(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4166939号
(P4166939)
(45)発行日　平成20年10月15日(2008.10.15)　(24)登録日　平成20年8月8日(2008.8.8)

(51)Int.Cl.　　F I
H04L 12/28 (2006.01)　　H04L 12/28　２００Ｍ
G06F 11/16 (2006.01)　　G06F 11/16　３１０Ｄ

請求項の数 24　(全 14 頁)

(21)出願番号	特願2000-502457 (P2000-502457)
(86)(22)出願日	平成10年7月13日 (1998.7.13)
(65)公表番号	特表2001-509657 (P2001-509657A)
(43)公表日	平成13年7月24日 (2001.7.24)
(86)国際出願番号	PCT/US1998/014451
(87)国際公開番号	WO1999/003038
(87)国際公開日	平成11年1月21日 (1999.1.21)
審査請求日	平成17年7月5日 (2005.7.5)
(31)優先権主張番号	08/891,539
(32)優先日	平成9年7月11日 (1997.7.11)
(33)優先権主張国	米国(US)

(73)特許権者　500020944
マラソン　テクノロジーズ　コーポレイション
アメリカ合衆国　マサチューセッツ州　０１７１９　ボックスボロ　マサチューセッツ　アベニュー　１３００
(74)代理人　100059959
弁理士　中村　稔
(74)代理人　100067013
弁理士　大塚　文昭
(74)代理人　100082005
弁理士　熊倉　禎男
(74)代理人　100065189
弁理士　宍戸　嘉一

最終頁に続く

(54)【発明の名称】能動的故障検出

(57)【特許請求の範囲】
【請求項１】
データ通信システムに結合された第１の入出力プロセッサ及び第２の入出力プロセッサを内含するフォールトトレラントコンピュータシステム内の故障を検出するための方法において、
－　第１のタイミング結果を生成するべく第１の入出力プロセッサにより処理されたデータ通信カテゴリに対してタイミング基準を適用する段階；
－　第２のタイミング結果を生成するべく第２の入出力プロセッサによって処理されたデータ通信カテゴリに対してタイミング基準を適用する段階；
－　第１のタイミング結果と第２のタイミング結果の間の関係を決定する段階；及び
－　決定された関係に基づいて故障が発生したか否かを検出する段階；
を含んで成る方法。
【請求項２】
故障が発生したか否かを検出する段階には、タイミング結果の間の差がしきい値を上回った時点で故障が発生したと決定する段階が含まれている、請求項１に記載の方法。
【請求項３】
タイミング基準が最後の送信又は受信の時間である、請求項１に記載の方法。
【請求項４】
データ通信カテゴリには、第１の入出力プロセッサを発信元とするメッセージが含まれる、請求項１に記載の方法。

【請求項５】
データ通信カテゴリには、第１の入出力プロセッサを発信元としかつ第２の入出力プロセッサが通常応答性をもつアドレスに対し導かれたメッセージが含まれる、請求項４に記載の方法。
【請求項６】
データ通信カテゴリには、第１の入出力プロセッサから送られかつデータ通信システムを通して第２の入出力プロセッサに導かれるメッセージが含まれる、請求項４に記載の方法。
【請求項７】
データ通信カテゴリには、コンピュータシステムの外部を発信元とするメッセージが含まれる、請求項１に記載の方法。
【請求項８】
メッセージが、データ通信システムに結合された第２のコンピュータを発信元としている、請求項７に記載の方法。
【請求項９】
データ通信カテゴリには、コンピュータシステムの外部を発信元としかつそのコンピュータシステムが一構成要素を成している１群のシステムにアドレス指定されたメッセージが含まれる、請求項７に記載の方法。
【請求項１０】
データ通信カテゴリには、入出力プロセッサとデータ通信状態にあるコンピュータシステムの第３の要素を発信元とするメッセージが含まれている、請求項１に記載の方法。
【請求項１１】
－　第１の入出力プロセッサから第２の入出力プロセッサに第１のタイミング結果を送信する段階をさらに含んで成り；
－　第１の入出力プロセッサにより処理されたデータ通信カテゴリに対しタイミング基準を適用する段階には、第１の入出力プロセッサにおいて第１の入出力プロセッサにより処理されたデータ通信カテゴリに対してタイミング基準を適用する段階が内含されており；
－　第２の入出力プロセッサにより処理されたデータ通信のカテゴリに対してタイミング基準を適用する段階には、第２の入出力プロセッサにおいて第２の入出力プロセッサにより処理されたデータ通信カテゴリに対しタイミング基準を適用する段階が含まれ；
－　タイミング結果の関係を決定する段階には、第２の入出力プロセッサにおけるタイミング結果の間の差を決定する段階が含まれている、請求項１に記載の方法。
【請求項１２】
第１のタイミング結果が第１の入出力プロセッサと第２の入出力プロセッサの間の専用通信チャネル上で送られる、請求項１１に記載の方法。
【請求項１３】
－　第１の入出力プロセッサにより処理されたデータ通信カテゴリに対しタイミング基準を適用する段階には、第１の複数のタイミング結果を生成するべく第１の入出力プロセッサにより処理された対応する複数のデータ通信カテゴリに対し複数のタイミング基準を適用する段階がさらに含まれ；
－　第２の入出力プロセッサにより処理されたデータ通信カテゴリに対しタイミング基準を適用する段階には、第２の複数のタイミング結果を生成するべく第２の入出力プロセッサにより処理された対応する複数のデータ通信カテゴリに対して複数のタイミング基準を適用する段階がさらに含まれ；
－　タイミング結果間の関係を決定する段階には、第１の複数のタイミング結果及び第２の複数のタイミング結果のうちの対応するタイミング結果の間の関係を決定する段階がさらに含まれる、請求項１に記載の方法。
【請求項１４】
データ通信システムに結合されたフォールトトレラントコンピュータシステムであって

、前記フォールトトレラントコンピュータシステムが、
－　データ通信カテゴリを処理しそのデータ通信カテゴリに対しタイミング基準を適用して第１のタイミング結果を生成するように構成された第１の入出力プロセッサ；及び
－　データ通信カテゴリを処理しそのデータ通信カテゴリに対しタイミング基準を適用して第２のタイミング結果を生成するように構成された第２の入出力プロセッサ；
を含んで成り、前記第１及び第２のタイミング結果間の関係を決定しその関係に基づいて故障が発生したか否かを決定するように構成されているコンピュータシステム。
【請求項１５】
タイミング基準が最後の送信又は受信の時間である、請求項１４に記載のシステム。
【請求項１６】
データ通信カテゴリには、第１の入出力プロセッサを発信元とするメッセージが含まれる、請求項１４に記載のシステム。
【請求項１７】
データ通信カテゴリには、第１の入出力プロセッサから送られデータ通信システムを通して第２の入出力プロセッサに対し導かれたメッセージが含まれる、請求項１６に記載のシステム。
【請求項１８】
データ通信カテゴリには、第１の入出力プロセッサを発信元としかつ第２の入出力プロセッサが通常応答性をもつアドレスに対し導かれたメッセージが含まれる、請求項１６に記載のシステム。
【請求項１９】
データ通信カテゴリには、フォールトトレラントシステムの外部を発信元とするメッセージが含まれる、請求項１４に記載のシステム。
【請求項２０】
メッセージがデータ通信システムに結合された第２のコンピュータを発信元としている、請求項１９に記載のシステム。
【請求項２１】
データ通信カテゴリには、フォールトトレラントシステムの外部を発信元としかつそのフォールトトレラントシステムが一構成要素を成している１群のシステムにアドレス指定されたメッセージが含まれる、請求項１９に記載のシステム。
【請求項２２】
入出力プロセッサとデータ通信状態にあるコンピュータシステムの第３の要素をさらに含んで成り、データ通信カテゴリには、該第３の要素を発信元とするメッセージが含まれる、請求項１４に記載のシステム。
【請求項２３】
第１の入出力プロセッサ及び第２の入出力プロセッサを結合する専用通信チャネルをさらに含んで成り、該通信チャネルは第１の入出力プロセッサから第２の入出力プロセッサに対し第１のタイミング結果を送るように構成されており、
－　第２の入出力プロセッサは、タイミング結果間の差を決定し、この差がしきい値を超えた時点で故障が発生したか否かを検出するように構成されている、請求項１４に記載のシステム。
【請求項２４】
－　第１の入出力プロセッサはさらに、第１の複数のタイミング結果を生成するべく第１の入出力プロセッサにより処理された対応する複数のデータ通信カテゴリに対し複数のタイミング基準を適用するように構成されており；
－　第２の入出力プロセッサは、さらに第２の複数のタイミング結果を生成するべく第２の入出力プロセッサにより処理された対応する複数のデータ通信カテゴリに対して複数のタイミング基準を適用するように構成されており；
－　コンピュータシステムは、さらに、第１及び第２の複数のタイミング結果のうちの対応するタイミング結果の間の関係を決定し、決定された関係に基づき故障が発生したか

否かを決定するように構成されている、請求項１４に記載のシステム。

【発明の詳細な説明】

【０００１】

本発明は、フォールトトレラント計算システム内での通信故障といったような故障の検出に関する。

【０００２】

フォールトトレラント計算システムにおいては一般に冗長なハードウェア要素が使用される。システムの個々の要素は標準的に、システム内のその他の要素によって生成されるか又はシステムの外部で生成された信号を監視することにより故障を検出しようと試みる。

【０００３】

さらに、システムの１つの要素は、要素の適切な作動を表示するいわゆる「ハートビート」信号を周期的に送信することができる。ハートビート信号がシステム内のもう１つの要素により受理されない場合、受信要素は、送信要素が作動状態にないことを疑うことができる。しかしながら、ハートビート信号の受信障害は、２つの要素間の通信経路内の故障の結果もたらされる可能性もある。一般に、故障の取扱いは、要素間の通信内の故障とシステムの１要素内の故障を区別しなくてはならない。

【０００４】

外部ネットワークとの信頼性が高く中断の無い通信を提供するために、冗長なネットワークインタフェースコントローラ（ＮＩＣ）が、フォールトトレラント計算システム内で使用される。一般的に、１つのＮＩＣは、それがネットワーク上のその他のデバイスとの通信を担当している一次的つまり活動状態モードで作動するのに対し、その他のＮＩＣは待機モードで作動する。

【０００５】

作動中、ＮＩＣは、１つのＮＩＣから外部ネットワークを通りそしてもう１つのＮＩＣに戻る１本の経路内の故障を検出するためにハートビートメッセージを交換することができる。ＮＩＣ間の経路内の故障は、ＮＩＣの入出力段、ＮＩＣと外部ネットワーク間の送信又は受信接続又は外部ネットワーク自体の中を含む数ヵ所で発生しうる。外部ネットワークへの接続点は一般に、ネットワークハブのポートにあり、このハブは多くのネットワークデバイスに接続された状態にある。各々のＮＩＣは、単一のハブが故障臨界点になるのを避けるため外部ネットワーク内の異なるハブに接続されていてよい。

要約

本発明は、データ通信システムに接続された２つの入出力プロセッサを内含するフォールトトレラントコンピュータシステム内の故障検出を提供する。該コンピュータシステムは、一次入出力プロセッサと、ならびに待機入出力プロセッサと結びつけられる故障を検出する能力をもつ。このシステムは同様に、入出力プロセッサの故障とデータ通信ネットワーク自体の中の通信障害を弁別することもできる。システムは、故障を検出するため「ハートビート様の」伝送以外のデータ通信カテゴリを分析する。システムは、プロセッサが同一のデータトラヒックを監視できるようにする共通のネットワークセグメント上に入出力プロセッサがあるときに故障を検出することができる。システムは同様に、スイッチ又は活動状態のハブといったネットワーク要素のろ過挙動の結果としてプロセッサが同一のデータトラヒックを監視できないかもしれない異なるセグメント上にプロセッサがあるときに故障を検出することもできる。

【０００６】

各々の入出力プロセッサにより処理されたデータ通信のカテゴリに対し１つのタイミング基準を適用することができ、故障を検出するために、各プロセッサについて得られた結果の間の関係を使用することができる。例えば、データ通信のタイミング間の差が１定のしきい値を上回った時点で、故障を表示することができる。このタイミング基準は、１つの通信カテゴリの最後の送信又は受信時間でありうる。データ通信カテゴリは例えば、データ通信システムに結合されたもう１つのコンピュータシステムからといったように、フォールトトレラントシステムの外部に発信元を有するメッセージであり得る。これらのメッ

セージは、フォールトトレラントシステムが一構成要素を成している一群のシステムに対してアドレス指定されうる。このカテゴリは同様に、他のプロセッサにアドレス指定されたメッセージを含めた、入出力プロセッサの１つを発信元とするメッセージ、又はシステムのいくつかのその他の要素を発信元とするメッセージを内含していてもよい。

【０００７】

１つの形態においては、一般に、本発明は、データ通信システムに結合された第１の入出力プロセッサ及び第２の入出力プロセッサを内含するフォールトトレラントコンピュータシステム内の故障を検出することを特長としている。第１及び第２のタイミング結果を生成するべく第１及び第２の入出力プロセッサにより処理されたデータ通信カテゴリにタイミング基準が適用される。タイミング結果間の関係が決定され、故障が発生したか否かが、この決定された関係に基づいて検出される。

【０００８】

本発明の実施形態は、以下の特長のうち単数又は複数のものを内含し得る。例えば、故障が発生したか否かを検出する段階には、タイミング結果間の差がしきい値を上回った時点で故障が発生したことを決定する段階が含まれていてよい。

【０００９】

タイミング基準は、最後の送信又は受信の時間でありうる。データ通信カテゴリは、第２の入出力プロセッサが通常応答性をもつアドレスに対して導かれたメッセージといったような第１の入出力プロセッサを発信元とするメッセージ、又は第１の入出力プロセッサから送られかつデータ通信システムを通って第２の入出力プロセッサに導かれたメッセージを内含することができる。データ通信カテゴリは同様に、データ通信システムに結合された第２のコンピュータを発信元とするメッセージといったコンピュータシステムの外部を発信元とするメッセージ又は、コンピュータシステムが一構成要素を成す一群のシステムにアドレス指定されたメッセージを内含することもできる。データ通信カテゴリは同様に、入出力プロセッサとデータ通信状態にあるコンピュータシステムの第３の要素を発信元とするメッセージを含むこともできる。

【００１０】

タイミング基準は、第１の入出力プロセッサにおいて第１の入出力プロセッサにより処理されたデータ通信カテゴリに対して適用され得、第１のタイミング結果は第１の入出力プロセッサから第２の入出力プロセッサまで送られ得る。第２の入出力プロセッサにおいて第２の入出力プロセッサにより処理されたデータ通信カテゴリに対し、タイミング基準を適用し、第２の入出力プロセッサにおいてタイミング結果間の関係を決定することができる。又、入出力プロセッサ間の専用通信チャネル上で第１のタイミング結果を送ることができる。

【００１１】

第１及び第２の複数のタイミング結果を生成するため、入出力プロセッサにより処理された対応する複数のデータ通信カテゴリに対し、複数のタイミング基準を適用することが可能である。第１の複数のタイミング結果及び第２の複数のタイミング結果のうちの対応するタイミング結果の間の関係を決定することができる。

【００１２】

本発明によって提供される１つの利点は、フォールトトレラントシステムからデータネットワークまでの通信経路の中の故障を識別でき、とくに、入出力プロセッサを結合するデータ経路内の故障をプロセッサ内の故障と全く異なるものとして検出することができる。

【００１３】

本発明のその他の特長及び利点は、図面を含む以下の記述及びクレームから明らかになることだろう。

記述

図１を参照すると、フォールトトレラントシステム１０は、２重冗長計算要素（ＣＥ）１２及び１４，２重冗長入出力プロセッサ（ＩＯＰｓ）２４及び２６，及び通信相互接続デバイス３６及び３８を内含している。ＣＥ１２及び１４は、並行動作シーケンスを実施す

る。各ＣＥは、両方のＩＯＰ２４及び２６と通信する。ＣＥ１２は、通信リンク５０及び５４上で通信し、一方ＣＥ１４は通信リンク５２及び５６上で通信する。通信リンクは、通信相互接続デバイス３６及び３８を通してルーティングされる。

【００１４】

各ＣＥは、中央処理ユニット（ＣＰＵ）１６又は１８及びインタフェースコントローラ（ＩＣ）２０又は２２を内含する。ＩＣは、ＣＰＵと通信リンクの間のインタフェースを提供する。例えば、ＣＰＵ１６によるＩ/Ｏ要求は、ＩＣ２０により通信リンク５０及び５４を通して伝送される。ＩＯＰは、その相互接続された通信構造により、正規の動作において各ＣＥから同一の指令シーケンスを受信するものと予測している。

【００１５】

各ＩＯＰは、ＣＰＵ２８又は３０及びネットワークインタフェースコントローラ（ＮＩＣ）３２又は３４を内含している。ネットワークプロバイダ２９及び３１は、ＣＰＵ２８及び３０上で実行するソフトウェアドライバである。ＮＩＣ３２及び３４は、ネットワークプロバイダがネットワーク接続６２及び６４を通してローカルエリアネットワーク（ＬＡＮ）４０上で通信できるようにしている。ＣＰＵ２８及び３０を接合する専用通信経路６０により、ネットワークプロバイダ２９及び３１は、ＬＡＮ４０を使用することなくメッセージを変換することが可能となる。

【００１６】

１つのネットワークプロバイダは、一次状態で動作し、一方もう１つのネットワークプロバイダは待機状態で動作する。一次状態で動作するネットワークプロバイダのみが、ＣＥ１２及び１４を発信元とするデータをＬＡＮ上のその他のデバイスにデータを伝送する。

【００１７】

ネットワーク接続６２及び６４はＮＩＣ３２及び３４に接続し、ＬＡＮ４０の通信ハブ４２及び４４のポートで終結する。ハブ４２及び４４はＬＡＮ４０のブリッジ６６を通して接続されている。ハブ４２及び４４は、いかなる通信もろ過しないが、一方ブリッジ６６は、ブリッジ上で特定のポートを通してアクセスされるデバイスに向けられていない通信をろ過する。従って、ハブ４２及び４４は、ＬＡＮ４０の異なるセグメント上にある。ＬＡＮ４０に接続されたデバイス４６といったその他のデバイスは、フォールトトレラントシステム１０と通信することができる。

【００１８】

ＮＩＣ３２及び３４の各々は、両方のＮＩＣについて同じであるように構成されているプログラミング可能な「論理」アドレス及び固定された一意的「物理」アドレスを有する。論理アドレスは、フォールトトレラントシステム１０とＬＡＮ４０上のデバイス例えばデバイス４６又はＬＡＮ４０からアクセス可能なデバイスの間での通信のために使用される。各々のＮＩＣは同様に、同報通信、マルチキャスト又は機能アドレスに送られるメッセージといったグループアドレス指定メッセージを受信するようにプログラミングされる。ＮＩＣにより送られたグループアドレス指定メッセージは、メッセージのソースとしてそのＮＩＣの一意的物理アドレスを特定する。かくして、グループアドレス指定メッセージの受信者は、どのＮＩＣがメッセージを送ったかを決定することができる。

【００１９】

ネットワーク接続６２及び６４は、ＬＡＮ４０の共通セグメント上又は異なるセグメント上で終結しうる。一般に、接続６２及び６４が共通のセグメントで終結する場合、両方のＮＩＣ共、そのセグメント上の全てのデータトラヒックを監視することができる。従って、１つのＮＩＣにより伝送されたデータは、たとえそのデータがそのＮＩＣにアドレス指定されていなくても、もう一方のＮＩＣにより受信され得る。これと対照的に、図１に示されているように、ネットワーク接続６２及び６４は、ハブ４２及び４４において異なるセグメント上で終結し得る。これらのハブは、それらがＬＡＮ４０の異なるセグメント上にくるようにブリッジ６６により接続されている。ブリッジ６６は、１つのセグメントの通信容量の不必要な利用を避けるためそのセグメントに伝送されたデータをろ過するように構成されている。ブリッジ６６に接続された特定のセグメント上にあるか又はそのセ

グメントを通して通信するデバイスのアドレステーブルが、そのセグメントから到着するデータを監視することによって、該ブリッジによって維持される。１つのセグメントについてのテーブル内にない特定のデバイスに対しアドレス指定されたメッセージ（すなわちグループアドレス指定メッセージではなく入出力指定メッセージ）は、ブリッジ６６によりそのセグメントに再度伝送される。一方、グループアドレス指定メッセージは、ろ過無しでＬＡＮの全てのセグメント上で再伝送される。

【００２０】

動作中、フォールトトレラントシステム１０は、ネットワーク接続６２及び６４がＬＡＮ４０の共通セグメントに接続されるか否か、従って両方のＮＩＣが同一のネットワークトラヒックを見ると予測すべきであるか否かを決定する。ＮＩＣが異なるセグメント上にあることをシステムが決定した場合には、システムはＮＩＣが同一のグループアドレス指定トラヒックのみを見ると予測すべきであることを決定する。

【００２１】

動作中、一次ネットワークプロバイダも待機ネットワークプロバイダも共に、故障が発生したか否かを決定するためデータ通信を監視する。故障が１つのＩＯＰのネットワーク接続を非機能的にした場合、適切な行動がとられる。活動状態のネットワークプロバイダがネットワーク連結性を失ない、待機ネットワークプロバイダがオンラインである場合、待機ネットワークプロバイダを新しい一次ネットワークプロバイダにするべく切換えが行なわれることになる。待機ネットワークプロバイダがネットワーク連結性を失なった場合、待機ネットワークプロバイダは、連結性が再度樹立されるまでオフライン状態に入ることになる。

【００２２】

ネットワークプロバイダは、データ通信カテゴリを監視し、各カテゴリ内の最後の通信が発生してからの時間を維持することにより故障を検出する。ネットワークプロバイダは、通信故障が発生したと疑っている場合、通信経路６０上でもう一方のネットワークプロバイダと状態メッセージを交換する。状態メッセージは、最後の通信の時間を内含する。各々のネットワークプロバイダは、受信した状態メッセージ内の時間と受信ネットワークプロバイダにより維持された時間を比較して、システム内に場合によって存在する故障を識別する。比較を行なう上で、ネットワークプロバイダは、時間が合致すべき許容誤差範囲を考慮する。この許容誤差は、転送時間の自然の可変性及び状態メッセージをアセンブルし伝送するのに必要な時間を説明する。

【００２３】

ネットワーク接続６２及び６４の状態ならびにＬＡＮ４０の状態を感知するため、ネットワークプロバイダ２９及び３１は、ＮＩＣ３２及び３４を通して、両方のＮＩＣにより監視されたグループアドレスにアドレス指定されたノイズパケットとして知られるグループアドレス指定メッセージを伝送する。グループアドレス指定メッセージのソースアドレスは、受信ＮＩＣがそのソースを決定できるように、送信ＮＩＣの一意的物理アドレスにセットされる。ネットワーク接続が共通セグメント上にあるとき、メッセージは直接ＮＩＣにより受信される。ソース及び受信ＮＩＣが別々のセグメント上にあるとき、グループアドレス指定パケットは、ＬＡＮ４０の正規の動作において１つのセグメントからもう１つのセグメントへと再度伝送される。

【００２４】

フォールトトレラントシステム１０が初期化された時点で、ネットワークプロバイダ２９及び３１は、３つの開始状態のシーケンスを通って進む。接合状態として識別される第１の状態では、両方のＩＯＰはすでにＬＡＮ４０との通信を樹立し、プロセッサ２８及び３０の間の通信経路６０上で通信を樹立している。次に、同期化状態では、ＣＰＵ１６及び１８上で実行する通信入出力先変更ソフトウェアがネットワークプロバイダ２９及び３１と同期化される。最後に、完全に初期化された状態では、ＣＰＵ１６及び１８上で実行された入出力要求を、ＬＡＮ４０との通信のため首尾よくネットワークプロバイダ２９及び３１に送ることができる。

【００２５】
図２を参照すると、両方のネットワークプロバイダ共、完全に初期化された時点でオンライン/待機状態７０に入る。１つのネットワークプロバイダはその後、オンライン/待機状態からオンライン/一次状態７２にされる。オンライン状態７０，７２の１つで動作するネットワークプロバイダが、自らネットワーク連結性を失なったかもしれないと疑っている場合にはつねに、このネットワークプロバイダは、周期的にネットワーク状態要求をもう１つのネットワークプロバイダに対し送る。ネットワーク連結の喪失が確認されたならば、そのネットワークプロバイダはオフライン状態７４へと移る。オンライン/一次状態７２で動作するネットワークプロバイダがネットワーク連結性の喪失を検出しオフライン状態７４に移行した場合、オンライン/待機状態７０で動作しているネットワークプロバイダはオンライン/一次状態７２へと移行する。ここで、オンライン/一次状態７２からオフライン状態７４への自動的遷移は、もう１つのネットワークプロバイダがオンライン/待機状態７０にあるときにのみ許容されるということに留意されたい。オフライン状態７４にある間、ネットワークプロバイダはリンク６０上でもう１方のネットワークプロバイダに対しネットワーク状態要求を周期的に送る。ネットワーク連結性が再度樹立されたならば、ネットワークプロバイダはオンライン/待機状態７０に戻る。

【００２６】
自動遷移に加えて、オペレータが制御するオンライン及びオフライン状態７０，７２及び７４から使用禁止状態への変更も発生しうる。状態７６にあるネットワークプロバイダが手動式で再度使用可能にされた時点で、ネットワークプロバイダはオンライン/待機状態７０に入り、次に、ネットワーク連結性が確認されない場合、オフライン状態７４まで直ちに遷移する。最後に、その他の検出機構が、ネットワークプロバイダの故障を決定することもでき、その結果ネットワークプロバイダは障害状態７８に入ることになる。

【００２７】
図３を参照すると、ネットワークプロバイダがオンライン又はオフライン状態７０，７２又は７４にあるとき、そのネットワークプロバイダは、ネットワーク状態要求が通信経路６０上でもう１方のネットワークプロバイダに送られるべきかを反復的に検査する。特にネットワークプロバイダは、NetworkStatusInterval と呼ばれる間隔が満了した時点でネットワーク状態要求を送るべきか否かを決定する。NetworkStatusInterval の省略時の値は１０００ミリセカンドである。その間隔が満了した時点で、ネットワークプロバイダは、３つの条件のうちの１つが満たされたか否かを決定する。ReceivePacketInterval と呼ばれる間隔内でいかなる非ノイズパケットも受信されなかった場合、第１の条件は真である（ステップ８０）。ReceicvePacketIntervalの省略時の値は、システム１０により受信されたパケット間の標準的最大間隔に対応する４０００ミリセカンドである。第２の条件は、先行する ReceivePacketInterval 内でもう１つのネットワークプロバイダからいかなるノイズパケットも受信されなかった場合に真である（ステップ８２）。第３の条件は、ネットワークプロバイダがオフライン状態７４にあるときいつでも真である（ステップ８４）。これらの条件のいずれかが満たされた場合、ネットワーク状態要求がもう１つのネットワークプロバイダに送られる（ステップ８６）。要求は、先行する要求に対する応答が受信されたか否かを考慮することなく伝送される。

【００２８】
ネットワーク状態要求を受信した時点で、ネットワークプロバイダが標準的に、以下の通信統計を含む応答メッセージを構築する。

TimeLastNoiseReceived:	① 最後のノイズパケットがもう１つのプロセッサから受信された時点以降の時間
TimeLastPacketReceived:	② 論理（システム）アドレスに対しアドレス指定された最後の非ノイズパケットが受信された時点以降の時間
TimeLastMulticastReceived:	③ 最後の非ノイズグループアドレス指定パケットが受信された時点以降の時間
TimeLastNoiseTransmitted:	④ 最後のメッセージが送られた時点以降の時間
TimeNetworkMonitored:	⑤ ネットワークプロバイダが統計を収集していた時間（すなわちシステムの動作可能時間）
CountTransmitFailures:	⑥ 現行の送信故障計数

【００２９】
両方のネットワークプロバイダが同時に故障を検出しオンライン/一次状態７２にいかなるネットワークプロバイダも残らないことがないようにするため、オンライン/一次状態７２にあるネットワークプロバイダは、オンライン/一次状態にないネットワークプロバイダとは異なる形でネットワーク状態要求に応答する。オンライン/一次状態７２にあるローカルネットワークプロバイダがネットワーク状態要求を送り、それ自体の要求に対する応答を受信する前にリモートネットワークプロバイダからのネットワーク状態要求を受信した場合、ローカルネットワークプロバイダは、応答を待つのではなくむしろその要求を満たすため受信要求の中の情報を使用する。ローカルネットワークプロバイダはこの受信要求に対し応答しない。オンライン/一次状態７２にないローカルネットワークプロバイダがネットワーク状態要求を送り、それ自体の要求に対する応答を受信する前にリモートネットワークプロバイダからのネットワーク状態要求を受信した場合、ローカルネットワークプロバイダは、その要求に応答し、その保留要求を満たすために受信要求内の情報を使用しない。
【００３０】
図４を参照すると、ネットワーク状態応答の受信時点（ステップ８７）で、ローカルネットワークプロバイダは、ローカル又はリポートネットワークプロバイダがまだ完全初期化状態にないか否か（ステップ８８），リモートネットワークプロバイダが障害状態７８又は使用不能状態７６にあるか否か（ステップ９０），又はローカルネットワークプロバイダが障害状態７８又は使用不能状態７６にある（すなわち、オンライン又はオフライン状態７０，７２又は７４にない）か否か（ステップ９２）を決定する。これらの条件のいずれも真でない場合、ローカルネットワークプロバイダは、ＩＯＰがネットワーク連結性を喪失したか否かを決定するべく、手順９４を実行する。いずれかの条件が真である場合、

応答は放棄され（ステップ９６），応答のいかなる処理も実施されない。

【００３１】

図５を参照すると、手順９４の最初のステップは、両方のＩＯＰが共通のトラヒックを受信しているか否かを決定することである（ステップ１００）。共通トラヒックの定義は、両方のＩＯＰがＬＡＮ４０の単一のセグメント上にあるか又は異なるＳＥＧ上にあることをローカルネットワークプロバイダが決定するか否かにより左右される。当初、両方のネットワークプロバイダは、ＩＯＰが異なるＳＥＧ上にあることを仮定している。この場合、共通トラヒックは、ＩＯＰにより受信された、ノイズパケット以外のグループアドレス指定パケットに対応する。両方のＩＯＰは、ネットワーク状態応答内の TimeLastMulticastReceived の値及びローカルネットワークプロバイダにおいて計算された値が ReceiveTolerance と呼ばれる許容誤差の範囲内にある場合、共通トラヒックを見る。ReceiveTolerance の省略時値は１０００ミリセカンドである。ネットワークプロバイダが、自らが同じセグメント上にあることを決定した場合、共通トラヒックは同様に、システム１０の論理アドレスに導かれたパケットをも内含する。従って、 TimeLastMulticastReceived を比較することに加えて、TimeLastPacketReceived が比較され、両方のＩＯＰは、時間が ReceiveTolerance の範囲内にある場合、共通トラヒックを見る。

【００３２】

両方のＩＯＰが共通トラヒックを見る場合、手順は、ローカルＩＯＰが共通トラヒックを受信しているか否かを決定する（ステップ１０６）ローカルネットワークプロバイダは、そのローカルＩＯＰについての Time Last Multicast Received 又は Time Last Packet Received の値が少なくとも Receive Tolerance による受信値よりも大きい場合（すなわち最後のグループアドレス指定メッセージが、リモートＩＯＰよりも少なくとも Receive Tolerance だけ早く局所的に受信された場合）ローカルＩＯＰが共通トラヒックを見ないということを決定する。ローカルＩＯＰが共通トラヒックを見ない場合には、ネットワークプロバイダは、ＬＡＮ４０からＩＯＰまでの受信経路内に故障が存在することを決定する。従ってネットワークプロバイダは、オフライン状態７４への遷移を行なう（又はその状態にとどまる）（ステップ１０８）。

【００３３】

両方のＩＯＰが共通トラヒックを受信した場合（ステップ１００），ネットワークプロバイダは、いずれのＩＯＰも、もう１方のＩＯＰのノイズパケットを受信していないか否かを決定する（ステップ１０２）。これは、各ネットワークプロバイダの TimeLastNoiseReceived が TimeLastNoiseTransmitted の局所的値よりも少なくとも ReceiveTolerance だけ大きく TimeLastNoiseReceived の局所的値が TimeLastNoiseTransmitted の受信値よりも少なくとも ReceiveTolerance だけ大きい場合に起こる。いずれのＩＯＰももう１方のＩＯＰのノイズパケットを受信していない場合（ステップ１０２），ローカルプロバイダは、それが送信障害を報告しているか否か検査する（ステップ１０９）。していない場合、状態値を説明するのに接続６２及び６４上の同時故障が必要となることから、ネットワークプロバイダはＬＡＮ４０内に障害があるにちがいないことを仮定する。これらの故障モードの両方が同時に発生することは、確率が低いものと仮定される。ネットワークプロバイダは、送信故障を報告している場合（ステップ１０９），オフライン状態へと進む（ステップ１０８）。

【００３４】

少なくとも１つのＩＯＰがもう１つのＩＯＰのノイズパケットを受信している場合、ネットワークプロバイダは、局所的に伝送されたノイズパケットがもう１つのＩＯＰにより受信されるか否かを決定する（ステップ１０４）。これは、TimeLastNoiseReceived の受信値が TimeLastNoiseTransmitted の局所的値を少なくとも ReceiveTolerance だけ上回っていない場合に起こる。局所的に伝送されたノイズパケットがもう１つのＩＯＰにより受信された場合、その局所的ＩＯＰについて障害は存在しない。一方、局所的に伝送されたノイズパケットがもう１つのＩＯＰにより受信されなかった場合、ネットワークプロバイダは、ローカルＩＯＰからＬＡＮ４０への送信経路内に障害が存在すると仮定し、ローカ

ルネットワークプロバイダはオフライン状態７４への遷移を行なう（ステップ１０８）。

【００３５】

全ての試験において、受信された時間が TimeNetworkMonitored よりも大きい場合、その時間は無効とみなれる。このメカニズムは、不正確な統計の使用を防ぐのに用いられる。状態変更に対するさらなる制約条件は、オフライン状態７４内のネットワークプロバイダがオンライン状態への遷移を行なう前にその状態にある間に少なくとも１つのパケットを受信しなくてはならないということにある。この制約条件により、いずれのＩＯＰにとっても可視的であるネットワークトラヒックが全く存在しないときに状態遷移が阻止される。

【００３６】

その他の実施形態も冒頭のクレームの範囲内に入る。例えば、上述したシステムは２重冗長ＩＯＰを使用する。さまざまな通信カテゴリの相対的時間を比較する類似の方法と共に、３つ以上のＩＯＰを使用することが可能である。３つ以上のＩＯＰが使用される場合、多数のＩＯＰからの応答を一緒に使用して通信障害を検出することができる。さらに、上述のもの以外のシステム事象カテゴリの相対的時間を、障害検出のために使用することもできる。さらに、通信事象の相対的タイミングを使用するアプローチを用いて、フォールトトレラントシステム自体の中の内部通信障害を検出することもできる。最後に、そのノイズパケットの適切な転送が使用可能にされた場合、異なるＬＡＮにＩＯＰをとりつけることも可能である。

【図面の簡単な説明】

【図１】　冗長計算要素及び入出力プロセッサを伴うフォールトトレラント計算システムのブロック図である。

【図２】　入出力プロセッサについての状態図である。

【図３】　入出力プロセッサにより実施される動作の流れ図である。

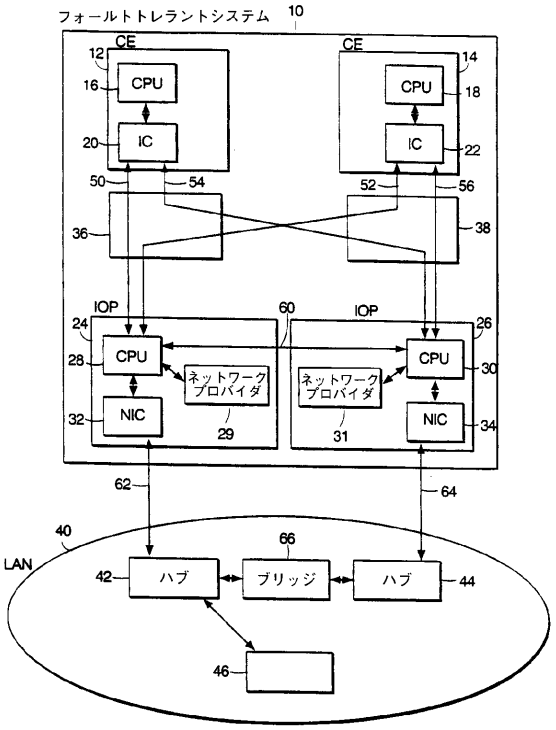
【図４】　入出力プロセッサにより実施される動作の流れ図である。

【図５】　入出力プロセッサにより実施される動作の流れ図である。

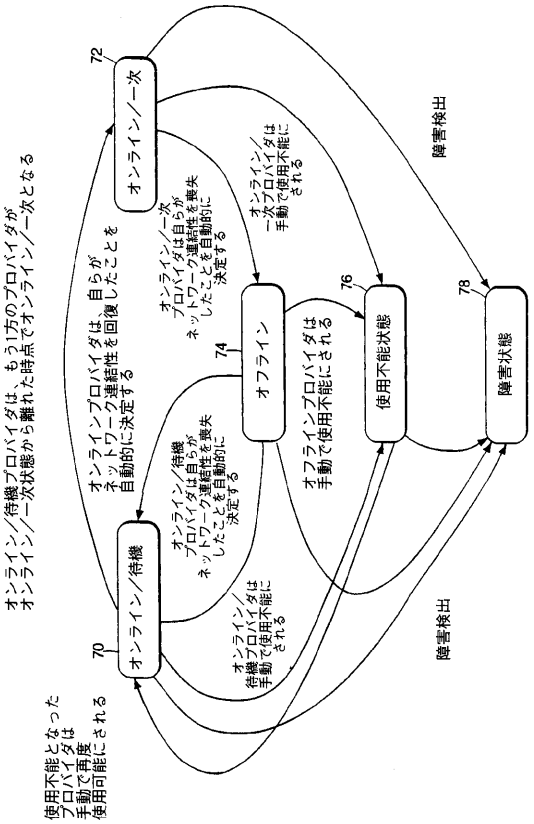
【符号の説明】

１０　フォールトトレラントシステム
１２，１４　２重冗長計算要素
１６，１８，２８，３０　中央処理ユニット
２０，２２　インタフェースコントローラ
２４，２６　２重冗長入出力プロセッサ
２９，３１　ネットワークプロバイダ
３２，３４　ネットワークインタフェースコントローラ
３６，３８　通信相互接続デバイス
４０　ローカルエリアネットワーク
４２，４４　通信ハブ
５０，５２，５４，５６　通信リンク
６２，６４　ネットワーク接続
６０　専用通信経路
６６　ブリッジ

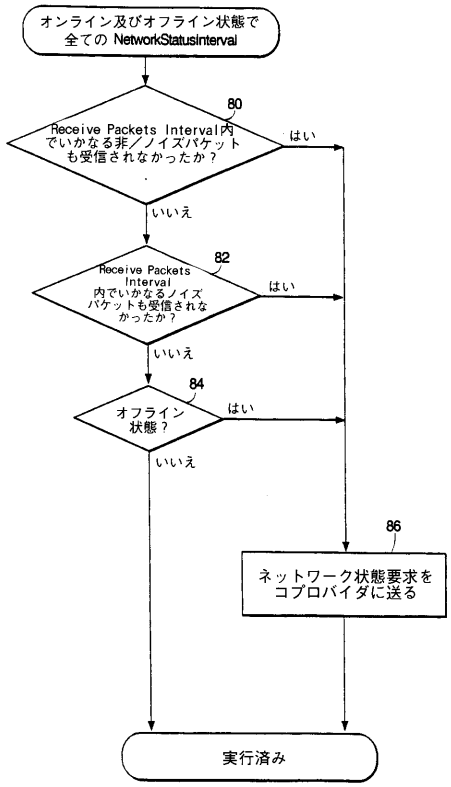
【図１】


フォールトトレラントシステム 10
CE
12
CPU
16
IC
20
CE
14
CPU
18
IC
22
50
54
52
56
36
38
IOP
24
CPU
28
ネットワーク
プロバイダ
29
NIC
32
60
IOP
26
CPU
30
ネットワーク
プロバイダ
31
NIC
34
62
64
40
LAN
66
42
ハブ
ブリッジ
ハブ
44
46

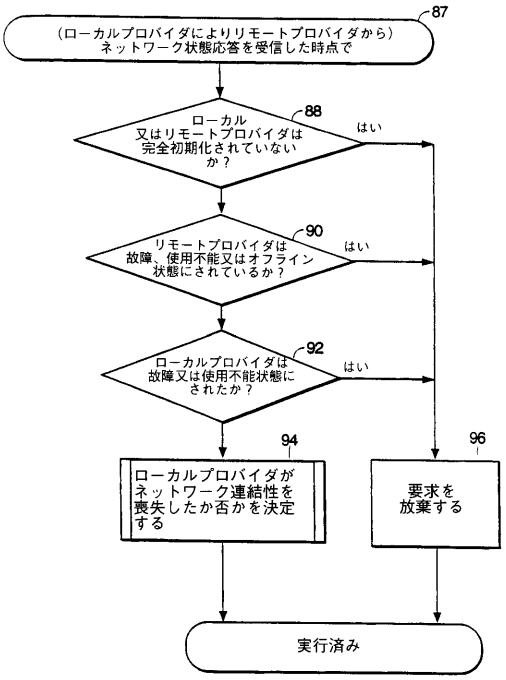
【図２】


72
オンライン／一次
障害検出
オンライン／待機プロバイダは、もう1方のプロバイダがオンライン／一次状態から離れた時点でオンライン／一次となる
オンライン／一次
プロバイダは自らがネットワーク連結性を喪失したことを自動的に決定する
オンライン／一次プロバイダは手動で使用不能にされる
オンラインプロバイダは、自らがネットワーク連結性を回復したことを自動的に決定する
74
オフライン
76
使用不能状態
78
障害状態
オンライン／待機
オンライン／待機プロバイダは自らがネットワーク連結性を喪失したことを自動的に決定する
オフラインプロバイダは手動で使用不能にされる
70
オンライン／待機
オンライン／待機プロバイダは手動で使用不能にされる
障害検出
使用不能となったプロバイダは手動で再度使用可能にされる

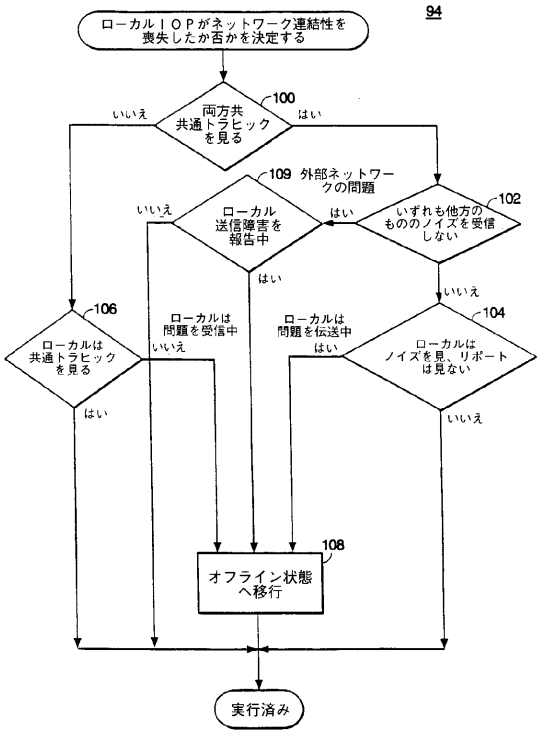
【図３】


オンライン及びオフライン状態で全ての NetworkStatusInterval
80
Receive Packets Interval内でいかなる非／ノイズパケットも受信されなかったか？
はい
いいえ
82
Receive Packets Interval内でいかなるノイズパケットも受信されなかったか？
はい
いいえ
84
オフライン状態？
はい
いいえ
86
ネットワーク状態要求をコプロバイダに送る
実行済み

【図４】


87
（ローカルプロバイダによりリモートプロバイダから）ネットワーク状態応答を受信した時点で
88
ローカル又はリモートプロバイダは完全初期化されていないか？
はい
90
リモートプロバイダは故障、使用不能又はオフライン状態にされているか？
はい
92
ローカルプロバイダは故障又は使用不能状態にされたか？
はい
94
ローカルプロバイダがネットワーク連結性を喪失したか否かを決定する
96
要求を放棄する
実行済み

【図５】


94
ローカルＩＯＰがネットワーク連結性を喪失したか否かを決定する
100
両方共共通トラヒックを見る
いいえ
はい
109 外部ネットワークの問題
ローカル送信障害を報告中
いいえ
はい
102
いずれも他方のもののノイズを受信しない
はい
いいえ
106
ローカルは共通トラヒックを見る
はい
ローカルは問題を受信中
いいえ
ローカルは問題を伝送中
はい
104
ローカルはノイズを見、リポートは見ない
いいえ
108
オフライン状態へ移行
実行済み

フロントページの続き

(74)代理人 100096194
弁理士 竹内 英人
(74)代理人 100074228
弁理士 今城 俊夫
(74)代理人 100084009
弁理士 小川 信夫
(74)代理人 100082821
弁理士 村社 厚夫
(74)代理人 100086771
弁理士 西島 孝喜
(74)代理人 100084663
弁理士 箱田 篤
(72)発明者 ロード クリストファー シー
アメリカ合衆国 マサチューセッツ州 ０１５４１ プリンストン ハーバードストン ロード ２９４
(72)発明者 シュウォルツ ディヴィッド ビー
アメリカ合衆国 マサチューセッツ州 ０１７０１ フレミンガム フラナガン ドライヴ ８８

審査官 脇水 佳弘

(56)参考文献 特開平０８－０４４６３７（ＪＰ，Ａ）
特開平０８－２５６１５８（ＪＰ，Ａ）
特開平０９－１６２９１０（ＪＰ，Ａ）
特開平０７－２３５９３３（ＪＰ，Ａ）
特開平０８－２６５３２０（ＪＰ，Ａ）
特開平０８－０９５８１６（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
H04L 12/00-66
G06F 11/16