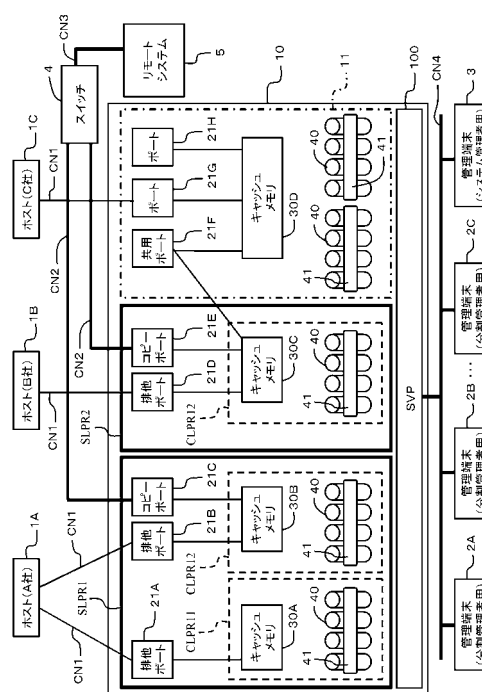




【特許請求の範囲】

【請求項 1】

通信ポートを介して上位装置及び外部装置とそれぞれ接続され、前記上位装置及び前記外部装置とのデータ授受を制御するチャネルアダプタと、

記憶デバイスとのデータ授受を制御するディスクアダプタと、

前記チャネルアダプタ及び前記ディスクアダプタによりそれぞれ使用されるキャッシュメモリと、

前記チャネルアダプタと前記ディスクアダプタと前記記憶デバイス及び前記キャッシュメモリがそれぞれ提供する資源を論理的に分割して構成される少なくとも 1 つ以上の論理的システム分割領域と、

前記論理的システム分割領域から前記外部装置へのデータ転送量を前記論理的システム分割領域毎に検出する転送量検出部と、

前記論理的システム分割領域毎に予め設定されている所定値と前記転送量検出部により検出されたデータ転送量とを比較し、前記データ転送量が前記所定値を超える場合は、前記論理的システム分割領域から前記外部装置へのデータ転送を制限する帯域制御を行うデータ転送制御部と、

を備えたディスクアレイ装置。

10

【請求項 2】

前記上位装置からの書込み要求に基づいて、前記外部装置へのデータ転送が行われる請求項 1 に記載のディスクアレイ装置。

20

【請求項 3】

前記所定値を可変に設定するための管理部を備えた請求項 1 に記載のディスクアレイ装置。

【請求項 4】

前記論理的システム分割領域内の設定変更を行う分割領域用管理部と、前記論理的システム分割領域への前記資源の割り当て及び前記所定値を可変に設定するためのシステム用管理部とを備えた請求項 1 に記載のディスクアレイ装置。

【請求項 5】

前記所定値は、前記論理的システム分割領域に割り当てられた前記資源の量に応じて設定可能である請求項 1 に記載のディスクアレイ装置。

30

【請求項 6】

前記データ転送制御部は、前記データ転送が非同期式で行われる場合と同期式で行われる場合とで、前記帯域制御を異ならせることができる請求項 1 に記載のディスクアレイ装置。

【請求項 7】

前記論理的システム分割領域から前記外部装置へのデータ転送には、初期データ転送と、差分データ転送とが含まれており、前記データ転送制御部は、前記初期データ転送を行う場合と前記差分データ転送を行う場合とで、前記帯域制御を異ならせることができる請求項 1 に記載のディスクアレイ装置。

【請求項 8】

前記データ転送制御部は、前記データ転送量が前記所定値を超える場合に、前記上位装置から前記論理的システム分割領域へのデータ書込み量を抑制することにより、前記データ転送を制限するものである請求項 1 に記載のディスクアレイ装置。

40

【請求項 9】

前記データ転送制御部は、前記データ転送量が前記所定値を超える場合に、前記上位装置からのデータ書込み要求に対する応答を予め設定された所定時間だけ遅延させることにより、前記データ転送を制限するものである請求項 1 に記載のディスクアレイ装置。

【請求項 10】

前記チャネルアダプタは、前記上位装置からのデータを受信するデータ受信部と、前記受信したデータを前記キャッシュメモリの所定領域に記憶させるキャッシュ制御部と、前

50

記キャッシュメモリから前記外部装置への転送対象となっているデータを取得するデータ取得部と、前記データ取得部により取得されたデータを前記外部装置に転送する転送処理部とを備えており、

前記データ転送制御部は、前記所定値と前記データ転送量とを比較する第1データ転送制御部と、前記データ転送量が前記所定値を超える場合は、前記論理的システム分割領域から前記外部装置へのデータ転送を制限する第2データ転送制御部とを備え、

前記転送量検出部は、前記転送処理部に設けられ、前記第1データ転送制御部は前記データ取得部に設けられ、前記第2データ転送制御部は前記データ受信部に設けられている請求項1に記載のディスクアレイ装置。

【請求項11】

前記論理的システム分割領域には、前記キャッシュメモリの記憶資源を分割して形成される少なくとも1つ以上のキャッシュ分割領域を設けることができる請求項1に記載のディスクアレイ装置。

【請求項12】

通信ポートを介して上位装置及び外部のディスクアレイ装置とそれぞれ接続され、前記上位装置及び前記外部のディスクアレイ装置とのデータ授受を制御するチャンネルアダプタと、

ディスクドライブとのデータ授受を制御するディスクアダプタと、

前記チャンネルアダプタ及び前記ディスクアダプタによりそれぞれ使用されるキャッシュメモリと、

前記チャンネルアダプタと前記ディスクアダプタと前記ディスクドライブに基づく論理デバイスと前記キャッシュメモリとがそれぞれ提供する資源を論理的に分割して構成される少なくとも1つ以上の論理的システム分割領域とを備え、

前記チャンネルアダプタは、

前記上位装置からのデータを受信するデータ受信部と、

前記受信したデータを前記キャッシュメモリの所定領域に記憶させるキャッシュ制御部と、

前記キャッシュメモリから前記外部のディスクアレイ装置への転送対象となっているデータを取得するデータ取得部と、

前記データ取得部により取得されたデータを前記外部のディスクアレイ装置に転送する転送処理部と、

前記論理的システム分割領域から前記外部のディスクアレイ装置へのデータ転送量を前記論理的システム分割領域毎に検出する転送量検出部と、

前記論理的システム分割領域毎に予め設定されている所定値と前記転送量検出部により検出されたデータ転送量とを比較する第1データ転送制御部と、

前記第1データ転送制御部により前記データ転送量が前記所定値を超えると判断された場合は、前記上位装置からのデータ書込み要求に対する応答を予め設定された所定時間だけ遅延させることにより、前記論理的システム分割領域から前記外部のディスクアレイ装置へのデータ転送を制限する第2データ転送制御部と、
を備えているディスクアレイ装置。

【請求項13】

通信ポートを介して上位装置及び外部のディスクアレイ装置とそれぞれ接続され、前記上位装置及び前記外部のディスクアレイ装置とのデータ授受を制御するチャンネルアダプタと、記憶デバイスとのデータ授受を制御するディスクアダプタと、前記チャンネルアダプタ及び前記ディスクアダプタによりそれぞれ使用されるキャッシュメモリと、前記チャンネルアダプタと前記ディスクアダプタと前記記憶デバイス及び前記キャッシュメモリがそれぞれ提供する資源を論理的に分割して構成される少なくとも1つ以上の論理的システム分割領域とを備えたディスクアレイ装置のリモートコピー制御方法であって、

前記論理的システム分割領域から前記外部のディスクアレイ装置へのリモートコピーを実行するか否かを判定するステップと、

10

20

30

40

50

前記リモートコピーを実行すると判定された場合は、リモートコピーを行う前記論理的システム分割領域を特定するステップと、

前記特定された論理的システム分割領域から前記外部のディスクアレイ装置へのデータ転送量を検出するステップと、

前記特定された論理的システム分割領域に予め設定されている最大転送量と前記検出されたデータ転送量とを比較するステップと、

前記データ転送量が前記最大転送量を超えると判定された場合は、前記上位装置からのデータ書込み要求に対する応答を予め設定された所定時間だけ遅延させることにより、前記特定された論理的システム分割領域から前記外部のディスクアレイ装置へのデータ転送を制限するステップと、

前記データ転送量が前記最大転送量以下であると判定された場合は、前記特定された論理的システム分割領域から前記外部のディスクアレイ装置へのデータ転送を制限することなく実行させるステップと、

を含んだディスクアレイ装置のリモートコピー制御方法。

【請求項 14】

通信ポートを介して上位装置及び外部装置とそれぞれ接続され、前記上位装置及び前記外部装置とのデータ授受を制御するチャネルアダプタと、記憶デバイスとのデータ授受を制御するディスクアダプタと、前記チャネルアダプタ及び前記ディスクアダプタによりそれぞれ使用されるキャッシュメモリと、前記チャネルアダプタと前記ディスクアダプタと前記記憶デバイス及び前記キャッシュメモリがそれぞれ提供する資源を論理的に分割して構成される少なくとも 1 つ以上の論理的システム分割領域とを備えたディスクアレイ装置の制御方法であって、

前記論理的システム分割領域から前記外部装置へのデータ転送量を前記論理的システム分割領域毎に検出するステップと、

前記論理的システム分割領域毎に予め設定されている所定値と前記検出されたデータ転送量とを比較するステップと、

前記データ転送量が前記所定値を超えると判定された場合は、前記論理的システム分割領域から前記外部装置へのデータ転送を制限するステップと、

を含んだディスクアレイ装置の制御方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ディスクアレイ装置及びディスクアレイ装置のリモートコピー制御方法に関する。

【背景技術】

【0002】

ディスクアレイ装置は、例えば、多数のディスクドライブをアレイ状に配設し、RAID (Redundant Array of Independent Inexpensive Disks) に基づいて構築されている。各ディスクドライブが有する物理的な記憶領域上には、論理的な記憶領域である論理ボリュームが形成されている。この論理ボリュームには LUN (Logical Unit Number) が予め対応付けられている。ホストコンピュータは、LUN 等を特定することにより、ディスクアレイ装置に対して所定形式の書込みコマンド又は読出しコマンドを発行する。これにより、ホストコンピュータは、ディスクアレイ装置に対して所望のデータの読み書きを行うことができる。

【0003】

ディスクアレイ装置には複数のホストコンピュータを接続可能である。ここで、あるホストコンピュータまたはアプリケーションプログラムが管理するデータ群を、他のホストコンピュータまたはアプリケーションプログラムから読み書き可能である場合、不都合を生じることがある。そこで、ゾーニングや LUN マスキング等の技術が用いられる。ゾーニングとは、ディスクアレイ装置に 1 つまたは複数のゾーンを設定し、ゾーンに属する特

10

20

30

40

50

定の通信ポートやWWN (World Wide Name) のみにデータ転送を許可する技術である。LUNマスキングとは、特定のホストコンピュータに対して特定のLUNへのアクセスを許可する技術である(特許文献1)。また、リモートコピーについては、後述の従来技術が知られている(特許文献2、特許文献3)。

【特許文献1】特開2003-30053号公報

【特許文献2】特開2002-334049号公報

【特許文献3】特開2003-032290号公報

【発明の開示】

【発明が解決しようとする課題】

【0004】

10

従来のゾーニングやLUNマスキングにより、ホストコンピュータからディスクアレイ装置へのアクセス制限を設定することができる。しかし、従来技術では、アクセス制限が可能となるだけで、セキュリティを考慮した資源の排他的使用や資源の分割を行うことができない。即ち、アクセス制限を設定している場合でも、管理端末のストレージ管理ソフトウェアから誤った構成変更を指示した場合は、他のホストコンピュータ用の資源にまで影響を及ぼす可能性がある。

【0005】

特に、ディスクアレイ装置の記憶領域を複数の企業等に割り当てて使用させるような場合は、各企業毎に使用される管理端末により、他の企業が管理する領域の構成が変更されないようにセキュリティを設定する必要があるが、従来技術ではこのような場合のセキュリティについて考慮されていない。

20

【0006】

ところで、ディスクアレイ装置では、データ保全等のために、データのバックアップを適宜行うようになっている。このバックアップ技術の1つとして、いわゆるリモートコピー(あるいはミラーリング)が知られている。リモートコピーでは、ローカルシステム(ローカルディスクアレイシステムであり、あるいはマスタシステムまたはプライマリシステムとも呼ばれる)に管理されているデータを、ローカルシステムから物理的に離れた場所に設置されるリモートシステム(リモートディスクアレイシステム)に転送する。これにより、ローカルシステムとリモートシステムとで同一内容のデータ群をそれぞれ保持することができる。

30

【0007】

1つのディスクアレイ装置の記憶資源を複数の企業等に割り当ててそれぞれ使用させる場合も、各企業の使用領域毎にそれぞれリモートコピーを行うことができる。ここで、ローカルシステムとリモートシステムとの間が、例えば高価な専用回線等で接続されている場合は、リモートコピーに使用する通信ネットワークの帯域が自ずと制限される。従って、限られた共通の通信資源を用いて複数の異なるリモートコピーを行う場合は、ある企業等の使用領域をリモートコピーすることにより、他の企業等が使用する領域のリモートコピーが影響を受ける可能性がある。即ち、例えば、あるリモートコピーが通信ネットワークの全帯域を長期間占有してリモートコピーを行っている間、他のリモートコピーを行うことができず、使い勝手等が低下する。

40

【0008】

そこで、特許文献2に記載の従来技術では、リモートコピーを行うサイドファイルの総量をディスクアレイ装置内で管理し、サイドファイルに占める各ホストコンピュータ毎のデータ量を個別に制御している。また、特許文献3に記載の従来技術では、ローカルシステムとリモートシステムとの間をチャネルエクステンダを介して接続し、ユーザポリシーに基づいて複数の通信ネットワークの中から自動的に通信ネットワークを選択するようにしている。

【0009】

しかし、ディスクアレイ装置内でサイドファイルの総量を制御した場合でも、ローカルシステムとリモートシステムとの間では、限られた共通の通信ネットワークを介してデー

50

タ転送が行われる。ローカルシステムとリモートシステム間の限られた通信資源を共用してリモートコピーを行うため、特許文献2に記載の技術では、各ホストコンピュータまたはアプリケーションプログラム間の影響が残り、改善の余地がある。

【0010】

特許文献3に記載の技術のように、チャンネルエクステンダを用いて通信ポート毎に使用帯域を割り当てることができる。しかし、この場合は、各ホストコンピュータ毎に予め通信ポートを割り当てておく必要があり、ディスクアレイ装置内の構成が変更された場合には、これに応じてチャンネルエクステンダの設定も変更する必要がある。従って、特許文献3の場合は、ディスクアレイ装置内の構成変更に対応することができず、使い勝手の点で改善の余地がある。

10

【0011】

そこで、本発明の1つの目的は、限られた通信資源を用いて複数のデータ転送を効率的に行うことが可能なディスクアレイ装置及びディスクアレイ装置のリモートコピー制御方法を提供することにある。本発明の1つの目的は、ディスクアレイ装置に設定される各論理的システム分割領域毎に、他の論理的システム分割領域におけるデータ転送の影響を低減して、データ転送を行うことができるディスクアレイ装置及びディスクアレイ装置のリモートコピー制御方法を提供することにある。本発明の他の目的は、後述する実施の形態の記載から明らかになるであろう。

【課題を解決するための手段】

【0012】

上記課題を解決すべく、本発明に従うディスクアレイ装置は、通信ポートを介して上位装置及び外部装置とそれぞれ接続され、上位装置及び外部装置とのデータ授受を制御するチャンネルアダプタと、記憶デバイスとのデータ授受を制御するディスクアダプタと、チャンネルアダプタ及びディスクアダプタによりそれぞれ使用されるキャッシュメモリと、チャンネルアダプタとディスクアダプタと記憶デバイス及びキャッシュメモリがそれぞれ提供する資源を論理的に分割して構成される少なくとも1つ以上の論理的システム分割領域と、論理的システム分割領域から外部装置へのデータ転送量を論理的システム分割領域毎に検出する転送量検出部と、論理的システム分割領域毎に予め設定されている所定値と転送量検出部により検出されたデータ転送量とを比較し、データ転送量が所定値を超える場合は、論理的システム分割領域から外部装置へのデータ転送を制限する帯域制御を行うデータ転送制御部と、備えている。

20

30

【0013】

論理的システム分割領域は、ディスクアレイ装置内に少なくとも1つ以上設定可能である。論理的システム分割領域は、チャンネルアダプタの提供する資源、ディスクアダプタの提供する資源、記憶デバイスの提供する資源、キャッシュメモリの提供する資源を論理的に分割することにより形成される。ここで、具体的な資源としては、例えば、チャンネルアダプタの有する通信ポート、記憶デバイスの記憶領域上に設定される論理デバイス（論理ボリューム）、キャッシュメモリ容量等のような物理的資源または論理的資源を挙げることができる。即ち、各論理的システム分割領域は、それぞれが独自のストレージシステム資源を占有している。そして、各論理的システム分割領域毎に、それぞれ少なくとも1つ以上の上位装置を接続することができる。さらに、論理的システム分割領域内には、キャッシュメモリの記憶資源を分割して形成される少なくとも1つ以上のキャッシュ分割領域を設けることもできる。

40

【0014】

各論理的システム分割領域毎のデータ転送に関して、予め所定値が設定されている。データ転送制御部は、転送量検出部により検出されたデータ転送量と所定値とを比較し、データ転送量が所定値を超える場合は、各論理的システム分割領域毎にデータ転送をそれぞれ制限させる。外部装置としては、例えば、ローカルサイト（プライマリサイト）のディスクアレイ装置と同様の構成を有するリモートサイト（セカンダリサイト）のディスクアレイ装置を挙げることができる。例えば、外部装置は、ローカルサイトに設置されるディ

50

スクアレィ装置のデータコピーを保持できれば足り、キャッシュメモリ量等が同一である必要はない。

【0015】

本発明の一態様では、上位装置からの書込み要求に基づいて、外部装置へのデータ転送が行われる。

本発明の一態様では、所定値を可変に設定するための管理部を備えている。

【0016】

本発明の一態様では、論理的システム分割領域内の設定変更を行う分割領域用管理部と、論理的システム分割領域への資源の割り当て及び所定値を可変に設定するためのシステム用管理部とを備えている。分割領域用管理部は、それぞれに割り当てられている論理的システム分割領域内の設定を変更する。システム用管理部は、ディスクアレィ装置全体の設定変更（資源の割り当てや所定値の設定等）を行う。各分割領域用管理部は、それぞれに関連付けられている論理的システム分割領域に対してのみ操作を行うことができ、他の論理的システム分割領域に対して操作を行うことはできない。これに対し、システム用管理部は、全ての論理的システム分割領域に対して操作可能である。

10

【0017】

本発明の一態様では、所定値は、論理的システム分割領域に割り当てられた資源の量に依りて設定可能である。例えば、論理的システム分割領域に割り当てられている論理ボリュームの量が大きい場合には、データ転送に使用する帯域が広くなるように制御することができる。あるいは、例えば、論理的システム分割領域に割り当てられているキャッシュメモリ量が大きい場合は、この論理的システム分割領域に割り当てる通信帯域を広くするように（所定値を大きくするように）制御することができる。

20

【0018】

本発明の一態様では、データ転送制御部は、データ転送が非同期式で行われる場合と同期式で行われる場合とで、帯域制御を異ならせることができる。ここで、同期式データ転送とは、上位装置からの書込み要求（データ更新操作）があった場合に、外部装置へのデータ転送を確認した後で、上位装置に対して書込み完了を報告するモードである。非同期式データ転送とは、上位装置からの書込み要求があった場合に、キャッシュメモリにデータを格納した時点で（つまり外部装置にデータを転送するよりも前に）、上位装置に対して書込み完了を報告するモードである。そして、例えば、同期式データ転送が選択される場合は、外部装置へのデータ転送の信頼性を求める場合であると考えることができ、非同期式データ転送が選択される場合は、応答速度の向上を求める場合であると考えることができる。従って、例えば、同期式データ転送が行われる場合は、データ転送の信頼性を確保しつつ応答性の低下を抑制するために、通信帯域が広くなるように（所定値を大きく設定するように）制御することができる。同期式データ転送を行う場合は、転送性能がサービス応答性に大きな影響を与えうるからである。

30

例えば、以下のような構成を採用可能である。即ち、同期式コピー転送用の帯域と非同期式コピー転送用の帯域とを、予め管理部を介して設定しておく。そして、データ転送が開始された場合に、同期式データ転送であるか非同期式データ転送であるかを判別し、同期式データ転送で転送されるデータを優先的に転送させる。同期式データ転送を優先する結果、非同期式データ転送で送信されるべきデータがキャッシュメモリに所定量以上溜まった場合は、非同期式データ転送に係わる上位装置に対して、データの流入制限を実施することができる。

40

なお、同期式データ転送または非同期式データ転送の帯域を設定する単位としては、例えば、転送対象（リモートコピー等）の論理ボリューム毎、または、複数の論理ボリュームをグループ化した各グループ毎、を挙げることができる。あるいは、ボリューム単位あるいは複数ボリュームから構成されるグループ単位で帯域を設定するのではなく、単純に、同期式データ転送と非同期式データ転送とで切り分けて、それぞれに対し帯域を設定することもできる。

【0019】

50

本発明の一態様では、論理的システム分割領域から外部装置へのデータ転送には、初期データ転送と、差分データ転送とが含まれており、データ転送制御部は、初期データ転送を行う場合と差分データ転送を行う場合とで、帯域制御を異ならせることができるようになっている。例えば、リモートコピーの場合は、ローカルシステムとリモートシステムとで同一のデータ群を保持する。各システムに同一のデータ群を形成する場合は、まず始めに、ある時点におけるローカルシステムの記憶内容をリモートシステムに丸ごとコピーする（初期データ転送あるいは初期コピー）。次に、初期コピー後に生じた差分データ（更新データ）をローカルシステムからリモートシステムに転送し、記憶内容を同期させる（差分データ転送あるいは差分コピー）。初期コピーが完了すると、ある時点におけるデータの複製がリモートシステムに構築され、リカバリ可能な状態となる。従って、例えば、リカバリ可能な状態への早期移行を望む場合は、初期コピーに広い通信帯域を割り当てることができる。 10

例えば、以下のような構成を採用可能である。初期コピー用の帯域と差分コピー用の帯域とを、予め管理部を介して設定しておく。帯域は、例えば、論理ボリューム毎、複数の論理ボリュームから構成されるグループ毎、あるいは、単純に初期コピーと差分コピーとで切り分ける等のようにして、設定可能である。そして、コピーが開始された場合は、初期コピーであるか差分コピーであるかを判別し、初期コピーであると判定された場合は、初期コピーのデータを優先的に転送させる。

【0020】

本発明の一態様では、データ転送制御部は、データ転送量が所定値を超える場合に、上位装置から論理的システム分割領域へのデータ書込み量を抑制することにより、データ転送を制限する。 20

【0021】

本発明の一態様では、データ転送制御部は、データ転送量が所定値を超える場合に、上位装置からのデータ書込み要求に対する応答を予め設定された所定時間だけ遅延させることにより、データ転送を制限する。ディスクアレイ装置からの応答を待って、上位装置は新たなデータ書込み要求を発行するため、応答を遅延させることにより、上位装置から書き込まれる単位時間あたりのデータ量を抑制することができる。

【発明を実施するための最良の形態】

【0022】

以下、図1～図12に基づき、本発明の実施の形態を説明する。本実施形態では、通信ポートを介して上位装置及び外部のディスクアレイ装置とそれぞれ接続され、上位装置及び外部のディスクアレイ装置とのデータ授受を制御するチャネルアダプタと、ディスクドライブとのデータ授受を制御するディスクアダプタと、チャネルアダプタ及びディスクアダプタによりそれぞれ使用されるキャッシュメモリと、チャネルアダプタとディスクアダプタとディスクドライブに基づく論理デバイスとキャッシュメモリとがそれぞれ提供する資源を論理的に分割して構成される少なくとも1つ以上の論理的システム分割領域とを備えたディスクアレイ装置が開示されている。さらに、チャネルアダプタは、上位装置からのデータを受信するデータ受信部と、受信したデータをキャッシュメモリの所定領域に記憶させるキャッシュ制御部と、キャッシュメモリから外部のディスクアレイ装置への転送対象となっているデータを取得するデータ取得部と、データ取得部により取得されたデータを外部のディスクアレイ装置に転送する転送処理部と、論理的システム分割領域から外部のディスクアレイ装置へのデータ転送量を論理的システム分割領域毎に検出する転送量検出部と、論理的システム分割領域毎に予め設定されている所定値と転送量検出部により検出されたデータ転送量とを比較する第1データ転送制御部と、第1データ転送制御部によりデータ転送量が所定値を超えると判断された場合は、上位装置からのデータ書込み要求に対する応答を予め設定された所定時間だけ遅延させることにより、論理的システム分割領域から外部のディスクアレイ装置へのデータ転送を制限する第2データ転送制御部と、を備えている。 40

【0023】

また、本実施形態には、少なくとも1つ以上の論理的システム分割領域を備えたディスクアレイ装置のリモートコピー制御方法も開示されている。このリモートコピー制御方法には、論理的システム分割領域から外部のディスクアレイ装置へのリモートコピーを実行するか否かを判定するステップと、リモートコピーを実行すると判定された場合は、リモートコピーを行う論理的システム分割領域を特定するステップと、特定された論理的システム分割領域から外部のディスクアレイ装置へのデータ転送量を検出するステップと、特定された論理的システム分割領域に予め設定されている最大転送量と検出されたデータ転送量とを比較するステップと、データ転送量が最大転送量を超えると判定された場合は、上位装置からのデータ書き込み要求に対する応答を予め設定された所定時間だけ遅延させることにより、特定された論理的システム分割領域から外部のディスクアレイ装置へのデータ転送を制限するステップと、データ転送量が最大転送量以下であると判定された場合は、特定された論理的システム分割領域から外部のディスクアレイ装置へのデータ転送を制限することなく実行させるステップと、を含んでいる。

10

20

30

40

50

【実施例1】

【0024】

図1は、ディスクアレイ装置10を含む記憶システムの全体概要を示すブロック図である。ディスクアレイ装置10は、ローカルサイトに設置されたローカルディスクアレイサブシステム（以下、ローカルシステムとも呼ぶ）である。ディスクアレイ装置10は、通信ネットワークCN1を介して、複数のホストコンピュータ1A～1C（以下、特に区別しない場合は、「ホストコンピュータ1」と呼ぶ）と双方向通信可能にそれぞれ接続されている。ここで、通信ネットワークCN1は、例えば、LAN（Local Area Network）、SAN（Storage Area Network）、インターネットあるいは専用回線等である。LANを用いる場合、ホストコンピュータ1とディスクアレイ装置10との間のデータ転送は、TCP/IP（Transmission Control Protocol/Internet Protocol）プロトコルに従って行われる。SANを用いる場合、ホストコンピュータ1とディスクアレイ装置10とは、ファイバチャネルプロトコルに従ってデータ転送を行う。また、ホストコンピュータ1がメインフレームの場合は、例えば、FICON（Fibre Connection：登録商標）、ESCON（Enterprise System Connection：登録商標）、ACONARC（Advanced Connection Architecture：登録商標）、FIBARC（Fibre Connection Architecture：登録商標）等の通信プロトコルに従ってデータ転送が行われる。

【0025】

各ホストコンピュータ1は、例えば、サーバ、パーソナルコンピュータ、ワークステーション、メインフレーム等として実現されるものである。例えば、各ホストコンピュータ1は、図外に位置する複数のクライアント端末と別の通信ネットワークを介して接続されている。各ホストコンピュータ1は、例えば、各クライアント端末からの要求に応じて、ディスクアレイ装置10にデータの読み書きを行うことにより、各クライアント端末へのサービスを提供する。後述のように、ディスクアレイ装置10内には、複数の仮想筐体（SLPR：Storage Logical Partition）が設けられている。そして、あるホストコンピュータ1Aは、A社に割り当てられたSLPR1のみにアクセス可能であり、他のホストコンピュータ1Bは、B社に割り当てられたSLPR2にのみアクセス可能である。さらに別のホストコンピュータ1Cは、非分割領域11内の各種資源を利用可能である。ホストコンピュータ1は、アクセス権限を有するSLPRのみにアクセス可能であり、他のSLPRを参照したり更新することはできず、また、未定義の論理デバイスにもアクセス不能である。

【0026】

ディスクアレイ装置10は、それぞれ後述するように、複数のチャネルアダプタ（以下、CHAと略記）20と、キャッシュメモリ30と、複数のディスクドライブ40と、複数のディスクアダプタ（以下、DKAと略記）50と、共有メモリ60と、スイッチ部70と、ハブ80と、SVP100とを備えている。

【0027】

ディスクアレイ装置10内には、複数のSLPR1，SLPR2が設定されている。ここで、SL

PRとは、ディスクアレイ装置10内の各種物理的資源及び論理的資源を各ユーザ毎に（あるいは各アプリケーションプログラム毎に）分割して割り当てることにより構成された領域である。即ち、例えば、各SLPRは、それぞれ専用の通信ポート（以下「ポート」）21と、それぞれ専用のキャッシュメモリ30と、それぞれ専用の論理デバイス（論理ボリューム）41とを備えている。具体的には、図示の例では、A社のみが使用可能なSLPR1は、3個のポート21A～21Cと、キャッシュメモリ30A、30Bと、論理デバイス41を備えている。また、B社のみが使用可能なSLPR2は、2個のポート21D、21Eと、キャッシュメモリ30Cと、論理デバイス41とを備えている。ディスクアレイ装置10は、各ユーザ毎に分割されたSLPR単位で管理されている。なお、C社のホストコンピュータ1Cは、専用の領域を持っておらず、非分割領域11を使用する。

10

【0028】

ディスクアレイ装置10内の各ポートには、次のような属性を設定可能である。例えば、排他ポートは、そのSLPRに占有されるポートであり、予め関連付けられたホストコンピュータ1以外に利用することができないポートであることを意味する。コピーポートは、外部のディスクアレイ装置であるリモートシステム5にデータを転送するためのポートであることを意味する。共用ポートとは、ディスクアレイ装置10内の全てのディスクドライブ40にアクセス可能なポートである。そのほか、SLPRを用いない場合の通常ポート、未使用ポート等の他の属性を設定することもできる。例えば、ポート21Gは通常ポート、ポート21Hは未使用ポートにそれぞれ該当する。

【0029】

各SLPRには、少なくとも1つ以上のキャッシュ分割領域（CLPR：Cache Logical Partition）を設けることができる。図示の例では、SLPR1内には、2個のCLPR11、CLPR12が設けられており、SLPR2内には、1個のCLPR21が設けられている。CLPRとは、キャッシュメモリ30を論理的に分割したものである。CLPR11にはキャッシュメモリ30Aが、CLPR12にはキャッシュメモリ30Bが、CLPR21にはキャッシュメモリ30Cが、それぞれ割り当てられている。SLPR内に割り当てられたキャッシュ容量をさらに複数のCLPRとして分割することにより、各CLPR毎に定められた容量のキャッシュメモリを各CLPRは独占的に利用することができる。従って、ある業務のアクセス量が増大した場合に、この業務用に大量のキャッシュが確保されてしまい、他の業務に影響を与えるというような事態を防止できる。

20

30

【0030】

SVP（Service Processor）100は、ディスクアレイ装置10の管理及び監視を行うためのコンピュータ装置であり、管理用のサーバ機能を提供する。SVP100は、例えば、ディスクアレイ装置10内に設けられたLAN等の内部ネットワークCN11（図2参照）を介して、各CHA20及び各DKA50等から各種の環境情報や性能情報等を収集する。SVP100が収集する情報としては、例えば、装置構成、電源アラーム、温度アラーム、入出力速度（IOPS）等が挙げられる。SVP100と各管理端末2A～2C、3とは、例えば、LAN等の通信ネットワークCN4を介して接続されている。管理者は、管理端末を介してSVP100にログインすることにより、権限のある範囲内において、例えば、RAID構成の設定、各種パッケージ（CHA、DKA、ディスクドライブ等）の閉塞処理、各種設定変更等を行うことができる。

40

【0031】

SVP100には、複数の管理端末2A～2C、3を接続可能である。ここで、管理端末2A～2Cは、各SLPR毎に設けられる端末であり、管理端末3は、ディスクアレイ装置10の全体を管理するために設けられる端末である。従って、管理端末2A～2Cは、各SLPRを管理する管理者（以下、分割管理者と呼ぶ）により操作される。管理端末3は、ディスクアレイ装置10の全体を管理するシステム管理者（あるいは全体管理者と呼ぶこともできる）により操作される。

【0032】

各SLPR毎の分割管理者は、自己が管理権限を有するSLPRについてのみ各種設定変更等を

50

行うことができ、他のSLPRの構成等を参照したり変更することは許可されない。これに対し、システム管理者は、各SLPRを含めてディスクアレイ装置10の全体の各種設定変更等を行うことができる。

【0033】

システム管理者は、管理端末3を介してSVP100にログインし、ディスクアレイ装置10の有する物理的資源及び論理的資源を適宜分割することにより、各ユーザ毎にSLPRを設定することができる。また、システム管理者は、各SLPRを管理する分割管理者に対して、ユーザID等を発行することもできる。分割管理者は、システム管理者から発行された専用のユーザIDを用いてSVP100にログインすることができる。分割管理者は、管理端末2を操作することにより、自己の管理下にあるSLPR内の設定を変更することができる。例えば、分割管理者は、業務形態に応じた構成を生成するために、ディスクドライブ40の割り当て、論理デバイス41の定義、キャッシュメモリ30の割り当て等を行うことができる。即ち、SLPRはシステム管理者が設定するが、各SLPR内のCLPRは分割管理者が設定する。

10

【0034】

ディスクアレイ装置10は、通信ネットワークCN3を介してスイッチ4に接続されている。そして、ディスクアレイ装置10は、スイッチ4から通信ネットワークCN3を介して、リモートディスクアレイサブシステム（以下、リモートシステムと略記）5に接続されている。スイッチ4は、例えば、チャネルエクステンダ等により構成される。スイッチ4の入力側には、各SLPR1, 2のコピーポート21C, 21Eが通信ネットワークCN2を介してそれぞれ接続されている。スイッチ4の出力側は、通信ネットワークCN3を介してリモートシステム5に接続されている。スイッチ4は、ローカルシステムであるディスクアレイ装置10からの信号を増幅し、リモートシステム5に送信する。

20

【0035】

各SLPR1, 2は、それぞれの管理するデータのうちリモートコピーすべきデータを、それぞれのコピーポート21C, 21E等を介して、リモートシステム5に送信し、リモートシステム5内にデータの複製を構築させる。これにより、万が一、ローカルシステムのデータが失われたり、機能を停止等した場合でも、リモートシステム5に記憶されているデータを用いてサービスを継続または早期に再開することができる。リモートシステム5は、データ複製に必要な記憶資源を備えていれば足り、ローカルシステムと同一の構成を備えている必要はない。図示は省略するが、一般的に、リモートシステム5は、CHAと、DKAと、キャッシュメモリと、ディスクドライブ等を備えることができる。

30

【0036】

図2は、ディスクアレイ装置10のハードウェア構成に着目したブロック図である。ディスクアレイ装置10内には、例えば、4個や8個等のように複数のCHA20を設けることができる。各CHA20は、それぞれポート21を備える。各CHA20は、例えば、オープン系用CHA、メインフレーム系用CHA等のように、ホストコンピュータ1の種類に応じて用意される。各CHA20は、ホストコンピュータ1との間のデータ転送を制御する。各CHA20は、それぞれプロセッサ部、データ通信部及びローカルメモリ部を備えている（いずれも不図示）。

40

【0037】

各CHA20は、それぞれに接続されたホストコンピュータ1から、データの読み書きを要求するコマンド及びデータを受信し、ホストコンピュータ1から受信したコマンドに従って動作する。DKA50の動作も含めて先に説明すると、例えば、CHA20は、ホストコンピュータ1からデータの読出し要求を受信すると、読出しコマンドを共有メモリ60に記憶させる。DKA50は、共有メモリ60を随時参照しており、未処理の読出しコマンドを発見すると、ディスクドライブ40からデータを読み出して、キャッシュメモリ30に記憶させる。CHA20は、キャッシュメモリ30に移されたデータを読み出し、コマンド発行元のホストコンピュータ1に送信する。

【0038】

50

また例えば、C H A 2 0 は、ホストコンピュータ 1 からデータの書込み要求を受信すると、書込みコマンドを共有メモリ 6 0 に記憶させると共に、受信したデータ（ユーザデータ）をキャッシュメモリ 3 0 に記憶させる。ここで、ホストコンピュータ 1 から書込みを要求されたデータは、ディスクドライブ 4 0 に書き込まれていない「ダーティデータ」であるため、例えば複数箇所にそれぞれ記憶されて多重化される。C H A 2 0 は、キャッシュメモリ 3 0 にデータを記憶した後、ホストコンピュータ 1 に対して書込み完了を報告する。そして、D K A 5 0 は、共有メモリ 6 0 に記憶された書込みコマンドに従って、キャッシュメモリ 3 0 に記憶されたデータを読み出し、所定のディスクドライブ 4 0 に記憶させる。ディスクドライブ 4 0 に書き込まれたデータは、「ダーティデータ」から「グリーンデータ」に属性が変化し、キャッシュメモリ 3 0 による多重管理から解放される。なお、10
本明細書において、「ダーティデータ」とは、ディスクドライブ 4 0 に書き込まれていないデータであって、多重管理が要請されるデータを意味する。また、「クリーンデータ」とは、ディスクドライブ 4 0 に書き込まれているデータであって、多重管理が要請されないデータを意味する。

【0039】

キャッシュメモリ 3 0 は、例えば、ユーザデータ等を記憶するものである。キャッシュメモリ 3 0 は、例えば不揮発メモリから構成される。キャッシュメモリ 3 0 は、複数のメモリから構成することができ、ダーティデータを多重管理することができる。

【0040】

ディスクアレイ装置 1 0 は、記憶装置 4 2 を備えている。記憶装置 4 2 は、多数のディスクドライブ 4 0 から構成される。各ディスクドライブ 4 0 は、例えば、ハードディスクドライブ（HDD）や半導体メモリ装置等として実現可能である。記憶装置 4 2 は、ディスクアレイ装置 1 0 の内部に設けられている必要はなく、外部に存在してもよい。また、20
ディスクアレイ装置 1 0 に固有の記憶装置 4 2 である必要はなく、旧型の記憶装置や他社の記憶装置であってもよい。記憶装置 4 2 が他社製品等の場合は、ディスクアレイ装置 1 0 の論理デバイス（LDEV）または論理デバイスの下に設けられる仮想デバイス（VDEV）に他社製の記憶装置 4 2 をマッピングし、あたかもディスクアレイ装置 1 0 自身の記憶装置であるかのようにして用いることができる。

【0041】

ここで、例えば、4 個のディスクドライブ 4 0 によって 1 つの R A I D グループを構成30
することができる。R A I D グループとは、例えば R A I D 5（R A I D 5 に限定されない）に従って、データの冗長記憶を実現するディスクグループである。各 R A I D グループにより提供される物理的な記憶領域の上には、論理的な記憶領域である論理デバイス 4 1（LUまたはLDEV）を少なくとも 1 つ以上設定可能である。

【0042】

各 D K A 5 0 は、ディスクアレイ装置 1 0 内に例えば 4 個や 8 個等のように複数個設けることができる。各 D K A 5 0 は、各ディスクドライブ 4 0 との間のデータ通信を制御するもので、それぞれプロセッサ部と、データ通信部と、ローカルメモリ等を備えている（いずれも不図示）。各 D K A 5 0 と各ディスクドライブ 4 0 とは、例えば、S A N 等の通信ネットワークを介して接続されており、ファイバチャネルプロトコルに従ってブロック40
単位 of データ転送を行う。各 D K A 5 0 は、ディスクドライブ 4 0 の状態を随時監視しており、この監視結果は内部ネットワークを介して S V P 1 0 0 に送信される。なお、各 C H A 2 0 及び各 D K A 5 0 は、例えば、プロセッサやメモリ等が実装されたプリント基板と、メモリに格納された制御プログラムとをそれぞれ備えており、これらのハードウェアとソフトウェアとの協働作業によって、後述する所定の機能を実現する。

【0043】

共有メモリ（あるいは制御メモリ）6 0 は、例えば不揮発メモリから構成される。共有メモリ 6 0 には、例えば、制御情報等が記憶される。なお、制御情報等の情報は、複数の共有メモリ 6 0 により多重管理することができる。共有メモリ 6 0 及びキャッシュメモリ 3 0 は、それぞれ複数個設けることができる。

【0044】

スイッチ部70は、各CHA20と、キャッシュメモリ30と、各DKA50と、共有メモリ60とをそれぞれ接続するものである。これにより、全てのCHA20、DKA50は、キャッシュメモリ30及び共有メモリ60にそれぞれアクセス可能である。

【0045】

SV P100は、ハブ80を介して各管理端末2、3にそれぞれ接続されている。SV P100は、共有メモリ60を介して、ディスクアレイ装置10内の各種構成情報を参照したり更新する。詳細は後述するが、SLPR構成情報管理テーブルT1は、SV P100及び共有メモリ60によりそれぞれ管理されている。

【0046】

次に、図3を参照して、各管理者が自分の管理するSLPRに関する情報を取得する概略を説明する。まず、分割管理者について説明する。分割管理者は、管理端末2を介してSV P100の有するSV Pサーバ部110にアクセスし、管理端末2の表示部に表示されたログイン画面にユーザIDを入力してログイン要求を出す(S1)。

【0047】

SV Pサーバ部110は、例えば、WWWサーバ機能を提供するもので、ユーザID管理テーブルT2を利用可能である。ユーザID管理テーブルT2は、例えば、ユーザ名(rotate、user1、user2等)と、各ユーザ(分割管理者)が管理するSLPR番号と、各ユーザの権限(SLPR管理者、システム管理者、参照のみ許可等)とを対応付けて管理しているものである。なお、ここでは、SLPRを特定するための情報として、SLPR番号を用いているが、番号以外の名称でもよい。

【0048】

SV Pサーバ部110は、管理端末2からのログイン要求を処理する(S2)。次に、SV Pサーバ部110は、管理端末2から入力されたユーザIDに基づいて、ユーザID管理テーブルT2を検索し、ログインしたユーザにより管理されているSLPR番号(図中、SLPR#)を取得する(S3)。SV Pサーバ部110は、この取得したSLPR番号をSV P構成情報管理部120に入力し、SLPR番号に対応するSLPR構成情報の取得を要求する(S4)。

【0049】

SV P構成情報管理部120は、例えば、SV P100内に設けられるもので、SLPR構成情報管理テーブルT1を利用可能である。SLPR構成情報管理テーブルT1は、例えば、SLPR番号と、各SLPR毎に割り当てられている物理的資源及び論理的資源に関する情報(資源分割情報)とが対応付けられている。資源分割情報としては、例えば、そのSLPRで利用可能な最大のキャッシュ容量、そのSLPRが使用可能なポート21の情報、そのSLPRが利用可能な論理デバイス(論理ボリューム)の情報、この論理デバイスに関連付けられているディスクドライブ群の情報を挙げることができる。SV P構成情報管理部120は、入力されたSLPR番号に基づいてSLPR構成情報管理テーブルT1を検索することにより、SLPR番号に対応するSLPRの構成情報を読み出し、SV Pサーバ部110に送信する(S5)。

【0050】

SV Pサーバ部110は、SV P構成情報管理部120から取得したSLPR構成情報を管理端末2に送信する(S6)。管理端末2は、SV Pサーバ部110から受信したSLPR構成情報を管理端末2の表示部に表示させる。これにより、分割管理者は、自己が管理するSLPRに関する情報のみを閲覧して操作することができる。分割管理者は、他のSLPRの参照が予め許可されていない場合、自己が管理するSLPR以外の他のSLPRに関する情報を参照等することができない。このようにして、ディスクアレイ装置10は、SLPR単位に分割されて管理される。

【0051】

システム管理者の場合も、分割管理者と同様の過程を経て、自己が管理するディスクアレイ装置10の構成情報を取得する。従って、詳細な説明を省略する。但し、分割管理者とは異なり、システム管理者は、ディスクアレイ装置10に含まれる全てのSLPRについて

10

20

30

40

50

構成情報を取得可能である。また、SVPサーバ部110やSVP構成情報管理部120は、所定のプログラムがSVP100上で実行されることにより、実現される。管理端末2, 3では、管理クライアントが実行されることにより、SLPR構成情報を取得することができるようにしている。

【0052】

図4は、SLPR構成情報管理テーブルT1の詳細を示す。SLPR構成情報は、例えば、SLPR管理テーブルT11と、CLPR管理テーブルT12と、ポート管理テーブルT13と、論理デバイス管理テーブルT14とによって管理することができる。

【0053】

SLPR管理テーブルT11には、例えば、各SLPR毎に、分割管理者が設定可能なSLPR名称と、そのSLPRにより管理されている各種情報とが対応付けられている。SLPR名称はディスクアレイ装置10内でユニークな名称を付けることができる。このSLPR名称により各SLPRが特定される。また、SLPRにより管理されている各種情報としては、例えば、SSID (Storage Subsystem ID) やCU (Control Unit) 等を挙げるができる。SSIDとは、メインフレームにおけるデバイスの制御単位の一つであり、論理デバイスのグルーピング単位を示す。CUとは、メインフレームにおけるデバイスの制御単位の一つであり、1つのCU内に例えば最大256個の論理デバイスを作成可能である。

【0054】

CLPR管理テーブルT12には、例えば、各CLPR毎に、分割管理者が設定可能なCLPR名称と、そのCLPRが属するSLPRを特定するためのSLPR番号 (SLPR名称) と、そのCLPRに割り当てられているキャッシュ容量とが対応付けられている。CLPR名称は、ディスクアレイ装置10内でユニークであれば足り、CLPR名称によって各CLPRを特定可能となっている。

【0055】

ポート管理テーブルT13には、各ポート毎に、そのポートが属するSLPR番号 (SLPR名称) と、そのポートに設定された属性とが対応付けられている。上述の通り、ポート属性としては、例えば、排他用ポート、共用ポート、未使用ポートがある。図1中に示すコピーポートは、排他用ポートのうちリモートコピーに使用されるポートを説明するために用いた説明の便宜上の属性であり、実際にコピーポートというポート属性が予め用意されているわけではない。

【0056】

論理デバイス管理テーブルT14には、例えば、各論理デバイス毎に、論理デバイスが属するCLPR番号 (CLPR名称) と、ボリュームサイズと、管理情報とが対応付けられている。管理情報としては、例えば、その論理デバイスが形成されるディスクドライブ群を特定するための情報等を挙げるができる。これらの各管理テーブルT11～T14により、SLPR構成情報を管理することができる。なお、テーブル構成は上述したものに限らない。

【0057】

次に、キャッシュメモリ30の管理方法について説明する。図5に示すように、キャッシュメモリ30は、スロットと呼ばれる論理的な分割単位で分割されている。各スロットは、さらにセグメントと呼ばれる論理的な分割単位で分割されている。各CLPRには、スロット単位でキャッシュ容量が割り当てられている。一方、図4に示すCLPR管理テーブルT12では、セグメント単位でキャッシュ容量を管理している。

【0058】

図6に示すように、キャッシュ管理テーブルT3には、各CLPRにおいて現在使用されているセグメント数と、未使用のセグメント数を示すフリーセグメント数と、使用中のセグメントのうちクリーンデータを記憶しているスロット数とが対応付けられている。既に述べた通り、クリーンデータとは、ホストコンピュータ1からの書込みデータのうちディスクドライブ40に書込み済のデータや、ディスクドライブ40から読み出されたデータであって、冗長管理の必要性がないデータである。従って、クリーンデータのみを記憶するスロットからクリーンデータを削除して、解放することができる。このキャッシュ管理テ

10

20

30

40

50

ーブル T 3 は、他のテーブル類と同様に共有メモリ 60 または共有メモリ 60 及び S V P 100 に記憶することができる。

【0059】

キャッシュ管理テーブル T 3 と CLPR 管理テーブル T 1 2 とは、互いに関連付けられており、各 CLPR 毎に使用可能なキャッシュ容量を簡単に求めることができるようになっている。

【0060】

図 7 は、各 CLPR にセグメントを割り当てるためのセグメント確保処理の概要を示すフローチャートである。本処理は、C H A 20 または D K A 50 に格納されているキャッシュ制御プログラムにより実行することができる。ホストコンピュータ 1 からの書込み要求に応じてセグメントを確保する場合は、C H A 20 が本処理を実行する。ホストコンピュータ 1 からの読出し要求に応じてセグメントを確保する場合は、D K A 50 が本処理を実行する。以下の説明では、C H A 20 により実行される場合を例に挙げる。

【0061】

ホストコンピュータ 1 からの書込み要求または読出し要求が発行されると、C H A 20 は、この要求の処理に必要なセグメント数を算出する (S 11)。次に、C H A 20 は、ホストコンピュータ 1 からアクセスが要求された論理デバイスの情報に基づいて、対応する CLPR を特定する (S 12)。

【0062】

CLPR が特定されると、次に、C H A 20 は、キャッシュ管理テーブル T 3 に管理されている使用中セグメント数と CLPR 管理テーブル T 1 2 に管理されているキャッシュ容量とを比較する (S 13)。使用中セグメント数の方が CLPR のキャッシュ容量よりも少ない場合は (使用セグメント数 < キャッシュ容量)、空いているセグメントが存在する場合である。そこで、C H A 20 は、S 11 で算出された要求セグメント数とフリーセグメント数とを比較する (S 14)。要求セグメント数がフリーセグメント数以下である場合 (要求セグメント数 ≤ フリーセグメント数)、C H A 20 は、要求セグメント数と同数のフリーセグメントを確保する (S 15)。そして、C H A 20 は、キャッシュ管理テーブル T 3 のフリーセグメント数から確保済セグメント数を減算してテーブルを更新させる (S 16)。また、同様に、C H A 20 は、キャッシュ管理テーブル T 3 の使用中セグメント数に確保済セグメント数を加算してテーブルを更新される。これにより、本処理は終了する。

【0063】

一方、要求セグメント数の方がフリーセグメント数よりも多い場合 (要求セグメント数 > フリーセグメント数)、C H A 20 は、その CLPR で使用可能な全てのフリーセグメントを確保する (S 17)。キャッシュ管理テーブル T 3 のフリーセグメント数は「0」に更新される (S 18)。

【0064】

次に、確保した全てのフリーセグメントだけでは足りない不足分のセグメントを確保する必要がある。そこで、C H A 20 は、不足分のセグメント数に応じた数だけクリーンスロット数を減算して (S 19)、必要な分以上のセグメントを解放する。C H A 20 は、解放されたセグメントの中から必要なセグメントを確保する (S 20)。C H A 20 は、解放されたクリーンスロットからセグメントを確保した際の余ったセグメントを、フリーセグメントに加算する (S 21)。そして、C H A 20 は、キャッシュ管理テーブル T 3 の使用中セグメント数に要求セグメント数を加算してテーブルを更新させ (S 22)、本処理を終了する。

【0065】

一方、ホストコンピュータ 1 からの要求を処理するのに必要な要求セグメント数がキャッシュ容量以上の場合 (要求セグメント数 ≥ キャッシュ容量)、C H A 20 は、要求セグメント数に応じた数だけクリーンスロットを解放し (S 19)、要求セグメント数と同数のセグメントを確保する (S 20)。そして、C H A 20 は、余ったセグメントの数をフリーセグメント数に加算し (S 21)、要求セグメント数だけ使用中セグメント数を加算

し（S22）、本処理を終了する。

【0066】

次に、本実施例によるリモートコピーの制御方法について説明する。まず最初に概要を簡単に説明すると、ディスクアレイ装置10からリモートシステム5にデータを転送する場合、ホストコンピュータ1から書き込まれたデータ及び管理情報は、キャッシュメモリ30に一時的に格納される。そして、このデータ及び管理情報は、キャッシュメモリ30からCHA20や通信ネットワークCN2、CN3を介してリモートシステム5に送信される。リモートシステム5は、ディスクアレイ装置10からデータを受信すると、この受信データをリモートシステム5内のキャッシュメモリに格納する。そして、リモートシステム5は、キャッシュメモリに格納したデータをディスクドライブに書込む。リモートシステム5は、キャッシュメモリにデータを格納した時点で、ディスクアレイ装置10に対し、書込み完了を報告することができる。リモートシステム5内のディスクドライブへのデータ書込みタイミングは、リモートシステム5内のキャッシュメモリへのデータ格納タイミングと非同期で行われる。

【0067】

同期式リモートコピーの場合、ホストコンピュータ1から書込みを要求されたデータをディスクアレイ装置10からリモートシステム5に転送し、リモートシステム5からの書込み完了報告を受信した時点で、ディスクアレイ装置10は、ホストコンピュータ1に書込み完了を報告する。従って、同期式リモートコピーの場合は、リモートシステム5から書込み完了報告を受信するまでの応答遅延が発生するので、リモートシステム5とローカルシステムであるディスクアレイ装置10との距離が比較的近い場合等に適している。一方、同期式リモートコピーの場合は、リモートシステム5によりデータが受信されたことを確認してからホストコンピュータ1に書込み完了を報告するため、信頼性は高い。

【0068】

これに対し、非同期式リモートコピーの場合、ホストコンピュータ1から書込みを要求されたデータを、ディスクアレイ装置10のキャッシュメモリ30に格納した時点で、ホストコンピュータ1に対し書込み完了を報告する。ディスクアレイ装置10は、ホストコンピュータ1に書込み完了を報告した後で、書込みデータをリモートシステム5に転送するようになっている。従って、非同期式リモートコピーの場合は、ホストコンピュータ1に直ちに書込み完了を報告できるため、同期式リモートコピーよりも応答性を高くすることができる。また、応答遅延の問題が生じないため、ディスクアレイ装置10とリモートシステム5との距離が離れている場合等に適している。しかし、非同期式リモートコピーの場合は、リモートシステム5にデータを転送する前に、ホストコンピュータ1に書込み完了を報告するため、同期式リモートコピーに比べると信頼性は低い。なお、同期式及び非同期式いずれの場合も、ホストコンピュータ1に処理負担をかけることなく、ディスクアレイ装置10からリモートシステム5に直接的にデータが転送される。

【0069】

図8は、リモートコピーに着目してディスクアレイ装置10の機能構成を示す概略ブロック図である。ディスクアレイ装置10からリモートシステム5へのリモートコピー実行に関して、ディスクアレイ装置10内では、データ受信部210と、キャッシュ制御部220と、物理ディスク制御部230と、スケジュール部240と、転送処理部250とが実行される。

【0070】

データ受信を制御するためのプログラムをCHA20内のプロセッサに実行させることにより、CHA20をデータ受信部210として機能させることができる。データ受信部210は、ホストコンピュータ1からの書込み要求を受信して処理する。データ受信部210内には、ホストコンピュータ1からの流入量を制限するための流入量制御部211が設けられている。後述のように、流入量制御部211は、ホストコンピュータ1への書込み完了報告の時期を予め設定された時間だけ意図的に遅らせることにより、ホストコンピュータ1からの書込み要求を抑制する。

【0071】

キャッシュ制御部220は、データ受信部210によって受信されたデータを、一時的にキャッシュメモリ30に格納させる処理を行う。キャッシュ制御部220は、ディスクドライブ40から読み出されたデータを一時的に記憶することもできる。キャッシュ制御プログラムをCHA20またはDKA50内のプロセッサに実行させることにより、CHA20またはDKA50をキャッシュ制御部220として機能させることができる。

【0072】

物理ディスク制御部230は、キャッシュメモリ30に格納されているデータを、論理アドレスから物理アドレスへの変換操作等をして、ディスクドライブ40の所定の位置に書き込む。また、物理ディスク制御部230は、ディスクドライブ40から読み出したデータを、物理アドレスから論理アドレスへの変換操作等をして、キャッシュメモリ30に格納させる。ディスク制御プログラムをDKA50内のプロセッサに実行させることにより、DKA50を物理ディスク制御部230として機能させることができる。

【0073】

スケジュール部240は、リモートコピーを実行する時期であるか否かを管理するものである。リモートコピースケジュール管理プログラムをCHA20のプロセッサに実行させることにより、CHA20をスケジュール部240として機能させることができる。スケジュール部240には、後述のように、帯域制御部241が設けられている。なお、ホストコンピュータ1から書き込み要求が出されるたびにリモートコピーを行ってもよいし、更新データの量が所定量に達した時点でリモートコピーを行ってもよいし、両者を組み合わせてリモートコピーを行ってもよい。

【0074】

転送処理部250は、通信ネットワークCN2等を介して、データをリモートシステム5に転送させる。転送処理プログラムをCHA20内のプロセッサに実行させることにより、CHA20を転送処理部250として機能させることができる。転送処理部250には、後述のように、転送量積算部251が設けられている。

【0075】

ここで、SLPR内の全ての論理デバイス41についてリモートコピーの実行を設定することもできるし、一部の論理デバイス41についてのみリモートコピーを実行するように設定することもできる。リモートコピーの設定は、SLPRを管理する分割管理者により行われる。

【0076】

さて、リモートコピーの対象となっている論理デバイス41に対し、ホストコンピュータ1から書き込み要求があった場合、データ受信部210は、キャッシュ制御部220に対し、リモートコピーの対象となっているデータであることを指示して、データをキャッシュ制御部220に渡す。キャッシュ制御部220は、書き込みを要求されたデータを、そのSLPRで使用可能なキャッシュ領域に格納させる。

【0077】

スケジュール部240は、定期的に起動して、キャッシュ制御部220に対し、キャッシュメモリ30に格納されたリモートコピー対象データの取得を要求する。キャッシュ制御部220は、リモートコピーの対象となるデータがキャッシュメモリ30に記憶されているか否かを調査し、リモートコピー対象データが存在する場合は、このデータをスケジュール部240に出力する。スケジュール部240は、キャッシュ制御部220から渡されたデータを転送処理部250に inputs。転送処理部250は、スケジュール部240から取得したデータを、所定のポートから通信ネットワークCN2等を介して、リモートシステム5に転送させる。

【0078】

転送処理部250には、転送量積算部251が設けられている。転送量積算部251は、単位時間内に転送されたデータ量を積算する。この転送量積算部251により、各SLPR毎に単位時間内転送量がそれぞれ管理される。後述のように、単位時間内転送量は、単位

時間内転送量管理情報 T 6 により管理される。

【0079】

スケジュール部 240 には、帯域制御部 241 が設けられている。帯域制御部 241 は、転送量積算部 251 により検出された各 SLPR 毎の単位時間内転送量と、予めシステム管理者により設定されている各 SLPR 毎の最大転送量とを比較し、両者の関係に基づいてリモートコピーに使用する通信資源の割り当てを各 SLPR 毎に制御させる。各 SLPR 毎に予め設定されており、システム管理者により変更可能な最大転送量は、最大転送量管理情報 T 5 により管理されている。帯域制御部 241 は、単位時間内転送量が最大転送量を超えている場合、転送を中止し、次のリモートコピー対象データについての処理を行う。また、帯域制御部 241 は、データ受信部 210 内に設けられた流入量制御部 211 に対して、流入制限の開始を指示する。転移時間内転送量が最大転送量を下回った場合、帯域制御部 241 は、流入制御部 211 に対して、流入制限の解除を指示する。

【0080】

流入量制御部 211 は、帯域制御部 241 から流入制限の開始を指示されると、流入制限が指示された SLPR について「流入制限中」である旨を、流入状態管理情報 T 4 に記録する。逆に、流入量制御部 211 は、帯域制御部 241 から流入制限の解除を通知されると、流入制限が設定されていた SLPR について、この流入制限を解除し、流入状態管理情報 T 4 を更新させる。

【0081】

データ受信部 210 は、ホストコンピュータ 1 から書込み要求を受信すると、流入状態管理情報 T 4 を参照し、書込み要求に対応する SLPR について流入制限が設定されているか否かを判断する。流入制限が設定されていない場合、データ受信部 210 は、通常の動作を行う。即ち、データ受信部 210 は、ホストコンピュータ 1 から受信したデータをキャッシュ制御部 220 を介してキャッシュメモリ 30 に書き込ませる。そして、データ受信部 210 は、キャッシュメモリ 30 にデータを格納した時点で、ホストコンピュータ 1 に対し、書込み完了を報告する。逆に、ホストコンピュータ 1 から書込みを要求された SLPR について流入制限が設定されている場合、データ受信部 210 は、予めシステム管理者により設定された応答遅延時間だけ、ホストコンピュータ 1 への書込み完了報告を遅らせる。ホストコンピュータ 1 への応答を遅らせることにより、ホストコンピュータ 1 から SLPR に流入するデータ量を抑制することができる。

【0082】

図 9 に基づいて各管理情報 T 4 ~ T 6 の構造を説明する。図 9 (a) に示すように、流入量制御部 211 によって使用される流入状態管理情報 T 4 は、例えば、各 SLPR の SLPR 番号 (SLPR 名称) と、各 SLPR 毎の流入制御 (帯域制御) に関するステータス情報と、応答遅延時間とを対応付けることにより構成することができる。ステータス情報としては、例えば「流入制限中」ステータスを挙げることができる。

【0083】

図 9 (b) に示すように、帯域制御部 241 により使用される最大転送量管理情報 T 5 は、例えば、各 SLPR の SLPR 番号と、各 SLPR 毎の最大転送量 (DTmax) とを対応付けることにより構成される。

【0084】

図 9 (c) に示すように、転送量積算部 251 により使用される転送量管理情報 T 6 は、例えば、各 SLPR の SLPR 番号と、各 SLPR 毎の単位時間内転送量 (DT) とを対応付けることにより構成される。

【0085】

図 10 は、帯域制御部 241 により実行される帯域制御処理の概要を示すフローチャートである。帯域制御部 241 は、キャッシュ制御部 220 に対して、リモートコピーの対象となっているデータの取得を要求する (S31)。リモートシステム 5 へ転送すべきデータが存在しない場合 (S32: N0)、帯域制御部 241 は、処理を終了する。

【0086】

10

20

30

40

50

転送対象のデータ（リモートコピー対象データ）が存在する場合（S32：YES）、帯域制御部241は、リモートコピーを行うSLPRを特定する（S33）。帯域制御部241は、特定されたSLPRにおける単位時間内転送量を転送量積算部251から取得し、この単位時間内転送量と、特定されたSLPRに予め設定されている最大転送量とを比較する（S34）。特定されたSLPRにおける単位時間内転送量が最大転送量以下である場合（単位時間内転送量 $\leq$ 最大転送量）、帯域制御部241は、転送処理部250に対してリモートシステム5へのデータ転送を指示し（S35）、次のリモートコピー対象データの処理に移る。即ち、この場合は、通常の動作によってリモートコピーが行われる。

【0087】

特定されたSLPRにおける単位時間内転送量が予め設定されている最大転送量よりも大きい場合（単位時間内転送量 $>$ 最大転送量）、帯域制御部241は、流入量制御部211に対して、特定されたSLPRに流入制限を設定するよう指示し（S36）、次のリモートコピー対象データの処理に移る。そして、リモートコピー対象データの全てについて処理を終えた場合は、S32で「NO」と判定され、本処理が終了する。

【0088】

図11は、データ受信部210（及び流入量制御部211）により実行される流入量制御処理の概略を示すフローチャートである。ホストコンピュータ1からの書込み要求が発生すると（S41：YES）、データ受信部210は、ホストコンピュータ1から受信したデータを、キャッシュ制御部220を介してキャッシュメモリ30に格納させ（S42）、書込み要求に対応するSLPRを特定する（S43）。次に、データ受信部210は、流入状態管理情報T4を参照し、特定されたSLPRについて流入制限が設定されているか否かを判定する（S44）。

【0089】

流入制限が設定されている場合（S44：YES）、データ受信部210は、そのSLPRに設定されている応答遅延時間を流入状態管理情報T4から取得し、応答遅延時間の計測を開始する（S45）。データ受信部210は、応答遅延時間が経過すると（S46：YES）、ホストコンピュータ1に対して書込み完了を報告する（S47）。一方、SLPRに流入制限が設定されていない場合（S44：NO）、データ受信部210は、応答を意図的に遅らせることなく、直ちにホストコンピュータ1に書込み完了を報告する（S47）。

【0090】

図12は、SLPR毎の最大転送量を設定するための処理の概要を示すフローチャートである。本処理は、例えば、システム管理者により使用される管理端末3を介して操作することができる。まず、システム管理者は、管理端末3を介して、最大転送量を設定するSLPRを特定する（S51）。次に、SVP100（または管理端末3）上で実行される最大転送量設定支援プログラムは、SLPRの構成情報と（S52）、リモートコピーのモードと（S53）、初期コピー及び差分コピーに関するユーザ指示と（S54）をそれぞれ取得する。

【0091】

ここで、リモートコピーのモードには、例えば、非同期式リモートコピーモードと同期式リモートコピーモードとが含まれる。初期コピー及び差分コピーに関するユーザ指示としては、例えば、初期コピーの早期終了によりリカバリ可能状態への早期移行を優先するか否か等の指示を含めることができる。

【0092】

最大転送量設定支援プログラムは、SLPRの構成情報、リモートコピーのモード、ユーザ指示等を考慮して、そのSLPRに設定する最大転送量を算出する（S55）。この最大転送量は、例えば、各パラメータに適切な重み付けを行う等により算出可能である。一例としては、SLPRに割り当てられている資源の量（論理デバイスのボリュームサイズやキャッシュ容量等）に応じて、資源量が大きくなるほど最大転送量が大きくなるように算出できる。また、例えば、同期式リモートコピーモードの場合は、できるだけ応答時間を短縮できるように、大きな最大転送量を算出する。さらに、例えば、初期コピーの早期完了がユー

10

20

30

40

50

ザにより指示されている場合は、初期コピーを行う際の最大転送量が大きくなるように算出することができる。

ここで、同期式リモートコピー／非同期式リモートコピー、あるいは、初期コピー／差分コピーのように、動作モード毎に最大転送量（帯域）をそれぞれ設定可能である。または、ボリューム単位や複数のボリュームから構成されるグループ単位で、最大転送量をそれぞれ設定することもできる。

【0093】

このようにして各パラメータを考慮して算出された最大転送量は、管理端末3の表示部に表示される。システム管理者は、算出された最大転送量を手動で調整することができる（S56）。そして、最終的にシステム管理者により承認された最大転送量は、最大転送量管理情報T5に登録される（S57）。なお、最大転送量の算出は、上記の例に限られない。例えば、上述した全てのパラメータ（SLPRへの資源割り当て量、リモートコピーモード、初期コピー等に関するユーザ指示）を考慮する必要はなく、いずれか1つまたは複数のパラメータに基づいて、最大転送量を算出し設定してもよい。

10

【0094】

このように構成される本実施例によれば、ディスクアレイ装置10を各ユーザ毎のSLPRに分割して各ユーザに提供することができ、さらに、分割単位（SLPR）毎にリモートコピーに使用する通信資源を割り当てることができる。

【0095】

従って、各SLPR間をリモートコピーについても分離することができ、各業務間（各ホストコンピュータ1間）の影響を低減することができる。即ち、例えば、ある業務に関するアクセス要求が増大し、そのSLPRからリモートシステム5に転送されるべきデータ量が増大した場合でも、そのSLPRによってリモートシステム5への通信ネットワークCN2が殆ど独占されてしまう事態を未然に防止できる。従って、スイッチ4に入力される前に、各SLPR毎に適切な帯域を割り当てることができ、各ユーザ間の干渉を防止して、より一層安定した業務環境を提供することができる。

20

【0096】

なお、本発明は、上述した実施の形態に限定されない。当業者であれば、本発明の範囲内で、種々の追加や変更等を行うことができる。例えば、上記実施例では、単位時間内転送量が予め設定された最大転送量を超えた場合に、ホストコンピュータからのデータ流入を所定の応答遅延時間だけ制限するものと述べた。しかし、帯域制御の方法としては、これに限らず、種々の方法を採用可能である。例えば、転送量の微分値と予め設定された基準値とを比較し、転送量の急激な増加が検出された場合は、ホストコンピュータからのデータ流入を制限することができる。

30

【図面の簡単な説明】

【0097】

【図1】本発明の実施例に係わるディスクアレイ装置の論理的な構成の全体概要を示すブロック図である。

【図2】ディスクアレイ装置の物理的な構成に着目したブロック図である。

【図3】各SLPR毎に資源を分割して管理するための様子を示す説明図である。

40

【図4】SLPR構成情報管理テーブルに関与する各管理テーブルの構成を示す説明図である。

【図5】キャッシュメモリの記憶構造の概略を示す説明図である。

【図6】各CLPR毎にキャッシュ容量を管理するための構成を示す説明図である。

【図7】セグメント確保処理を示すフローチャートである。

【図8】ディスクアレイ装置のリモートコピーに関する各機能の構成を示す説明図である。

【図9】リモートコピーに使用される各管理情報を示す説明図であって、（a）は流入状態管理情報を、（b）は最大転送量管理情報を、（c）は単位時間内転送量管理情報を、それぞれ示す。

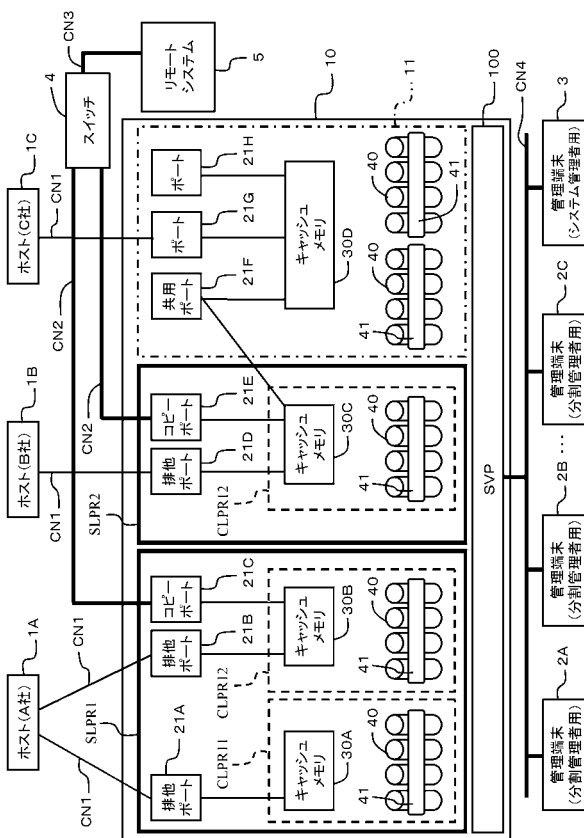
50

- 【図 10】 帯域制御処理を示すフローチャートである。
 【図 11】 流入量制御処理を示すフローチャートである。
 【図 12】 最大転送量設定処理を示すフローチャートである。
 【符号の説明】
 【0098】

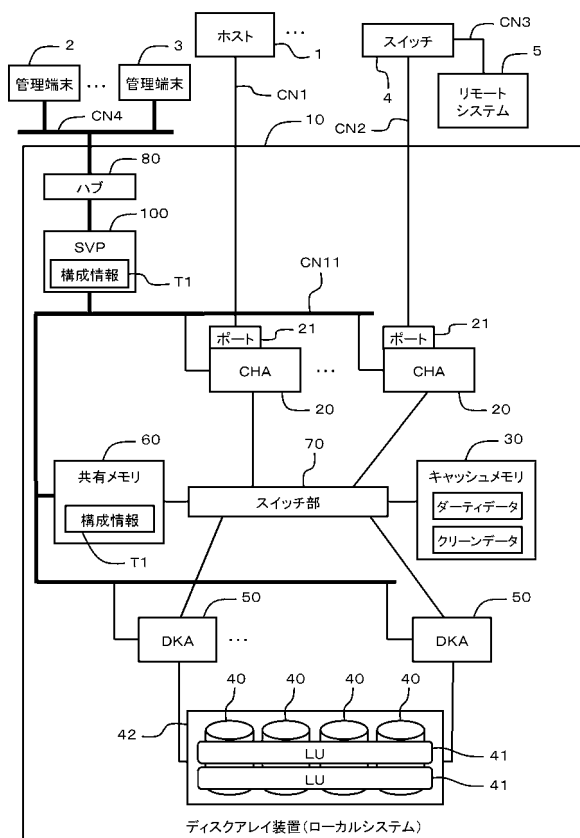
1 A～1 C…ホストコンピュータ、2 A～2 C…管理端末、3…管理端末、4…スイッチ、5…リモートシステム、10…ディスクアレイ装置、11…非分割領域、20…CHA、21 A～21 H…ポート、30 A～30 D…キャッシュメモリ、40…ディスクドライブ、41…論理デバイス、50…DKA、60…共有メモリ、70…スイッチ部、80…ハブ、100…SVP、110…SVPサーバ部、120…構成情報管理部、210…データ受信部、211…流入制御部、211…流入量制御部、220…キャッシュ制御部、230…物理ディスク制御部、240…スケジュール部、241…帯域制御部、250…転送処理部、251…転送量積算部、CN…通信ネットワーク

10

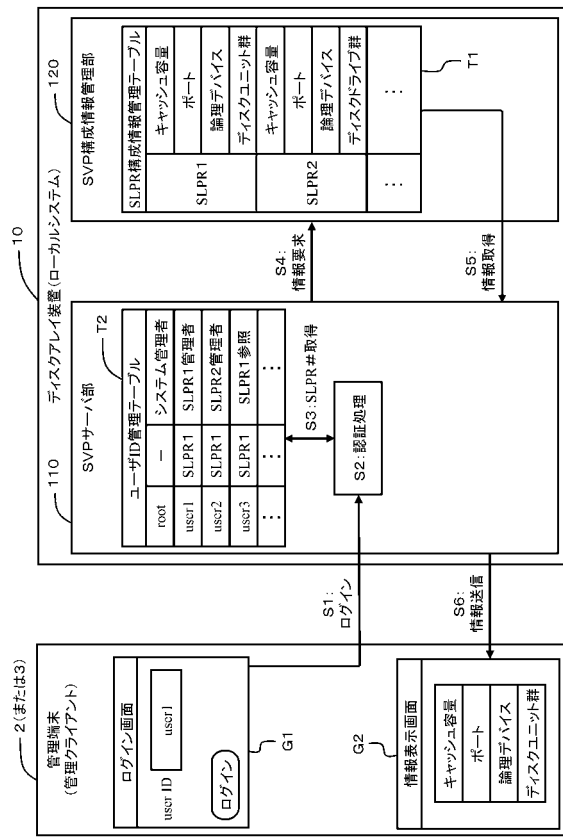
【図 1】



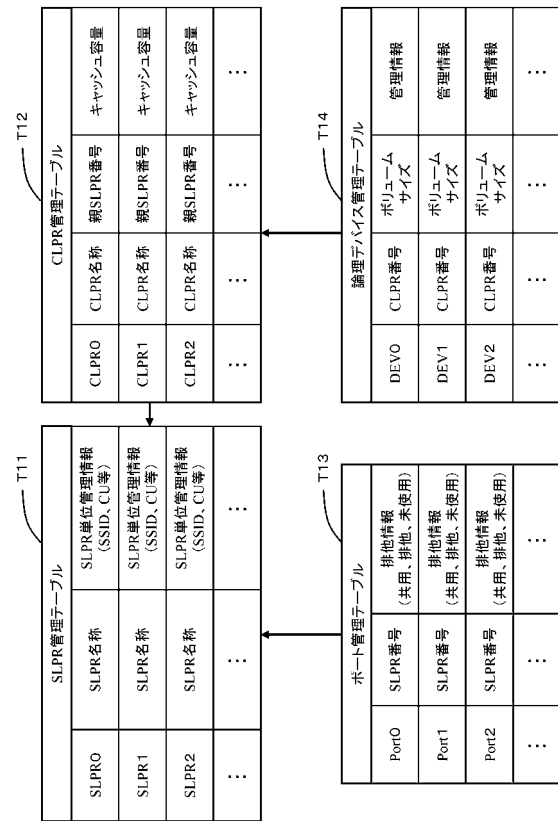
【図 2】



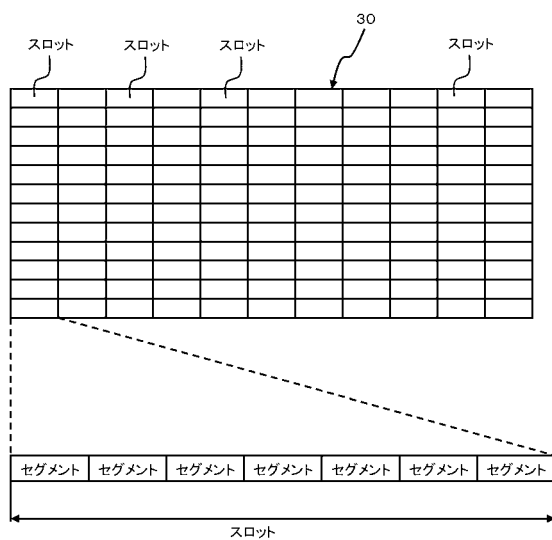
【図 3】



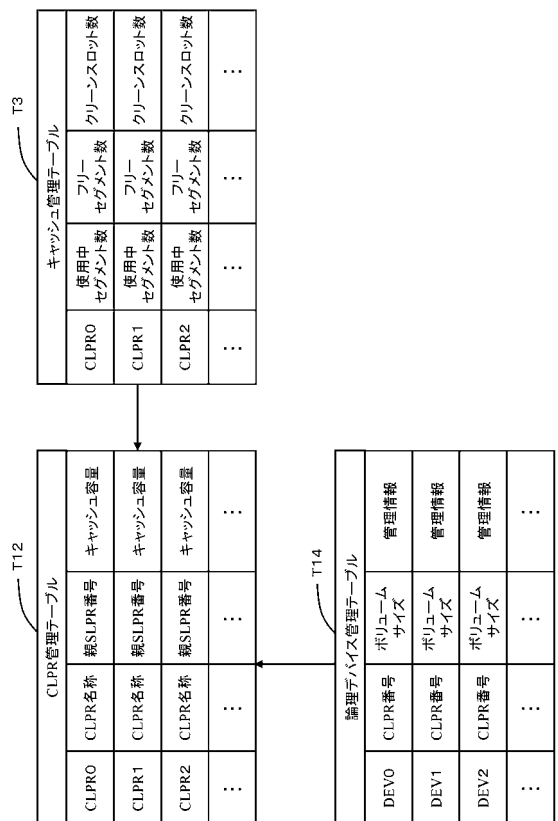
【図 4】



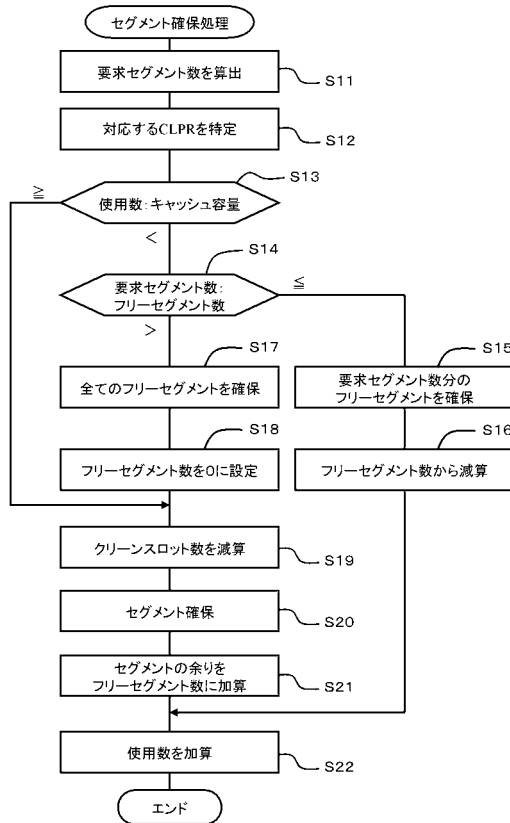
【図 5】



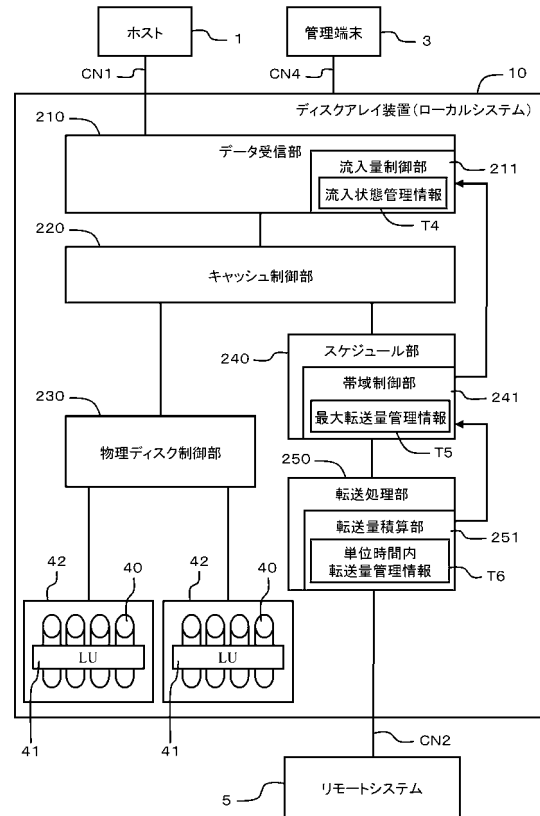
【図 6】



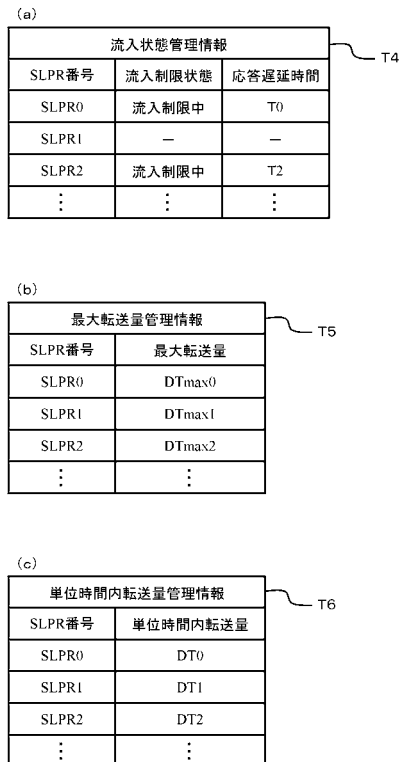
【図 7】



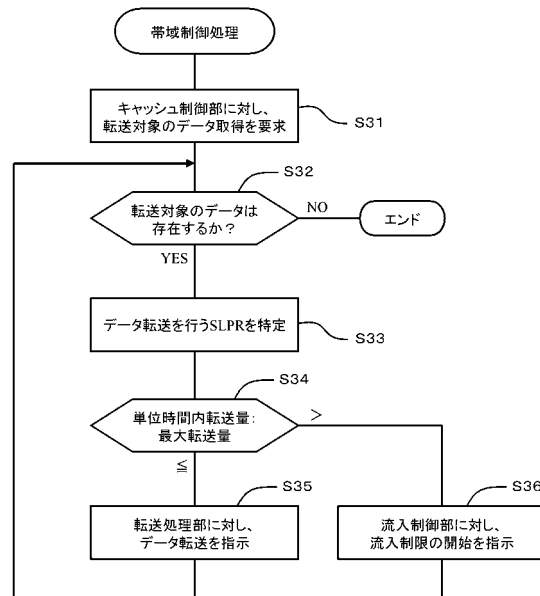
【図 8】



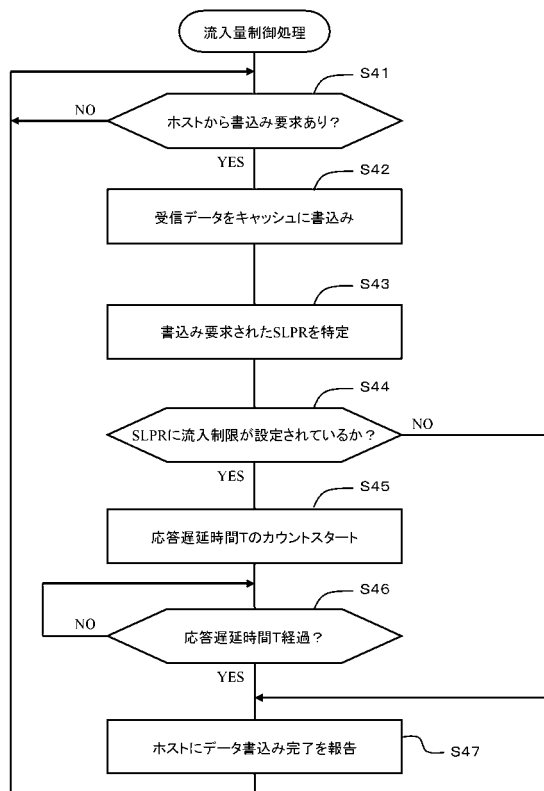
【図 9】



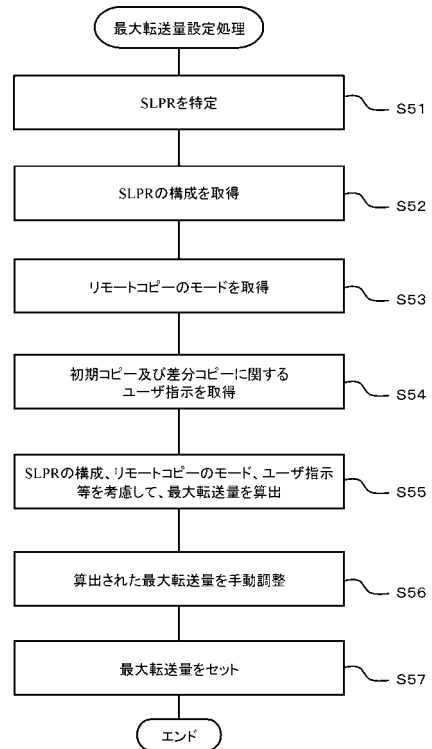
【図 10】



【図 1 1】



【図 1 2】



フロントページの続き

(51)Int.Cl. ⁷	F I	テーマコード (参考)
	G 0 6 F 12/08	5 4 3 B
	G 0 6 F 12/08	5 5 1 Z
	G 0 6 F 12/08	5 5 7

(72)発明者 平川 裕介

神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I D システム事業部内

F ターム (参考) 5B005 JJ13 MM12 UU41 VV03

5B065 BA01 CA12 CA30 CE01 CE22 CH01 EA02 EA12 EA34

5B082 DE07