(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4779948号
(P4779948)
(45)発行日　平成23年9月28日(2011.9.28)　(24)登録日　平成23年7月15日(2011.7.15)

(51)Int.Cl.　F I
H04L 29/14 (2006.01)　H04L 13/00　311

請求項の数 5　(全 8 頁)

(21)出願番号　特願2006-319464 (P2006-319464)
(22)出願日　平成18年11月28日 (2006.11.28)
(65)公開番号　特開2008-135897 (P2008-135897A)
(43)公開日　平成20年6月12日 (2008.6.12)
審査請求日　平成21年2月16日 (2009.2.16)

(73)特許権者　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　100100310
弁理士　井上　学
(72)発明者　渡辺恭司
神奈川県秦野市堀山下１番地　株式会社日立製作所　エンタープライズサーバ事業部内
(72)発明者　軸屋孝之
神奈川県秦野市堀山下１番地　株式会社日立製作所　エンタープライズサーバ事業部内

最終頁に続く

(54)【発明の名称】サーバシステム

(57)【特許請求の範囲】
【請求項１】
複数のサーバモジュールと複数のネットワークスイッチモジュールと前記複数のサーバモジュールと前記複数のネットワークスイッチモジュールとを接続する内部ネットワークから成り、各サーバモジュールは外部ネットワークとの接続制御を行うネットワークコントローラを備え、各ネットワークスイッチモジュールは内部ネットワークと外部ネットワークでデータの交換を行うスイッチ部と両ネットワークと接続するための複数のポートを備えたサーバシステムにおいて、
各ネットワークスイッチモジュールは自ネットワークスイッチモジュールの状態を示すスイッチ状態信号を前記複数のサーバモジュールのネットワークコントローラに送出するプロセッサを備えたことを特徴とするサーバシステム。
【請求項２】
前記ネットワークスイッチモジュールは各ポートのリンク状態を検出する手段を備え、前記プロセッサは前記スイッチ状態信号として検出された各ポートのリンク状態を示す信号を対応するサーバモジュールのネットワークコントローラに送出することを特徴とする請求項１記載のサーバシステム。
【請求項３】
前記ネットワークスイッチモジュールはスイッチ部の障害を検出する手段を備え、前記プロセッサは前記スイッチ状態信号として検出された前記スイッチ部の障害を示す信号を全てのサーバモジュールのネットワークコントローラに送出することを特徴とする請求項

１記載のサーバシステム。
【請求項４】
各サーバモジュールは複数のネットワークコントローラを備え、内部ネットワークを多重構成とし、前記ネットワークスイッチモジュールからの前記スイッチ状態信号が障害を示した場合に、ネットワークパスを予備系ネットワークパスに切替えることを特徴とする請求項１乃至３記載のサーバシステム。
【請求項５】
前記ネットワークスイッチモジュールのプロセッサには管理インタフェースが接続され、前記プロセッサは外部ポート経由の指示により前記スイッチ状態信号の値を操作することを特徴とする請求項１乃至４記載のサーバシステム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、計算機とネットワークスイッチとが搭載され、それらを基板配線で接続して一体化したサーバシステムに係り、そのシステム内ネットワークの障害検知方法に関し、システム内ネットワーク障害発生時の多重化ネットワークパスの切り替え(フェイルオーバー)方法に関する。
【背景技術】
【０００２】
従来、例えばブレードサーバシステムなどの複数の計算機(サーバモジュール)とネットワークスイッチを搭載して一体化したシステムの内部ネットワーク接続は、一般的な外部ローカルエリアネットワーク(LAN)ケーブルによる接続とは異なり、ミッドプレーン上で内蔵ネットワークスイッチのポートとサーバモジュール上のネットワークコントローラの各ポート間を基板配線するので、ケーブルの抜けや半抜け状態や断線などの障害が発生しない事を前提としていて、ネットワークスイッチのポートリンク状態情報をネットワークコントローラが直接検出する仕組みをもたない。
【０００３】
そして、サーバモジュール上のネットワークコントローラは、内蔵ネットワークスイッチ経由で外部と接続しているため、外部ネットワークがダウンした際のリンク状態を認識することができない。
【０００４】
また、一般的にネットワークスイッチはユーザが多重化またはチーミング機能などによりバンド幅を広げることができるように、２個１組として用いられることが多く、ネットワーク多重化しているシステムの場合に系切り替えを行う際は、データ送受信用の信号線を通してポートリンクのアップ又はダウン状態を検知して系の切り替えを行う方法が一般的であるが、サーバモジュール間のネットワークスイッチにてポートリンクダウンした場合も、その障害をサーバモジュールへ伝えることができない。
【０００５】
そのためサーバモジュールは通信タイムアウトによって障害検出するか、データ信号にオートネゴシエーション失敗のようなリンクダウンを示すデータを載せてサーバモジュールに伝える必要があるが、いずれも複雑なソフトウェア処理が必要な上に検出するまでに時間がかかる。
【０００６】
【特許文献１】特開2000－269967
【非特許文献１】BladeServer Base Specification For Processor Blade Subsystems 2/3/2006 IBM and Intel
【発明の開示】
【発明が解決しようとする課題】
【０００７】
上述の従来のタイムアウト検出によるネットワーク障害認識方法では、ネットワークス

イッチ側で発生した障害を速やか検出できない問題があり、更に、ネットワークパスを多重化したシステムの場合では、ネットワーク障害をタイムアウト発生により認識した後、主系ネットワークパスから予備系に切り替える処理を行うことになり、予備系ネットワークに切り替わるまで時間が長くなる。また、ネットワークスイッチを介して接続する外部ネットワークに障害が発生した場合は、外部ネットワークのリンク状態を認識できないため、チーミングによる冗長化機能が使用できない。

【０００８】

本発明の目的は、ブレードサーバシステム内のネットワーク障害検知の可用性向上に関し、システムのネットワークスイッチのポートに障害が発生してもサーバのソフトウェアに極力意識させずネットワークスイッチの障害を速やかに検知させ、かつ多重に備えたネットワーク接続の場合では、複雑な処理や切り替え制御ソフトウェアの介入なしで速やかに交替が可能な方法を提供するという課題を解決しようとするものである。

【課題を解決するための手段】

【０００９】

本発明は、複数のサーバモジュールと複数のネットワークスイッチモジュールと前記複数のサーバモジュールと前記複数のネットワークスイッチモジュールとを接続する内部ネットワークから成り、各サーバモジュールは外部ネットワークとの接続制御を行うネットワークコントローラを備え、各ネットワークスイッチモジュールは内部ネットワークと外部ネットワークでデータの交換を行うスイッチ部と両ネットワークと接続するための複数のポートを備えたサーバシステムにおいて、各ネットワークスイッチモジュールは自ネットワークスイッチモジュールの状態を示すスイッチ状態信号を前記複数のサーバモジュールのネットワークコントローラに送出するプロセッサを備えたことを特徴とする。

【発明の効果】

【００１０】

本発明によれば、ネットワークスイッチのポートリンク状態情報およびネットワークスイッチ状態情報を計算機上のネットワークコントローラが直接検出できるので、OSやデバイスドライバはポーリングまたは割り込みによりネットワークスイッチの障害およびポートリンク状態を速やかに検知できるようになる。

【００１１】

従って、OSやデバイスドライバは該当ネットワークパスの障害発生を他のソフトウェアの介在や複雑な処理を必要とせず、かつ速やかにネットワークスイッチの障害およびリンクアップまたはダウン状態を検知できるようになり、タイムアウトを待たず主系ネットワークから予備系ネットワークへ切り替えることができる効果がある。

【発明を実施するための最良の形態】

【００１２】

次に、本発明の実施の形態について図に基づいて説明する。

【実施例１】

【００１３】

図１は本発明の一実施の形態に係るネットワークスイッチモジュールのブロック図である。ネットワークスイッチモジュール(110)内部には、ネットワークスイッチモジュール全体を管理・制御するマイコン(113)とポート(1)～(n)の通信を送受信するPHY IC(112)と、PHY IC(112)が受けたデータをスイッチングして、ポート(1)～ポート(n)へ送信するスイッチIC(111)と、機器内の温度監視を行う温度監視IC(114)、冷却ファンの制御および異常検出を行うファン制御IC(115)、電圧監視を行う電圧監視IC(116)が設置され、それぞれがI2Cバスによりマイコン(113)と接続される。なお、ポートには外部ネットワークに接続される外部ポートと内部ネットワークに接続される内部ポートがある。

【００１４】

PHY IC(112)はポート(1)～(n)の通信の異常を検知する機能も持ち、ポート(1)～(n)のどれかで異常を検出した場合、ネットワークスイッチ内マイコン(113)に接続された割り込み線(119)で割り込み信号を送ることにより、ポートの異常を通知する。割り込み線(11

9)を有さないPHY ICの場合はI2Cバス(120)を介してマイコン(113)がPHY IC(119)の内部状態レジスタを定期的に観察することで、異常を検出できるようにするのでも良い。

【００１５】

スイッチIC(111)はマイコン(113)とPCI割り込み線を含むPCIバス(121)で接続され、スイッチIC(111)はPHY IC(112)からのデータにエラーを検出した時などの異常を検出した場合、PCI割り込み信号を送ることにより、異常をマイコン(113)は通知することが出来る。

【００１６】

温度監視IC(114)とファン制御IC(115)と電圧監視IC(116)はマイコン(113)からI2Cバスで内部レジスタを定期的に監視され、機器の温度、電圧またはファンの異常をマイコンが検出することが出来るようになっている。機器の異常は前述のケースに限らず、使用環境に応じて監視したい環境監視ICを追加することで、機器の動作状態監視を細かくすることができる。

【００１７】

さらにマイコン(113)からポート(1)～(n)の接続・切断状態（リンクアップ・リンクダウン）を示すSW_READY出力信号(118)が設置されていて、マイコン(113)が内部のGPIOレジスタ(117)の値を書き換えることによりSW_READY信号が示す接続または切断状態を切り替えることができる。

例えば、GPIOレジスタ(117)の値に000…0を書き込んだ場合は、ポート(1)～(n)全てが接続状態、111…1を書き込んだ場合はポート(1)～(n)全てが切断状態となっていることを示す。

【００１８】

マイコン(113)が、ネットワークスイッチモジュール内部のマイコン(113)またはマイコン(113)上で動作するソフトウェアの異常、スイッチIC(111)の異常、PHY IC(112)の異常、温度監視IC(115)の異常、ファン制御IC(114)または電圧監視ICの異常の何れかの異常を検知した際に、GPIOレジスタ(117)に切断を示す値を書き込むことで、SW_READY出力信号(118)で外部接続機器（サーバモジュール等）にポートの切断を通知することができる。

【００１９】

なお、マイコン(113)はネットワークスイッチモジュール内部の異常を検出したときだけでなく、ユーザからのアクセスにより、GPIOレジスタ(117)を制御して、切断状態とする機能も含む。

【００２０】

さらにGPIOレジスタ(117)の1ビットだけを全てのポートに対応するSW_READY出力信号(118)へ振り分けることでも良く、この場合はネットワークスイッチモジュールの障害を全ポートに通知するので、ネットワークスイッチモジュールが通信可能であるか否かを通知する事になる。

【００２１】

また、マイコン(113)には管理インタフェースが接続され（図示せず）、外部から外部ポートを介してマイコン(113)へアクセスすることができ、GPIOレジスタ(117)の値を操作することもできる。

【実施例２】

【００２２】

次に図２を参照して、本発明の一実施形態としての、ブレードサーバシステム内部ネットワークシステムの説明をする。ブレードサーバシステム(210)は、サーバモジュール1(300)～サーバモジュールn(310)をn台と内蔵ネットワークスイッチモジュール１(240)と内蔵ネットワークスイッチモジュール２(250)を備え、ミッドプレーン(260)が各サーバモジュール間および内蔵ネットワークスイッチモジュール間を接続し、複数のサーバモジュール、ネットワークスイッチモジュールを一つの筐体内に一体化したシステムである。

【００２３】

サーバモジュール１(300)はCPU(301)、メモリコントローラ(302)、メモリ(303)、I/Oコントローラ(304)、記憶装置(305)、ネットワークコントローラLSI1(306)を備える。ネッ

トワークコントローラLSIは、ネットワークコントローラLSI1(306)とネットワークコントローラLSI2(307)が備えられ、冗長化されている。更に、ネットワークコントローラLSI1(306)はポートA(306a)とポートB(306b)を備え、一つのLSIで２つのネットワークと接続することができる構造を持つ。

【００２４】

サーバモジュール1(300)の２つのネットワークコントローラLSIの合計４つのポート(306a)から(307b)をミッドプレーン260を介して図２の様にたすきがけ状に接続することで、システム内ネットワークを冗長化し、各サーバモジュールとシステム外部ネットワーク間の接続の信頼性を高めている。

【００２５】

更に、実施例１で記述したネットワークスイッチモジュール１(240)の異常を通知するための出力信号SW_READY(248)をミッドプレーン(260)を介してネットワークポートに対応したネットワークコントローラLSIのSIGDET入力と接続されている。

この接続形態により、各ネットワークスイッチモジュールの装置障害やポート通信障害によるリンクダウン状態をSW_READY出力信号(245)により直接、対応ネットワークコントローラへ通知できるので、サーバモジュール１(300)～サーバモジュールn(310)上で動作しているOSやデバイスドライバはネットワークコントローラLSI (306,307,316,317)からの割り込みまたはポーリングによってSW_READY信号(248～258)の状態を速やかに検知でき、速やかに主系ネットワークパスから予備系への交替処理が行われるようになり、さらに、他のネットワーク監視/切り替えソフトウェアの導入と複雑な処理も必要とせず予備系ネットワークへの交替が可能となる。

【００２６】

例えば、内蔵ネットワークスイッチモジュール１(240)のスイッチIC(241)に動作不能となる障害が起きた場合、マイコン(244)がバス(243)を介してスイッチIC(241)の障害を検知し、マイコン(244)内部のGPIOレジスタ(245)をダウン状態値に変える。この値は、SW_READY出力信号(248)として各サーバモジュールの対応するポート(306a、307a、316a、317a)のSIGDET入力へ伝わる。各サーバモジュール上のCPU(301、311)で動作するOSまたはデバイスドライバは、ネットワークコントローラLSIからの割り込みまたはポーリングで速やかにSIGDET信号から内蔵スイッチモジュール１(240)の障害を検知し、障害を検知した後に速やかに予備系ポート(306b、307b、316b、317b)へ切り替えることができる。

【００２７】

このように本発明により、ネットワークスイッチモジュールから直接サーバモジュールへネットワークダウン状態を通知し、サーバモジュール上のOSまたはデバイスドライバが、それを検知した後に速やかに、他のネットワーク監視または切り替えソフトの導入と複雑な処理も必要とせずに、予備系ポートへ切り替えることができる。

る。

【００２８】

さらに発明によれば、ネットワークスイッチモジュールの障害時に限らず外部からの操作でネットワークパスを交替させることもでき、ネットワークスイッチモジュール障害の程度に応じ、全ポート一括でネットワークパスを交替させたり、選択したポートだけを交替させたりすることもできる。

【図面の簡単な説明】

【００２９】

【図１】本発明の一実施の形態に係るネットワークスイッチモジュールのブロック図である。

【図２】本発明の一実施形態としての、ブレードサーバシステム内部ネットワークシステムのブロック図である。

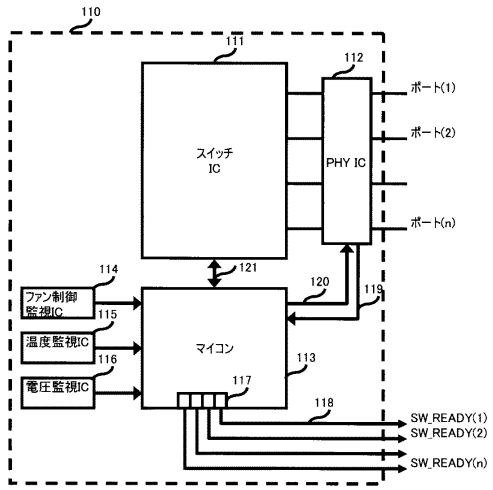
【符号の説明】

【００３０】

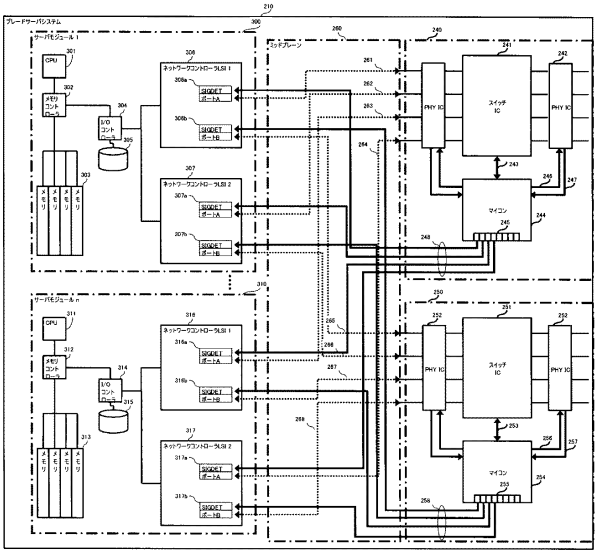
110…本発明の実施例のネットワークスイッチモジュール

111…スイッチIC
112…PHY IC
113…マイコンLSI
114…ファン制御・監視IC
115…温度監視IC
116…電圧監視IC
117…GPIOレジスタ1～n
118…SW_READY出力信号1～n
119…割り込み線
120…I2Cバス
121…PCIバス
210…ブレードサーバシステム
240、250…本発明の実施例のネットワークスイッチモジュール
241、251…スイッチIC
242、252…PHY IC
243、253…PCIバス
244、254…マイコン
245、255…GPIOレジスタ
246、256…I2Cバス
247、257…割り込み線
248、258…SW_READY出力信号
260…ミッドプレーン
261、262、263、264、265、266、267、268…内部ネットワーク
300、310…サーバモジュール
301、311…CPU
302、312…メモリコントローラLSI
303、313…メモリ
304、314…I/Oコントローラ LSI
305、315…ストレージデバイス
306、307、316、317…ネットワークコントローラLSI
306a、306b、307a、307b、316a、316b、317a、317b…SIGDET信号入力

【図１】


110
111
112
スイッチ
IC
PHY IC
ポート(1)
ポート(2)
ポート(n)
121
114
ファン制御
監視IC
115
温度監視IC
116
電圧監視IC
マイコン
120
119
113
117
118
SW_READY(1)
SW_READY(2)
SW_READY(n)

【図２】



フロントページの続き

(72)発明者　森本成重
神奈川県秦野市堀山下１番地　株式会社日立製作所　エンタープライズサーバ事業部内
(72)発明者　矢田浩勝
神奈川県秦野市堀山下１番地　株式会社日立製作所　エンタープライズサーバ事業部内

審査官　阿部　弘

(56)参考文献　特開平１１－０８８４０１（ＪＰ，Ａ）
特開２００１－０９４５８５（ＪＰ，Ａ）
特開平０９－３０５５１１（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｈ０４Ｌ　２９／１４