

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4495248号
(P4495248)

(45) 発行日 平成22年6月30日 (2010. 6. 30)

(24) 登録日 平成22年4月16日 (2010. 4. 16)

(51) Int. Cl. F 1
G 0 6 F 9/46 (2006. 01) G O 6 F 9/46 3 5 0
G 0 6 F 11/30 (2006. 01) G O 6 F 11/30 3 2 0 C

請求項の数 6 (全 10 頁)

(21) 出願番号	特願2009-507374 (P2009-507374)	(73) 特許権者	000005223 富士通株式会社 神奈川県川崎市中原区上小田中4丁目1番1号
(86) (22) 出願日	平成19年3月29日 (2007. 3. 29)	(74) 代理人	100101856 弁理士 赤澤 日出夫
(86) 国際出願番号	PCT/JP2007/056957	(72) 発明者	村上 大士 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
(87) 国際公開番号	W02008/120383	審査官	久保 正典
(87) 国際公開日	平成20年10月9日 (2008. 10. 9)		
審査請求日	平成21年7月23日 (2009. 7. 23)		

最終頁に続く

(54) 【発明の名称】 情報処理装置、障害処理方法

(57) 【特許請求の範囲】

【請求項 1】

ハードウェア資源を分割してそれぞれを異なるパーティションで利用する物理分割を行うことができる情報処理装置であって、

複数のパーティションに割り当てられた少なくとも一つのハードウェア資源に接続され、前記ハードウェア資源の管理を行うハードウェア管理部であって、前記複数のパーティションのいずれかで障害が発生したことを検知した場合、前記障害が発生したパーティションを障害発生パーティションとし、前記障害の発生を報知するハードウェア管理部と、

前記パーティション毎に備えられたパーティション管理部であって、物理分割の状態であり、且つ前記ハードウェア管理部による報知に基づいて前記障害の発生を認識し、且つ自己のパーティションが前記障害発生パーティションと異なり、且つ前記障害に関する情報である障害情報の取得が必要であると判定した場合、前記障害情報の取得を行うパーティション管理部と

を備える情報処理装置。

【請求項 2】

請求項 1 に記載の情報処理装置において、

前記パーティション管理部は、取得した前記障害情報の解析を行い、該解析の結果を障害解析情報として出力し、

更に、前記パーティション管理部により出力された障害解析情報を記録する障害情報記録部を備える情報処理装置。

【請求項 3】

請求項 1 または請求項 2 に記載の情報処理装置において、

前記パーティション管理部は、前記障害発生パーティションにおいて前記障害の情報である障害情報の取得が失敗したと認識した場合、前記障害情報の取得が必要であると判定する情報処理装置。

【請求項 4】

請求項 1 乃至請求項 3 のいずれかに記載の情報処理装置において、

前記ハードウェア管理部は、前記パーティション毎に前記障害情報を保持し、

前記パーティション管理部は、物理分割の状態であり、且つ前記ハードウェア管理部による報知に基づいて前記障害の発生を認識し、且つ自己のパーティションが前記障害発生パーティションと異なり、且つ前記障害情報の取得が必要であると判定した場合、前記ハードウェア管理部から前記障害発生パーティションの前記障害情報を取得する情報処理装置。

10

【請求項 5】

請求項 4 に記載の情報処理装置において、

前記ハードウェア管理部は、前記障害発生パーティションの障害情報を他のパーティションから隠蔽しておき、

前記パーティション管理部は、物理分割の状態であり、且つ前記ハードウェア管理部による報知に基づいて前記障害の発生を認識し、且つ自己のパーティションが前記障害発生パーティションと異なり、且つ前記障害に関する情報である障害情報の取得が必要であると判定した場合、前記ハードウェア管理部による前記隠蔽を解除させ、前記障害情報の取得を行う情報処理装置。

20

【請求項 6】

ハードウェア資源を分割してそれぞれを異なるパーティションで利用することができる情報処理装置の障害処理方法であって、

複数のパーティションに割り当てられた少なくとも 1 つのハードウェア資源の管理を行い、

前記複数のパーティションのいずれかで障害が発生したことを検知した場合、前記障害が発生したパーティションを障害発生パーティションとし、前記障害の発生の報知を行い、

30

前記複数のパーティションの一つである特定パーティションについて、物理分割の状態であり、且つ前記報知に基づいて前記障害の発生を認識し、且つ前記特定パーティションが前記障害発生パーティションと異なり、且つ前記障害に関する情報である障害情報の取得が必要であると判定した場合、前記障害情報の取得を行う

障害処理方法。

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、ハードウェア資源を分割してそれぞれを異なるパーティションで利用する物理分割を行うことができる情報処理装置、障害処理方法に関するものである。

40

【背景技術】**【0002】**

基幹システムで運用されるサーバには、高い可用性や柔軟なリソース運用が要求される。この要求に対し、ASIC (Application Specific Integrated Circuit) および Firmware (以下 F/W) の機能として、従来は Hardware (以下 H/W) (例えば、プロセッサやメモリ等) に固定であったリソース配分を n 個のパーティションに分割し、それぞれを別の OS (Operating System) で使用する物理分割機能が存在する。この機能により、H/W に制限されない柔軟なリソース運用が可能となる。

【0003】

物理分割機能を使用している場合も使用しない場合と同様に、正確な障害情報の解析と

50

通知の機能が必要である。そのために、物理分割機能を使用しない場合と同等の障害検知機能の他に、障害を管理する機能が実装される。物理分割機能使用時の障害を管理する方法として、大きく分けて以下の3つの方法がある。

【0004】

(管理方法1) H/Wリソースの配分、および情報の振り分け機能を全てASICに実装する。

(管理方法2) H/Wリソースの配分、および情報の振り分け機能をASICおよびF/Wそれぞれの特性に合わせて実装し、連携させる。

(管理方法3) H/Wリソースの配分、および情報の振り分け機能を全てF/Wに実装する。(=仮想化)

10

【0005】

ここでは、信頼性、実装、コスト、および他機能との融合性を考慮し、H/Wの障害によるパーティションへの影響が少なく、より柔軟な機能拡張が可能な、管理方法2を対象とする。このとき、ASICはH/Wを管理し、H/Wリソースを分割し、障害情報を含むリソース管理情報をF/Wへ提供する。F/Wはそのリソース管理情報を必要に応じて解析し、OSなどの上位のレイヤにパーティションの障害情報を提供する。

【0006】

なお、本発明の関連ある従来技術として、第1及び第2のデータ処理装置群にそれぞれ対応する障害処理装置が、第1及び第2のデータ処理装置群からの障害情報を切り替えて入力する障害処理システムがある(例えば、特許文献1参照)。

20

【特許文献1】特開平1-50135号公報

【発明の開示】

【発明が解決しようとする課題】

【0007】

サーバの運用時に障害が発生した場合、ほとんどの場合は障害情報が格納され、障害対策時間を短縮する役割を果たす。しかし、多重障害の発生や、障害報告パスの障害、設計時に予期し得なかったASIC、F/W障害等により障害情報が格納されない場合も存在する。

【0008】

障害情報が格納されない場合とは、F/Wが利用するメモリのアドレス障害や、プロセス情報を退避する領域の固定障害等の場合である。

30

【0009】

本発明は上述した問題点を解決するためになされたものであり、物理分割機能の使用時に障害情報が取得できる確率を向上させる情報処理装置、障害処理方法を提供することを目的とする。

【課題を解決するための手段】

【0010】

上述した課題を解決するため、本発明は、ハードウェア資源を分割してそれぞれを異なるパーティションで利用する物理分割を行うことができる情報処理装置であって、複数のパーティションに割り当てられた少なくとも1つのハードウェア資源に接続され、前記ハードウェア資源の管理を行うハードウェア管理部であって、前記複数のパーティションのいずれかで障害が発生したことを検知した場合、前記障害が発生したパーティションを障害発生パーティションとし、前記障害の発生を報知するハードウェア管理部と、前記パーティション毎に備えられたパーティション管理部であって、物理分割の状態であり、且つ前記ハードウェア管理部による報知に基づいて前記障害の発生を認識し、且つ自己のパーティションが前記障害発生パーティションと異なり、且つ前記障害に関する情報である障害情報の取得が必要であると判定した場合、前記障害情報の取得を行うパーティション管理部とを備える。

40

【0011】

また、本発明は、ハードウェア資源を分割してそれぞれを異なるパーティションで利用

50

することができる情報処理装置の障害処理方法であって、複数のパーティションに割り当てられた少なくとも1つのハードウェア資源の管理を行い、前記複数のパーティションのいずれかで障害が発生したことを検知した場合、前記障害が発生したパーティションを障害発生パーティションとし、前記障害の発生を報知するハードウェア管理ステップと、パーティション毎に実行され、物理分割の状態であり、且つ前記ハードウェア管理ステップによる報知に基づいて前記障害の発生を認識し、且つ自己のパーティションが前記障害発生パーティションと異なり、且つ前記障害に関する情報である障害情報の取得が必要であると判定した場合、前記障害情報の取得を行うパーティション管理ステップとを実行する。

【図面の簡単な説明】

10

【0012】

【図1】本実施の形態に係るサーバの構成の一例を示すブロック図である。

【図2】本実施の形態に係るサーバの障害発生時の第1の動作を示すフローチャートである。

【図3】本実施の形態に係るサーバの障害発生時の第2の動作を示すフローチャートである。

【図4】本実施の形態に係るサーバにおける物理分割及びパーティションの構成の一例を示すブロック図である。

【発明を実施するための最良の形態】

【0013】

20

以下、本発明の実施の形態について図面を参照しつつ説明する。

【0014】

本実施の形態においては、本発明の情報処理装置を適用したサーバについて説明する。

【0015】

まず、本実施の形態に係るサーバの構成について説明する。

【0016】

図1は、本実施の形態に係るサーバの構成の一例を示すブロック図である。このサーバは、筐体管理装置11、H/W12a, 12b, 12c, 12d、ASIC13a、F/W14a, 14b、OS15a, 15bを備える。また、このサーバには、パーティション16a, 16bが設定される。パーティション16aには、OS15a, F/W14a, ASIC13a, H/W12a, 12bが割り当てられ、パーティション16bには、OS15b, F/W14b, ASIC13a, H/W12c, 12dが割り当てられる。F/W14a, 14b、OS15a, 15bは、サーバのCPU上で動作するソフトウェアであり、F/W及びOSは、パーティション毎にそれぞれ1つずつ動作する。

30

【0017】

物理分割により分割されたH/W12a, 12b, 12c, 12dについて、設定された情報及び取得した障害情報は管理するASIC13aの記憶領域に保存されており、特別なプログラムやハードウェアを仲介せずにF/W14a, 14bから取得することが可能である。

【0018】

40

ASIC13aは、通信の制御を行うチップであり、例えば、CPUとインターフェースとクロスバとを接続及び制御するNorth-Bridge、メモリとインターフェースとクロスバとを接続及び制御するMLDS (Memory And Local Data Switch)、I/Oデバイスとバスとクロスバとを接続及び制御するSouth-Bridge等である。

【0019】

筐体管理装置11は、F/W14a, 14bから受信したサーバの障害情報の格納を行い、外部からの要求により外部へ障害情報を出力することができる。

【0020】

次に、本実施の形態に係るサーバにおける障害発生時の動作の概要について説明する。

【0021】

50

図1におけるS1～S3は、サーバにおける障害発生時の動作の概要を示す。

【0022】

(S1)ASIC13aは、自らの管理下のH/W12a, 12bで障害が起きた場合、その内容を全てのパーティションへ通知する。

【0023】

(S2)F/W14a, 14bは、動作モードを判定し、物理分割時には障害報告にタイムアウト監視する。

【0024】

(S3)F/W14a, 14bは、タイムアウトが発生した場合、ASIC13aの反対側のパーティションの障害情報をASIC13aから採取して解析し、筐体管理装置11へ通知する。この図の例において、F/W14a側の多重障害によりF/W14aの障害処理が中断し、F/W14bは、タイムアウトにより、パーティション16aの障害情報をASIC13aから採取して解析し、筐体管理装置11へ通知する。

10

【0025】

次に、本実施の形態に係るサーバにおける障害発生時の動作の詳細について説明する。

【0026】

ASIC13aの記憶領域は、物理分割モードであることを示す動作モード情報(動作情報)、障害が発生したか否かを示す障害発生情報、障害が発生した物理分割位置を示す物理分割位置情報(位置情報)、障害が発生した部位(物理分割の位置)を示す障害部位情報、障害の詳細を示す障害詳細情報、異なるパーティションからのアクセスに対する障害情報(障害部位情報、障害詳細情報)の隠蔽を解除するか否かを設定する解除情報、障害処理が正常終了したか否かを示す正常終了情報を有する。解除情報、正常終了情報、障害情報は、パーティション毎に保持される。

20

【0027】

また、ASIC13aは、誤動作防止、信頼性及びセキュリティ確保のために障害情報の隠蔽機能を有する。通常は解除情報のフラグがリセットされた状態であり、この状態でASIC13aに保持された特定のパーティションの障害情報に対して、他のパーティションからのアクセスがあった場合、隠蔽機能は障害情報の代わりに所定のデータを返す。また、特定のパーティションの解除情報のフラグがセットされると隠蔽が解除され、そのパーティションの障害情報は他のパーティションから読み出すことができる。

30

【0028】

図2は、本実施の形態に係るサーバの障害発生時の第1の動作を示すフローチャートである。図3は、本実施の形態に係るサーバの障害発生時の第2の動作を示すフローチャートである。まず、ASIC13aが管理下のパーティションにおける障害を検知する(S11)。図1の例において、パーティション16aに所属するH/W12bで障害が発生したとする。H/W12bを管理するASIC13aは、検知した障害情報を保持すると共に、自己に割り当てられた全てのパーティション16a, 16bへ障害発生を通知(ブロードキャスト)する(S12)。図1の例において、ASIC13aは、障害発生情報のフラグをセットすると共に、物理分割位置情報に障害の発生した物理分割位置(パーティション16aのH/W12b)を記録する。

40

【0029】

次に、F/W14a, 14bは、障害割り込みを受け付ける(S13)。ここで、F/W14a, 14bは、ASIC13aの障害発生情報を監視しており、障害発生情報のフラグにより障害発生を認識する。次に、F/W14a, 14bは、ASIC13aの動作モード情報を参照し、物理分割モードか否かの判断を行う(S14)。物理分割モードでない場合(S14, N)、処理S16へ移行する。物理分割モードである場合(S14, Y)、F/W14a, 14bは、物理分割位置情報を参照して障害が発生したパーティション(障害発生パーティション)を認識し(S15)、自己のパーティションが障害発生パーティションであるか否かの判断を行う(S16)。図1の例において、F/W14a, 14bは、障害発生パーティションがパーティション16aであることを認識する。

50

【 0 0 3 0 】

障害発生パーティションである場合 (S 1 6 , Y)、F / W 1 4 a , 1 4 b は、A S I C 1 3 a から自己のパーティションの障害部位情報を収集し (S 1 7)、処理 S 3 1 へ移行する。図 1 の例において、障害発生パーティションに所属する F / W 1 4 a は、パーティション 1 6 a の障害部位情報を収集する。

【 0 0 3 1 】

障害の起きたパーティションでない場合 (S 1 6 , N)、F / W 1 4 a , 1 4 b は、A S I C 1 3 a における他のパーティションの正常終了情報のフラグのタイムアウト監視を行い、タイムアウトしたか否かの判断を行う (S 2 1)。図 1 の例において、障害発生パーティションに所属しない F / W 1 4 b は、このタイムアウト監視を行う。

10

【 0 0 3 2 】

タイムアウトしなかった場合 (S 2 1 , N)、即ち、所定の時間内に A S I C 1 3 a における障害発生パーティションの正常終了情報のフラグがセットされた場合、このフローは終了する。

【 0 0 3 3 】

タイムアウトした場合 (S 2 1 , Y)、F / W 1 4 a , 1 4 b は、障害発生パーティションの解除情報を設定する (S 2 3)。ここで、F / W 1 4 a , 1 4 b は、解除情報のフラグをセットすることにより、A S I C 1 3 a に隠蔽の解除の要求を行う。解除情報のフラグがセットされた A S I C 1 3 a は、障害発生パーティションの障害情報の隠蔽を解除する (S 2 4)。次に、F / W 1 4 a , 1 4 b は、物理分割位置情報と所定の算出式により、障害情報のアドレスを算出し (S 2 5)、A S I C 1 3 a から障害部位情報を収集する (S 2 6)。図 1 の例において、障害発生パーティションに所属しない F / W 1 4 b は、障害発生パーティションの正常終了情報のタイムアウト監視を行い、これがタイムアウトすると、パーティション 1 6 a の解除情報のフラグをセットし、パーティション 1 6 a の障害部位情報を収集する。

20

【 0 0 3 4 】

次に、F / W 1 4 a , 1 4 b は、障害詳細情報が必要であるか否かの判断を行う (S 3 1)。障害詳細情報が必要でない場合 (S 3 1 , N)、処理 S 3 3 へ移行する。障害詳細情報が必要である場合 (S 3 1 , Y)、F / W 1 4 a , 1 4 b は、取得した障害部位情報に対応する障害詳細情報を取得する (S 3 2)。次に、F / W 1 4 a , 1 4 b は、取得した障害情報 (障害部位情報、障害詳細情報) のマージを行い (S 3 3)、障害情報を解析することにより、その障害の影響を受けるパーティションと影響を与える障害情報とを特定し (S 3 4)、障害解析が正常終了したか否かの判断を行う (S 3 5)。

30

【 0 0 3 5 】

正常終了しない場合 (S 3 5 , N)、このフローは終了する。正常終了しない場合とは、多重障害により障害情報の収集ができなかった場合等である。図 1 の例において、F / W 1 4 a は、多重障害により障害処理が中断し、パーティション 1 6 a の正常終了情報のフラグはセットされない。

【 0 0 3 6 】

正常終了した場合 (S 3 5 , Y)、F / W 1 4 a , 1 4 b は、A S I C 1 3 a の影響範囲のうち自己の所属するパーティション以外へ障害処理の終了報告を行う (S 4 1)。ここで、F / W 1 4 a , 1 4 b は、A S I C 1 3 a における自己のパーティションの正常終了情報のフラグをセットし、他のパーティションの F / W は正常終了情報のフラグを監視することにより正常終了を認識する。図 1 の例において、F / W 1 4 b は、パーティション 1 6 b の正常終了情報のフラグをセットする。

40

【 0 0 3 7 】

次に、F / W 1 4 a , 1 4 b は、自己のパーティションが障害発生パーティションであるか否かの判断を行う (S 4 2)。障害発生パーティションである場合 (S 4 2 , Y)、F / W 1 4 a , 1 4 b は、障害解析結果を筐体管理装置 1 1 及び自己のパーティションの OS へ送信し (S 4 3)、このフローは終了する。障害発生パーティションでない場合 (

50

S 4 2 , N)、 F / W 1 4 a , 1 4 b は、障害解析結果を筐体管理装置 1 1 へ送信し (S 4 4)、このフローは終了する。通常、筐体管理装置 1 1 へ送信される障害解析結果は、障害発生パーティションからの報告である。処理 S 4 3 における障害解析結果は、障害発生パーティション以外からの報告であるが、障害発生パーティションからの報告として表される。

【 0 0 3 8 】

なお、本実施の形態において、 A S I C 1 3 a は、 2 つのパーティションについて H / W の管理を行うとしたが、 3 つ以上のパーティションについて H / W の管理を行う場合でも本発明を適用することができる。

【 0 0 3 9 】

次に、本実施の形態の効果について、物理分割及びパーティションの具体例を用いて説明する。

【 0 0 4 0 】

図 4 は、本実施の形態に係るサーバにおける物理分割及びパーティションの構成の一例を示すブロック図である。この図のサーバは、ハードウェアとして、システムボード (S B) 3 0 , 3 1 , 3 2 , 3 3 と、 I O ユニット (I O U) 4 0 , 4 1 , 4 2 , 4 3 を備える。また、パーティション (P) 5 0 , 5 1 , 5 2 , 5 3 が設定される。 S B 3 0 , 3 1 , 3 2 , 3 3 のそれぞれにおいて、上述した A S I C が 1 つずつ備えられる。また、 P 5 0 , 5 1 , 5 2 , 5 3 のそれぞれにおいて、 F / W 及び O S が 1 つずつ動作する。

【 0 0 4 1 】

S B 3 0 , 3 3 は、物理分割されず、それぞれ 1 つの S B (P S B) 3 0 0 , 3 3 0 として利用される。 S B 3 1 , 3 2 は、物理分割され、それぞれ 2 つの分割された S B (X S B) 3 1 0 及び 3 1 1、 3 2 0 及び 3 2 1 として利用される。

【 0 0 4 2 】

I O U 4 1 , 4 3 は、物理分割されず、それぞれ 1 つの I O U 4 1 0 , 4 3 0 として利用される。 I O U 4 0 , 4 2 は、物理分割され、それぞれ 2 つの分割された I O U (L I O U) 4 0 0 及び 4 0 1、 4 2 0 及び 4 2 1 として利用される。

【 0 0 4 3 】

P 5 0 には、 P S B 3 0 0、 L I O U 4 0 0 が割り当てられている。 P 5 1 には、 X S B 3 1 0、 L I O U 4 0 1 が割り当てられている。 P 5 2 には、 X S B 3 1 1、 I O U 4 1 0 が割り当てられている。 P 5 3 には、 X S B 3 2 0 , 3 2 1、 P S B 3 3 0、 L I O U 4 2 0 , 4 2 1、 I O U 4 3 0 が割り当てられている。

【 0 0 4 4 】

S B 3 1 のいずれかの X S B において、 X S B 内を影響範囲とする障害が発生した場合、何らかの原因で障害処理が継続出来なければ、もう片方の X S B が属するパーティションで障害処理が続行される。このように、ハードウェアが自らのパーティション以外の障害を処理する場合、本発明の効果により障害情報の取得及び解析を行うことができる。

【 0 0 4 5 】

また、 S B 3 2 のいずれかの X S B で障害が発生した場合、この X S B が属している P 5 3 はパーティションダウンしてしまう。このように、物理分割しているハードウェアが同一のパーティションに属している場合、本発明の効果が得られない。

【 0 0 4 6 】

本実施の形態によれば、従来、多重障害等により障害情報を採取できなかった場合でも、障害情報を採取することができ、障害対策時間を短縮することができる。

【 0 0 4 7 】

なお、ハードウェア管理部は、実施の形態における A S I C に対応する。また、パーティション管理部は、実施の形態における F / W に対応する。また、障害情報記録部は、実施の形態における筐体管理装置に対応する。ハードウェア管理ステップは、実施の形態における A S I C の処理に対応する。また、パーティション管理ステップは、実施の形態における F / W の処理に対応する。また、障害情報記録ステップは、実施の形態における筐

10

20

30

40

50

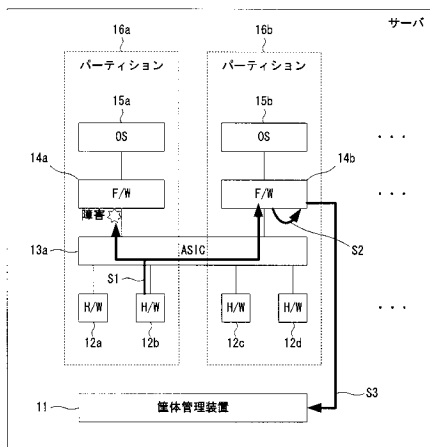
体管理装置の処理に対応する。

【産業上の利用可能性】

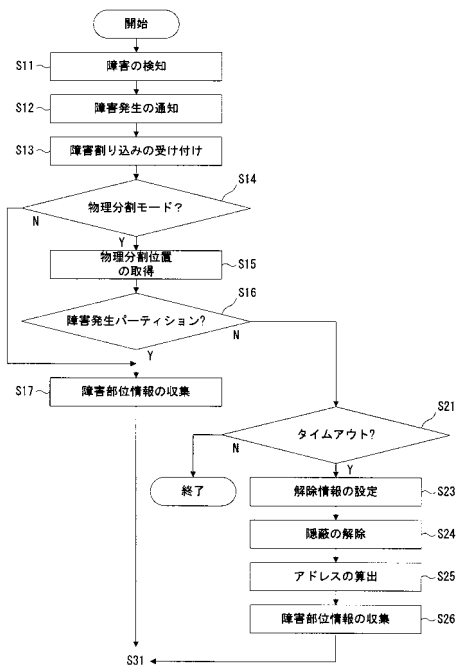
【0048】

以上説明したように、物理分割機能の使用時に障害情報が取得できる確率を向上させることができる。

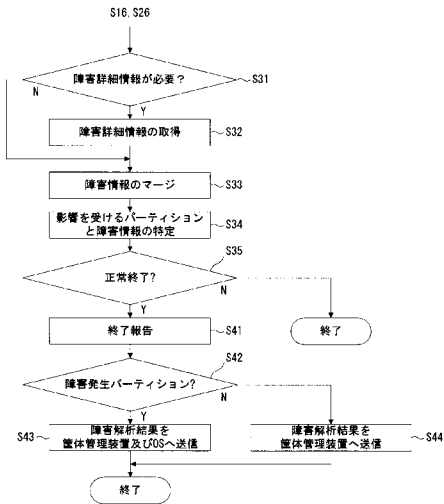
【図1】



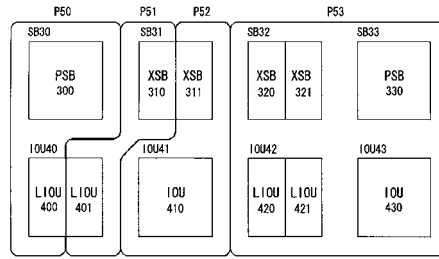
【図2】



【図3】



【図4】



フロントページの続き

- (56)参考文献 特開2004-342109(JP,A)
特開2005-122229(JP,A)
特開2002-229811(JP,A)
特開2004-213178(JP,A)
特開2004-62535(JP,A)
特開2003-76671(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F9/46-9/54
G06F11/28-11/36
G06F15/16-15/177