　作田　康夫
(74)代理人　100100310
弁理士　井上　学
(72)発明者　志賀　賢太
神奈川県川崎市麻生区王禅寺１０９９番地
株式会社日立製作所システム開発研究所内
(72)発明者　熊谷　敦也
神奈川県川崎市麻生区王禅寺１０９９番地
株式会社日立製作所システム開発研究所内

最終頁に続く

(54)【発明の名称】パス制御方法

(57)【要約】
【課題】
計算機あるいはストレージ装置が、複数の経路を用いてストレージ装置へアクセスするシステムにおいて、経路上で輻輳が発生した時の、スループットの低下を防止する。
【解決手段】
計算機、ストレージ装置およびストレージ管理装置がネットワークで接続され、計算機あるいは第一のストレージ装置が、ネットワーク内の複数の経路を用いて、第二のストレージ装置へアクセスし、かつ計算機あるいは第一のストレージ装置が、あらかじめ設定された割合に基づき複数の経路間の負荷分散を行う計算機システムにおいて、経路上で輻輳が発生したことを検出し、当該経路の輻輳時の比重を計算し、輻輳時の割合と設定された割合との差分を、計算機あるいは第一のストレージ装置と第二のストレージ装置との間の、その他の経路の割合へ割り振る。
【選択図】　図２

図 2



【特許請求の範囲】
【請求項１】
第一の装置と、
第二の装置と、
前記第一の装置と前記第二の装置とを接続する複数のパスと、
前記第一の装置と接続される第三の装置とを有し、
前記第一の装置は、所定の割合で前記複数のパスを使用して前記第二の装置へデータを転送し、
前記第三の装置は、前記複数のパスの輻輳を検出して前記第一の装置へ通知し、
前記第一の装置は、前記通知に基づいて、前記所定のパス間の割合を変更して前記複数のパスを使用して前記第二の装置へデータを転送することを特徴とするシステム。
【請求項２】
前記第一の装置及び前記第二の装置はストレージ装置であることを特徴とする請求項１記載のシステム。
【請求項３】
前記第一の装置は計算機であり、前記第二の装置はストレージ装置であることを特徴とする請求項１記載のシステム。
【請求項４】
前記複数のパスの各々は前記第一の装置と前記第二の装置とを接続するためのネットワーク装置を有し、
前記第三の装置は、前記ネットワーク装置とネットワークを介して接続されており、
前記第三の装置は、前記ネットワーク装置から前記ネットワークを介して前記ネットワーク装置での輻輳の発生の通知を受け取ることを特徴とする請求項２及び３のうちいずれか一つに記載のシステム。
【請求項５】
前記通知とは、ＳＮＭＰ　Ｔｒａｐに基づく通知であることを特徴とする請求項４記載のシステム。
【請求項６】
前記複数のパスの各々は前記第一の装置と前記第二の装置とを接続するためのネットワーク装置を有し、
前記第三の装置は、前記ネットワーク装置とネットワークを介して接続されており、
前記第三の装置は、前記ネットワーク装置から前記ネットワークを介して前記ネットワーク装置での廃棄パケットに関する情報を受け取り、前記廃棄パケットの情報に基づいて前記複数のパスの輻輳を判断することを特徴とする請求項２及び３のうちいずれか一つに記載のシステム。
【請求項７】
前記第三の装置は、前記ネットワーク装置から受信する廃棄パケット数が、以前受信した廃棄パケット数を上回る場合に前記ネットワーク装置を有する前記複数のパスに輻輳が発生したと判断することを特徴とする請求項６記載のシステム。
【請求項８】
前記第三の装置は、前記所定の割合及び前記所定の割合の変更率に関する情報を有し、前記複数のパスの輻輳を検出した場合、変更後の前記所定のパス間の割合を前記変更率に基づいて計算して、前記第一の装置に前記変更後の所定の割合の情報を送信し、
前記第一の装置は、受信した前記変更後の所定のパス間の割合に基づいて、前記複数のパスを用いて前記第二の装置へデータを転送することを特徴とする請求項１記載のシステム。
【請求項９】
前記第三の装置は、前記複数のパスの輻輳からの回復を検知し、前記所定のパス間の割合の情報を前記第一の装置に送信し、
前記第一の装置は、受信した前記所定のパス間の割合に基づいて、前記複数のパスを用

いて前記第二の装置へデータを転送することを特徴とする請求項８記載のシステム。
【請求項１０】
第一の装置と、
第二の装置と、
前記第一の装置と前記第二の装置とを接続する複数のパスとを有し、
前記第一の装置は、所定の割合で前記複数のパスを使用して前記第二の装置へデータを転送し、
前記第一の装置は、前記複数のパスの輻輳を検出し、
前記第一の装置は、前記複数のパスの輻輳の検出に応じて、前記所定のパス間の割合を変更して前記複数のパスを使用して前記第二の装置へデータを転送することを特徴とするシステム。
【請求項１１】
前記第一の装置及び前記第二の装置はストレージ装置であることを特徴とする請求項１０記載のシステム。
【請求項１２】
前記第一の装置は計算機、前記第二の装置はストレージ装置であることを特徴とする請求項１０記載のシステム。
【請求項１３】
前記複数のパスの各々は前記第一の装置と前記第二の装置とを接続するためのネットワーク装置を有し、
前記第一の装置は、前記ネットワーク装置から前記複数のパスを介して前記ネットワーク装置での輻輳の発生の通知を受け取ることを特徴とする請求項１１及び１２のうちいずれか一つに記載のシステム。
【請求項１４】
前記通知とは、ＥＣＮに基づくフラグであることを特徴とする請求項１３記載のシステム。
【請求項１５】
前記第一の装置は、前記第二の装置から所定の期間応答が帰ってこない場合に前記複数のパスに輻輳が発生したと判断することを特徴とする請求項１１及び１２のうちいずれか一つに記載のシステム。
【請求項１６】
前記第一の装置は、前記第二の装置へ送信したデータの受信確認応答を重複して受信した場合に、前記複数のパスに輻輳が発生したと判断することを特徴とする請求項１１及び１２のうちいずれか一つに記載のシステム。
【請求項１７】
前記第一の装置は、前記所定の割合及び前記所定の割合の変更率に関する情報を有し、前記複数のパスの輻輳を検出した場合、変更後の前記所定のパス間の割合を前記変更率に基づいて計算し、前記変更後の所定のパス間の割合に基づいて、前記複数のパスを用いて前記第二の装置へデータを転送することを特徴とする請求項１０記載のシステム。
【請求項１８】
前記第一の装置は、輻輳発生後、輻輳が発生した前記複数のパスに送信できるデータサイズがあらかじめ設定された値を超過したときに、輻輳が回復したと判断することを特徴とする請求項１７記載のシステム。
【請求項１９】
制御部と、
前記制御部に接続されるディスク装置と、
ネットワークに接続されるインターフェースとを有し、
前記インターフェースは、前記ネットワーク内の複数のパスで他の装置と接続され、
前記制御部は所定のパス間の割合で前記複数のパスを用いて前記他の装置に前記ディスク装置に格納されたデータをパケットにして送信し、

前記制御部は、前記他の装置に送信したパケットに対する受領報告が一定期間受信されない場合に、前記複数のパスに輻輳が発生したと判断し、
前記制御部は、前記輻輳の発生に応じて、前記所定のパス間の割合を変更し、前記変更されたパス間の割合で前記他の装置へのパケット転送を行うことを特徴とするストレージ装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、ネットワークを介して接続されたストレージ装置と情報処理装置を有するシステムに関する。
【背景技術】
【０００２】
近年、計算機（以下「ホスト」ともいう）にストレージ装置を直結する形態の代わりに、一つ又は複数のストレージ装置を、ネットワークを介して複数のホストと接続する形態が盛んに用いられるようになっている。このネットワークを用いたストレージ装置の接続形態のことをストレージエリアネットワーク（以下「ＳＡＮ」）という。これまでＳＡＮはファイバチャネル（以下「ＦＣ」）技術を用いて構築されていた。以下、ＦＣベースのＳＡＮをＦＣ－ＳＡＮと称する。
【０００３】
ＦＣ－ＳＡＮを構築する場合、システム全体の可用性を向上させるために、ホストとストレージ装置との間は複数の物理的な通信経路（以下「パス」ともいう）で接続される。ホストは、ストレージ装置へ送信するコマンドやデータを各々のパスを使用して送信することで、パスに掛かる負荷を分散させる。複数のパスの使用割合を決めるアルゴリズムとして、送信されるコマンドやデータを各々のパスに異なる割合（例えば送信されるコマンド数の６割：４割）で割り振る重み付けラウンドロビンが挙げられる。またホストは、ＦＣ－ＳＡＮを構成するスイッチの故障やネットワークの切断等の原因により１つのパスが使用不能になったことを検出した場合、残りのパスを使用してストレージ装置との通信を継続する。
【０００４】
上記のようなパス間の負荷分散は、ホストとストレージ装置との間の通信（以下「ホストストレージ装置間通信」）だけではなく、ストレージ装置と他のストレージ装置との間の通信（以下「ストレージ装置間通信」）でも実施されている。ストレージ装置間通信の例としてリモートコピーが挙げられる。リモートコピーとは、ストレージ装置に格納されるデータを、地理的に離れた拠点に設置された他のストレージ装置へ転送してデータを複製保存する技術である。
【０００５】
上述したホストストレージ装置間通信におけるパス制御技術の１つが特許文献１に開示されている。
【０００６】
一方、ＦＣよりも導入コストが低いインターネットプロトコル（以下「ＩＰ」）を用いたネットワーク（以下「ＩＰネットワーク」）を用いて構築されたＳＡＮであるＩＰ－ＳＡＮが注目を集めている。ホストがＩＰ－ＳＡＮを介してストレージ装置と通信する場合、ＳＣＳＩプロトコルをＴＣＰ／ＩＰでカプセル化したプロトコルであるｉＳＣＳＩが主に用いられる。
【０００７】
【特許文献１】特開２０００－３３０９２４号公報
【発明の開示】
【発明が解決しようとする課題】
【０００８】
ＩＰ－ＳＡＮにおいてホストストレージ装置間通信又はストレージ装置間通信をする場

合、特許文献１に開示されたパス制御技術では以下の問題が発生する。
上述したように、特許文献１ではネットワークとしてＦＣを使用する。ＦＣでは、通信におけるフロー制御にバッファクレジットと呼ばれる技術が用いられる。バッファクレジットでは、ホストやストレージ装置などのネットワークの末端（以下「エンドポイント」）と、そのエンドポイントが接続されているスイッチとが、受信したパケットを一時的に保存するためのバッファの空きサイズの情報をお互いに交換する。同様に、隣接したスイッチ同士もバッファの空きサイズの情報をお互いに交換する。これによりネットワーク間でバッファの空きサイズ以上のパケットを送信しない様に各装置がデータを転送する。このため、ＦＣでは基本的にネットワーク内での輻輳が発生しない。したがって、特許文献１ではそもそも輻輳に基づいてパスを制御することは全く考慮されていない。

【０００９】
一方、ＴＣＰ／ＩＰでは、フロー制御にスライディングウィンドウと呼ばれる技術が用いられる。スライディングウィンドウでは、エンドポイント同士がバッファの空きサイズの情報をお互いに交換する。ここで、エンドポイントはネットワークを構成するスイッチのバッファの空きサイズを考慮せずにパケットを送信する。したがって、ネットワークに送出されたパケットがスイッチのバッファの許容量を超過した場合、輻輳が発生する。輻輳が発生すると、スイッチは受信したパケットを廃棄してしまう。更にエンドポイントは、タイムアウト等の手段で輻輳の発生を検出すると、輻輳を回復させるためにパケットの送信レートを大幅に下げる。この仕組みは、ＲＦＣ２５８１により規定されている。

【００１０】
従って、輻輳が発生しないＦＣを前提とした従来のパス制御技術をＩＰ－ＳＡＮに適用すると、ＴＣＰ／ＩＰの仕組みそのままに、輻輳が発生したパスの転送レートが大幅に下がってしまう。しかし、特許文献１等で開示された従来のパス制御技術では輻輳を考慮しないので、各エンドポイントはその送信レートが低下したパスを同じ使用率で使用し続ける。したがってシステム全体のパケット送信のスループットが大幅に下がってしまう。特にストレージ装置へのデータの入出力の際にはスループットが低下することは大きな問題（書き込み応答の遅延等）となる。

【課題を解決するための手段】
【００１１】
そこで、本発明の一実施形態として、第一の装置と、第二の装置と、第一の装置と前記第二の装置とを接続する複数のパスと、第一の装置と接続される第三の装置とを有するシステムで、第一の装置は、所定の割合で複数のパスを使用して第二の装置へデータを転送し、第三の装置は、複数のパスの輻輳を検出して第一の装置へ通知し、第一の装置は、通知に基づいて、所定のパス間の割合を変更して複数のパスを使用して第二の装置へデータを転送する構成とする。

【００１２】
尚、本発明の実施態様として、第一の装置は計算機でもストレージ装置でも良い。又、第三の装置は第一の装置に含まれていても良い。

【００１３】
更に、本発明の実施態様として、第三の装置で所定の割合及び所定の割合の変更率に関する情報を有し、複数のパスの輻輳を検出した場合、変更後の前記所定のパス間の割合を変更率に基づいて計算し、変更後の所定のパス間の割合の情報を第一の装置に送信し、第一の装置は、変更後の所定のパス間の割合に基づいて、複数のパスを用いて第二の装置へデータを転送する構成でも良い。

【００１４】
更に、本発明の実施態様として、複数のパスは第一の装置と第二の装置を接続するネットワーク装置を有し、第三の装置はネットワーク装置から輻輳の発生や輻輳からの回復についての情報を得る構成も考えられる。その他の実施態様については、以下の記載で明らかにされる。

【発明の効果】

【００１５】
本発明により、システム全体のパケット送信スループット、具体的にはホストストレージ装置間又はストレージ装置間のパケット転送スループットの低下を最小限に抑えることができる。
【発明を実施するための最良の形態】
【００１６】
以下、本発明の実施形態について図面を用いて説明する。以下の図中、同一の部分には同一の符号を付加する。ただし本発明が実施形態に制限されることは無く、本発明の思想に合致するあらゆる応用例が本発明に該当する。また、特に限定しない限り、各構成要素は複数でも単数でも構わない。
【実施例１】
【００１７】
第一の実施形態は、ストレージ装置間で実行されるｉＳＣＳＩを用いたリモートコピーに本発明を適用したシステムに関する。第一の実施形態では、一方のストレージ装置（以下「マスタストレージ装置」）が、他方のストレージ装置（以下「リモートストレージ装置」）と１つ以上のＴＣＰコネクションから成る１つのｉＳＣＳＩセッションを確立してリモートコピーを実行する。それぞれのＴＣＰコネクションを用いて送信されるパケットが通る経路が１つのパスに相当する。
【００１８】
図１は、第一の実施形態のシステムの構成例を示した図である。システムは、マスタサイト１１、リモートサイト１２およびこれらのサイト間を接続する１つ以上のＷＡＮ（Ｗｉｄｅ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）８を有する。なお「サイト」とは装置が設置される場所や装置群を指し、一つのビル等を指す。また本実施形態では、マスタサイト１１とリモートサイト１２とはある程度距離が離れているもの（例えば東京と大阪等）とする。
【００１９】
マスタサイト１１は、マスタストレージ装置１、マスタストレージ装置１を管理するストレージ管理装置２、ホスト４、端末６、ホスト４、マスタストレージ装置１およびストレージ管理装置２を接続するＩＰネットワークであるＩＰ－ＳＡＮ７並びに端末６とホスト４とを接続するＩＰネットワークであるＬＡＮ９を有する。また、マスタストレージ装置１、ストレージ管理装置２およびホスト４とＩＰ－ＳＡＮ７とは、ＵＴＰ（Ｕｎｓｈｉｅｌｄｅｄ　Ｔｗｉｓｔｅｄ　Ｐａｉｒ）ケーブルや光ファイバケーブル等の通信線１０で接続されている。同様に、端末６およびホスト４とＬＡＮ９とは、通信線１０で接続されている。
【００２０】
なお、端末６等の装置と、ＩＰ－ＳＡＮ７やＬＡＮ９等のＩＰネットワークとを、無線通信技術を使用して接続する場合、通信線１０は不要である。また、本実施形態ではＩＰ－ＳＡＮ７とＬＡＮ９とが各々別である例を示すが、ＩＰ－ＳＡＮ７がＬＡＮ９を兼ねる、すなわち全ての装置がＩＰ－ＳＡＮ７に接続される構成としても良い。この場合にはシステム設置のコストは抑えられる。しかしストレージ装置間通信用のパケットと端末間での通信に使用されるパケットが一つのネットワークに混在してネットワークが混雑するという問題がある。これを解消するためには本実施形態の構成が好ましい。
【００２１】
リモートサイト１２は、リモートストレージ装置５、ホスト４、端末６、ホスト４とリモートストレージ装置５とを接続するＩＰ－ＳＡＮ７、端末６とホスト４とを接続するＬＡＮ９を有する。リモートストレージ装置５およびホスト４とＩＰ－ＳＡＮ７とは、通信線１０で接続されている。同様に、端末６およびホスト４とＬＡＮ９とは、通信線１０で接続されている。リモートサイト１２でも、ネットワークが一つである構成も考えられる。
【００２２】
ホスト４および端末６は一般的な計算機であり、ＣＰＵ、主メモリおよび入出力装置等

を有する。また、ホスト４および端末６は、他の装置と通信線１０を介して接続するインタフェースであるネットワークインタフェース（以下「ＮＩＦ」）を有する。

ＷＡＮ８は、受信したパケットを転送する１つ以上のネットワーク装置３を有する。

【００２３】

図２（ａ）は、マスタストレージ装置１の構成例を示した図である。マスタストレージ装置１は、単体の記憶装置あるいは複数の記憶装置を有する記憶装置システムである。なお、記憶装置は、ハードディスクドライブやＤＶＤと言った、不揮発性の記憶媒体を用いた装置が含まれる。また、記憶装置システムでは、ＲＡＩＤ構成が採用されていてもよい。マスタストレージ装置１は、記憶装置（以下「ディスク装置」）１０３、ディスク装置１０３に対するデータの書き込みや読み出しを制御する制御装置１０７および制御装置１０７とディスク装置１０３とを接続する通信線１０６を有する。

【００２４】

制御装置１０７は、揮発性のメモリ（以下「主メモリ」）１０１、バスなどの通信線１０２、中央演算装置（以下「ＣＰＵ」）１０４、制御装置１０７と通信線１０６とを接続するインタフェースであるＩＯインタフェース（以下「ＩＯ　ＩＦ」）１０５および制御装置１０７と通信線１０とを接続するＮＩＦ１０８を有する。

【００２５】

主メモリ１０１には、ディスク装置１０３から読み出されたデータ又はホスト等から受信したデータを記憶するキャッシュ領域１１０、リモートコピーを実行する際にＣＰＵ１０４で実行されるリモートコピー実行プログラム１１１、ストレージ管理装置２から割合変更要求を受信した際にＣＰＵ１０４で実行される割合変更要求受信プログラム１１２が格納される。

【００２６】

図２（ｂ）は、ストレージ管理装置２の構成例を示した図である。ストレージ管理装置２は、主メモリ２０１、通信線２０２、ディスク装置２０３、ＣＰＵ２０４、表示装置などの出力装置（以下「ディスプレイ」）２０５、マウスなどのポインティング装置２０６、キーボードなどの文字入力装置２０７およびＮＩＦ２０８を有する計算機である。

【００２７】

主メモリ２０１には、ネットワーク装置３から輻輳発生通知や輻輳回復通知を受信した際にＣＰＵ２０４で実行される輻輳受信プログラム２１０、パス間の負荷分散の割合を計算する際にＣＰＵ２０４で実行される割合計算プログラム２１１、マスタストレージ装置１等へ割合変更要求を送信する際にＣＰＵ２０４で実行される割合変更プログラム２１２およびシステム管理者等にグラフィカルなユーザインターフェースを提供する際にＣＰＵ２０４で実行されるＧＵＩ制御プログラム２１３が格納される。

【００２８】

なお、ここで「パス間の負荷分散の割合」（以下単に「パス間の割合」又は「パスの割合」ということもある）とは、ホストストレージ装置間通信やストレージ装置間通信で使用されるパス群でどのようにコマンドやデータの転送量（あるいは数）を分担するかを示す数値である。具体的には、ある一つの処理において３本のパスを使用する場合、「パス間の負荷分散の割合」とは、当該処理に含まれる転送パケット数のうちの３割を第一のパスで、３割を第二のパスで、４割を第三のパスで送信する場合の「３割、３割、４割」の組又はその組の個々の数値を示す。

【００２９】

また、ディスク装置２０３には、ＷＡＮ８を構成するネットワーク装置３の情報を記憶するネットワーク管理テーブル２２０、マスタストレージ装置１がリモートコピーの個々のタスクを実行するのに必要な情報を記憶するタスク管理テーブル２２１およびマスタストレージ装置１とリモートストレージ装置５との間で使用されるパスの情報を記憶するパス管理テーブル２２２が格納される。ここで、タスクとは、マスタストレージ装置１が有する論理ユニット（Ｌｏｇｉｃａｌ　Ｕｎｉｔ：ＬＵ）から、リモートストレージ装置２が有する論理ユニットへのリモートコピーに関わるデータ転送処理を指す。なお、ＬＵと

はディスク装置が有する物理的な記憶領域から構成される論理的な記憶領域である。ＬＵは一つのディスク装置が有する記憶領域から構成されてもよく、複数のディスク装置の個々の記憶領域の集合体として定義されても良い。

【００３０】

なお、第一の実施形態では、ストレージ管理装置２にて障害が発生した場合でも各テーブルが記憶する情報が喪失されないようにディスク装置２０３に各テーブルが格納されることとしたが、各テーブルの情報が主メモリ２０１に格納されても良い。

【００３１】

なお、リモートストレージ装置５は、主メモリ１０１に割合変更要求受信プログラム１１２が格納されない点を除き、マスタストレージ装置１と同様の構成を有する。

【００３２】

図３は、ネットワーク装置３の構成例を示した図である。ネットワーク装置３は、主メモリ３０１、通信線３０２、ＣＰＵ３０４、ＮＩＦ３０８およびＮＩＦ３０８が受信したパケットを別のＮＩＦ３０８を介して他の装置へ送信するパケット転送装置３０９を有する。主メモリ３０１には、転送待ちパケットを一時的に格納するバッファ領域３１０およびバッファ領域３１０の容量が不足した時にストレージ管理装置２へ輻輳発生通知を送信する際やバッファ領域３１０の容量が空いた時にストレージ管理装置２へ輻輳回復通知を送信する際にＣＰＵ３０４で実行される輻輳状態通知プログラム３１１が格納される。

【００３３】

なお、上述のプログラムは、あらかじめ、または可搬型記録媒体からの読み込み、または他の計算機からのネットワーク経由のダウンロードにより、ディスク装置やメモリに格納される。これらのプログラムは、必要に応じて主メモリに転送され、ＣＰＵで実行される。

【００３４】

次に、ストレージ管理装置２のディスク装置２０３に格納される各種テーブルのデータ構造について説明する。

【００３５】

ネットワーク管理テーブル２２０、タスク管理テーブル２２１およびパス管理テーブル２２２は配列構造を成し、１つ以上のレコードを格納可能である。ただし、データ構造が配列構造に限定されることはない。

【００３６】

図４（ａ）は、ネットワーク管理テーブル２２０のデータ構造例を示す図である。ネットワーク管理テーブル２２０は、マスタストレージ装置１とリモートストレージ装置５との通信に関わるネットワーク装置３の個数分のレコードを有する。ネットワーク管理テーブル２２０の各レコードは、対応するネットワーク装置３が属するネットワークを識別する為の識別子であるネットワークＩＤが登録されるエントリ２２０１、ネットワーク装置３を識別するための識別子である装置ＩＤが登録されるエントリ２２０２およびネットワーク装置３が送信する輻輳発生通知や輻輳回復通知のソースＩＰアドレスが登録されるエントリ２２０３を有する。

【００３７】

図４（ｂ）は、タスク管理テーブル２２１のデータ構造例を示す図である。タスク管理テーブル２２１は、リモートコピーのタスクごとに１つのレコードを有する。タスク管理テーブル２２１の各レコードは、個々のリモートコピータスクを識別するための識別子であるタスクＩＤが登録されるエントリ２２１１、ｉＳＣＳＩイニシエータとして動作するマスタストレージ装置１が有するｉＳＣＳＩ名であるイニシエータｉＳＣＳＩ名が登録されるエントリ２２１２、コピー対象となる論理ユニットを識別するための識別子であるイニシエータＬＵ番号（ＬＵＮ）が登録されるエントリ２２１３、ｉＳＣＳＩターゲットとして動作するリモートストレージ装置５が有するｉＳＣＳＩ名であるターゲットｉＳＣＳＩ名が登録されるエントリ２２１４、コピー先のＬＵを識別するためのターゲットＬＵＮが登録されるエントリ２２１５、輻輳が発生した時にパス間の負荷分散の割合を変更する

か否かを示す割合変更フラグが登録されるエントリ２２１６および輻輳が発生したパスの負荷分散の割合の変更率を示す割合変更率が登録されるエントリ２２１７を有する。

【００３８】

第一の実施形態では、エントリ２２１６の「０」は輻輳時に割合変更を行わないことを、「１」は輻輳時に割合変更を行うことをそれぞれ表すものとする。また、第一の実施形態では、エントリ２２１７には、割合変更率を百分率で表した値が登録されるものとする。例えば、エントリ２２１７に登録された値が「５０」の場合、輻輳が発生したパスの負荷分散の割合を、以前の割合の５０％とする。

【００３９】

図５は、パス管理テーブル２２２のデータ構造例を示す図である。パス管理テーブル２２２は、ホストストレージ装置間通信又はストレージ装置間通信で使用されるパスごとに１つのレコードを有する。パス管理テーブル２２２の各レコードは、レコードに対応する個々のパスを識別する為の識別子であるパスＩＤが登録されるエントリ２２２１、当該レコードに対応するパスを使用するリモートコピータスクのタスクＩＤが登録されるエントリ２２２２、当該パスが経由するネットワークのネットワークＩＤが登録されるエントリ２２２３、当該パス上で輻輳が発生していない時点での割合であるデフォルト割合が登録されるエントリ２２２４、当該パスの現時点での割合が登録されるエントリ２２２５、当該パス上で発生した輻輳の回数が登録されるエントリ２２２６、当該パスのマスタストレージ装置１側のＩＰアドレスであるイニシエータＩＰアドレスが登録されるエントリ２２２７および当該パスのリモートストレージ装置５側のＩＰアドレスであるターゲットＩＰアドレスが登録されるエントリ２２２８を有する。

【００４０】

なお第一の実施形態では、リモートストレージ装置５が、ウェルノウンポートである「３２６０」以外のＴＣＰポート番号をｉＳＣＳＩ通信の通信ポートに使用する場合、エントリ２２２８には、ＩＰアドレスと前記ＴＣＰポート番号とを「：」でつなげた文字列がターゲットＩＰアドレスとして登録されるものとする。ここで、ウェルノウンポートとは、ＩＡＮＡ（Ｉｎｔｅｒｎｅｔ Ａｓｓｉｇｎｅｄ Ｎｕｍｂｅｒｓ Ａｕｔｈｏｒｉｔｙ）が、ｉＳＣＳＩ等のアプリケーションレイヤプロトコルに割り当てたＴＣＰポート番号である。

【００４１】

本実施形態では、ストレージ管理装置２がストレージ装置間通信で使用されているパスの状況を監視する。ここで、パスを構成するネットワーク装置３は、輻輳が発生すると、輻輳の発生をストレージ管理装置２へ通知する。輻輳の発生を通知されたストレージ管理装置２は、あらかじめ設定された変更率でマスタストレージ装置１が使用するパスの割合の変更後の値を計算し、その結果をマスタストレージ装置１へ送信する。結果を受信したマスタストレージ装置１は、受け取った情報に基づいてパスの割合を変更してストレージ装置間通信を継続する。

【００４２】

まず、システム管理者等があらかじめストレージ装置間通信のパス、そのパスの使用率、輻輳時のパスの割合の変更率等の情報をストレージ管理装置２及びマスタストレージ装置１に設定する際の処理手順について説明する。

【００４３】

最初に、第一の実施形態で用いられるグラフィカルユーザインタフェース（以下「ＧＵＩ」）について説明する。これらのＧＵＩは、ＣＰＵ２０４がＧＵＩ制御プログラム２１３を実行することによってディスプレイ２０５に表示される。システム管理者等は、文字入力装置２０７およびポインティング装置２０６を用いて、表示されたＧＵＩ上で各パラメータを設定する。

【００４４】

なお、ディスプレイ２０５、文字入力装置２０７およびポインティング装置２０６は、ストレージ管理装置２とは別の計算機が有していてもよい。例えば、ストレージ管理装置

２とＩＰ－ＳＡＮ７あるいはシリアルケーブルを介して接続されるコンソール用端末がディスプレイ２０５等を有していても良い。この場合、ＣＰＵ２０４は、ＧＵＩ制御プログラム２１３を実行して画面データをコンソール用端末へ送信し、コンソール用端末がディスプレイ２０５にＧＵＩを表示する。さらに、コンソール用端末は、システム管理者等が文字入力装置２０７とポインティング装置２０６とを用いて設定した各パラメータをストレージ管理装置２へ送信する。

【００４５】

また、ストレージ管理装置２は、第一の実施形態で説明するＧＵＩのかわりに、そのＧＵＩと同等の機能を有するコマンドラインインタフェースを備えてもよい。

【００４６】

図６は、システム管理者等が、ネットワーク装置３の情報をストレージ管理装置２に登録するために使用するネットワーク登録画面６００の表示例を示す図である。ネットワーク登録画面６００は、ネットワークＩＤが入力される領域６０１、ネットワークを構成するネットワーク装置３の情報が入力される領域６０２、領域６０２で入力された情報をネットワーク管理テーブル２２０に追加する際に使用されるボタン６０５、領域６０７を用いて指定されたネットワーク装置３に相当するレコードをネットワーク管理テーブル２２０から削除する際に使用されるボタン６０６、既にネットワーク管理テーブル２２０に登録されたネットワーク装置３のリストを表示する領域６０７およびネットワーク登録画面６００を閉じる際に使用されるボタン６１８を有する。

【００４７】

さらに、領域６０２は、ネットワーク装置３の装置ＩＤが入力される領域６０３およびネットワーク装置３が送信する輻輳発生通知や輻輳回復通知のソースＩＰアドレスが入力される領域６０４を有する。

【００４８】

図７は、システム管理者等が、リモートコピータスクの情報およびリモートコピータスクが使用する１以上のパスの情報をストレージ管理装置２に登録するために使用するリモートコピータスク登録画面７００の表示例を示す図である。リモートコピータスク登録画面７００は、タスクＩＤが入力される領域７０１、イニシエータｉＳＣＳＩ名が入力される領域７０２、イニシエータＬＵＮが入力される領域７０３、ターゲットｉＳＣＳＩ名が入力される領域７０４、ターゲットＬＵＮが入力される領域７０５、パスの情報を設定する領域７０６、パス間の負荷分散の割合の設定を行う領域７１９、これらの領域で指定された情報を登録する際に使用されるボタン７２８および登録を取り消す際に使用されるボタン７２９を有する。

【００４９】

さらに、領域７０６は、パスのパスＩＤが入力される領域７０７、当該パスが経由するネットワークのネットワークＩＤの一覧を表示してその内の一つを選択する際に使用されるボタン７０９、ボタン７０９で選択されたネットワークＩＤを表示する領域７０８、当該パスのイニシエータＩＰアドレスの一覧を表示してその内の一つを選択する際に使用されるボタン７１１、ボタン７１１で選択されたＩＰアドレスを表示する領域７１０、当該パスのターゲットＩＰアドレスの一覧を表示してその内の一つを選択する際に使用されるボタン７１３、管理者によって入力された又はボタン７１３で選択されたターゲットＩＰアドレスを表示する領域７１２、領域７０４に入力されたｉＳＣＳＩ名に対応するＩＰアドレスとポート番号とをｉＳＮＳ（ｉｎｔｅｒｎｅｔ Ｓｉｍｐｌｅ Ｎａｍｉｎｇ Ｓｅｒｖｉｃｅ）サーバやＳＬＰ ＤＡ（Ｓｅｒｖｉｃｅ Ｌｏｃａｔｉｏｎ Ｐｒｏｔｏｃｏｌ Ｄｉｒｅｃｔｏｒｙ Ａｇｅｎｔ）サーバ等のネーム管理サーバから読み出し、ボタン７１３で選択可能にする際に使用されるボタン７１４、当該パスの負荷分散の割合が入力される領域７１５、領域７０７から領域７１５までで指定された情報をパス管理テーブル２２２に追加する際に使用されるボタン７１６、領域７１８を用いて指定されたパスに相当するレコードをパス管理テーブル２２２から削除する際に使用されるボタン７１７および既にパス管理テーブル２２２に登録されたパスのリストを表示する領域７１８を

有する。

【００５０】

また、領域７１９は、輻輳時にパス間の負荷分散の割合を変更することを指定する際に使用されるボタン７２０およびボタン７２０が選択された時に入力可能となり、輻輳したパスの割合変更率が百分率で入力される領域７２１を有する。

【００５１】

なお、ボタン７１４が指定された時にストレージ管理装置２が通信を行うネーム管理サーバのＩＰアドレスは、ストレージ管理装置２にあらかじめ設定されているものとする。

【００５２】

以下、ＧＵＩ操作で各種情報が登録される時のストレージ管理装置２の処理について説明する。なお、以下の処理は、ＣＰＵ２０４で実行されるＧＵＩ制御プログラム２１３によって実行される。

【００５３】

図８（ａ）は、システム管理者等が、ネットワーク登録画面６００のボタン６０５あるいはボタン６０６を指定した時に、ストレージ管理装置２が実行するネットワーク登録処理の動作手順を示すフローチャートである。この処理によって、システムの管理者等は、ストレージ管理装置２にネットワーク装置３の情報を追加（又は削除）する。

【００５４】

まず、ストレージ管理装置２は、指定されたのがボタン６０５（追加）かボタン６０６（削除）かを判定する（Ｓ８０１）。ボタン６０５が指定された場合、ストレージ管理装置２は、ＧＵＩで入力された情報に基づいてネットワーク管理テーブル２２０にレコードを追加する。追加されるレコードのエントリ２２０１、エントリ２２０２およびエントリ２２０３には、ボタン６０５が指定された時の領域６０１、領域６０３および領域６０４の内容がそれぞれ登録される（Ｓ８０２）。一方、ボタン６０６が指定された場合、ストレージ管理装置２は、ボタン６０６が指定された時に、領域６０７で指定されたネットワーク装置３の装置ＩＤがエントリ２２０２の内容と一致する条件で、ネットワーク管理テーブル２２０を検索する（Ｓ８０３）。そして、ストレージ管理装置２は、見つかったレコードをネットワーク管理テーブル２２０から削除する（Ｓ８０４）。以上で、ストレージ管理装置２はネットワーク登録処理を終了する。

【００５５】

図８（ｂ）は、システム管理者等が、リモートコピータスク登録画面７００のボタン７１６あるいはボタン７１７を指定した時に、ストレージ管理装置２が実行するパス登録処理の動作手順を示すフローチャートである。システム管理者等は、この処理によって、リモートコピーの１つのタスクで使用する複数のパス、パスの割合、輻輳時のパスの割合の変更の要否等の情報をストレージ管理装置２に追加する。又、管理者等は、この処理によって、一旦設定したリモートコピーに使用するパスの本数等の変更（パス数の増加又は減少、パスの割合の変更等）をストレージ管理装置２に登録することが出来る。

【００５６】

まず、ストレージ管理装置２は、指定されたのがボタン７１６（追加）かボタン７１７（削除）かを判定する（Ｓ８１１）。

【００５７】

ボタン７１６が指定された場合、ストレージ管理装置２は、ＧＵＩで入力された情報に基づいてパス管理テーブル２２２にレコードを追加する。ここで追加するレコードのエントリ２２２１、エントリ２２２２、エントリ２２２３およびエントリ２２２４には、ボタン７１６が指定された時の領域７０７、領域７０１、領域７０８及び領域７１５の内容がそれぞれ登録される。エントリ２２２５にも、ボタン７１６が指定された時の領域７１５の内容が登録される。エントリ２２２６には「０」が登録される。エントリ２２２７およびエントリ２２２８には、ボタン７１６が指定された時の領域７１０および領域７１２の内容がそれぞれ登録される（Ｓ８１２）。

【００５８】

一方、ボタン７１７が指定された場合、ストレージ管理装置２は、ボタン７１７が指定された時に、領域７１８で指定されたパスのタスクＩＤおよびパスＩＤが、それぞれエントリ２２２２およびエントリ２２２１の内容と一致する条件で、パス管理テーブル２２２を検索する（Ｓ８１３）。そして、ストレージ管理装置２は、見つかったレコードをパス管理テーブル２２２から削除する（Ｓ８１４）。以上で、ストレージ管理装置２はパス登録処理を終了する。

【００５９】

図８（ｃ）は、システム管理者等が、リモートコピータスク登録画面７００を用いて、１つのリモートコピータスクの情報を登録した時に、ストレージ管理装置２が実行するリモートコピータスク登録処理の動作手順を示すフローチャートである。

【００６０】

この処理によって、システム管理者等は、１つのタスクについて、図８（ｂ）で示すパスの登録が正しく（パスの割合が１００％）設定されたかを確認して、そのタスクの情報をストレージ管理装置２へ登録することができる。

【００６１】

まず、ストレージ管理装置２は、領域７０１で入力されたタスクＩＤがエントリ２２２２の内容と一致する条件で、パス管理テーブル２２２を検索する（Ｓ８２１）。そして、見つかった全てのレコードのエントリ２２２５（割合）を取り出し、割合の和が１００であることを確認する（Ｓ８２２）。全ての割合の和が１００ではない場合、ストレージ管理装置２は、ディスプレイ２０５にエラーを伝えるＧＵＩを表示して、リモートコピータスク登録処理を終了する。この場合、システム管理者等は、再度パスの登録を行う（Ｓ８２５）。

【００６２】

全ての割合の和が１００の場合、ストレージ管理装置２は、タスク管理テーブル２２１にレコードを追加する。ここで追加するレコードのエントリ２２１１には領域７０１の内容が、エントリ２２１２には領域７０２の内容が、エントリ２２１３には領域７０３の内容が、エントリ２２１４には領域７０４の内容が、エントリ２２１５には領域７０５の内容が、エントリ２２１６にはボタン７２０がオフの場合には「０」が、ボタン７２０がオンの場合には「１」が、エントリ２２１７にはボタン７２０がオフの場合には「０」が、ボタン７２０がオンの場合には領域７２１の内容が、それぞれ登録される（Ｓ８２３）。

【００６３】

最後に、ストレージ管理装置２は、Ｓ８２３で登録されたリモートコピータスクの情報（タスクＩＤ、イニシエータｉＳＣＳＩ名、イニシエータＬＵＮ、ターゲットｉＳＣＳＩ名およびターゲットＬＵＮ）と、このリモートコピータスクで使用される全てのパスの情報（パスＩＤ、割合、イニシエータＩＰアドレスおよびターゲットＩＰアドレス）とを含むリモートコピー初期化要求を作成し、マスタストレージ装置１へ送信する（Ｓ８２４）。以上で、ストレージ管理装置２は、リモートコピータスク登録処理を終了する。　その後、リモートコピー初期化要求を受信したマスタストレージ装置１は、ＣＰＵ１０４でリモートコピー実行プログラム１１１を実行し、リモートコピー初期化処理を行う。

【００６４】

まず、マスタストレージ装置１は、受信したリモートコピー初期化要求からリモートコピータスクの情報とパスの情報とをディスク装置１０３に記憶する。そして、マスタストレージ装置１は、リモートストレージ装置５との間でｉＳＣＳＩセッションを確立する。このｉＳＣＳＩセッションの、イニシエータのｉＳＣＳＩ名は、受信したリモートコピー初期化要求から取り出したイニシエータｉＳＣＳＩ名とし、ターゲットのｉＳＣＳＩ名は、受信したリモートコピー初期化要求から取り出したターゲットｉＳＣＳＩ名とする。

【００６５】

また、マスタストレージ装置１は、受信したリモートコピー初期化要求から取り出したパスの情報ごとに、イニシエータＩＰアドレスをソースＩＰアドレスとし、ターゲットＩＰアドレスをデスティネーションＩＰアドレスとするＴＣＰコネクションをリモートスト

レージ装置５との間で確立し、すでに確立したｉＳＣＳＩセッションに割り当てる。

【００６６】

最後に、マスタストレージ装置１は、確立したｉＳＣＳＩセッションを用いて初期コピーを実行する。この初期コピーでは、マスタストレージ装置１は、受信したリモートコピー初期化要求から取り出したイニシエータＬＵＮで識別されるマスタストレージ装置１の論理ユニットに格納されたデータを、受信したリモートコピー初期化要求から取り出したターゲットＬＵＮで識別されるリモートストレージ装置５の論理ユニットに複製する。

【００６７】

尚、本実施形態では、リモートコピー初期化要求に基づいてストレージ管理装置２がマスタストレージ装置１にリモートコピーのタスクやパスの情報の登録とリモートコピーの初期化（初期コピー）及び開始の指示をまとめて行う形態とした。しかし、ストレージ管理装置２は、システム管理者等の情報登録の際に、リモートコピーのタスクやパスの情報のマスタストレージ装置２への送信のみを行い、その後、システム管理者（あるいはホスト４の使用者）からのリモートコピーの開始（又は準備の開始）指示を受け取った際に、ストレージ管理装置２（又はホスト４）が、マスタストレージ装置１へリモートコピーの初期化処理を指示しても良い。この場合、マスタストレージ装置１は、リモートコピーのバスの情報を受け取った際にはその情報をディスク装置１０３に格納する処理だけを行う。その後、マスタストレージ装置１は、リモートコピーの初期化処理を指示された際に、格納された情報に基づいてセッションの確立、初期化コピー等の処理を行う。

【００６８】

リモートコピー初期化要求に基づく処理の終了後（又はリモートコピー初期化要求の受領後）、マスタストレージ装置１はリモートコピーを開始する。リモートコピー実行中に、ストレージ管理装置２は、リモートコピーで使用されるパスの状態を監視している。

【００６９】

次に、第一の実施形態においてマスタストレージ装置１がリモートコピーを実行している際にネットワーク装置３で輻輳が発生した場合の各装置間の通信シーケンスについて説明する。

【００７０】

図９は、マスタストレージ装置１とリモートストレージ装置５との間でリモートコピーを実行中にネットワーク装置３にて輻輳が発生し、その後輻輳が回復するまでの、装置間での通信シーケンス例を示す図である。

【００７１】

まず、マスタストレージ装置１が、ホスト４からデータ書き込み要求を受信すると、書き込まれたデータをリモートストレージ装置５へ転送する。ここで、マスタストレージ装置１は、リモートコピー初期化処理においてストレージ管理装置２から受信したパスの割合に従って、データ転送に使用するＴＣＰコネクションを選択する。例えば、マスタストレージ装置１が、リモートストレージ装置５との通信に、二つのパスを使用し、かつそれぞれの割合がどちらも５０％であった場合、マスタストレージ装置１は、Ｓ９０１のデータ転送を第一のパスに対応するＴＣＰコネクションを用いて実行し、Ｓ９０２のデータ転送を第二のパスに対応するＴＣＰコネクションを用いて実行する（Ｓ９０１、Ｓ９０２）。

【００７２】

その後ネットワーク装置３にて輻輳が発生すると（Ｓ９０３）、そのネットワーク装置３は輻輳発生通知をストレージ管理装置２へ送信する（Ｓ９０４）。

【００７３】

輻輳発生通知を受信したストレージ管理装置２は、パス管理テーブル更新処理および割合計算処理を実行し、変更後のパス間の割合を得る。そして、ストレージ管理装置２は、マスタストレージ装置１へパスの割合変更要求を送信する（Ｓ９０５）。割合変更要求を受信したマスタストレージ装置１は、割合変更要求から変更後のパス間の割合を取り出してディスク装置１０３に記憶することで変更前のパス間の割合の値を変更後の値に変更し

た後、割合変更応答をストレージ管理装置２へ送信する（Ｓ９０６）。

【００７４】

その後、マスタストレージ装置１は、変更後のパス間の割合を用いて、リモートストレージ装置５へコピーするデータを転送する。例えば、パス間の割合が７５％と２５％とに変更されたら、マスタストレージ装置１は、データ転送の際、第一のパスに対応するＴＣＰコネクションを４回に３回の割合で使用し、第二のパスに対応するＴＣＰコネクションを４回に１回の割合で使用する（Ｓ９０７）。

【００７５】

その後、ネットワーク装置３にて発生した輻輳が回復すると（Ｓ９０８）、ネットワーク装置３は輻輳回復通知をストレージ管理装置２へ送信する（Ｓ９０９）。輻輳回復通知を受信したストレージ管理装置２は、パス管理テーブル更新処理および割合計算処理を実行し、変更後のパス間の割合を得る。

【００７６】

そして、ストレージ管理装置２は、マスタストレージ装置１へパスの割合変更要求を送信する（Ｓ９１０）。割合変更要求を受信したマスタストレージ装置１は、パス間の割合を変更した後、割合変更応答をストレージ管理装置２へ送信する（Ｓ９１１）。そして、マスタストレージ装置１は、変更後のパス間の割合を用いて、リモートストレージ装置５へデータを転送する。

【００７７】

第一の実施形態では、ネットワーク装置３は、ＳＮＭＰ（Ｓｉｍｐｌｅ Ｎｅｔｗｏｒｋ Ｍａｎａｇｅｍｅｎｔ Ｐｒｏｔｏｃｏｌ） Ｔｒａｐを用いて、輻輳発生通知や輻輳回復通知を、ＷＡＮ８およびＩＰ－ＳＡＮ７を介してストレージ管理装置２へ送信することを想定している。ここで、ＳＮＭＰ Ｔｒａｐを用いた輻輳発生通知および輻輳回復通知には、それぞれ輻輳発生および輻輳回復を表すＯＩＤ（Ｏｂｊｅｃｔ Ｉｄｅｎｔｉｆｉｅｒ）が含まれる。しかしネットワーク装置３はこれ以外のプロトコルを用いて輻輳の発生等をストレージ管理装置２に通知してもよい。

【００７８】

次に、図９で説明したパス管理テーブル更新処理や割合計算処理の詳細な処理手順について説明する。ここでの処理手順の概略は以下のとおりとなる。まず、ストレージ管理装置２が輻輳の発生したネットワーク装置３に関連するパスを特定する。次に、ストレージ管理装置２は特定されたパスを使用するタスクを選択する。その後、ストレージ管理装置２は、選択されたタスクでパス間の割合変更が必要な場合に変更後の値を計算し、その結果をタスクを実行しているストレージ装置へ割合変更要求として送信する。

【００７９】

図１０および図１１は、ストレージ管理装置２におけるパス管理テーブル更新処理の動作手順を示すフローチャートである。なお、以下の処理は、ＣＰＵ２０４で割合計算プログラム２１１が実行されることによって実行される。

【００８０】

パス管理テーブル更新処理を開始したストレージ管理装置２は、まず、パス管理テーブル２２０のエントリ２２０３に登録された値が受信した輻輳発生通知を送信したネットワーク装置３のＩＰアドレスと一致するという条件でネットワーク管理テーブル２２０を検索する。そして、ストレージ管理装置２は、該当するレコードのネットワークＩＤ（エントリ２２０１）を読み出す（Ｓ１００１）。

【００８１】

次にストレージ管理装置２は、エントリ２２２３に登録された値がＳ１００１で読み出したネットワークＩＤと一致するという条件でパス管理テーブル２２２を検索する。そして、ストレージ管理装置２は、検索条件に合致するするレコードを全て読み出す（Ｓ１００２）。

【００８２】

次にストレージ管理装置２は、受信した通知から取り出したＯＩＤから、通知された内

容が輻輳発生か輻輳回復かを判断する（Ｓ１００３）。輻輳発生の場合、ストレージ管理装置２はＳ１００２で読み出した全てのレコードの輻輳回数（エントリ２２２６）の値に１を加算する（Ｓ１００４）。一方、輻輳回復の場合、ストレージ管理装置２はＳ１００２で読み出した全てのレコードの輻輳回数（エントリ２２２６）の値から１を減算する（Ｓ１００５）。

【００８３】

輻輳回数の加減算をした後、ストレージ管理装置２はＳ１００２で読み出した各レコードのタスクＩＤ（エントリ２２２２）を読み出し、タスクＩＤのリストを作成する。この際、ストレージ管理装置２は重複しているタスクＩＤをリストから除外する（Ｓ１００６）。

【００８４】

（ここから図１１）次にストレージ管理装置２は、Ｓ１００６で作成したタスクＩＤリストの先頭のタスクＩＤを選択する（Ｓ１１０１）。そしてストレージ管理装置２は、エントリ２２１１に登録された値がＳ１１０１で選択したタスクＩＤと一致するという条件で、タスク管理テーブル２２１を検索する。その後ストレージ管理装置２は、検索条件に合致するレコードの割合変更フラグ（エントリ２２１６）と割合変更率（エントリ２２１７）とを読み出す（Ｓ１１０２）。

【００８５】

次にストレージ管理装置２は、読み出した割合変更フラグの値が「０」か「１」かを判定する（Ｓ１１０３）。割合変更フラグの値が「１」の場合、ストレージ管理装置２は、エントリ２２２２に登録された値がＳ１１０１で読み出したタスクＩＤと一致するという条件でパス管理テーブル２２２を検索する。そしてストレージ管理装置２は、検索条件に合致するレコードを全て読み出し（Ｓ１１０４）、図１２で説明する割合計算処理を実行する（Ｓ１１０５）　この割合計算処理の結果に基づき、ストレージ管理装置２は、Ｓ１１０４で読み出した各レコードの割合（エントリ２２２５の値）を更新する（Ｓ１１０６）。Ｓ１１０６の処理の終了後又はＳ１１０３で割合変更フラグが「０」と判断された場合、ストレージ管理装置２は、Ｓ１１０２からＳ１１０６までの処理をタスクＩＤリストに含まれる全てのタスクＩＤについて実行し（Ｓ１１０７、Ｓ１１０８）、パス管理テーブル更新処理を終了する。

【００８６】

図１２は、ストレージ管理装置２における割合計算処理の動作手順を示すフローチャートである。なお、以下の処理は、ＣＰＵ２０４で割合計算プログラム２１１が実行されることによって実行される。割合計算処理では、輻輳が発生したパスに割り当てられているパス間の割合を指定された変更率で減少させ、その減少分を他のパスに振り分ける処理を行う。また、全てのパスの輻輳回数が一致した場合（タスクで使用される全てのパスで同じ回数の輻輳が起こった場合に該当する）および全てのパスで輻輳が回復した場合には、割合計算処理ではパス間の割合をデフォルト値に戻す処理が行われる。

【００８７】

まずストレージ管理装置２は、図１１のＳ１１０４で読み出した全てのレコードの輻輳回数（エントリ２２２６）の値が一致するか否かを調べる（Ｓ１２０１）。全ての輻輳回数が一致する場合、ストレージ管理装置２は図１１のＳ１１０４で読み出した全てのレコードのパス間の割合（エントリ２２２５）をデフォルト割合（エントリ２２２４）の内容で上書きし（Ｓ１２０２）、割合計算処理を終了する。

【００８８】

一方Ｓ１１０４で読み出した全てのレコードの中に一つでも異なる輻輳回数の値がある場合、ストレージ管理装置２は全てのレコードの中の輻輳回数の最小値を調べる。この結果得た値を最小輻輳回数という（Ｓ１２０３）。次にストレージ管理装置２は、図１１のＳ１１０４で読み出した各レコードに対応するパスの相対輻輳回数を、以下の式に基づいて計算する（Ｓ１２０４）。

相対輻輳回数＝輻輳回数－最小輻輳回数

その後ストレージ管理装置２は、相対輻輳回数が０より大きいパスについて、パス間の割合と差分を以下の式に基づいて計算する（Ｓ１２０５）。ただし、「＾」はべき乗演算を表す。

パス間の割合＝デフォルト割合×（割合変更率＾相対輻輳回数）

差分＝パス間の割合－デフォルト割合

またストレージ管理装置２は、相対輻輳回数が０であるパスについて、パス間の割合を以下の式に基づいて計算する（Ｓ１２０６）。

割合の和＝相対輻輳回数が０のパスのデフォルト割合の和

パス間の割合＝デフォルト割合＋Ｓ１２０５で計算した全ての差分の和×デフォルト割合／割合の和

上述の割合等を計算したストレージ管理装置２は、割合計算処理を終了する。

【００８９】

上記の割合計算処理について、例えば、ストレージ装置間で行われる１つのリモートコピータスクで３つのパスが使用されており、かつ第一、第二および第三のパスの割合がそれぞれ４０（％）、３０（％）および３０（％）であり、輻輳時の割合変更率が５０（％）の場合を考える。この場合、第一のパス上で輻輳が発生したら、上記の割合計算処理を行うと、第一のパスの割合は２０（％）、残りのパスの割合はそれぞれ４０（％）となる。

【００９０】

以上で第一の実施形態を説明した。第一の実施形態によると、マスタストレージ装置１がリモートストレージ装置５との間の複数のパスを用いてｉＳＣＳＩを用いたリモートコピーを実行する際に、あるパス上のネットワーク装置３で輻輳が発生した時、ネットワーク装置３がストレージ管理装置２へ輻輳の発生を通知する。通知されたストレージ管理装置２は、それを契機に輻輳発生の結果として送信レートが低下したパスの使用率（ここでは割合と称している）を下げるよう、マスタストレージ装置１に指示する。これにより、例えばリモートコピー等におけるストレージ装置間のデータ転送のスループット低下を最小限に抑えることができる。

【００９１】

なお、第一の実施形態では、マスタサイト１１にあるマスタストレージ装置１と、リモートサイト１２にあるリモートストレージ装置５との間のリモートコピーに本発明を適用したが、１つのサイト（例えば、マスタサイト１１）にある二つのストレージ装置間のストレージ装置間通信（例えばリモートコピー）にも本発明は適用可能である。この場合、ネットワーク装置３はＩＰ－ＳＡＮ７を構成する機器となる。

【実施例２】

【００９２】

次に、第二の実施形態について第一の実施形態と相違する部分に限って説明する。第一の実施形態では、ネットワーク装置３がストレージ管理装置２にＳＮＭＰ　Ｔｒａｐ等のプロトコルを用いて輻輳発生通知と輻輳回復通知とを送信していた。しかし、この通知を行うためにはＳＮＭＰの標準仕様に手を加える必要があり、必ずしも全てのネットワーク装置３が実施可能とは限らない。

【００９３】

一方、ネットワーク装置３は一般に、廃棄したパケット数を記憶し外部からの要求に応じてその廃棄パケット数の情報を外部の装置に送信することができる。そこで本実施形態では、ストレージ管理装置２が定期的にネットワーク装置３から廃棄パケット数の情報を読み出し、前回読み出した値と比較して廃棄パケット数が増えていたら輻輳が発生したとみなす。一方、輻輳発生とみなしてから一定期間経過した後、ストレージ管理装置２が廃棄パケット数を読み出し、前回読み出した値と比較して廃棄パケット数が増えていなかったら輻輳が回復したとみなす。これにより、輻輳発生通知および輻輳回復通知を送信できないネットワーク装置３を有するシステムで、輻輳の発生によるパス間の割合の変更を行うようにする。

【００９４】
図１３（ａ）は、第二の実施形態におけるストレージ管理装置２の構成例を示した図である。ストレージ管理装置２のディスク装置２０３には、第一の実施形態のネットワーク管理テーブル２２０の代わりにネットワーク管理テーブル２２３が格納される。ネットワーク管理テーブル２２３の各レコードは、第一の実施形態で説明した各エントリに加え、ネットワーク装置３がバッファ領域不足のために廃棄したパケット数を表す輻輳廃棄パケット数が登録されるエントリ２２０４を有する。
【００９５】
ストレージ管理装置２の主メモリ２０１には、第一の実施形態で説明した輻輳受信プログラム２１０の代わりに、ネットワーク管理テーブル２２３に登録された全てのネットワーク装置３から定期的に廃棄パケット数の情報を含む統計データを読み出す際にＣＰＵ２０４で実行される統計データ読出プログラム２１４が格納される。
【００９６】
図１３（ｂ）は、第二の実施形態におけるネットワーク装置３の構成例を示した図である。ネットワーク装置３の主メモリ３０１には、第一の実施形態で説明した輻輳状態通知プログラム３１１の代わりに、各種統計データを記憶する統計データ記憶領域３１３および各種統計データを統計データ記憶領域３１３に書き込む際にＣＰＵ３０４で実行される統計データ記憶プログラム３１２が格納される。
【００９７】
統計データ記憶領域３１３に記憶される統計データには、ネットワーク装置３が受信したパケットのうち廃棄したパケットの総数（以下「廃棄パケット総数」という）、廃棄パケット総数のうちパケットのフォーマットエラーのために廃棄したパケット数（以下「エラー廃棄パケット数」という）および廃棄パケット総数のうちネットワーク装置３が対応できないプロトコルのパケットであったため廃棄したパケット数（以下「プロトコル不正廃棄パケット数」という）が挙げられる。
【００９８】
輻輳廃棄パケット数は以下の式で表される。
輻輳廃棄パケット数＝廃棄パケット総数－エラー廃棄パケット数－プロトコル不正廃棄パケット数
本実施形態におけるネットワーク登録処理は、ネットワーク管理テーブル２２０がネットワーク管理テーブル２２３に置き換わる点および図８（ａ）のＳ８０２にて、エントリ２２０４にレコード追加時点での輻輳廃棄パケット数が登録される点を除き、第一の実施形態と同様である。ここで、ストレージ管理装置２は、追加するレコードに対応するネットワーク装置３へ統計データ要求を送信し、廃棄パケット総数、エラー廃棄パケット数およびプロトコル不正廃棄パケット数といった統計データを要求する。そして、ストレージ管理装置２は、ネットワーク装置３から要求された統計データを含む統計データ応答を受信し、その統計データ応答の内容に基づき輻輳廃棄パケット数を計算する。これにより、ストレージ管理装置２は、レコード追加時点での輻輳廃棄パケット数を得る。
【００９９】
図１４は、第二の実施形態におけるシステム内の各装置間の通信シーケンス例を示す図である。第一の実施形態と同様、マスタストレージ装置１はストレージ管理装置２からの情報に基づいてリモートコピーを行っているとする。まず、マスタストレージ装置１が、デフォルトのパス間の割合に従って、リモートストレージ装置５へコピーするデータの転送を行う（Ｓ１４０１、Ｓ１４０２）。ストレージ管理装置２は、定期的にネットワーク装置３に対して統計データ要求を送信し、廃棄パケット総数、エラー廃棄パケット数およびプロトコル不正廃棄パケット数といった統計データを要求する（Ｓ１４０３）。
【０１００】
統計データ要求を受信したネットワーク装置３は、要求された統計データを含む統計データ応答を送信する（Ｓ１４０４）。統計データ応答を受信したストレージ管理装置２は、統計データ応答の内容に基づき輻輳廃棄パケット数を計算し、ネットワーク管理テーブ

ル２２３の当該ネットワーク装置３に対応するレコードのエントリ２２０４に登録された値と比較する。これらの値が一致した場合、ストレージ管理装置２は、そのネットワーク装置３で輻輳は発生していないとみなす。本シーケンスでは、この時点ではこれらの値が一致したとする。

【０１０１】

その後ネットワーク装置３で輻輳が発生するとする（Ｓ１４０５）。ストレージ管理装置２は輻輳発生後の所定のタイミング（又は一定間隔）で統計データ要求を送信し（Ｓ１４０６）、統計データ応答を受信する（Ｓ１４０７）。その後、ストレージ管理装置２は、統計データ応答の内容に基づき輻輳廃棄パケット数を計算し、ネットワーク管理テーブル２２３の当該ネットワーク装置３に対応するレコードのエントリ２２０４に登録された値と比較する。この時、輻輳が既に発生しているので、輻輳廃棄パケット数がエントリ２２０４に登録された値を上回る。

【０１０２】

この場合、ストレージ管理装置２は、そのネットワーク装置３に輻輳が発生したとみなし、ネットワーク管理テーブル２２３の当該レコードのエントリ２２０４に輻輳廃棄パケット数を書き込む。更にストレージ管理装置２は、図１０、図１１および図１２で説明したパス管理テーブル更新処理および割合計算処理を行う。その結果、パスの変更後の割合を得たストレージ管理装置２は、マスタストレージ装置１へパスの割合変更要求を送信し（Ｓ１４０８）、割合変更応答を受信する（Ｓ１４０９）。その後、マスタストレージ装置１は変更後のパス間の割合を用いて、リモートストレージ装置５へコピーするデータを送信する（Ｓ１４１０）。

【０１０３】

その後、ネットワーク装置３にて発生していた輻輳が回復するとする（Ｓ１４１１）。所定のタイミング（例えば定期的に）でストレージ管理装置２が統計データ要求を送信し（Ｓ１４１２）、統計データ応答を受信する（Ｓ１４１３）。統計データ応答を受信したストレージ管理装置２は、統計データ応答の内容に基づき輻輳廃棄パケット数を計算し、ネットワーク管理テーブル２２３の当該ネットワーク装置に対応するレコードのエントリ２２０４に登録された値と比較する。この場合、既に輻輳が回復しているので、これらの値が一致する。この場合輻輳が回復したとみなし、ストレージ管理装置２は、図１０、図１１および図１２で説明したパス管理テーブル更新処理、割合計算処理を行う。この処理の結果パスの変更後のパス間の割合を得たストレージ管理装置２は、マスタストレージ装置１へパスの割合変更要求を送信し（Ｓ１４１４）、割合変更応答を受信する（Ｓ１４１５）。

【０１０４】

なお、上記の通信シーケンス例では、ストレージ管理装置２がＳ１４０７で統計データ応答を受信し輻輳が発生したとみなした後、Ｓ１４１３で統計データ応答を受信し輻輳廃棄パケット数とエントリ２２０４に登録された値とを比較して一致した時に輻輳が回復したとみなしたが、輻輳廃棄パケット数とエントリ２２０４に登録された値とが連続して何回一致すれば、輻輳が回復したとみなすか、についてはシステム管理者等が任意に設定できるようにしてもよい。この回数は、例えば、ＷＡＮ８で輻輳が発生してから回復するまでの平均時間を、ストレージ管理装置２が統計データ要求を送信する間隔で割った結果を切り上げた値とすればよい。

【０１０５】

以上で第二の実施形態を説明した。第二の実施形態によると、ストレージ管理装置２が定期的にネットワーク装置３から廃棄パケットに関する統計データを読み出して輻輳発生や輻輳回復を判断することにより、割合の変更の指示等を第一の実施形態と同様に行うことができる。

【実施例３】

【０１０６】

次に、第三の実施形態を第一の実施形態と相違する部分に限って説明する。第三の実施

形態では、ネットワーク装置３は、ＥＣＮ（Ｅｘｐｌｉｃｉｔ Ｃｏｎｇｅｓｔｉｏｎ Ｎｏｔｉｆｉｃａｔｉｏｎ）というＲＦＣ３１６８で規定された標準仕様を用いて、マスタストレージ装置１に輻輳発生と輻輳回復とを直接通知する。従って、第一および第二の実施形態と比較して、ストレージ管理装置２と各装置間での通信量が削減されるという利点がある。なおネットワーク装置３とマスタストレージ装置１との間で使用される通信の仕様はＥＣＮに限る必要は無い。

【０１０７】

図１５（ａ）は、第三の実施形態におけるマスタストレージ装置１の構成例を示した図である。マスタストレージ装置１の主メモリ１０１には、第一の実施形態で説明した割合変更要求受信プログラム１１１の代わりに、ストレージ管理装置２からシステム管理者等が入力した設定内容を受信した際にＣＰＵ１０４で実行される設定受信プログラム１１３、ネットワーク装置３からＥＣＮによる輻輳通知や輻輳回復を受信した際にＣＰＵ１０４で実行される輻輳受信プログラム１１４およびパス間の負荷分散の割合を計算する際にＣＰＵ１０４で実行される割合計算プログラム１１５が格納される。

【０１０８】

また、ディスク装置１０３には、タスク管理テーブル２２１およびパス管理テーブル２２２が格納される。つまり、第一および第二の実施形態においてストレージ管理装置２で行われていた割合計算等の処理がマスタストレージ装置１にて行われることになる。

【０１０９】

図１５（ｂ）は、第三の実施形態におけるストレージ管理装置２の構成例を示した図である。本実施形態においては、ストレージ管理装置２は単なる設定情報の入力端末に過ぎない。従って、ストレージ管理装置２の主メモリ２０１には、ＧＵＩ制御プログラム２１３およびシステム管理者等が入力した設定内容をマスタストレージ装置１へ送信する際にＣＰＵ２０４で実行される設定送信プログラム２１５が格納される。ディスク装置２０３には、本発明に関わるいずれのテーブルも格納されない。

【０１１０】

第三の実施形態では、システム管理者等がディスプレイ２０５に表示されたリモートコピータスク登録画面７００上で、文字入力装置２０７およびポインティング装置２０６を用いて各パラメータを設定する。その後、システム管理者等がボタン７１６、ボタン７１７あるいはボタン７２８を指定すると、設定送信プログラム２１５がＣＰＵ２０４で実行され、設定内容がマスタストレージ装置１に送信される。なお、第三の実施形態では、リモートコピータスク登録画面７００の領域７０８およびボタン７０９は不要である。

【０１１１】

一方、設定内容を受信したマスタストレージ装置１は、設定受信プログラム１１３をＣＰＵ１０４で実行し、ストレージ管理装置２から受信した設定内容に基づき、タスク管理テーブル２２１およびパス管理テーブル２２２にレコードを登録あるいはテーブルからレコードを削除する。登録されるレコードの内容は、パス管理テーブル２２２のエントリ２２２３には何も登録されない点を除き、図８（ｂ）および（ｃ）で説明したそれぞれパス登録処理およびリモートコピータスク登録処理と同様である。なお、第三の実施形態ではネットワーク装置３をストレージ管理装置２で管理する必要が無いため、ネットワーク登録画面６００は不要である。

【０１１２】

図１６は、第三の実施形態における各装置間の通信シーケンス例を示す図である。ここでは、マスタストレージ装置１からリモートストレージ装置５へリモートコピーが実施されているとする。尚、リモートコピーの初期化要求や開始指示等は第一の実施形態と同じである。又、第三の実施形態で使用するＥＣＮはＴＣＰの拡張仕様であるため、図１６には、ＴＣＰレイヤの通信まで記す。

【０１１３】

まず、マスタストレージ装置１が、デフォルトのパス間の割合に従って、リモートストレージ装置５へコピーするデータを含むパケットを送信する（Ｓ１６０１）。なお、マス

タストレージ装置１は、ＥＣＮを用いることを示すために、Ｓ１６０１で送信するパケットのＩＰヘッダのＥＣＴ（ＥＣＮ Ｃａｐａｂｌｅ Ｔｒａｎｓｐｏｒｔ）フラグをＯＮにする。以降、第三の実施形態において、マスタストレージ装置１およびリモートストレージ装置５は、送信する全てのパケットのＩＰヘッダのＥＣＴフラグを、同様にＯＮにする。

【０１１４】

ネットワーク装置３は、パケットを受信すると、それをそのままリモートストレージ装置５へ転送する。パケットを受信したリモートストレージ装置５は、受信確認応答（以下「ＡＣＫ」という）をマスタストレージ装置１へ送信する（Ｓ１６０２）。ネットワーク装置３は、ＡＣＫを受信すると、そのままＡＣＫをマスタストレージ装置１へ転送する。

【０１１５】

その後、ネットワーク装置３にて輻輳が発生した場合（Ｓ１６０３）、マスタストレージ装置１がリモートストレージ装置５へコピーするデータを含むパケットを送信すると（Ｓ１６０４）、そのパケットを受信したネットワーク装置３は、受信したパケットのＩＰヘッダのＥＣＴフラグがＯＮであることを確認し、そのパケットのＣＥ（Ｃｏｎｇｅｓｔｉｏｎ Ｅｘｐｅｒｉｅｎｃｅｄ）フラグをＯＮにした上で、そのパケットをリモートストレージ装置５へ転送する。ここで、ＣＥフラグとは、ネットワーク装置３が輻輳の有無をエンドポイントに伝えるためのＩＰヘッダのフラグであり、ＲＦＣ３１６８で規定されている（Ｓ１６０５）。

【０１１６】

ＣＥフラグがＯＮになったパケットを受信したリモートストレージ装置５は、ＥＣＥ（ＥＣＮ Ｅｃｈｏ）フラグをＯＮにしたＡＣＫをマスタストレージ装置１へ送信する。ここで、ＥＣＥフラグとは、ＴＣＰ／ＩＰ通信を行うエンドポイント間で輻輳の有無に関する情報を交換するためのＴＣＰヘッダのフラグであり、ＲＦＣ３１６８で規定されている（Ｓ１６０６）。ネットワーク装置３は、そのＡＣＫをマスタストレージ装置１へ転送する。

【０１１７】

ＴＣＰヘッダのＥＣＥフラグがＯＮになったＡＣＫを受信したマスタストレージ装置１は、ＷＡＮ８内で輻輳が発生したとみなし、図１７で説明するパス管理テーブル更新処理および図１２で説明した割合計算処理を実行する。その結果パスの変更後の割合を得たマスタストレージ装置１は、パス間の割合を変更した上で、リモートストレージ装置５へ、コピーするデータを含むパケットを送信する（Ｓ１６０７）。

【０１１８】

この際、マスタストレージ装置１は、Ｓ１６０７で送信するパケットのＴＣＰヘッダのＣＷＲ（Ｃｏｎｇｅｓｔｉｏｎ Ｗｉｎｄｏｗ Ｒｅｄｕｃｅｄ）フラグをＯＮにする。ここで、ＣＷＲフラグとは、ＴＣＰ／ＩＰ通信を行うエンドポイント間で輻輳への対策を実行したことを伝えるためのＴＣＰヘッダのフラグであり、ＲＦＣ３１６８で規定されている。

【０１１９】

このパケットを受信したネットワーク装置３は、まだ輻輳が継続しているため、ＣＥフラグをＯＮにしてパケットをリモートストレージ装置５へ転送する（Ｓ１６０８）。リモートストレージ装置５が、ＣＷＲフラグ及びＣＥフラグとがＯＮであるパケットを受信すると、Ｓ１６０６と同様に、ＥＣＥフラグをＯＮにしたＡＣＫをマスタストレージ装置１に送信する（Ｓ１６０９）。

【０１２０】

その後、ネットワーク装置３にて発生していた輻輳が回復し（Ｓ１６１０）、マスタストレージ装置１がリモートストレージ装置５へコピーするデータを含むパケット送信すると（Ｓ１６１１）、そのパケットを受信したネットワーク装置３は、ＣＥフラグをＯＦＦのまま、そのパケットをリモートストレージ装置５へ転送する（Ｓ１６１２）。

【０１２１】

リモートストレージ装置５は、ＣＥフラグがＯＦＦのパケットを受信すると、ＥＣＥフラグがＯＦＦのＡＣＫをマスタストレージ装置１へ送信する（Ｓ１６１３）。ＥＣＥフラグがＯＦＦのＡＣＫを受信したマスタストレージ装置１は、ＷＡＮ８の輻輳が回復したとみなし、図１７で説明するパス管理テーブル更新処理および図１２で説明した割合計算処理を実行する。その結果、マスタストレージ装置１は、パス間の割合を元に戻してデータ転送を行う。

【０１２２】

図１７は、第三の実施形態におけるマスタストレージ装置１のパス管理テーブル更新処理の動作手順を示すフローチャートである。なお、以下の処理は、ＣＰＵ１０４で割合計算プログラム１１５が実行されることによって実行される。

【０１２３】

マスタストレージ装置１は、まずエントリ２２２７に登録された値が受信したパケットの宛先（デスティネーション）ＩＰアドレスに一致し、かつエントリ２２２８に登録された値が受信したパケットのソースＩＰアドレスに一致するという条件でパス管理テーブル２２２を検索し、検索条件に該当するレコードを読み出す。ここで、受信したパケットの宛先ＩＰアドレスはマスタストレージ装置１のＩＰアドレスであり、ソースＩＰアドレスはリモートストレージ装置５のＩＰアドレスである（Ｓ１７０１）。

【０１２４】

次に、マスタストレージ装置１は、受信したパケットの内容が輻輳発生を示すか輻輳回復を示すかを判定する。具体的には、受信したパケットのＴＣＰヘッダのＥＣＥフラグがＯＮであれば輻輳発生を示し、あるいはＯＦＦであれば輻輳回復であると判定する（Ｓ１７０２）。輻輳発生の場合、マスタストレージ装置１は、Ｓ１７０１で読み出したレコードの輻輳回数（エントリ２２２６）の値に１加算する（Ｓ１７０３）。一方輻輳回復の場合、マスタストレージ装置１は、Ｓ１７０１で読み出したレコードの輻輳回数（エントリ２２２６）の値から１減算する（Ｓ１７０４）。

【０１２５】

輻輳回数の加減算の後、マスタストレージ装置１は、当該レコードからタスクＩＤ（エントリ２２２２）を読み出す（Ｓ１７０５）。その後、マスタストレージ装置１は、エントリ２２１１に登録された値がＳ１７０５で読み出したタスクＩＤと一致するという条件でタスク管理テーブル２２１を検索し、検索条件に合致するレコードの割合変更フラグ（エントリ２２１６）と割合変更率（エントリ２２１７）とを読み出す（Ｓ１７０６）。

【０１２６】

次に、マスタストレージ装置１は、読み出した割合変更フラグの値が「０」か「１」かを判定する（Ｓ１７０７）。割合変更フラグが「０」の場合、マスタストレージ装置１は、パス管理テーブル更新処理を終了する。一方、割合変更フラグが「１」の場合、マスタストレージ装置１は、エントリ２２２２に登録された値がＳ１７０５で読み出したタスクＩＤと一致するという条件でパス管理テーブル２２２を検索し、検索条件に合致したレコードを全て読み出す（Ｓ１７０８）。

【０１２７】

その後、マスタストレージ装置１は、図１２で説明した割合計算処理を実行する（Ｓ１７０９）。この割合計算処理の結果に基づき、マスタストレージ装置１はＳ１７０８で読み出した各レコードの割合（エントリ２２２５）の値を更新し（Ｓ１７１０）、パス管理テーブル更新処理を終了する。

【０１２８】

以上で第三の実施形態を説明した。第三の実施形態によると、ネットワーク装置３がマスタストレージ装置１へ輻輳発生と輻輳回復とを直接通知することにより、第一の実施形態と比較してストレージ管理装置２のＣＰＵ負荷やストレージ管理装置２と他の装置との間のトラフィック量の増大を防止できる。

【実施例４】

【０１２９】

次に、第四の実施形態を第三の実施形態と相違する部分に限って説明する。第三の実施形態では、ネットワーク装置３が、ＥＣＮを用いてマスタストレージ装置１へ輻輳発生と輻輳回復とを通知する。本実施形態では、マスタストレージ装置１自体が輻輳発生や輻輳回復を検出する構成とする。

【０１３０】

第四の実施形態では、マスタストレージ装置１が、リモートコピーに使用するＴＣＰコネクションにおいてリモートストレージ装置５から送信されるＡＣＫがタイムアウト、すなわち一定期間経過しても返ってこない場合や、同じシーケンス番号を持つＡＣＫを３つ以上受信した場合に輻輳発生とみなす。これらの動作は、それぞれＲＦＣ７９３およびＲＦＣ２５８１が規定するＴＣＰの仕様に基づいている。

【０１３１】

マスタストレージ装置１は、上記の動作により輻輳発生を検出した場合、図１０、図１１および図１２で説明したパス管理テーブル更新処理および割合計算処理を行い、その結果得た変更後のパス間の割合を用いて、リモートコピーを実行する。

【０１３２】

一方、どのような契機を用いて輻輳回復とみなすかは、ＲＦＣで規定されていない。そこで、第四の実施形態では、例えば、輻輳発生からあらかじめシステム管理者等が設定した時間が経過した時に輻輳回復とみなすこととする。あるいは、ＴＣＰコネクションの輻輳ウィンドウサイズがあらかじめシステム管理者等が設定したサイズを超過した時に輻輳回復とみなしてもよい。マスタストレージ装置１は、上記の動作により輻輳回復を検出した場合、図１０、図１１および図１２で説明したパス管理テーブル更新処理および割合計算処理を行い、その結果得た変更後のパス間の割合を用いて、リモートコピーを実行する。

【０１３３】

以上で、第四の実施形態を説明した。第四の実施形態によると、マスタストレージ装置１が、自律的に輻輳発生および輻輳回復を検出することにより、ネットワーク装置３に依存せずにパス間の割合を変更することができる。

【実施例５】

【０１３４】

次に、第五の実施形態を第一の実施形態と相違する部分に限って説明する。第五の実施形態は、ホストストレージ装置間通信、すなわちホストとストレージ装置との間で実行されるデータの読み出しや書き込み（以下「ホストＩ／Ｏ」ともいう）に本発明を適用したシステムに関する。第五の実施形態では、ホストが、ストレージ装置と１つ以上のＴＣＰコネクションから成る１つのｉＳＣＳＩセッションを確立してホストＩ／Ｏを実行するものとする。ＴＣＰコネクションを用いて送信されるパケットが通る経路がパスに相当する。

【０１３５】

図１８は、第五の実施形態のシステム構成例を示した図である。システムは、データを記憶するストレージ装置１５、ストレージ装置１５からデータを読み出したり、データを書き込むホスト１６、端末６、ストレージ管理装置２、ホスト１６とストレージ装置１５とを接続するネットワーク装置３を有するＩＰ－ＳＡＮ７、端末６とホスト１６とを接続するＬＡＮ９、ホスト１６、ストレージ装置１５並びにストレージ管理装置２およびネットワーク装置３とを接続する管理用ネットワーク１３を有する。

【０１３６】

また、ホスト１６およびストレージ装置１５とＩＰ－ＳＡＮ７とは、通信線１０で接続されている。同様に、端末６およびホスト１６とＬＡＮ９とは、通信線１０で接続されている。さらに、ホスト１６、ストレージ装置１５、ストレージ管理装置２およびネットワーク装置３と管理用ネットワーク１３とは、通信線１４で接続されている。なお、第一の実施形態でも述べた通り、ＩＰ－ＳＡＮ７はＬＡＮ９を兼ねても良い。

【０１３７】

図１９はホスト１６の構成例を示した図である。ホスト１６は、主メモリ６０１、通信線６０２、ディスク装置６０３、ＣＰＵ６０４、ディスプレイ６０５、ポインティング装置６０６、文字入力装置６０７および複数のＮＩＦ６０８を有する。１つの以上のＮＩＦ６０８は通信線１４経由で管理用ネットワーク１３に、残りの１つ以上のＮＩＦ６０８は通信線１０経由でＩＰ－ＳＡＮ７に接続される。

【０１３８】

主メモリ６０１には、ストレージ管理装置２から割合変更要求を受信する際にＣＰＵ６０４で実行される割合変更要求受信プログラム６１０およびストレージ装置１５からデータを読み出したり、データを書き込む時に使用するパスを決定する際にＣＰＵ６０４で実行されるマルチパス制御プログラム６１１が格納される。

【０１３９】

ストレージ装置１５は、単体の記憶装置又は複数の記憶装置を有する記憶装置システムである。

ストレージ管理装置２およびネットワーク装置３の構成は第一の実施形態と同様である。また、端末６は一般的な計算機であり、ＣＰＵ、主メモリ、入出力装置および他の装置と通信線１０を介して接続するためのＮＩＦ等を有する。

【０１４０】

第五の実施形態における各種テーブルのデータ構造、ＧＵＩおよびＧＵＩ操作で各種情報が登録される際のストレージ管理装置２の処理は第一の実施形態と同様である。

【０１４１】

第五の実施形態における各装置間の通信シーケンスは、マスタストレージ装置１がホスト１６に置き換わる点、リモートストレージ装置５がストレージ装置１５に置き換わる点とを除き、第一の実施形態と同様である。

【０１４２】

つまり、マスタストレージ装置１に代わってホスト１６がストレージ管理装置２からパスの割合変更要求を受けて割合の変更を行うこととなる。

【０１４３】

第五の実施形態におけるパス管理テーブル更新処理および割合計算処理の動作手順は、第一の実施形態と同様である。

【０１４４】

以上で、第五の実施形態を説明した。第五の実施形態によると、ホスト１６が、ストレージ装置１５との間の複数のパスを用いてｉＳＣＳＩプロトコルに基づいたホストＩ／Ｏを実行するシステムにおいて、あるパス上のネットワーク装置３で輻輳が発生した時、ネットワーク装置３、ストレージ管理装置２およびホスト１６とが輻輳の情報をやり取りすることで、輻輳発生の結果送信レートが低下したパスの負荷分散の割合を下げることができる。これにより、第一の実施形態と同様に、通信に係るスループット低下を最小限に抑えることができる。

【実施例６】

【０１４５】

次に、第六の実施形態について第五の実施形態と相違する部分に限って説明する。第六の実施形態では、第二の実施形態と同様に、ストレージ管理装置２が定期的にネットワーク装置３から廃棄パケット数を読み出すことで、輻輳発生および輻輳回復を検出する。

【０１４６】

第六の実施形態におけるストレージ管理装置２およびネットワーク装置３の構成は、第二の実施形態と同様である。

【０１４７】

第六の実施形態における各装置間の通信シーケンスは、マスタストレージ装置１がホスト１６に置き換わる点、リモートストレージ装置５がストレージ装置１５に置き換わる点とを除き、第二の実施形態と同様である。

【０１４８】

以上で、第六の実施形態を説明した。第六の実施形態によると、ホスト１６が、ストレージ装置１５との間の複数のパスを用いてｉＳＣＳＩプロトコルに基づいたデータの書き込みや読み出しを実行するシステムにおいて、あるパス上のネットワーク装置３で輻輳が発生した時、ストレージ管理装置２から割合変更要求を受けたホストが輻輳発生の結果送信レートが低下したパスの割合を下げることができる。これにより、第一の実施形態と同様に、輻輳発生通知および輻輳回復通知を送信する機能を持たないネットワーク装置３を有するシステムでも、通信に係るスループット低下を最小限に抑えることができる。

【実施例７】

【０１４９】

次に、第七の実施形態を第五の実施形態と相違する部分に限って説明する。第七の実施形態では、第三の実施形態と同様に、ネットワーク装置３がＥＣＮ等を用いてホスト１６に輻輳発生および輻輳回復を通知する。

【０１５０】

第七の実施形態におけるホスト１６の構成は、主メモリ６０１に、設定受信プログラム６１２、輻輳受信プログラム６１３および割合計算プログラム６１４が格納される点およびディスク装置６０３に、タスク管理テーブル２２１およびパス管理テーブル２２２が格納される点以外は、第五の実施形態と同様である。

【０１５１】

第七の実施形態におけるストレージ管理装置２の構成は、第三の実施形態と同様である。

【０１５２】

第七の実施形態では、第三の実施形態と同様に、システム管理者等がリモートコピータスク登録画面７００上で設定した各パラメータを、ＣＰＵ２０４で実行される設定送信プログラム２１５がホスト１６へ送信する。そしてホスト１６のＣＰＵ６０４で実行される設定受信プログラム６１２が前記パラメータを受信し、その内容を各種テーブルのレコードに登録する。

【０１５３】

第七の実施形態における各装置間の通信シーケンスは、マスタストレージ装置１がホスト１６に置き換わる点、リモートストレージ装置５がストレージ装置１５に置き換わる点とを除き、第三の実施形態と同様である。

【０１５４】

第七の実施形態におけるホスト１６のパス管理テーブル更新処理の動作手順は、第三の実施形態と同様である。

【０１５５】

以上で、本発明の第七の実施形態を説明した。第七の実施形態によると、ネットワーク装置３が、ホスト１６へＥＣＮ等を用いて輻輳発生や輻輳回復を直接通知することにより、ストレージ管理装置２の機能が制限された装置を有するシステムでも、通信に係るスループット低下を最小限に抑えることができる。

【実施例８】

【０１５６】

次に、第八の実施形態を第七の実施形態と相違する部分に限って説明する。第八の実施形態では、第四の実施形態と同様に、ホスト１６自体が輻輳の発生や輻輳の回復を検出する。

【０１５７】

第八の実施形態では、第四の実施形態と同様に、ホスト１６が、データの読み出し又は書き込みに使用するＴＣＰコネクションにおいてＡＣＫがタイムアウトした場合や、同じシーケンス番号を持つＡＣＫを３つ以上受信した場合に、輻輳発生とみなす。

【０１５８】

一方、輻輳発生からあらかじめシステム管理者等が設定した時間が経過した時あるいはＴＣＰコネクションの輻輳ウィンドウサイズが、あらかじめシステム管理者等が設定した

サイズを超過した時に輻輳回復とみなす。

【０１５９】

以上で、本発明の第八の実施形態を説明した。第八の実施形態によると、ホスト１６が自律的に輻輳発生および輻輳回復を検出することにより、ネットワーク装置３に依存せず通信に係るスループット低下を最小限に抑えることができる。

【図面の簡単な説明】

【０１６０】

【図１】第一の実施形態におけるシステム構成例を示す図である。

【図２】第一の実施形態におけるマスタストレージ装置およびストレージ管理装置の構成例を示す図である。

【図３】第一の実施形態におけるネットワーク装置の構成例を示す図である。

【図４】ネットワーク管理テーブルおよびタスク管理テーブルのデータ構造例を示す図である。

【図５】パス管理テーブルのデータ構造例を示す図である。

【図６】ネットワーク登録画面の表示例を示す図である。

【図７】リモートコピータスク登録画面の表示例を示す図である。

【図８】ネットワーク登録処理、パス登録処理およびリモートコピータスク登録処理の動作を示すフローチャートである。

【図９】第一の実施形態における通信シーケンス例を示す図である。

【図１０】第一の実施形態におけるパス管理テーブル更新処理の動作を示すフローチャートである。

【図１１】第一の実施形態におけるパス管理テーブル更新処理の動作を示すフローチャートである。

【図１２】割合計算処理の動作を示すフローチャートである。

【図１３】第二の実施形態におけるストレージ管理装置およびネットワーク装置の構成例を示す図である。

【図１４】第二の実施形態における通信シーケンス例を示す図である。

【図１５】第三の実施形態におけるマスタストレージ装置およびストレージ管理装置の構成例を示す図である。

【図１６】第三の実施形態における通信シーケンス例を示す図である。

【図１７】第三の実施形態におけるパス管理テーブル更新処理の動作を示すフローチャートである。

【図１８】第五の実施形態におけるシステム構成例を示す図である。

【図１９】第五の実施形態におけるホストの構成例を示す図である。

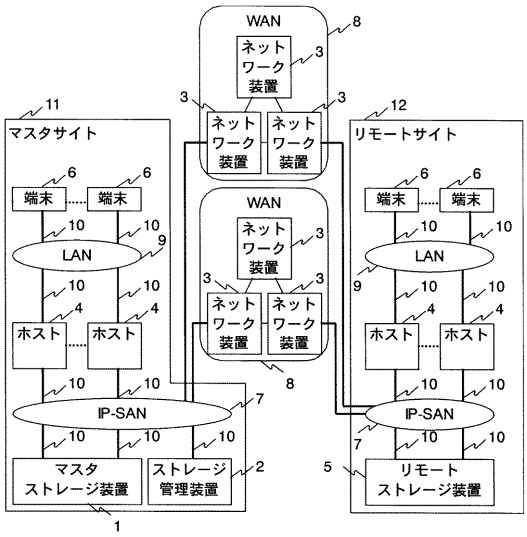
【符号の説明】

【０１６１】

１…マスタストレージ装置、２…ストレージ管理装置、３…ネットワーク装置、４…ホスト、５…リモートストレージ装置、６…端末、７…ＩＰ－ＳＡＮ、８…ＷＡＮ、９…ＬＡＮ、１０…通信線、１１…マスタサイト、１２…リモートサイト、１３…管理用ネットワーク、１４…通信線、１５…ストレージ装置、１６…ホスト。

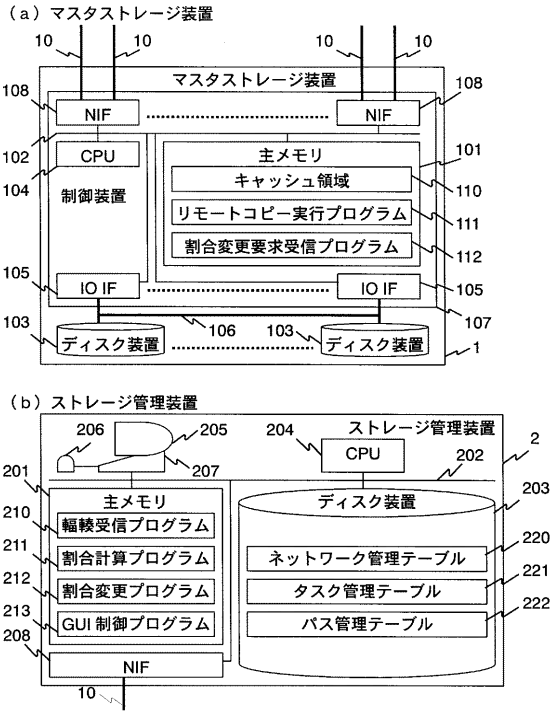
【図１】

図 1



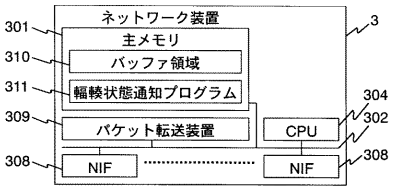
【図２】

図 2



【図３】

図 3



【図４】

図 4

（a）ネットワーク管理テーブル 220

2201 ネットワーク ID	2202 装置 ID	2203 IP アドレス
WAN01	DEV01	12.34.56.78
WAN01	DEV02	12.34.56.80
・・・	・・・	・・・

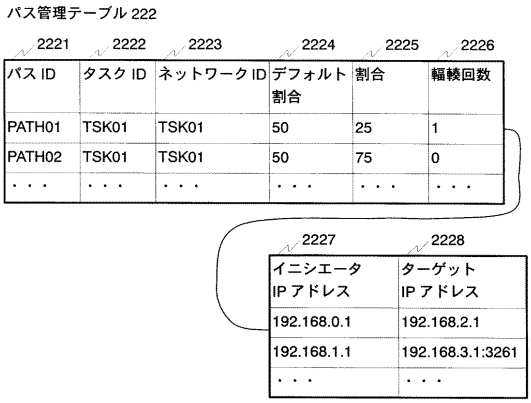
（b）タスク管理テーブル 221

2211 タスク ID	2212 イニシエータ iSCSI 名	2213 イニシエータ LUN	2214 ターゲット iSCSI 名	2215 ターゲット LUN
TSK01	mcu0101	mcu0101	rcu0101	mcu0101
TSK02	mcu0102	mcu0102	rcu0102	mcu0102
・・・	・・・	・・・	・・・	・・・

2216 割合変更フラグ	2217 割合変更率
1	50
1	50
・・・	・・・

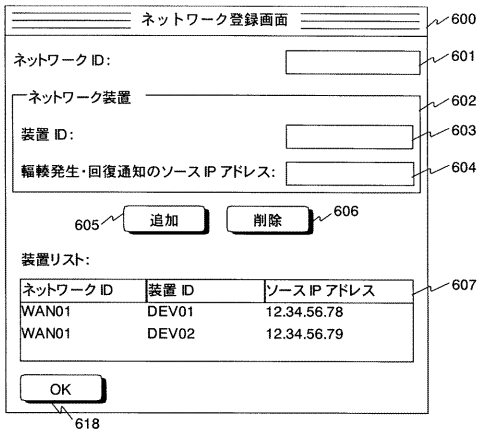
【図５】

図５



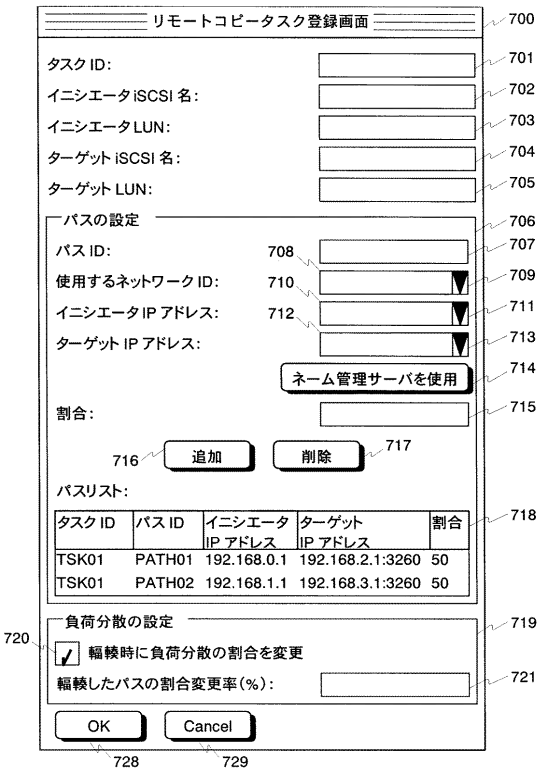
【図６】

図６



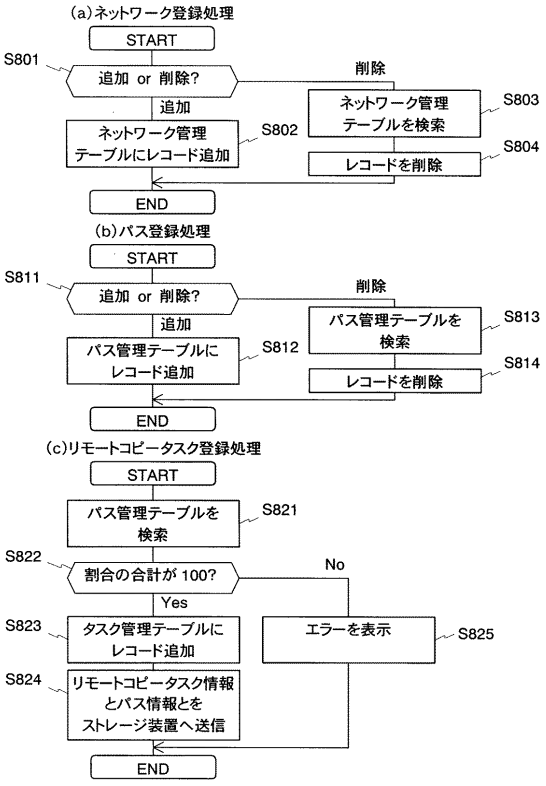
【図７】

図７



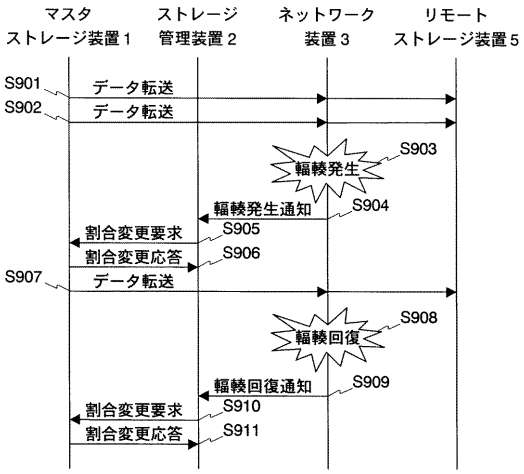
【図８】

図８



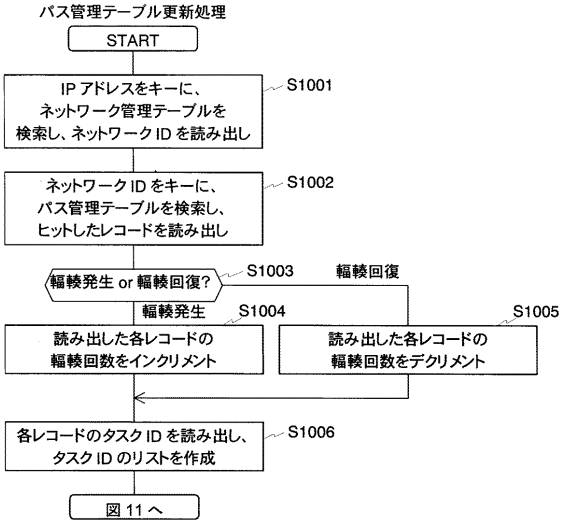
【図９】

図９



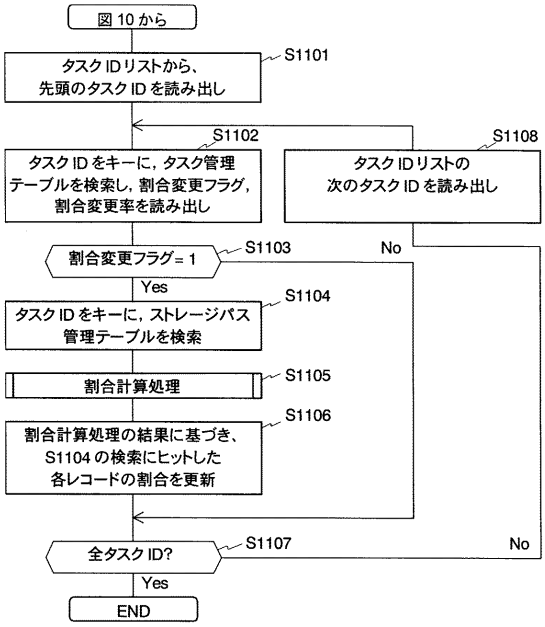
【図１０】

図１０



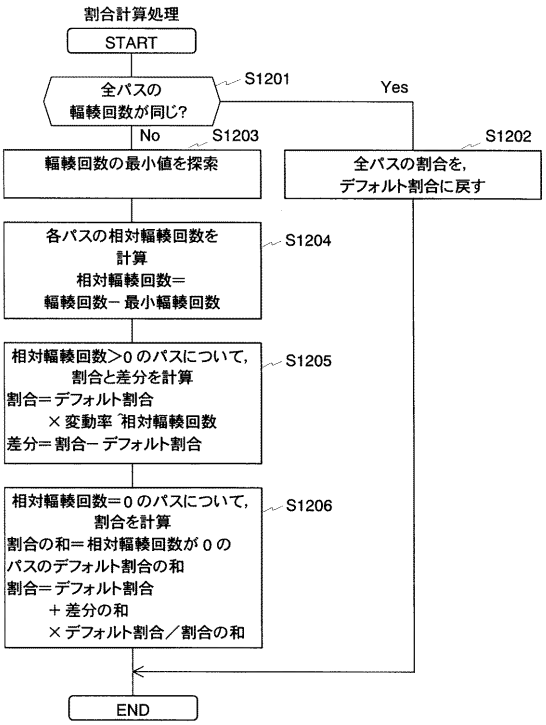
【図１１】

図１１



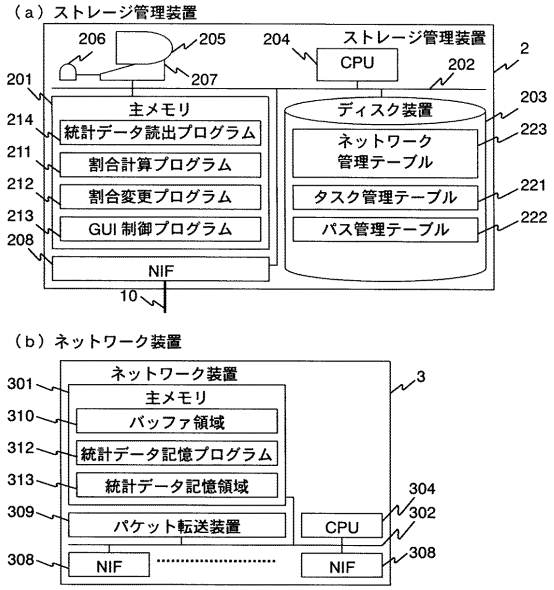
【図１２】

図１２



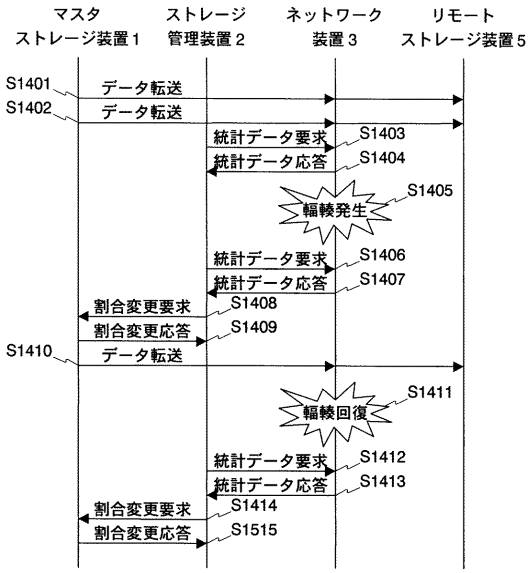
【図１３】

図１３



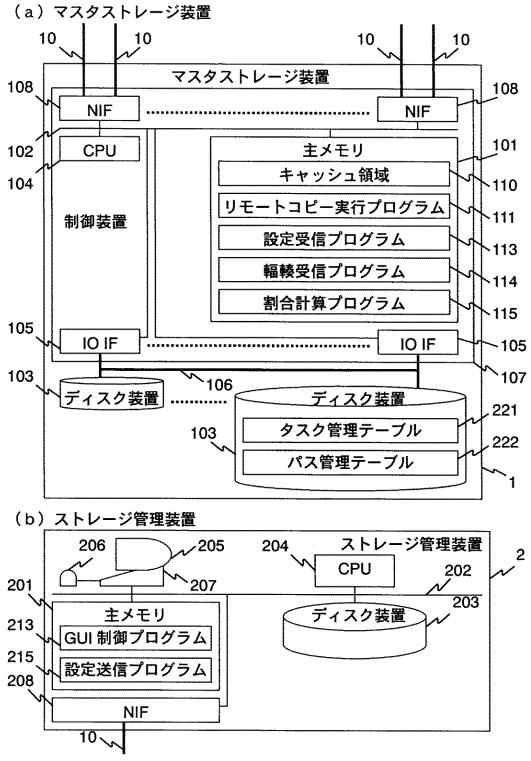
【図１４】

図１４



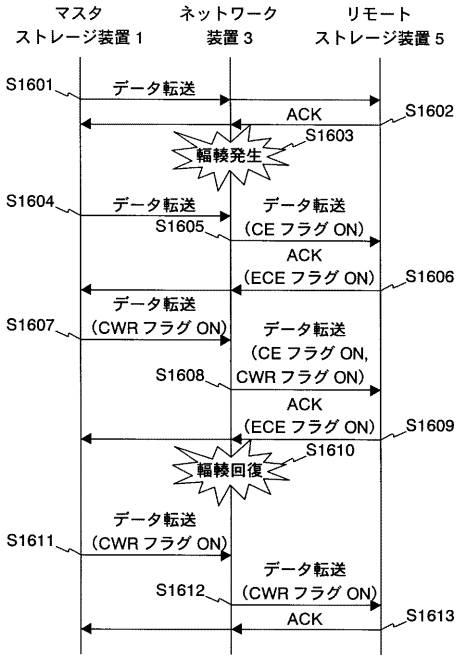
【図１５】

図１５



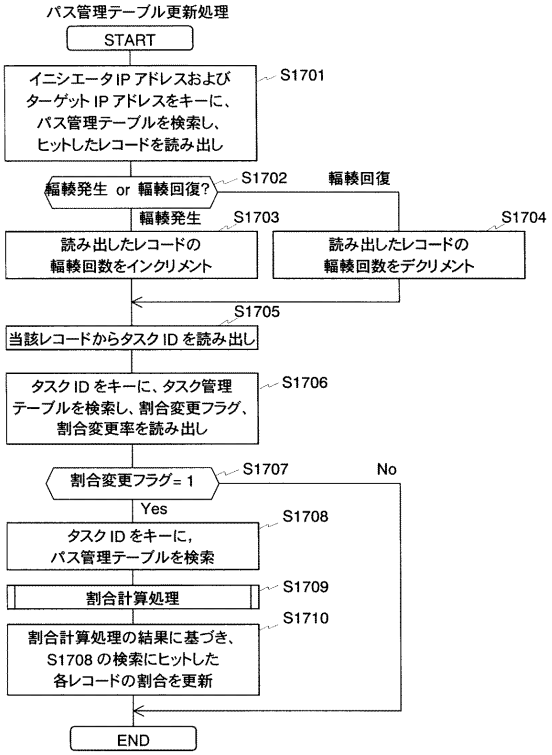
【図１６】

図１６



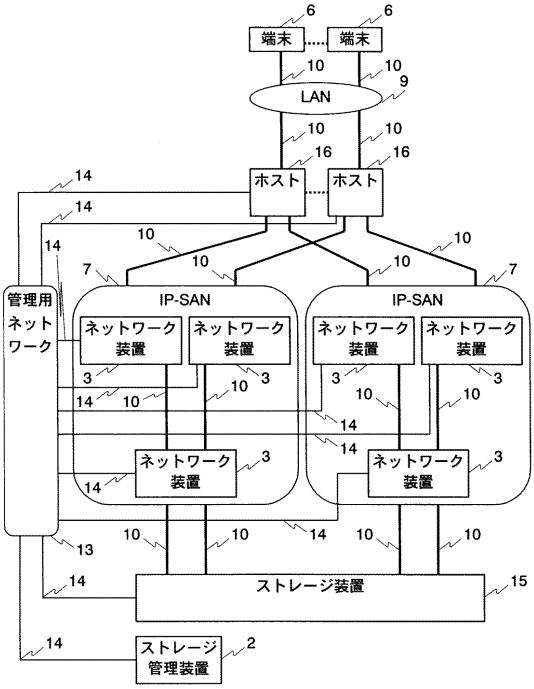
【図１７】

図１７



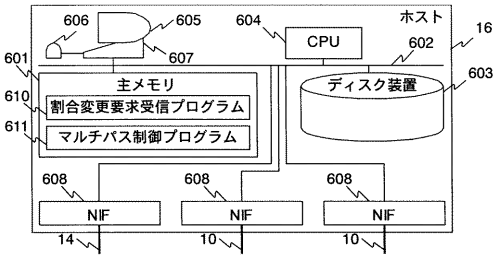
【図１８】

図１８



【図１９】

図１９



フロントページの続き

(72)発明者　藤原　啓成
神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所システム開発研究所内

Ｆターム(参考) 5B014 EB04
5B065 CA50
5K030 GA13 HA08 KA05 LB08 LE03
5K033 AA03 BA04 CB06 CB08 DB16