

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6031992号
(P6031992)

(45) 発行日 平成28年11月24日 (2016.11.24)

(24) 登録日 平成28年11月4日 (2016.11.4)

(51) Int.Cl.	F I
G 0 6 F 3 / 0 6 (2006.01)	G 0 6 F 3 / 0 6 3 0 4 P
	G 0 6 F 3 / 0 6 3 0 5 K
	G 0 6 F 3 / 0 6 3 0 6 K

請求項の数 7 (全 28 頁)

(21) 出願番号	特願2012-277129 (P2012-277129)	(73) 特許権者	000005223
(22) 出願日	平成24年12月19日 (2012.12.19)		富士通株式会社
(65) 公開番号	特開2014-120125 (P2014-120125A)		神奈川県川崎市中原区上小田中4丁目1番1号
(43) 公開日	平成26年6月30日 (2014.6.30)	(74) 代理人	100092978
審査請求日	平成27年8月4日 (2015.8.4)		弁理士 真田 有
		(74) 代理人	100112678
			弁理士 山本 雅久
		(72) 発明者	小橋 究
			神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
		審査官	田中 啓介

最終頁に続く

(54) 【発明の名称】 情報処理装置、情報処理システム、入出力方法及び入出力プログラム

(57) 【特許請求の範囲】

【請求項 1】

ストレージ装置と複数の経路を介して接続され、
 前記複数の経路の各経路のタイムアウト時間に関する管理情報を記憶する記憶部と、
 前記ストレージ装置に対する入出力の発行先の経路及び他の経路について、前記記憶部から該経路に応じたタイムアウト時間を取得する取得部と、
 前記管理情報を収集し、前記記憶部に記憶する収集部と、
 前記取得部によって取得された前記タイムアウト時間を用いて前記ストレージ装置に対して入出力を発行する発行部と、

をそなえ、

前記収集部は、前記複数の経路に対する複数の入出力タイムアウト時間を前記管理情報として前記ストレージ装置から収集し、

前記取得部は、前記複数の経路のうちの第1の経路に対する第1の入出力タイムアウト時間を取得し、

前記第1の経路への前記入出力でタイムアウトエラーが発生した場合に、前記複数の経路のうちの前記第1の経路とは異なる第2の経路に対する、前記第1の入出力タイムアウト時間よりも長い第2の入出力タイムアウト時間を取得することを特徴とする情報処理装置。

【請求項 2】

前記第1の入出力タイムアウト時間は前記複数の経路のうちの最終経路以外の経路に対

10

20

するタイムアウト時間であり、

前記第 2 の入出力タイムアウト時間は前記複数の経路のうちの最終経路に対するタイムアウト時間であることを特徴とする請求項 1 記載の情報処理装置。

【請求項 3】

前記取得部は、前記第 1 の経路への前記入出力でタイムアウトエラーが発生した場合に、前記第 2 の経路に前記入出力を発行する前に、前記複数の経路のうちの前記第 1 及び第 2 の経路とは異なる第 3 の経路に対する、前記第 1 の入出力タイムアウト時間よりも長く前記第 2 の入出力タイムアウト時間よりも短い第 3 の入出力タイムアウト時間を取得することを特徴とする請求項 2 記載の情報処理装置。

【請求項 4】

前記収集部は、前記ストレージ装置において前記複数の入出力タイムアウト時間が増え変わると、変更後の前記複数の入出力タイムアウト時間を前記管理情報として収集することを特徴とする請求項 1 ～ 3 のいずれか 1 項に記載の情報処理装置。

【請求項 5】

情報処理装置と、
前記情報処理装置と複数の経路を介して接続されたストレージ装置と、
前記複数の経路の各経路のタイムアウト時間に関する管理情報を記憶する記憶部と、
前記情報処理装置から前記ストレージ装置に対する入出力の発行先の経路及び他の経路について、前記記憶部から該経路に応じたタイムアウト時間を取得する取得部と、
前記管理情報を収集し、前記記憶部に記憶する収集部と、
前記取得部によって取得された前記タイムアウト時間を用いて前記ストレージ装置に対して入出力を発行する発行部と、

をそなえ、

前記収集部は、前記複数の経路に対する複数の入出力タイムアウト時間を前記管理情報として前記ストレージ装置から収集し、

前記取得部は、前記複数の経路のうちの第 1 の経路に対する第 1 の入出力タイムアウト時間を取得し、

前記第 1 の経路への前記入出力でタイムアウトエラーが発生した場合に、前記複数の経路のうちの前記第 1 の経路とは異なる第 2 の経路に対する、前記第 1 の入出力タイムアウト時間よりも長い第 2 の入出力タイムアウト時間を取得することを特徴とする情報処理システム。

【請求項 6】

情報処理装置に複数の経路を介して接続されたストレージ装置における入出力方法であって、

前記複数の経路の各経路のタイムアウト時間に関する管理情報を記憶部に記憶し、
前記ストレージ装置に対する入出力の発行先の経路及び他の経路について、前記記憶部から該経路に応じたタイムアウト時間を取得し、

前記管理情報を収集して、前記記憶部に記憶し、
取得された前記タイムアウト時間を用いて前記ストレージ装置に対して入出力を発行し、

前記複数の経路に対する複数の入出力タイムアウト時間を前記管理情報として前記ストレージ装置から収集し、

前記複数の経路のうちの第 1 の経路に対する第 1 の入出力タイムアウト時間を取得し、

前記第 1 の経路への前記入出力でタイムアウトエラーが発生した場合に、前記複数の経路のうちの前記第 1 の経路とは異なる第 2 の経路に対する、前記第 1 の入出力タイムアウト時間よりも長い第 2 の入出力タイムアウト時間を取得することを特徴とする入出力方法。

【請求項 7】

情報処理装置に複数の経路を介して接続されたストレージ装置における入出力プログラムであって、

前記複数の経路の各経路のタイムアウト時間に関する管理情報を記憶部に記憶し、
前記ストレージ装置に対する入出力の発行先の経路及び他の経路について、前記記憶部から該経路に応じたタイムアウト時間を取得し、
前記管理情報を収集して、前記記憶部に記憶し、
取得された前記タイムアウト時間を用いて前記ストレージ装置に対して入出力を発行し

、
前記複数の経路に対する複数の入出力タイムアウト時間を前記管理情報として前記ストレージ装置から収集し、

前記複数の経路のうちの第1の経路に対する第1の入出力タイムアウト時間を取得し、
前記第1の経路への前記入出力でタイムアウトエラーが発生した場合に、前記複数の経路のうちの前記第1の経路とは異なる第2の経路に対する、前記第1の入出力タイムアウト時間よりも長い第2の入出力タイムアウト時間を取得する

処理を前記情報処理装置に実行させることを特徴とする入出力プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本件は、情報処理装置、情報処理システム、入出力方法及び入出力プログラムに関する。

【背景技術】

【0002】

近年、技術の進歩により、物理動作のないソリッドステートドライブ (solid state drive : S S D) や高性能なストレージが開発されている。

又、新しい情報は高速なディスクに、古い情報は低速の低価格なディスクに置くという情報ライフサイクル管理 (Information Lifetime management : I L M)、階層化や仮想化などの技術の登場により、ストレージ装置内に複数の種類のディスクを混在させることが一般的となっている。このため、ストレージ装置に、例えば S S D などの高速アクセス可能なストレージと、アクセス速度の低いハードディスクドライブ (hard disk drive ; H D D) とが共存しているケースが多い。

【0003】

図13は、従来のストレージシステム101を示す図である。

このストレージシステム101は、図13に示すように、サーバ102に、バス (経路) 130 - 1 ~ 130 - 6 を介して複数 (図13に示す例では2つ) のストレージ装置110, 120 が通信可能に接続されている。

サーバ102は、接続されたストレージ装置110, 120 のボリュームにデータの書き込みや読み出しを行なう情報処理装置である。例えば、サーバ102はストレージ装置110, 120 のボリュームに対してリードやライトのデータアクセス要求を行なう。ストレージ装置110, 120 は、このデータアクセス要求に応じてボリュームに対するデータアクセスを行ない、サーバ102に対して応答を行なう。サーバ2は、例えば、図示しないCentral Processing Unit (C P U) やRandom Access Memory (R A M)、Read Only Memory (R O M) 等をそなえる。又、サーバ102は、マルチバスドライバ部103、Logical Unit Number (L U N) 情報107 - 1 ~ 107 - 6、及びホストバスアダプタ (Host Bus Adaptor : H B A) 109 - 1 ~ 109 - 6 をそなえる。

【0004】

マルチバスドライバ部103は、ストレージ装置110, 120 へのバスが複数存在するマルチバス環境において、I O (Input/Output) タイムアウトの発生時にアクセスバスを切り替えてI O をリトライする。I O タイムアウトとは、I O を発行してからそのI O に対する応答が返されるまでの最大待機時間であるタイムアウト時間 (タイムアウト値とも呼ぶ) を超えることを指す。

【0005】

L U N 情報107 - 1 ~ 107 - 6 は、サーバ102に接続されているストレージ装置

10

20

30

40

50

110, 120のLUNに関する情報である。LUN情報107-1~107-6には、LUN毎に、例えば、当該LUNのディスク種別(例えば、“SSD”)や装置識別情報(例えば、“A4000-6A0298”)などが格納される。

HBA109-1~109-6は、サーバ102とストレージ機器等の外部機器とを接続するアダプタである。HBA109-1~109-6は、例えばSCSIアダプタ、ファイバチャネル(FC)アダプタ、シリアルATAアダプタなどである。或いは、HBA109-1~109-6は、IDE、イーサネット(登録商標)、FireWire(登録商標)、Universal Serial Bus(USB)などに準拠した機器を接続するデバイスであってもよい。

【0006】

ストレージ装置110は、サーバ102に対して記憶領域を提供するものであり、例えば、RAID装置である。ストレージ装置110は、図中の例では、装置名として“A4000”、IDとして“6A0298”を有する。ストレージ装置110は、Controller Module(CM)111-1, 111-2、及びディスク113-1~113-4をそなえる。

【0007】

CM111-1は、Channel Adapter(CA)112-1, 112-2のほか不図示のCPUやメモリ、キャッシュメモリ等をそなえ、IO処理を行なう。

CM111-2は、CA112-3, 112-4のほか不図示のCPUやメモリ、キャッシュメモリ等をそなえ、IO処理を行なう。

ここで、CM111-1とCM111-2とは、同様に構成されている。また、CA112-1~CA112-4は同様に構成されている。

【0008】

CA112-1~112-4は、サーバ102と通信可能に接続するインタフェースコントローラであり、例えば、ファイバチャネルアダプタである。CA112-1, 112-2, 112-4は、それぞれパス130-1~130-3経由でサーバのHBA109-1~109-3と接続されている。

パス130-1~130-3は、サーバ102とストレージ装置110とを接続するパスであり、例えばFCである。

【0009】

ディスク113-1~113-4は、情報を記憶可能な記憶装置であり、例えば、ニアラインドライブ(以下、ニアラインと呼ぶ)113-1、HDD113-2, 113-3、及びSSD113-4をそなえる。SSDは厳密にはディスクではないが、以下、説明の便宜上SSDもニアラインやHDDのようにディスクとして扱う。

ニアライン113-1は、磁性体を塗布した円盤を記録媒体として使用し、磁気ヘッドを移動させ、高速回転している円盤に対して情報の読み書きを行なうストレージドライブである。ニアライン113-1は、一般的にHDD113-2, 113-3よりも大容量であるが、アクセスが低速である。

【0010】

HDD113-2, 113-3は、磁性体を塗布した円盤を記録媒体として使用し、磁気ヘッドを移動させ、高速回転している円盤に対して情報の読み書きを行なうストレージドライブである。

SSD113-4、半導体素子メモリを記憶装置として使用するストレージドライブであり、シリコンディスクドライブ、半導体ディスクドライブともよばれる。一般に、SSD33は、ニアライン113-1やHDD113-2, 113-3のような磁気ヘッドの移動に伴うヘッドシークタイムがないため、これらよりも高速なランダムアクセスが可能である。その一方、半導体メモリを使用するので、一般的にニアライン113-1やHDD113-2, 113-3よりも高価である。

【0011】

ストレージ装置120も、サーバ102に対して記憶領域を提供するものであり、例え

10

20

30

40

50

ば、RAID装置である。ストレージ装置120は、図中の例では、装置名として“A3000”、IDとして“4A0290”を有する。ストレージ装置120は、CM121-1, 121-2及びディスク123-1, 123-2をそなえる。

CM121-1は、CA122-1, 122-2のほか不図示のCPUやメモリ、キャッシュメモリ等をそなえ、IO処理を行なう。

【0012】

CM121-2は、CA122-3, 122-4のほか不図示のCPUやメモリ、キャッシュメモリ等をそなえ、IO処理を行なう。

CA122-1~122-4は、サーバ102と通信可能に接続するインタフェースコントローラであり、例えば、ファイバチャネルアダプタである。CA122-1, 122-2, 122-4は、それぞれバス130-4~130-6経由でサーバのHBA109-4~109-6と接続されている。

10

【0013】

ここで、CM121-1とCM121-2とは、同様に構成されている。また、CA122-1~CA122-4は同様に構成されている。

バス130-4~130-6は、サーバ102とストレージ装置120とを接続するバスであり、例えばFCである。

ディスク123-1, 123-2は、情報を記憶可能な記憶装置であり、例えば、HDD123-1, 123-2である。

【0014】

20

HDD123-1, 123-2は、磁性体を塗布した円盤を記録媒体として使用し、磁気ヘッドを移動させ、高速回転している円盤に対して情報の読み書きを行なうストレージドライブである。

ストレージシステム101においては、サーバ102からストレージ装置110, 120に対して、IOアクセスが発生する。

【0015】

このような構成においては、サーバ102側で、サーバ102がIOを発行後、ストレージ装置110, 120からの応答を待つ最大時間であるタイムアウト時間（以下、このタイムアウト時間を「IOタイムアウト時間」と呼ぶ）を設定している。

ストレージ装置110, 120の故障等により、IOタイムアウト時間を超えてもストレージ装置110, 120から応答がない場合、サーバ102は、タイムアウトエラーが発生したとして、エラー処理を行なう。

30

【0016】

なお、サーバ102がストレージ装置110, 120にIOを発行してから、ストレージ装置110, 120が応答するまでのレイテンシは、IO要求転送時間、ディスクシークタイム、read/write時間、IO転送時間などの総計となる。

サーバ102でIOタイムアウト時間を短く設定すれば、ストレージ装置110, 120の故障時のエラー検出やバス切替時間を短くできる。一方、IOタイムアウト時間が短い場合、ストレージ装置110, 120が正常に動作している場合でも、ストレージ装置110, 120の負荷が一時的に高くなるなどの原因によりレイテンシ値が通常より長くなると、タイムアウトが検出されてしまう。

40

【0017】

サーバ102は、IOタイムアウト時間を1つ設定できる。例えば、Solaris（登録商標）標準ドライバではIOタイムアウト時間のデフォルト値が60秒である。この値を変更するには、ユーザが定義を書き換え、サーバ102を再起動する。

【先行技術文献】

【特許文献】

【0018】

【特許文献1】特開2006-235843号公報

【発明の概要】

50

【発明が解決しようとする課題】

【0019】

SSD 113 - 4 はアクセス時間が短いので、I/O タイムアウト時間を短い値に設定することができる。

しかし、前述の如く、従来のストレージ装置 110, 120 においては、設定される I/O タイムアウト値は 1 つである。従って、最新のディスク 113、旧型のディスク 113 のいずれにも同じ I/O タイムアウト値が使用され、SSD に合わせて I/O タイムアウト時間を 60 秒未満の値に設定すると、低速のディスクで I/O タイムアウトが頻発するおそれがある。

【0020】

さらに、I/O の切り替え先バスが存在する場合でも、タイムアウトするまで一定時間待ってからバスの切り替えが行なわれるので、別バスに切り替えて I/O を直ちに再発行することもできない。

1 つの側面では、本発明は、ストレージ装置に対する I/O 発行時の応答時間を短縮することを目的とする。

【0021】

なお、前記目的に限らず、後述する発明を実施するための形態に示す各構成により導かれる作用効果であって、従来の技術によっては得られない作用効果を奏することも本発明の他の目的の 1 つとして位置付けることができる。

【課題を解決するための手段】

【0022】

このため、情報処理装置は、ストレージ装置と複数の経路を介して接続され、前記複数の経路の各経路のタイムアウト時間に関する管理情報を記憶する記憶部と、前記ストレージ装置に対する入出力の発行先の経路及び他の経路について、前記記憶部から該経路に応じたタイムアウト時間を取得する取得部と、前記管理情報を収集し、前記記憶部に記憶する収集部と、前記取得部によって取得された前記タイムアウト時間を用いて前記ストレージ装置に対して入出力を発行する発行部と、をそなえ、前記収集部は、前記複数の経路に対する複数の入出力タイムアウト時間を前記管理情報として前記ストレージ装置から収集し、前記取得部は、前記複数の経路のうちの第 1 の経路に対する第 1 の入出力タイムアウト時間を取得し、前記第 1 の経路への前記入出力でタイムアウトエラーが発生した場合に、前記複数の経路のうちの前記第 1 の経路とは異なる第 2 の経路に対する、前記第 1 の入出力タイムアウト時間よりも長い第 2 の入出力タイムアウト時間を取得する。

又、本発明に関連するストレージ装置は、情報処理装置と複数の経路を介して接続され、前記複数の経路に対する複数の入出力タイムアウト時間を格納している格納部と、前記情報処理装置から前記複数の入出力タイムアウト時間を要求されると、前記情報処理装置に前記複数の入出力タイムアウト時間を送信するコントローラと、をそなえる。

【0023】

さらに、情報処理システムは、情報処理装置と、前記情報処理装置と複数の経路を介して接続されたストレージ装置と、前記複数の経路の各経路のタイムアウト時間に関する管理情報を記憶する記憶部と、前記情報処理装置から前記ストレージ装置に対する入出力の発行先の経路及び他の経路について、前記記憶部から該経路に応じたタイムアウト時間を取得する取得部と、前記管理情報を収集し、前記記憶部に記憶する収集部と、前記取得部によって取得された前記タイムアウト時間を用いて前記ストレージ装置に対して入出力を発行する発行部と、をそなえ、前記収集部は、前記複数の経路に対する複数の入出力タイムアウト時間を前記管理情報として前記ストレージ装置から収集し、前記取得部は、前記複数の経路のうちの第 1 の経路に対する第 1 の入出力タイムアウト時間を取得し、前記第 1 の経路への前記入出力でタイムアウトエラーが発生した場合に、前記複数の経路のうちの前記第 1 の経路とは異なる第 2 の経路に対する、前記第 1 の入出力タイムアウト時間よりも長い第 2 の入出力タイムアウト時間を取得する。

【0024】

なおさらに、情報処理装置に複数の経路を介して接続されたストレージ装置における入出力方法は、前記複数の経路の各経路のタイムアウト時間に関する管理情報を記憶部に記憶し、前記ストレージ装置に対する入出力の発行先の経路及び他の経路について、前記記憶部から該経路に応じたタイムアウト時間を取得し、前記管理情報を収集して、前記記憶部に記憶し、取得された前記タイムアウト時間を用いて前記ストレージ装置に対して入出力を発行し、前記複数の経路に対する複数の入出力タイムアウト時間を前記管理情報として前記ストレージ装置から収集し、前記複数の経路のうちの第1の経路に対する第1の入出力タイムアウト時間を取得し、前記第1の経路への前記入出力でタイムアウトエラーが発生した場合に、前記複数の経路のうちの前記第1の経路とは異なる第2の経路に対する、前記第1の入出力タイムアウト時間よりも長い第2の入出力タイムアウト時間を取得する。

10

又、情報処理装置に複数の経路を介して接続されたストレージ装置における入出力プログラムは、前記複数の経路の各経路のタイムアウト時間に関する管理情報を記憶部に記憶し、前記ストレージ装置に対する入出力の発行先の経路及び他の経路について、前記記憶部から該経路に応じたタイムアウト時間を取得し、前記管理情報を収集して、前記記憶部に記憶し、取得された前記タイムアウト時間を用いて前記ストレージ装置に対して入出力を発行し、前記複数の経路に対する複数の入出力タイムアウト時間を前記管理情報として前記ストレージ装置から収集し、前記複数の経路のうちの第1の経路に対する第1の入出力タイムアウト時間を取得し、前記第1の経路への前記入出力でタイムアウトエラーが発生した場合に、前記複数の経路のうちの前記第1の経路とは異なる第2の経路に対する、前記第1の入出力タイムアウト時間よりも長い第2の入出力タイムアウト時間を取得する処理を前記情報処理装置に実行させる。

20

【発明の効果】

【0025】

開示の技術によれば、ストレージ装置に対するI/O発行時の応答時間を短縮することができる。

【図面の簡単な説明】

【0026】

【図1】実施形態の一例としてのストレージシステムの構成を示す図である。

【図2】実施形態の一例としてのサーバ側タイムアウトテーブルの初期状態を例示する図である。

30

【図3】実施形態の一例としてのストレージ装置の識別後のサーバ側タイムアウトテーブルを例示する図である。

【図4】図3のサーバ側タイムアウトテーブルにレコードをさらに追加した例を示す図である。

【図5】実施形態の一例としてのストレージ装置側のタイムアウトテーブルを例示する図である。

【図6】実施形態の一例としてのタイムアウト上限テーブルを例示する図である。

【図7】実施形態の一例としての初期化処理を示すフローチャートである。

【図8】実施形態の一例としてのI/O発行及び応答時の処理を示すフローチャートである。

40

【図9】実施形態の一例としてのタイムアウト値の更新処理を示すフローチャートである。

【図10】実施形態の第1変形例におけるサーバ側タイムアウトテーブルを例示する図である。

【図11】実施形態の第4変形例におけるサーバ側タイムアウトテーブルを例示する図である。

【図12】実施形態の第5変形例におけるLUN管理テーブルを例示する図である。

【図13】従来のストレージシステムを示す図である。

【発明を実施するための形態】

50

【 0 0 2 7 】

以下、図面を参照して本情報処理装置、情報処理システム、入出力方法、及び入出力プログラムに係る実施の形態を説明する。ただし、以下に示す実施形態はあくまでも例示に過ぎず、実施形態で明示しない種々の変形例や技術の適用を排除する意図はない。すなわち、本実施形態を、その趣旨を逸脱しない範囲で種々変形して実施することができる。

【 0 0 2 8 】

(A) システム構成

図 1 は実施形態の一例としてのストレージシステム（情報処理システム）1 の構成を示す図である。

実施形態の一例としてのストレージシステム 1 は、図 1 に示すように、サーバ（情報処理装置）2 に、パス（経路）3 0 - 1 ~ 3 0 - 6 を介して複数（図 1 に示す例では 2 つ）のストレージ装置 1 0 , 2 0 が通信可能に接続されている。

【 0 0 2 9 】

サーバ 2 は、接続されたストレージ装置 1 0 , 2 0 のボリュームにデータの書き込みや読み出しを行なう情報処理装置である。例えば、サーバ 2 はストレージ装置 1 0 , 2 0 のボリュームに対してリードやライトのデータアクセス要求を行なう。ストレージ装置 1 0 , 2 0 は、このデータアクセス要求に応じてボリュームに対するデータアクセスを行ない、サーバ 2 に対して応答を行なう。サーバ 2 は、例えば、図示しない C P U や R A M 、 R O M 等をそなえる。又、サーバ 2 は、マルチバスドライバ部 3、サーバ側タイムアウトテーブル（記憶部）4、L U N 情報 7 - 1 ~ 7 - 6、及び H B A 9 - 1 ~ 9 - 6 をそなえる。

【 0 0 3 0 】

マルチバスドライバ部 3 は、ストレージ装置 1 0 からタイムアウト情報を取得すると共に、マルチバス環境において、I O タイムアウトの発生時にアクセスパスを切り替えて I O をリトライする。マルチバスドライバ部 3 は、タイムアウトテーブル収集部（収集部）5、タイムアウト時間取得部（取得部）6、及び I O 発行部（発行部）8 をそなえる。

タイムアウトテーブル収集部 5 は、外部から I O タイムアウト値を設定可能なストレージ装置 1 0 から、後述するタイムアウトテーブル 1 4 を取得する。そして、取得したタイムアウトテーブル 1 4 のデータを、サーバ側タイムアウトテーブル 4 に記録する。このとき、タイムアウトテーブル収集部 5 は、例えば、ストレージ装置 1 0 のモデル名、シリアル番号等の情報を、装置識別情報としてサーバ側タイムアウトテーブル 4 に追加する。

【 0 0 3 1 】

なお、以下、サーバ 2 等の外部から I O タイムアウト値を設定可能なストレージ装置を「タイムアウト対応ストレージ装置」と呼ぶ。一方、タイムアウト値を設定できないストレージ装置を「タイムアウト非対応ストレージ装置」と呼ぶ。図 1 に示す例においては、ストレージ装置 1 0 がタイムアウト対応ストレージ装置、ストレージ装置 2 0 がタイムアウト非対応ストレージ装置である。

【 0 0 3 2 】

タイムアウト時間取得部 6 は、ストレージ装置 1 0 , 2 0 に I O を発行する際に、後述のサーバ側タイムアウトテーブル 4 を参照して、当該 I O に設定するタイムアウト時間を読み出して取得する。タイムアウト非対応のストレージ装置 2 0 の場合、タイムアウト時間取得部 6 は、デフォルトのタイムアウト時間（例えば 4 0 秒）を読み出す。

I O 発行部 8 は、タイムアウト時間取得部 6 が取得したタイムアウト値を使用して I O を発行する。又、I O 発行部 8 は、図 8 を用いて説明するように、I O 発行後に当該 I O がタイムアウトとなったか、正常終了したか、或いはタイムアウト以外の原因で異常終了したかを判定する。

【 0 0 3 3 】

サーバ側タイムアウトテーブル 4 は、サーバ 2 に接続されているストレージ装置 1 0 のディスク毎のタイムアウト時間を記憶するテーブルである。サーバ側タイムアウトテーブル 4 は、例えば、サーバ 2 の不図示のメモリに格納される。このサーバ側タイムアウトテ

ーブル4については、図2～図4を用いて後述する。

LUN情報7-1～7-6は、サーバ2に接続されているストレージ装置10, 20のLUNに関する情報である。LUN情報7-1～7-6には、LUN毎に、例えば、当該LUNのディスク種別（例えば、“SSD”）や装置識別情報（例えば、“E4000-6A0299”）などが格納される。

【0034】

HBA9-1～9-6は、サーバ2とストレージ機器等の外部機器とを接続するアダプタである。HBA9-1～9-6は、例えばSCSIアダプタ、ファイバチャネル（FC）アダプタ、シリアルATAアダプタなどである。或いは、HBA9-1～9-6は、IDE、イーサネット、FireWire、USBなどに準拠した機器を接続するデバイス

10

【0035】

なお、以下、LUN情報を示す符号としては、複数のLUN情報のうち1つを特定する必要があるときには符号7-1～7-6を用いるが、任意のLUN情報を指すときには符号7を用いる。

ストレージ装置10は、サーバ2に対して記憶領域を提供するものであり、例えば、RAID装置である。ストレージ装置10は、外部からI/Oタイムアウト値を設定可能なタイプのストレージ装置である。又、図中の例では、装置名として“E4000”、IDとして“6A0299”を有する。ストレージ装置10は、CM（送信部）11-1, 11-2、ディスク13-1～13-4、タイムアウトテーブル（格納部）14、タイムアウト時間変更部（変更部）15、及びタイムアウト上限テーブル16をそなえる。

20

【0036】

CM11-1は、CA12-1, 12-2のほか不図示のCPUやメモリ、キャッシュメモリ等をそなえ、I/O処理を行なう。

CM11-2は、CA12-3, 12-4のほか不図示のCPUやメモリ、キャッシュメモリ等をそなえ、I/O処理を行なう。

CA12-1～12-4は、サーバ2と通信可能に接続するインタフェースコントローラであり、例えば、ファイバチャネルアダプタである。CA12-1, 12-2, 12-4は、それぞれパス30-1～30-3経由でサーバのHBA9-1～9-3と接続されている。

30

【0037】

ここで、CM11-1とCM11-2とは、同様に構成されている。また、CA12-1～CA12-4は同様に構成されている。

パス30-1～30-3は、サーバ2とストレージ装置10とを接続するパスであり、例えばFCである。

ディスク13-1～13-4は、情報を記憶可能な記憶装置であり、例えば、ニアライン13-1、HDD13-2, 13-3、及びSSD13-4をそなえる。

【0038】

ニアライン13-1は、磁性体を塗布した円盤を記録媒体として使用し、磁気ヘッドを移動させ、高速回転している円盤に対して情報の読み書きを行なうストレージドライブである。ニアライン13-1は、一般的にHDD13-2, 13-3よりも大容量であるが、アクセスが低速である。

40

HDD13-2, 13-3は、磁性体を塗布した円盤を記録媒体として使用し、磁気ヘッドを移動させ、高速回転している円盤に対して情報の読み書きを行なうストレージドライブである。

【0039】

SSD13-4、半導体素子メモリを記憶装置として使用するストレージドライブであり、シリコンディスクドライブ、半導体ディスクドライブともよばれる。一般に、SSD33は、ニアライン13-1やHDD13-2, 13-3のような磁気ヘッドの移動に伴うヘッドシークタイムがないため、これらよりも高速なランダムアクセスが可能である。

50

その一方、半導体メモリを使用するので、一般的にニアライン 13 - 1 や HDD 13 - 2 , 13 - 3 よりも高価である。

【0040】

タイムアウトテーブル 14 は、ストレージ装置 10 にそなえられているディスク 13 毎のタイムアウト時間を記憶するテーブルである。タイムアウトテーブル 14 は、例えば、CM 11 - 1 , 11 - 2 の不図示のメモリに格納される。このタイムアウトテーブル 14 については、図 5 を用いて後述する。

なお、ストレージ装置 10 の高負荷などの原因により処理遅延が発生しているときに、I/O タイムアウトエラーが頻発するのは好ましくない。

【0041】

10

そこで、タイムアウト時間変更部 15 は、ストレージ装置 10 の負荷状態を監視し、ストレージ装置 10 の保守や高負荷などによりタイムアウトの発生が予測される場合に、タイムアウトテーブル 14 に設定されているタイムアウト時間の値を上げる。このとき、タイムアウト時間変更部 15 は、後述するタイムアウト上限テーブル 16 に設定されている上限値を上限としてタイムアウト時間を各パス 30 に対して設定する。又、タイムアウト時間変更部 15 は、sense を用いてタイムアウトテーブル 14 の更新をサーバ 2 に通知する。

【0042】

ここで、sense とは、ストレージ装置 10 , 20 が、サーバ 2 からの I/O に対して、応答する際にサーバ 2 に通知するために用いられるメッセージである。sense は当

20

業界において周知の手法であるためその説明は省略する。
又、タイムアウト時間変更部 15 は、ストレージ装置 10 の高負荷状態が解消されたときやファームウェアの更新時にも、タイムアウトテーブル 14 に設定されている値を変更する。このときも、タイムアウト時間変更部 15 は、タイムアウトテーブル 14 に更新が行なわれた旨を、sense を用いてサーバ 2 に通知する。

【0043】

これらのタイムアウト時間の変更及びサーバ 2 への通知処理については図 9 を用いて後述する。

タイムアウト上限テーブル 16 は、ストレージ装置 10 にそなえられているディスク 13 及びパス 30 毎に設定可能なタイムアウト値の上限値を記憶するテーブルである。このタイムアウト上限テーブル 16 については、図 6 を用いて後述する。

30

【0044】

ストレージ装置 20 も、サーバ 2 に対して記憶領域を提供するものであり、例えば、RAID 装置である。ストレージ装置 20 は、外部からタイムアウト値を設定できないタイプのストレージ装置である。又、図中の例では、装置名として“E3000”、ID として“4A0291”を有する。ストレージ装置 20 は、CM 21 - 1 , 21 - 2 及びディスク 23 - 1 , 23 - 2 をそなえる。

【0045】

CM 21 - 1 は、CA 22 - 1 , 22 - 2 のほか不図示の CPU やメモリ、キャッシュメモリ等をそなえ、I/O 処理を行なう。

40

CM 21 - 2 は、CA 22 - 3 , 22 - 4 のほか不図示の CPU やメモリ、キャッシュメモリ等をそなえ、I/O 処理を行なう。

CA 22 - 1 ~ 22 - 4 は、サーバ 2 と通信可能に接続するインタフェースコントローラであり、例えば、ファイバチャネルアダプタである。CA 22 - 1 , 22 - 2 , 22 - 4 は、それぞれパス 30 - 4 ~ 30 - 6 経由でサーバの HBA 9 - 4 ~ 9 - 6 と接続されている。

【0046】

ここで、CM 21 - 1 と CM 21 - 2 とは、同様に構成されている。また、CA 22 - 1 ~ CA 22 - 4 は同様に構成されている。

パス 30 - 4 ~ 30 - 6 は、サーバ 2 とストレージ装置 20 とを接続するパスであり、

50

例えば F C である。

ディスク 2 3 - 1 , 2 3 - 2 は、情報を記憶可能な記憶装置であり、例えば、H D D 2 3 - 1 , 2 3 - 2 である。

【 0 0 4 7 】

H D D 2 3 - 1 , 2 3 - 2 は、磁性体を塗布した円盤を記録媒体として使用し、磁気ヘッドを移動させ、高速回転している円盤に対して情報の読み書きを行なうストレージドライブである。

なお、以下、ストレージ装置 1 0 にそなえられた C M を示す符号としては、複数の C M のうち 1 つを特定する必要があるときには符号 1 1 - 1 , 1 1 - 2 を用いるが、任意の C M を指すときには符号 1 1 を用いる。

10

【 0 0 4 8 】

又、ストレージ装置 1 0 にそなえられた C A を示す符号としては、複数の C A のうち 1 つを特定する必要があるときには符号 1 2 - 1 ~ 1 2 - 4 を用いるが、任意の C A を指すときには符号 1 2 を用いる。

さらに、ストレージ装置 1 0 にそなえられたディスクを示す符号としては、複数のディスクのうち 1 つを特定する必要があるときには符号 1 3 - 1 ~ 1 3 - 4 を用いるが、任意のディスクを指すときには符号 1 3 を用いる。

【 0 0 4 9 】

さらに、ストレージ装置 2 0 にそなえられた C M を示す符号としては、複数の C M のうち 1 つを特定する必要があるときには符号 2 1 - 1 , 2 1 - 2 を用いるが、任意の C M を指すときには符号 2 1 を用いる。

20

又、ストレージ装置 2 0 にそなえられた C A を示す符号としては、複数の C A のうち 1 つを特定する必要があるときには符号 2 2 - 1 ~ 2 2 - 4 を用いるが、任意の C A を指すときには符号 2 2 を用いる。

【 0 0 5 0 】

さらに、ストレージ装置 2 0 にそなえられたディスクを示す符号としては、複数のディスクのうち 1 つを特定する必要があるときには符号 2 3 - 1 , 2 3 - 2 を用いるが、任意のディスクを指すときには符号 2 3 を用いる。

さらに、パスを示す符号としては、複数のパスのうち 1 つを特定する必要があるときには符号 3 0 - 1 ~ 3 0 - 6 を用いるが、任意のパスを指すときには符号 3 0 を用いる。

30

【 0 0 5 1 】

なお、図 1 に示す例においては、ストレージ装置 1 0 に 2 つの C M 1 2 がそなえられているが、これに限定されるものではなく、3 以上の C M 1 2 をそなえてもよい。

又、ストレージ装置 1 0 に 4 つのディスク 1 3 がそなえられているが、これに限定されるものではなく、3 以下又は 5 以上のディスク 1 3 をそなえてもよい。

さらに、図 1 に示す例においては、ストレージ装置 2 0 に 2 つの C M 2 2 がそなえられているが、これに限定されるものではなく、3 以上の C M 2 2 をそなえてもよい。

【 0 0 5 2 】

又、ストレージ装置 2 0 に 2 つのディスク 2 3 がそなえられているが、これに限定されるものではなく、3 以上のディスク 2 3 をそなえてもよい。

40

次に、図 2 ~ 図 4 を参照して、サーバ 2 にそなえられているサーバ側タイムアウトテーブル 4 について説明する。

図 2 は、本実施形態の一例としてのサーバ側タイムアウトテーブル 4 の初期状態を例示する図である。又、図 3 は、本実施形態の一例としてのストレージ装置 1 0 , 2 0 の識別後のサーバ側タイムアウトテーブル 4 を例示する図である。図 4 は、図 3 のサーバ側タイムアウトテーブル 4 にレコードをさらに追加した例を示す図である。

【 0 0 5 3 】

サーバ側タイムアウトテーブル 4 は、装置識別情報 4 1、ディスク種別 4 2、C M 係数 4 3、先頭パスのタイムアウト時間 4 4、中間パスのタイムアウト時間 4 5、最終パスのタイムアウト時間 4 6、及び上限 4 7 の各フィールドを有する。

50

装置識別情報フィールド 4 1 は、ストレージ装置 1 0 を識別する情報を示す。例えば、装置識別情報フィールド 4 1 には、ストレージ装置 1 0 のモデル名とシリアル番号とを連結した文字列が格納される。

【 0 0 5 4 】

ディスク種別フィールド 4 2 は、ストレージ装置 1 0 にそなえられた各ディスク 1 3 の種別を示す。例えば、ディスク種別フィールド 4 2 には、図 3 に示すように、“ H D D 1 0 0 0 0 ”、“ H D D 1 5 0 0 0 ”、“ S S D ”、“ N E A R L I N E ”などのディスクの性能や特性を表わす文字列が格納される。“ H D D 1 0 0 0 0 ”は、回転速度が 1 0 , 0 0 0 R P M の H D D を示し、“ H D D 1 5 0 0 0 ”は、回転速度が 1 5 , 0 0 0 R P M の H D D を示す。“ S S D ”は S S D を示し、“ N E A R L I N E ”はニアラインを示す。

10

【 0 0 5 5 】

先頭パスのタイムアウト時間フィールド 4 4 は、マルチパス環境の場合に先頭パスに使用するタイムアウト時間を格納している。

中間パスのタイムアウト時間フィールド 4 5 は、マルチパス環境の場合に先頭パス以外であり、且つ最終パス以外のパス（以降、このようなパスを「中間パス」と呼ぶ）におけるタイムアウト時間を格納している。

【 0 0 5 6 】

最終パスのタイムアウト時間フィールド 4 6 は、マルチパス環境の場合に最終パスに使用するタイムアウト時間を格納している。

20

なお、ここでの「先頭パス」、「中間パス」、「最終パス」は、パスの物理的な番号と必ずしも一致しない。ここでは、I O 発行時に最初にアクセスされるパスを「先頭パス」、先頭パスがエラーとなり、リトライ時に使用されるパスを「中間パス」、最終リトライ時（最大リトライ回数でのリトライ）に使用されるパスを「最終パス」（又は最終経路）と呼ぶ。また、先頭パスと中間パスとを「最終パス以外のパス」（又は最終経路以外の経路）と呼ぶ。

【 0 0 5 7 】

なお、一般に、I O アクセス時には、I O 発行時点で最も負荷の低いパスが選択されてそのパスに I O が発行される。

例えば、物理パス # 1 ~ # 4 の 4 パス構成の場合を考える。

30

（ 1 ）初回の I O は、物理パス # 1 ~ # 4 のうち、そのとき負荷が最も低い物理パス # 2 に発行される。

（ 2 ）物理パス # 2 がエラーとなり、パスを切り替えて 1 回目（初回）のリトライを行なう。物理パス # 1 , # 3 , # 4 のうち、そのとき負荷の少ない物理パス # 1 が選択される。

（ 3 ）物理パス # 1 もエラーとなり、2 回目のリトライでは物理パス # 4 が選択される。

（ 4 ）物理パス # 4 もエラーとなり、3 回目のリトライは物理パス # 3 が選択される。

【 0 0 5 8 】

上記の例は 4 パス構成であるので、物理パス # 3 でのリトライが最後となる。

この場合、先頭パスは物理パス # 2、中間パスは、1 回目及び 2 回目のリトライで選択された物理パス # 1 及び # 4、最終パスは最終リトライで選択された物理パス # 3 となる。

40

なお、最大リトライ回数よりもパス 3 0 の数のほうが多い場合は、たとえまだリトライしていないパス 3 0 が残存していても、最大リトライ回数でのリトライが最終リトライとなる。

【 0 0 5 9 】

タイムアウト時間フィールド 4 4 ~ 4 6 に、先頭パス、中間パス及び最終パスのタイムアウト時間を設定する理由を以下に説明する。

I O のタイムアウトは、パス 3 0 の高負荷や故障によるものと、ストレージ装置 1 0 の高負荷や故障など、ストレージ装置 1 0 に共通に発生するものがある。

50

パス 30 に依存するタイムアウトが発生した場合はパスを切り替えるほうが好ましい。これに対し、ストレージ装置 10 の高負荷によるタイムアウトが発生した場合は待機するほうが好ましい。ストレージ装置 10 の故障の場合は、I/O は成功しないので、直ちに I/O エラーとするほうが好ましい。

【0060】

しかし、タイムアウトの原因が、パスであるか、或いはストレージ装置 10 の高負荷であるか故障であるかを、サーバ 2 からは判断することができない。

そこで、本実施形態においては、ストレージ装置 10 のタイムアウトテーブル 14 に、1 つのタイムアウト時間を設定する代わりに、先頭パス、中間パス、最終パスのそれぞれに対して個別にタイムアウト時間 T_1 , T_2 , T_3 を設定する。その際、先頭パス、中間パス、最終パスのそれぞれのタイムアウト時間の間に、 $T_1 \geq T_2 \geq T_3$ の関係が成立するように、それぞれのタイムアウト時間を設定する。

10

【0061】

詳細には、先頭パスのタイムアウト時間 T_1 は、パスの異常を考慮して短い値に設定して、パスを短時間で切り替える。

中間パスのタイムアウト時間 T_2 は、2 パス目以降でもストレージ装置 10 に起因するタイムアウトの可能性を考慮して短い値に設定にする。これにより、ストレージ装置 10 の高負荷時に、先頭パスと 2 パス目との合計時間応答を待機することになり、I/O エラーの発生を回避できる可能性が高くなる。ストレージ装置 10 の故障の場合でも、タイムアウト時間が短く設定されるので、待ち時間全体を短縮することができる。

20

【0062】

最終パスのタイムアウト時間 T_3 は、高負荷状態を想定し、もう少し待機すれば I/O が成功する可能性があるので T_1 , T_2 よりも長い値に設定する。

上限フィールド 47 は、ストレージ装置 10 , 20 のタイムアウト時間の上限値を示す。

C/M 係数フィールド 43 は、ストレージ装置 10 が担当 C/M 型の場合に、待機 C/M を経由するパスに I/O を発行する際に使用する C/M 係数 (1) を格納している。

【0063】

ここで、ストレージ装置 10 , 20 には、各 LUN に対する通常のアクセスパスが一方の C/M 11 , 21 に固定されているストレージ装置と、固定されていないストレージ装置とが存在する。このようにアクセスパスが固定されているストレージ装置を「担当 C/M 型」、固定されていないストレージ装置を「非担当 C/M 型」と呼ぶ。

30

担当 C/M 型のストレージ装置 10 , 20 では、一方の C/M 11 , 21 に接続されたパスがアクティブ状態となり、他方の C/M 11 , 21 に接続されたパスは待機状態となる。通常のアクセスはアクティブ状態のパスのみが使用され、待機状態のパスは使用されない。アクティブ状態となるパスに接続される C/M を「担当 C/M」、待機状態となるパスに接続される C/M を「待機 C/M」と呼ぶ。

【0064】

これに対し、非担当 C/M 型のストレージ装置 10 , 20 では、全てのパスがアクティブ状態となり、通常アクセスに全パスが使用される。

40

このように、担当 C/M 型のストレージ装置 10 , 20 では、C/M 11 , 21 が担当 C/M か待機 C/M かによって性能に差が生じるため、タイムアウト時間についても、C/M 11 , 21 が担当 C/M か否かを考慮する。

【0065】

この性能差はストレージ装置 10 , 20 毎に異なるが、予め既知の情報であるので、その性能差の割合に応じた C/M 係数を C/M 係数フィールド 43 に格納する。

サーバ 2 は、アクセスするパスが担当 C/M 経由か否かがわかる。このため、待機 C/M に I/O を発行するときは、後述するタイムアウトフィールド 44 ~ 46 のタイムアウト時間に C/M 係数を乗算して、担当 C/M 経由のパスよりもタイムアウト時間を長めの値に設定する。

50

【 0 0 6 6 】

例えば、4パス構成を有し、うち2パスが担当CM11に接続され、残り2パスが待機CM11に接続され、SSD13-4にアクセスする場合を考える。

図4の5行目を参照すると、SSDでは、先頭パスのタイムアウト時間には10秒、2パス目のタイムアウト時間には15秒が使用される。

担当CM11に接続されている先頭パス、2パス目いずれもIOエラーが発生した場合、非担当CM11に接続されている3パス目が使用される。

【 0 0 6 7 】

このとき、待機CM側の3パス目のタイムアウト時間には、図4の中間パスのタイムアウト時間フィールド45の値“15”とCM係数フィールド43の値“1.4”とから、 $15 \times 1.4 = 21$ 秒が設定される。最終パスである4パス目のタイムアウト時間は、 $20 \times 1.4 = 28$ 秒であるが、上限フィールド47の値“25”を超えているので、上限値の25秒が設定される。

10

【 0 0 6 8 】

サーバ2が起動されると、タイムアウトテーブル収集部5により、図2に示すサーバ側タイムアウトテーブル4が、サーバ2の不図示のメモリに確保される。初期状態のサーバ側タイムアウトテーブル4には、タイムアウト非対応のストレージ装置20に使用されるデフォルトのタイムアウト時間である“40”が全パスに対して設定される。

その後、ストレージ装置10, 20が認識されると、タイムアウトテーブル収集部5は、ストレージ装置10の認識時にinquiryコマンド等を使用してストレージ装置10からタイムアウトテーブル14を取得し、データをサーバ側タイムアウトテーブル4に追加する。

20

【 0 0 6 9 】

ここで、inquiryコマンドとは、SCSI装置の種類と構成をSCSI装置に問い合わせるために使用されるSCSIコマンドであり、その機能は公知であるため、ここでの説明を省略する。なお、マルチパスドライバ部3がほかのコマンドを使用してストレージ装置10, 20に対して問い合わせを行なってもよい。

前述のように、このとき、タイムアウトテーブル収集部5は、例えば、ストレージ装置10のモデル名、シリアル番号の情報を、サーバ側タイムアウトテーブル4に装置識別情報として追加する。例えば、タイムアウトテーブル収集部5により、サーバ側タイムアウトテーブル4に、図2に示すようなレコードが追加される。

30

【 0 0 7 0 】

例えば、その後別のストレージ装置（不図示）が認識されると、サーバ側タイムアウトテーブル4に図2に示すようなレコードが追加される。

又、前述のように、ストレージ装置10側にもタイムアウトテーブル14がそなえられている。

図5は、本実施形態の一例としてのストレージ装置10側のタイムアウトテーブル14を例示する図である。

【 0 0 7 1 】

タイムアウトテーブル14は、ディスク種別142、CM係数143、先頭パスのタイムアウト時間144、中間パスのタイムアウト時間145、最終パスのタイムアウト時間146、及び上限147の各フィールドを有する。

40

ディスク種別フィールド142は、ストレージ装置10内のディスクの種別を示す。例えば、ディスク種別フィールド142には、図5に示すように、“HDD10000”、“HDD15000”、“SSD”、“NEARLINE”などのディスクの性能や特性を表わす文字列が格納される。“HDD10000”は、回転速度が10,000RPMのHDDを示し、“HDD15000”は、回転速度が15,000RPMのHDDを示す。“SSD”はSSDを示し、“NEARLINE”はニアラインを示す。

【 0 0 7 2 】

CM係数フィールド143は、ストレージ装置10が担当CM型の場合に、待機CMに

50

ＩＯを発行する際に使用するＣＭ係数を格納している。

先頭パスのタイムアウト時間フィールド１４４は、マルチパス環境の場合に先頭パスに使用するタイムアウト時間を格納している。

中間パスのタイムアウト時間フィールド１４５は、マルチパス環境の場合に先頭パスにおけるタイムアウト時間を格納している。

【００７３】

最終パスのタイムアウト時間フィールド１４６は、マルチパス環境の場合に最終パスに使用するタイムアウト時間を格納している。

上限フィールド１４７は、ストレージ装置１０の高負荷などの際に、ストレージ装置１０のタイムアウト時間を変更可能な範囲の上限値を示す。

10

上記のフィールド１４４～１４７には、ストレージ装置１０の性能等の特性に合わせ、ストレージ装置１０の工場出荷時に最適な値がそれぞれ事前に設定されている。又、フィールド１４４～１４７の各値は、後からオペレータ等が変更可能としてもよい。

【００７４】

図６は、本実施形態の一例としてのタイムアウト上限テーブル１６を例示する図である。

タイムアウトテーブル１６は、ディスク種別フィールド１６２と上限フィールド１６７とを有する。

ディスク種別フィールド１６２は、ストレージ装置１０内のディスクの種別を示す。例えば、ディスク種別フィールド１６２には、図６に示すように、“ＨＤＤ１００００”、“ＨＤＤ１５０００”、“ＳＳＤ”、“ＮＥＡＲＬＩＮＥ”などのディスクの性能や特性を表わす文字列が格納される。“ＨＤＤ１００００”は、回転速度が１０，０００ＲＰＭのＨＤＤを示し、“ＨＤＤ１５０００”は、回転速度が１５，０００ＲＰＭのＨＤＤを示す。“ＳＳＤ”はＳＳＤを示し、“ＮＥＡＲＬＩＮＥ”はニアラインを示す。

20

【００７５】

上限フィールド１６７は、ストレージ装置１０の高負荷などの際に、ストレージ装置１０のタイムアウト時間を変更可能な範囲の上限値を示す。

上記の上限フィールド１６７には、ストレージ装置１０の性能等の特性に合わせ、ストレージ装置１０の工場出荷時に最適な値が事前に設定されている。又、上限フィールド１６７は、後からオペレータ等が変更可能としてもよい。

30

【００７６】

なお、サーバ２において、マルチパスドライバ部３、タイムアウトテーブル収集部５、タイムアウト時間取得部６、及びＩＯ発行部８としての機能を実現するためのプログラムは、例えばフレキシブルディスク、ＣＤ（ＣＤ－ＲＯＭ、ＣＤ－Ｒ、ＣＤ－ＲＷ等）、ＤＶＤ（ＤＶＤ－ＲＯＭ、ＤＶＤ－ＲＡＭ、ＤＶＤ－Ｒ、ＤＶＤ＋Ｒ、ＤＶＤ－ＲＷ、ＤＶＤ＋ＲＷ、ＨＤ－ＤＶＤ等）、ブルーレイディスク、磁気ディスク、光ディスク、光磁気ディスク等の、コンピュータ読取可能な記録媒体に記録された形態で提供される。そして、コンピュータはその記録媒体からプログラムを読み取って内部記憶装置又は外部記憶装置に転送し格納して用いる。又、そのプログラムを、例えば磁気ディスク、光ディスク、光磁気ディスク等の記憶装置（記録媒体）に記録しておき、その記憶装置から通信経路を介してコンピュータに提供するようにしてもよい。

40

【００７７】

又、ストレージ装置１０において、タイムアウト時間変更部１５としての機能を実現するためのプログラムは、例えばフレキシブルディスク、ＣＤ（ＣＤ－ＲＯＭ、ＣＤ－Ｒ、ＣＤ－ＲＷ等）、ＤＶＤ（ＤＶＤ－ＲＯＭ、ＤＶＤ－ＲＡＭ、ＤＶＤ－Ｒ、ＤＶＤ＋Ｒ、ＤＶＤ－ＲＷ、ＤＶＤ＋ＲＷ、ＨＤ－ＤＶＤ等）、ブルーレイディスク、磁気ディスク、光ディスク、光磁気ディスク等の、コンピュータ読取可能な記録媒体に記録された形態で提供される。そして、コンピュータはその記録媒体からプログラムを読み取って内部記憶装置又は外部記憶装置に転送し格納して用いる。又、そのプログラムを、例えば磁気ディスク、光ディスク、光磁気ディスク等の記憶装置（記録媒体）に記録しておき、その記憶

50

装置から通信経路を介してコンピュータに提供するようにしてもよい。

【0078】

マルチパスドライバ部3、タイムアウトテーブル収集部5、タイムアウト時間取得部6、及びI/O発行部8としての機能を実現する際には、内部記憶装置（本実施形態ではサーバ2の不図示のRAMやROM）に格納されたプログラム（入出力プログラム）がコンピュータのマイクロプロセッサ（本実施形態ではサーバ2の不図示のCPU）によって実行される。このとき、記録媒体に記録されたプログラムをコンピュータが読み取って実行するようにしてもよい。

【0079】

又、タイムアウト時間変更部15としての機能を実現する際には、内部記憶装置（本実施形態ではストレージ装置10の不図示のRAMやROM）に格納されたプログラムがコンピュータのマイクロプロセッサ（本実施形態ではストレージ装置10のCM11）によって実行される。このとき、記録媒体に記録されたプログラムをコンピュータが読み取って実行するようにしてもよい。

10

【0080】

なお、本実施形態において、コンピュータとは、ハードウェアとOSとを含む概念であり、OSの制御の下で動作するハードウェアを意味している。又、OSが不要でアプリケーションプログラム単独でハードウェアを動作させるような場合には、そのハードウェア自体がコンピュータに相当する。ハードウェアは、少なくとも、CPU等のマイクロプロセッサと、記録媒体に記録されたコンピュータプログラムを読み取るための手段とをそなえており、本実施形態においては、サーバ2及びCM11がコンピュータとしての機能を有しているのである。

20

【0081】

（B）動作

次に、図7～図9を参照して、ストレージシステム1の動作について説明する。

まず、図7を参照して、マルチパスドライバ部3の起動時の処理について説明する。

図7は、本実施形態の一例としての初期化処理を示すフローチャート（ステップSA1からSA11，SB1，SB2）である。

【0082】

ステップSA1において、サーバ2が起動されて、マルチパスドライバ部3がロードされ、例えばサーバ2の不図示のメモリにサーバ側タイムアウトテーブル4が確保される。

30

次にステップSA2において、マルチパスドライバ部3は、ストレージ装置10，20のLUNの認識とマルチパスの構築とを行なう。このとき、マルチパスドライバ部3は、ストレージ装置10，20のディスク種別をinquiryコマンドなどで取得する。そして、LUN情報7（図1参照）に、取得したディスク種別を記録する。

【0083】

次に、タイムアウトテーブル収集部5は、以下のステップSA3～SA11の処理を、サーバ2に接続されているストレージ装置10，20のそれぞれに対して実行する。

まず、ステップSA4において、タイムアウトテーブル収集部5は、ストレージ装置10，20に、例えばinquiryコマンドを発行し、ストレージ装置10，20がタイムアウト対応かどうかを問い合わせる。

40

【0084】

ステップSB1において、ストレージ装置10は、inquiryコマンドの応答として、タイムアウト対応であることを示す情報を返す。一方、タイムアウト非対応のストレージ装置20はこのような情報を返さない。

ステップSA5において、タイムアウトテーブル収集部5は、ストレージ装置10，20からタイムアウト対応を示す応答が返されたかどうかを判定する。

【0085】

タイムアウト対応を示す応答が返された場合（ステップSA5のYESルート参照）、ステップSA6において、タイムアウトテーブル収集部5は、タイムアウトテーブル14

50

を取得するためのコマンドをストレージ装置 10 に対して発行する。

ステップ S B 2 において、ストレージ装置 10 はサーバ 2 にタイムアウトテーブル 14 を送信する。

【0086】

ステップ S A 7 において、タイムアウトテーブル収集部 5 はストレージ装置 10 からタイムアウトテーブル 14 を受け取り、取得したタイムアウトテーブル 14 のデータを、ステップ S A 1 で確保したサーバ側タイムアウトテーブル 4 に追加する。

次にステップ S A 8 において、タイムアウトテーブル収集部 5 は、ストレージ装置 10 の各 L U N に対し *i n q u i r y* コマンドを発行し、装置名とディスク種別とを取得する。

10

【0087】

ステップ S A 9 において、タイムアウトテーブル収集部 5 は、ステップ S A 8 で取得した装置名とディスク種別とをタイムアウトテーブル収集部 5 の L U N 情報 7 に追加する。

一方、タイムアウト対応を示す応答が返されない場合（ステップ S A 5 の N O ルート参照）、非対応のストレージ装置 20 であるので、ステップ S A 10 において、ストレージ装置 20 にデフォルトのタイムアウト値を使うように L U N 情報 7 にフラグを立てる。

【0088】

タイムアウトテーブル収集部 5 は、上記のステップ S A 3 ~ S A 11 の処理を、サーバ 3 に接続されている全ストレージ装置 10 , 20 に実行して、全ストレージ装置 10 , 20 の情報を取得する。

20

次に、図 8 を参照して、I O 発行及び応答時の処理について説明する。

図 8 は、本実施形態の一例としての I O 発行及び応答時の処理を示すフローチャート（ステップ S C 1 ~ S C 16）である。

【0089】

ステップ S C 1 において、タイムアウト時間取得部 6 は、I O を発行しようとしているストレージ装置 10 , 20 が、タイムアウト対応ストレージ装置かどうかを判別する。その際、タイムアウト時間取得部 6 は、L U N 情報 7 に装置情報とディスク種別とが存在するかどうかを調べ、これらが存在する場合はタイムアウト対応ストレージ装置であると判定する。

【0090】

30

タイムアウト対応ストレージ装置の場合（ステップ S C 1 の Y E S ルート参照）、ステップ S C 2 において、タイムアウト時間取得部 6 は、サーバ側タイムアウトテーブル 4 を参照して、タイムアウト値を取得する。詳細には、タイムアウト時間取得部 6 は、L U N 情報 7 のストレージ装置 10 及び L U N のディスク 13 の情報に基づいて、サーバ側タイムアウトテーブル 4 のレコードを参照する。例えば、I O の発行先の L U N が “ E 4 0 0 0 - 6 A 0 2 9 9 ” であり、ディスク種別が “ S S D ” である場合、タイムアウト時間取得部 6 は、図 4 のサーバ側タイムアウトテーブル 4 を参照して、初回の I O のタイムアウト時間として “ 10 秒 ” を取得する。

【0091】

一方、タイムアウトテーブル 14 を持たず、外部からタイムアウト値を設定できないストレージ装置 20 の場合（ステップ S C 1 の N O ルート参照）、ステップ S C 3 において、タイムアウト時間取得部 6 は、デフォルト用のレコードを参照し、そのタイムアウト値を取得する。例えば、タイムアウト時間取得部 6 は、タイムアウト時間として “ 40 秒 ” を取得する。

40

【0092】

次に、ステップ S C 4 において、タイムアウト時間取得部 6 は、I O を発行可能なパスが他に存在するかどうかを確認する。

I O を発行可能なパスが他に存在する場合（ステップ S C 4 の Y E S ルート参照）、ステップ S C 5 において、タイムアウト時間取得部 6 は、I O が初回かリトライかを判別する。

50

【 0 0 9 3 】

ＩＯが初回の場合（ステップＳＣ５のＹＥＳルート参照）、ステップＳＣ６において、タイムアウト時間取得部６は、ステップＳＣ２又はステップＳＣ３で参照したレコードの、先頭パスのタイムアウト時間をＩＯに設定する。前述のように、最初のＩＯを発行するときは、タイムアウト時間取得部６は、先頭パスの短いタイムアウト時間を使用する。このタイムアウト時間は、短い、通常のＩＯ性能ではタイムアウトと誤検出しにない程度に長い値である。

【 0 0 9 4 】

ＩＯがリトライの場合（ステップＳＣ５のＮＯルート参照）、ステップＳＣ７において、タイムアウト時間取得部６は、ステップＳＣ２又はステップＳＣ３で参照したレコードの、中間パスのタイムアウト時間をＩＯに設定する。このとき、ＩＯの発行先のパス３０が待機ＣＭ１１を経由するパスの場合、タイムアウトテーブル１４のＣＭ係数フィールド４７の値をタイムアウト時間に乗算する。

10

【 0 0 9 5 】

一方、ＳＣ４でＩＯを発行可能なパスが他にない、つまり最終パスの場合（ステップＳＣ４のＮＯルート参照）、ステップＳＣ８において、タイムアウト時間取得部６は、ＳＣ２又はＳＣ３で選択したレコードの、最終パスのタイムアウト値をＩＯに設定する。このときも、ＩＯの発行先のパス３０が待機ＣＭ１１を経由するパスの場合、タイムアウトテーブル１４のＣＭ係数フィールド４７の値をタイムアウト時間に乗算する。

【 0 0 9 6 】

ステップＳＣ９において、ＩＯ発行部８がＩＯを発行する。

ステップＳＣ１０において、ＩＯ発行部８は、ＩＯの応答が返され、タイムアウト以外のエラーが発生したかどうかを判定する。

タイムアウト以外のエラーの場合（ステップＳＣ１０のＹＥＳルート参照）、ステップＳＣ１１において、ＩＯ発行部８は、必要に応じてＩＯをリトライする。

20

【 0 0 9 7 】

一方、タイムアウトエラーの場合（ステップＳＣ１０のＮＯルート参照）、ステップＳＣ１２において、ＩＯ発行部８は、タイムアウトしたかどうかを判定する。

タイムアウトした場合（ステップＳＣ１２のＹＥＳルート参照）、ステップＳＣ１３において、ＩＯ発行部８は、代替パスがあり、かつ最大リトライ回数を超えていない（リトライアウトしていない）かどうかを判定する。

30

【 0 0 9 8 】

代替パスがあり、かつリトライ回数を超えていない場合（ステップＳＣ１３のＹＥＳルート参照）、ステップＳＣ１４において、ＩＯ発行部８は、パスを換えてＩＯをリトライするためにステップＳＣ４に戻る。

一方、ステップＳＣ１３で代替パスがないか、或いはリトライ回数を超えている場合（ステップＳＣ１３のＮＯルート参照）、ステップＳＣ１６において、ＩＯ発行部８は、ＩＯがエラー終了したと判定する。

【 0 0 9 9 】

ステップＳＣ１２でタイムアウトしていない場合（ステップＳＣ１２のＹＥＳルート参照）、ステップＳＣ１５において、ＩＯ発行部８はＩＯが正常終了したと判定する。

40

次に、図９を参照して、ディスク１３，２３のタイムアウト値が更新され、ストレージ装置１０，２０のタイムアウトテーブル１４が更新されたときの処理について説明する。

前述のように、タイムアウトテーブル１４の更新は、図７を参照して説明したストレージ装置１０の認識時以外にも行なわれることがある。このようなケースとして、ストレージ装置１０がファームアップされて性能が向上した場合などが挙げられる。

【 0 1 0 0 】

このような場合、ストレージ装置１０は、特定のｓｅｎｓｅ情報などを用いて、タイムアウトテーブル１４の更新をマルチパストライバ部３に通知する。マルチパストライバ部３はｉｎｑｕｉｒｙコマンド等を用いて、更新されたタイムアウトテーブル１４を取得し

50

、サーバ側タイムアウトテーブル 4 を更新する。

図 9 は、本実施形態の一例としてのタイムアウト値の更新処理を示すフローチャート（ステップ S D 1 ～ S D 4 , S E 1 ～ S E 6 ）である。

【 0 1 0 1 】

ステップ S D 1 において、ストレージ装置 1 0 のタイムアウト時間変更部 1 5 は、ファームウェアの更新や、ディスク 1 3 , 2 3 の高負荷又は負荷の解消等を検出し、タイムアウトテーブル 1 4 のタイムアウト値を変更する。

ステップ S D 2 において、タイムアウト時間変更部 1 5 は、タイムアウト時間の更新をサーバ 2 に通知するための *s e n s e* をセット（準備）する。

【 0 1 0 2 】

前述のように、*s e n s e* とは、ストレージ装置 1 0 , 2 0 が、サーバ 2 からの I O に対して、応答する際にサーバ 2 に通知するために用いられるメッセージである。このため、サーバ 2 から I O が発行されない限り、ストレージ装置 1 0 のタイムアウト時間変更部 1 5 がサーバ 2 にタイムアウトテーブル 1 4 の更新を通知することはない。

ステップ S E 1 において、サーバ 2 のアプリケーション等から通常 I O が発行される。

【 0 1 0 3 】

ステップ S D 3 において、ストレージ装置 1 0 のタイムアウト時間変更部 1 5 は、ステップ S D 2 で設定した *s e n s e* をサーバ I O に応答する。

ステップ S E 2 において、サーバ 2 の I O 発行部 8 は、ステップ S E 1 で発行した I O の応答と共に *s e n s e* メッセージを受信する。

ステップ S E 3 において、サーバ 2 の I O 発行部 8 は、I O が正常応答したと判定し、I O の処理を終了する。

【 0 1 0 4 】

ステップ S E 4 において、サーバ 2 のタイムアウトテーブル収集部 5 は、タイムアウトテーブル取得コマンドをストレージ装置 1 0 に対して発行する。

ステップ S D 4 において、ストレージ装置 1 0 のタイムアウト時間変更部 1 5 は更新後のタイムアウトテーブル 1 4 をサーバ 2 に送信する。

ステップ S E 5 において、サーバ 2 のタイムアウトテーブル収集部 5 は、タイムアウトテーブル 1 4 を取得し、取得したタイムアウトテーブル 1 4 を用いて、サーバ側タイムアウトテーブル 4 を更新する。

【 0 1 0 5 】

ステップ S E 6 において、サーバ 2 の I O 発行部 8 は、更新されたタイムアウト値を使用して I O を発行する。

（ C ）変形例

（ C - 1 ）第 1 変形例

上記の実施形態においては、先頭パス、中間パス、最終パスのそれぞれに対して、タイムアウト時間を設定していた。

【 0 1 0 6 】

しかし、上記の実施形態の第 1 の変形例として、図 2 ～ 図 4 に示した先頭パスと中間パスとを「代替パスありの場合」としてまとめて、処理を簡略化してもよい。

図 1 0 に、先頭パスと中間パスとを「代替パスありの場合」としてまとめた場合のサーバ側タイムアウトテーブル 2 0 4 を示す。

図 1 0 は、本実施形態の第 1 変形例におけるサーバ側タイムアウトテーブル 2 0 4 を例示する図である。

【 0 1 0 7 】

サーバ側タイムアウトテーブル 2 0 4 は、上記の実施形態におけるサーバ側タイムアウトテーブル 4 の先頭パスのタイムアウト時間 4 4 及び中間パスのタイムアウト時間 4 5 に代えて、代替パスありの場合のタイムアウト時間フィールド 4 8 を有する。

代替パスありの場合のタイムアウト時間フィールド 4 8 は、マルチパス環境の場合に最終パス以外のパスに使用するタイムアウト時間を格納している。

10

20

30

40

50

【 0 1 0 8 】

なお、サーバ側タイムアウトテーブル 2 0 4 のその他のフィールドについては前述の実施形態と同じであるため、その説明を省略する。

又、特に図示はしないが、上記の実施形態の第 1 の変形例においては、ストレージ装置 1 0 側のタイムアウトテーブル 1 4 も、タイムアウト時間フィールド 1 4 4 ~ 1 4 5 に代えて、代替パスありの場合のタイムアウト時間フィールドを有する。

【 0 1 0 9 】

タイムアウト時間取得部 6 は、サーバ側タイムアウトテーブル 2 0 4 を参照して、最終パス以外のパスに対しては、タイムアウト時間フィールド 4 8 に設定されている値を設定して I O を発行する。最終パスの場合は、最終パスのタイムアウト時間フィールド 4 6 に設定されている値を使用して I O を発行する。

10

(C - 2) 第 2 変形例

或いは、本実施形態の第 2 変形例として、タイムアウト時間取得部 6 が、図 1 0 のサーバ側タイムアウトテーブル 2 0 4 を使用して、先頭パスにはタイムアウト時間フィールド 4 8 に設定されている値を設定して I O を発行してもよい。そして、2 パス目、3 パス目 ... と少しずつ長めの値を設定していてもよい。

(C - 3) 第 3 変形例

或いは、本実施形態の第 3 変形例として、サーバ側タイムアウトテーブル 4 に複数のタイムアウト時間を設定する代わりに、サーバ側タイムアウトテーブル 4 のタイムアウト時間を 1 つだけにしてもよい。この場合、タイムアウト時間取得部 6 は、リトライ可能な残パス数に応じた係数 (> 1) を掛けて、タイムアウト時間を短くしてもよい。この係数はマルチパストライバ部 3 が独自に持ってもよい。

20

(C - 4) 第 4 変形例

或いは、本実施形態の第 4 変形例として、図 1 1 に示すように、サーバ側タイムアウトテーブル 2 0 4 に、先頭パス、2 パス目、...、N パス目 (N は 2 以上の整数) と、全パスのタイムアウト時間を設定してもよい。

【 0 1 1 0 】

図 1 1 に、全てのパスのタイムアウト時間を設定したサーバ側タイムアウトテーブル 3 0 4 を示す。

図 1 1 は、本実施形態の第 4 変形例におけるサーバ側タイムアウトテーブル 3 0 4 を例示する図である。

30

サーバ側タイムアウトテーブル 3 0 4 は、先頭パス、2 パス目、...、N パス目のそれぞれに対するタイムアウト時間フィールド 4 4、4 5 - 1 ~ 4 5 - (N - 2)、4 6 を有する。

(C - 5) 第 5 変形例

又、ストレージ装置 1 0 によっては、L U N が複数種類のストレージで構築されている場合もある。このような例として、L U N コンカチネーションやシンプロビジョニングがある。L U N コンカチネーションやシンプロビジョニングでは、1 つの L U N の中に複数の R A I D グループやディスク種別が混在している。

【 0 1 1 1 】

40

そこで、本実施形態の第 5 変形例として、マルチパストライバ部 3 が、i n q u i r y コマンドを用いてストレージ装置 1 0、2 0 から L U N 単位のディスク種別と共に、論理ブロックアドレス (Logical Block Address : L B A) 範囲も取得する。そして、マルチパストライバ部 3 は、取得したこれらの情報を、図 1 2 に示すような L U N 管理テーブル 5 0 に記憶する。この L U N 管理テーブル 5 0 は L U N 毎に作成される。

【 0 1 1 2 】

図 1 2 は、実施形態の第 5 変形例における L U N 管理テーブル 5 0 を例示する図である。

L U N 管理テーブル 5 0 は、装置識別情報フィールド 5 1、ディスク種別フィールド 5 2、開始 L B A フィールド 5 3、及び終了 L B A フィールド 5 4 を有する。

50

装置識別情報フィールド 5 1 は、ストレージ装置 1 0 を識別する情報を示す。例えば、装置識別情報フィールド 5 1 には、ストレージ装置 1 0 のモデル名とシリアル番号とを連結した文字列が格納される。

【 0 1 1 3 】

ディスク種別フィールド 5 2 は、ストレージ装置 1 0 内のディスクの種別を示す。例えば、ディスク種別フィールド 5 2 には、“ H D D 1 0 0 0 0 ”、“ H D D 1 5 0 0 0 ”、“ S S D ”、“ N E A R L I N E ”などのディスクの性能や特性を表わす文字列が格納される。“ H D D 1 0 0 0 0 ”は、回転速度が 1 0 , 0 0 0 R P M の H D D を示し、“ H D D 1 5 0 0 0 ”は、回転速度が 1 5 , 0 0 0 R P M の H D D を示す。“ S S D ”は S S D を示し、“ N E A R L I N E ”はニアラインを示す。

10

【 0 1 1 4 】

開始 L B A フィールド 5 3 は、ディスク種別フィールド 5 2 のディスクの先頭の L B A を示す。

終了 L B A フィールド 5 3 は、ディスク種別フィールド 5 2 のディスクの末尾の L B A を示す。

そして、タイムアウト時間取得部 6 は、タイムアウトテーブル 4 と L U N 管理テーブル 5 0 とを参照して、I O の発行位置によってタイムアウト時間を変更させる。

【 0 1 1 5 】

例えば、図 1 2 に示す L U N は、1 0 0 0 0 L B A 毎に 2 種類のディスク 1 3 , 2 3 でコンカチネーションされている例を示す。

20

この例の場合、タイムアウト時間取得部 6 は、L B A 0 ~ 9 9 9 9 の範囲に I O アクセスする場合は、H D D 1 5 0 0 0 に設定されているタイムアウト値を使用する。一方、L B A 1 0 0 0 0 ~ 1 9 9 9 9 の範囲にアクセスがあった場合は、H D D 1 0 0 0 0 に設定されているタイムアウト値を使用する。

【 0 1 1 6 】

又、L B A 9 9 9 9 と 1 0 0 0 0 との間をまたぐ I O の場合は、タイムアウト時間取得部 6 は、タイムアウト時間が長い方の H D D 1 0 0 0 0 に設定されているタイムアウト値を使用して、H D D 1 0 0 0 0 が多少遅れてもタイムアウトが発生しないようにする。

(D) 効果

前述の如く、本実施形態の一例及びその各変形例においては、マルチパス環境において代替パスが存在する場合、サーバからストレージ装置への I O 発行時に短いタイムアウト時間を設定して I O を発行する。当該 I O でタイムアウトが発生した場合、パスを切り替えて前回よりも長いタイムアウト時間を設定して I O をリトライすることで、I O の応答時間を短縮することができる。これにより、異常が発生したときエラー応答やパス切替が発生するまでの時間を短くでき、最終的に I O 応答時間を短縮することが可能となる。

30

【 0 1 1 7 】

詳細には、本実施形態の一例及びその各変形例においては、ストレージ装置 1 0 にディスク 1 3 の種類や最適な I O タイムアウト時間の情報を予め記憶させておき、サーバ 2 がストレージ装置 1 0 を認識した際にその情報がサーバ 2 に渡される。そして、マルチパスドライバ部 3 のタイムアウト時間取得部 6 は、I O 発行時に、ストレージ装置 1 0 から受け取ったタイムアウトテーブル 1 4 に基づいて、最適な I O タイムアウト時間を取得する。これにより、マルチパスドライバ部 3 は、ストレージ装置 1 0 のディスク 1 3 の種別及び残りパス数を考慮して最適な I O タイムアウト時間を設定することができる。

40

【 0 1 1 8 】

さらに、本実施形態の一例及びその各変形例においては、タイムアウトテーブル 4 , 1 4 に C M 係数 4 3 , 1 4 3 を持っているので、ストレージ装置 1 0 の C M の性能 (担当 C M か否か) を考慮して適切なタイムアウト時間を設定することができる。

なおさらに、本実施形態の一例及びその各変形例においては、ストレージ装置 1 0 のタイムアウト時間変更部 1 5 が高負荷や故障等を検出して、タイムアウトテーブル 1 4 のタイムアウト時間を長めに設定する。これにより、I O 応答時間を短縮することが可能とな

50

る。

【 0 1 1 9 】

その際、タイムアウト時間変更部 1 5 は *s e n s e* を用いてサーバ 2 にタイムアウト値の変更を通知するため、ストレージ装置 1 0 側でのタイムアウト値の更新時にネットワーク負荷が上昇するのを抑えることができる。

(E) その他

そして、開示の技術は上述した実施形態に限定されるものではなく、本実施形態の趣旨を逸脱しない範囲で種々変形して実施することができる。

【 0 1 2 0 】

例えば、上記の実施形態及びその各変形例においては、*i n q u i r y* コマンドを使用して、ストレージ装置 1 0 , 2 0 からタイムアウト対応 / 非対応の状況やタイムアウトテーブル 1 4 を取得している。しかし、ほかのコマンドを使用してこれらの情報を取得してもよい。

10

或いは、上記の実施形態及びその各変形例においては、C M が担当 C M、非担当 C M かに応じて C M 係数を設定していたが、パスが、サーバ 2 からみて担当パスか非担当パスかに応じて「パス係数」を設定してもよい。

【 0 1 2 1 】

又、上記の実施形態及びその各変形例においては、ストレージ装置 1 0 のタイムアウト時間変更部 1 5 がサーバ 2 に対して *s e n s e* を用いてタイムアウトテーブル 1 4 が更新された旨を通知していた。しかし、ほかの手法を用いて更新の通知を行なってもよい。又、ストレージ装置 1 0 がタイムアウトテーブル 1 4 を更新すると直ちに、ストレージ装置 1 0 はサーバ 2 にタイムアウトテーブル 1 4 を送信してもよい。

20

(F) 付記

以上の実施形態に関し、更に以下の付記を開示する。

【 0 1 2 2 】

(付記 1)

ストレージ装置と複数の経路を介して接続され、

前記複数の経路の各経路のタイムアウト時間に関する管理情報を記憶する記憶部と、

前記ストレージ装置に対する入出力の発行先の経路及び他の経路について、前記記憶部から該経路に応じたタイムアウト時間を取得する取得部と、

30

をそなえることを特徴とする情報処理装置。

【 0 1 2 3 】

(付記 2)

前記管理情報を収集し、前記記憶部に記憶する収集部と、

前記取得部によって取得されたタイムアウト時間を用いて前記ストレージ装置に対して入出力を発行する発行部と、をそなえることを特徴とする付記 1 記載の情報処理装置。

(付記 3)

前記収集部は、前記複数の経路に対する複数の入出力タイムアウト時間を前記管理情報として前記ストレージ装置から収集し、

前記取得部は、前記複数の経路のうちの第 1 の経路に対する第 1 の入出力タイムアウト時間を取得し、

40

前記第 1 の経路への前記入出力でタイムアウトエラーが発生した場合に、前記複数の経路のうちの前記第 1 の経路とは異なる第 2 の経路に対する、前記第 1 の入出力タイムアウト時間よりも長い第 2 の入出力タイムアウト時間を取得することを特徴とする付記 2 記載の情報処理装置。

【 0 1 2 4 】

(付記 4)

前記第 1 の入出力タイムアウト時間は前記複数の経路のうちの最終経路以外の経路に対するタイムアウト時間であり、

前記第 2 の入出力タイムアウト時間は前記複数の経路のうちの最終経路に対するタイム

50

アウト時間であることを特徴とする付記 3 記載の情報処理装置。

【 0 1 2 5 】

(付記 5)

前記取得部は、前記第 1 の経路への前記入出力でタイムアウトエラーが発生した場合に、前記第 2 の経路に前記入出力を発行する前に、前記複数の経路のうちの前記第 1 及び第 2 の経路とは異なる第 3 の経路に対する、前記第 1 の入出力タイムアウト時間よりも長く前記第 2 の入出力タイムアウト時間よりも短い第 3 の入出力タイムアウト時間を取得することを特徴とする付記 4 記載の情報処理装置。

【 0 1 2 6 】

(付記 6)

前記収集部は、前記ストレージ装置において前記複数の入出力タイムアウト時間が変更されると、変更後の前記複数の入出力タイムアウト時間を前記管理情報として収集することを特徴とする付記 2 ～ 5 のいずれか 1 項に記載の情報処理装置。

(付記 7)

情報処理装置と複数の経路を介して接続され、
前記複数の経路に対する複数の入出力タイムアウト時間を格納している格納部と、
前記情報処理装置から前記複数の入出力タイムアウト時間を要求されると、前記情報処理装置に前記複数の入出力タイムアウト時間を送信する送信部と、
をそなえることを特徴とするストレージ装置。

【 0 1 2 7 】

(付記 8)

前記複数の経路の状況に応じて前記複数の入出力タイムアウト時間を変更する変更部をさらにそなえ、

前記送信部は、前記情報処理装置から入出力を受けると、前記変更部によって変更された前記複数の入出力タイムアウト時間を前記情報処理装置に送信することを特徴とする付記 7 記載のストレージ装置。

【 0 1 2 8 】

(付記 9)

情報処理装置と、
前記情報処理装置と複数の経路を介して接続されたストレージ装置と、
前記複数の経路の各経路のタイムアウト時間に関する管理情報を記憶する記憶部と、
前記情報処理装置から前記ストレージ装置に対する入出力の発行先の経路及び他の経路について、前記記憶部から該経路に応じたタイムアウト時間を取得する取得部と、
をそなえることを特徴とする情報処理システム。

【 0 1 2 9 】

(付記 1 0)

前記管理情報を収集し、前記記憶部に記憶する収集部と、
前記取得部によって取得されたタイムアウト時間を用いて前記情報処理装置から前記ストレージ装置に対して入出力を発行する発行部と、をそなえることを特徴とする付記 9 記載の情報処理システム。

【 0 1 3 0 】

(付記 1 1)

前記収集部は、前記複数の経路に対する複数の入出力タイムアウト時間を前記管理情報として前記ストレージ装置から収集し、

前記取得部は、前記複数の経路のうちの第 1 の経路に対する第 1 の入出力タイムアウト時間を取得し、

前記第 1 の経路への前記入出力でタイムアウトエラーが発生した場合に、前記複数の経路のうちの前記第 1 の経路とは異なる第 2 の経路に対する、前記第 1 の入出力タイムアウト時間よりも長い第 2 の入出力タイムアウト時間を取得することを特徴とする付記 1 0 記載の情報処理システム。

10

20

30

40

50

【 0 1 3 1 】

(付 記 1 2)

前記第 1 の入出力タイムアウト時間は前記複数の経路のうちの最終経路以外の経路に対するタイムアウト時間であり、

前記第 2 の入出力タイムアウト時間は前記複数の経路のうちの最終経路に対するタイムアウト時間であることを特徴とする付記 1 1 記載の情報処理システム。

【 0 1 3 2 】

(付 記 1 3)

前記取得部は、前記第 1 の経路への前記入出力でタイムアウトエラーが発生した場合に、前記第 2 の経路に前記入出力を発行する前に、前記複数の経路のうちの前記第 1 及び第 2 の経路とは異なる第 3 の経路に対する、前記第 1 の入出力タイムアウト時間よりも長く前記第 2 の入出力タイムアウト時間よりも短い第 3 の入出力タイムアウト時間を取得することを特徴とする付記 1 2 記載の情報処理システム。

10

【 0 1 3 3 】

(付 記 1 4)

前記収集部は、前記ストレージ装置において前記複数の入出力タイムアウト時間が変更されると、変更後の前記複数の入出力タイムアウト時間を前記管理情報として収集することを特徴とする付記 1 0 ~ 1 3 のいずれか 1 項に記載の情報処理システム。

(付 記 1 5)

情報処理装置に複数の経路を介して接続されたストレージ装置における入出力方法であって、

20

前記複数の経路の各経路のタイムアウト時間に関する管理情報を記憶し、

前記ストレージ装置に対する入出力の発行先の経路及び他の経路について、前記記憶部から該経路に応じたタイムアウト時間を取得することを特徴とする入出力方法。

【 0 1 3 4 】

(付 記 1 6)

情報処理装置に複数の経路を介して接続されたストレージ装置における入出力プログラムであって、

前記複数の経路の各経路のタイムアウト時間に関する管理情報を記憶し、

前記ストレージ装置に対する入出力の発行先の経路及び他の経路について、前記記憶部から該経路に応じたタイムアウト時間を取得する
処理を前記情報処理装置に実行させることを特徴とする入出力プログラム。

30

【 符号の説明 】

【 0 1 3 5 】

1 ストレージシステム (情報処理システム)

1 0 ストレージ装置

1 1 - 1 , 1 1 - 2 (1 1) C M (送信部)

1 2 - 1 ~ 1 2 - 4 (1 2) C A

1 3 - 1 ~ 1 3 - 4 (1 3) ディスク

1 4 タイムアウトテーブル (格納部)

40

1 5 タイムアウト時間変更部 (変更部)

1 6 タイムアウト上限テーブル

2 サーバ (情報処理装置)

2 0 ストレージ装置

3 マルチパスドライバ部

3 0 - 1 ~ 3 0 - 6 (3 0) パス (経路)

4 サーバ側タイムアウトテーブル (記憶部)

5 タイムアウトテーブル収集部 (収集部)

5 0 L U N 管理テーブル

6 タイムアウト時間取得部 (取得部)

50

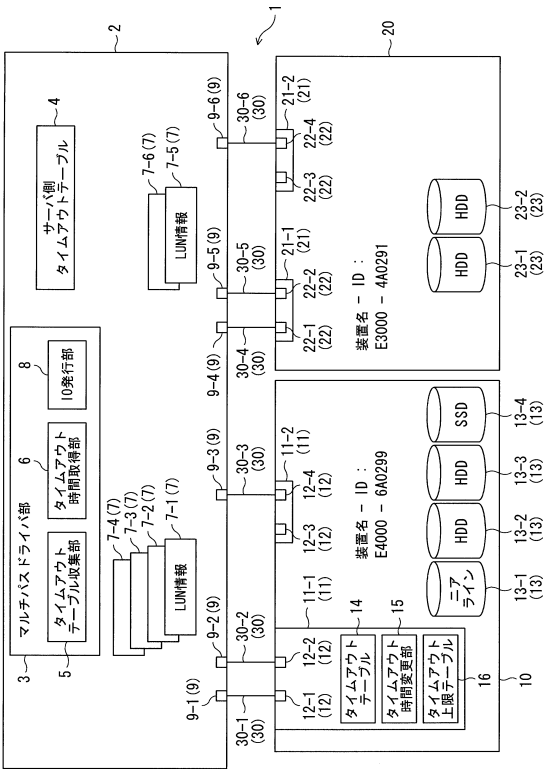
7

8

LUN情報

I O発行部（発行部）

【図 1】



【図 2】

41	42	43	44	45	46	47
装置識別情報	ディスク種別	CM係数	先頭バスのタイムアウト時間	中間バスのタイムアウト時間	最終バスのタイムアウト時間	上限
デフォルト	デフォルト	1.0	40	40	40	40

【図 3】

41	42	43	44	45	46	47
装置識別情報	ディスク種別	CM係数	先頭バスのタイムアウト時間	中間バスのタイムアウト時間	最終バスのタイムアウト時間	上限
デフォルト	デフォルト	1.0	40	40	40	40
E4000 - 6A0299	NEARLINE	1.1	35	35	40	40
E4000 - 6A0299	HDD10000	1.2	30	35	40	40
E4000 - 6A0299	HDD15000	1.3	20	25	30	35
E4000 - 6A0299	SSD	1.4	10	15	20	25

【図 4】

41	42	43	44	45	46	47
装置識別情報	ディスク種別	CM係数	先頭バスのタイムアウト時間	中間バスのタイムアウト時間	最終バスのタイムアウト時間	上限
デフォルト	デフォルト	1.0	40	40	40	40
E4000 - 6A0299	NEARLINE	1.1	35	35	40	40
E4000 - 6A0299	HDD10000	1.2	30	35	40	40
E4000 - 6A0299	HDD15000	1.3	20	25	30	35
E4000 - 6A0299	SSD	1.4	10	15	20	25
DX900 - 2D3334	NEARLINE	1.0	30	35	35	35
DX900 - 2D3334	HDD15000	1.0	15	20	35	35

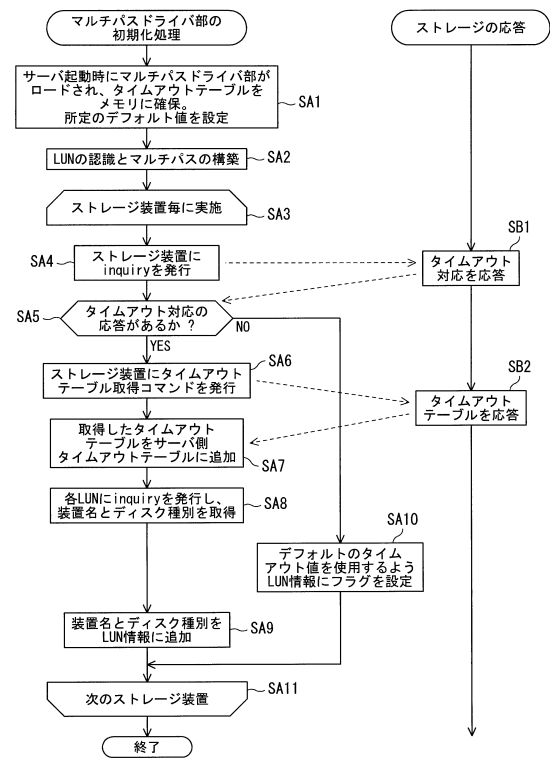
【図 5】

ディスク 種別	CM 係数	先頭バスの タイムアウト 時間	中間バスの タイムアウト 時間	最終バスの タイムアウト 時間	上限
NEARLINE	1.1	35	35	40	40
HDD10000	1.2	30	35	40	40
HDD15000	1.3	20	25	30	35
SSD	1.4	10	15	20	25

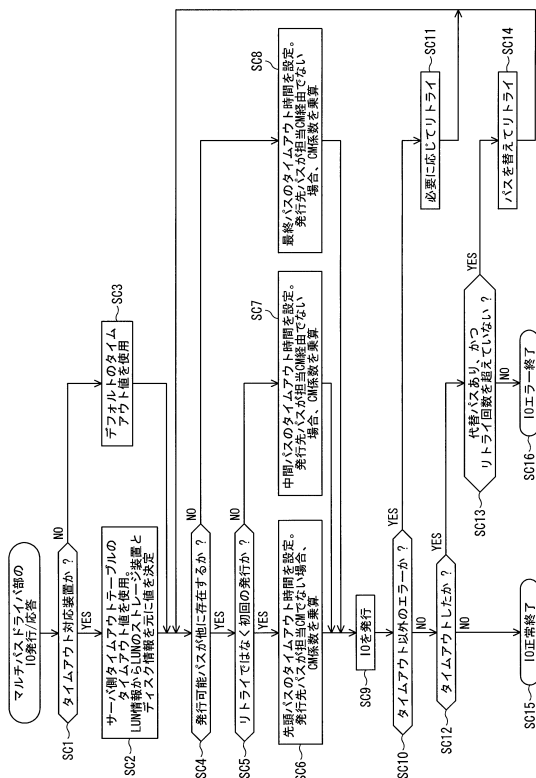
【図 6】

ディスク 種別	...	上限
NEARLINE	...	40
HDD10000	...	40
HDD15000	...	40
SSD	...	35

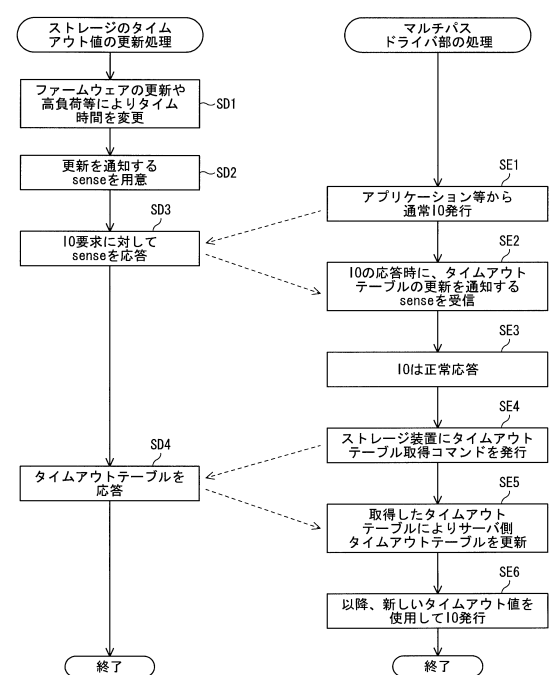
【図 7】



【図 8】



【図 9】



【図 10】

装置識別情報	ディスク 種別	CM 係数	代替バス有りの場合の タイムアウト 時間	最終バスの タイムアウト 時間	上限
デフォルト	デフォルト	1.0	40	40	40
E4000 - 6A0299	NEARLINE	1.1	35	40	40
E4000 - 6A0299	HDD10000	1.2	30	40	40
E4000 - 6A0299	HDD15000	1.3	20	30	35
E4000 - 6A0299	SSD	1.4	10	20	25

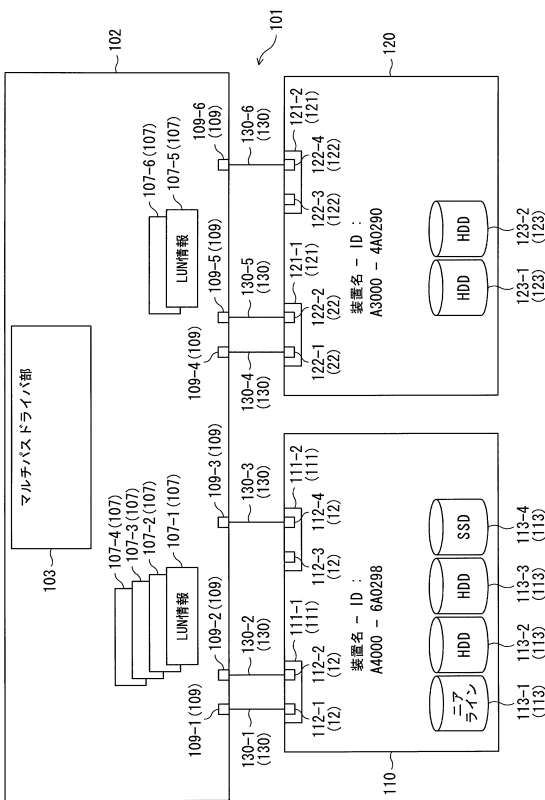
【図 11】

装置識別情報	ディスク 種別	CM 係数	先頭バスの タイムアウト 時間	2バス目の タイムアウト 時間	...	最終バスの タイムアウト 時間	上限
デフォルト	デフォルト	1.0	40	40		40	40
E4000 - 6A0299	NEARLINE	1.1	35	35		40	40
E4000 - 6A0299	HDD10000	1.2	30	35		40	40
E4000 - 6A0299	HDD15000	1.3	20	25		30	35
E4000 - 6A0299	SSD	1.4	10	15		20	25

【図 12】

装置識別情報	ディスク種別	開始LBA	終了LBA
E4000 - 6A0299	HDD10000	0	9999
E4000 - 6A0299	HDD15000	10000	19999

【図 13】



フロントページの続き

(56)参考文献 特開2012-073983(JP,A)
特開平08-328759(JP,A)
特開2008-084066(JP,A)
米国特許出願公開第2009/0313401(US,A1)
特開2009-223702(JP,A)
特開2005-318073(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F3/06-3/08
11/07、11/28-11/34
11/36、13/00-13/14