

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5204603号
(P5204603)

(45) 発行日 平成25年6月5日(2013.6.5)

(24) 登録日 平成25年2月22日(2013.2.22)

(51) Int.Cl.

F I

G 0 6 F 11/18 (2006.01)

G 0 6 F 11/18 3 1 0 F

請求項の数 10 (全 17 頁)

(21) 出願番号	特願2008-249707 (P2008-249707)	(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
(22) 出願日	平成20年9月29日(2008.9.29)	(74) 代理人	100064414 弁理士 磯野 道造
(65) 公開番号	特開2010-79789 (P2010-79789A)	(74) 代理人	100111545 弁理士 多田 悦夫
(43) 公開日	平成22年4月8日(2010.4.8)	(72) 発明者	西村 卓真 茨城県日立市大みか町五丁目2番1号 株式会社日立製作所 情報制御システム事業部内
審査請求日	平成22年11月11日(2010.11.11)	(72) 発明者	中野 義弘 茨城県日立市大みか町五丁目2番1号 株式会社日立製作所 情報制御システム事業部内

最終頁に続く

(54) 【発明の名称】 4重化コンピュータシステムおよび2重化リングネットワーク

(57) 【特許請求の範囲】

【請求項1】

ストアアンドフォワード型の2重のリングネットワークによって相互に接続され、所定の同期化データを相互に送受信して処理の同期化を行う4台のコンピュータを含んで構成された4重化コンピュータシステムであって、

前記2重のリングネットワークは、互いに独立に動作する第1のリングネットワークと第2のリングネットワークとにより構成され、

前記第1のリングネットワークは、前記4台のコンピュータを4辺形の各頂点に配置したとしたとき、その各頂点に配置された各コンピュータを、前記4辺形の各辺に沿って形成された通信路によって相互に接続するように構成され、

前記第2のリングネットワークは、前記4辺形の各頂点に配置された各コンピュータを、前記4辺形の4辺のうち互いに対向する2辺に沿って形成された通信路、および、前記4辺形の2つの対角線に沿って形成された通信路によって相互に接続するように構成され、

前記第2のリングネットワークのブロッキングは、前記第2のリングネットワークの、前記4辺形の前記互いに対向する2辺に沿って形成された通信路のいずれか一方の通信路部分に設けられ、

前記第1のリングネットワークのブロッキングは、前記第2のリングネットワークのブロッキングが設けられた通信路が沿う前記4辺形の辺に対向する辺に沿って形成された、前記第1のリングネットワークの通信路部分に設けられること

10

20

を特徴とする4重化コンピュータシステム。

【請求項2】

前記4台のコンピュータのそれぞれは、前記処理同期化を行う場合には、あらかじめ定められた同期化データを前記第1のリングネットワークおよび前記第2のリングネットワークの両方へ送出するとともに、

前記4台のコンピュータのうち自身を除く他のコンピュータのそれぞれから、前記2重のリングネットワークを介してそれぞれ送信されてくる前記同期化データのうち、先に着信した前記同期化データのみを受信し、

前記自身を除く他のコンピュータのすべてから前記同期化データを受信し、前記受信した同期化データと、自身が生成した同期化データとが同じ、または、所定の関係を満たすとき、前記処理の同期化を行うこと

を特徴とする請求項1に記載の4重化コンピュータシステム。

【請求項3】

前記4辺形の前記互いに対向する2辺のうちの1辺に沿って形成された、前記第2のリングネットワークの通信路部分、および、前記1辺に対向する辺に沿って形成された、前記第1のリングネットワークの通信路部分の少なくとも一方の通信路部分に、前記4台のコンピュータとは別のコンピュータまたはネットワークスイッチからなる1台以上のノード装置が接続されること

を特徴とする請求項1に記載の4重化コンピュータシステム。

【請求項4】

ストアアンドフォワード型の2重のリングネットワークによって相互に接続され、所定の同期化データを相互に送受信して処理の同期化を行う4台のコンピュータを含んで構成された4重化コンピュータシステムであって、

前記4台のコンピュータは、第1のコンピュータと第2のコンピュータと第3のコンピュータと第4のコンピュータとにより構成され、

前記2重のリングネットワークは、互いに独立に動作する第1のリングネットワークと第2のリングネットワークとにより構成され、

前記第1のリングネットワークは、前記第1のコンピュータと前記第2のコンピュータとを接続する通信路、前記第2のコンピュータと前記第3のコンピュータとを接続する通信路、前記第3のコンピュータと前記第4のコンピュータとを接続する通信路、および、前記第4のコンピュータと前記第1のコンピュータとを接続する通信路によって構成され、

前記第2のリングネットワークは、前記第2のコンピュータと前記第4のコンピュータとを接続する通信路、前記第4のコンピュータと前記第1のコンピュータとを接続する通信路、前記第1のコンピュータと前記第3のコンピュータとを接続する通信路、および、前記第3のコンピュータと前記第2のコンピュータとを接続する通信路によって構成され、

前記第1のリングネットワークのブロッキングは、前記第1のリングネットワークの、前記第4のコンピュータと前記第1のコンピュータとを接続する通信路部分に設けられ、

前記第2のリングネットワークのブロッキングは、前記第2のリングネットワークの、前記第3のコンピュータと前記第2のコンピュータとを接続する通信路部分に設けられること

を特徴とする4重化コンピュータシステム。

【請求項5】

前記4台のコンピュータのそれぞれは、前記処理同期化を行う場合には、あらかじめ定められた同期化データを前記第1のリングネットワークおよび前記第2のリングネットワークの両方へ送出するとともに、

前記4台のコンピュータのうち自身を除く他のコンピュータのそれぞれから、前記2重のリングネットワークを介してそれぞれ送信されてくる前記同期化データのうち、先に着信した前記同期化データのみを受信し、

前記自身を除く他のコンピュータのすべてから前記同期化データを受信し、前記受信した同期化データと、自身が生成した同期化データとが同じ、または、所定の関係を満たすとき、前記処理の同期化を行うこと

を特徴とする請求項 4 に記載の 4 重化コンピュータシステム。

【請求項 6】

前記第 1 のリングネットワークの、前記第 4 のコンピュータと前記第 1 のコンピュータとを接続する通信路部分、および、前記第 2 のリングネットワークの、前記第 3 のコンピュータと前記第 2 のコンピュータとを接続する通信路部分、の少なくとも一方の通信路部分に、前記 4 台のコンピュータとは別のコンピュータまたはネットワークスイッチからなる 1 台以上のノード装置が接続されること

10

を特徴とする請求項 4 に記載の 4 重化コンピュータシステム。

【請求項 7】

所定の同期化データを相互に送受信して処理の同期化を行う 4 台のコンピュータを含んで構成された 4 重化コンピュータシステムにおける、前記 4 台のコンピュータを相互に接続するストアアンドフォワード型の 2 重化リングネットワークであって、

前記 2 重化リングネットワークは、互いに独立に動作する第 1 のリングネットワークと第 2 のリングネットワークとにより構成され、

前記第 1 のリングネットワークは、前記 4 台のコンピュータを 4 辺形の各頂点に配置したとしたとき、その各頂点に配置された各コンピュータを、前記 4 辺形の各辺に沿って形成された通信路によって相互に接続するように構成され、

20

前記第 2 のリングネットワークは、前記 4 辺形の各頂点に配置された各コンピュータを、前記 4 辺形の 4 辺のうち互いに対向する 2 辺に沿って形成された通信路、および、前記 4 辺形の 2 つの対角線に沿って形成された通信路によって相互に接続するように構成され

前記第 2 のリングネットワークのブロッキングは、前記第 2 のリングネットワークの、前記 4 辺形の前記互いに対向する 2 辺に沿って形成された通信路のいずれか一方の通信路部分に設けられ、

前記第 1 のリングネットワークのブロッキングは、前記第 2 のリングネットワークのブロッキングが設けられた通信路が沿う前記 4 辺形の辺に対向する辺に沿って形成された、前記第 1 のリングネットワークの通信路部分に設けられること

30

を特徴とする 2 重化リングネットワーク。

【請求項 8】

前記 4 辺形の前記互いに対向する 2 辺のうちの 1 辺に沿って形成された、前記第 2 のリングネットワークの通信路部分、および、前記 1 辺に対向する辺に沿って形成された、前記第 1 のリングネットワークの通信路部分の少なくとも一方の通信路部分に、前記 4 台のコンピュータとは別のコンピュータまたはネットワークスイッチからなる 1 台以上のノード装置が接続されること

を特徴とする請求項 7 に記載の 2 重化リングネットワーク。

【請求項 9】

所定の同期化データを相互に送受信して処理の同期化を行う 4 台のコンピュータを含んで構成された 4 重化コンピュータシステムにおける、前記 4 台のコンピュータを相互に接続するストアアンドフォワード型の 2 重化リングネットワークであって、

40

前記 4 台のコンピュータは、第 1 のコンピュータと第 2 のコンピュータと第 3 のコンピュータと第 4 のコンピュータとにより構成され、

前記 2 重化リングネットワークは、互いに独立に動作する第 1 のリングネットワークと第 2 のリングネットワークとにより構成され、

前記第 1 のリングネットワークは、前記第 1 のコンピュータと前記第 2 のコンピュータとを接続する通信路、前記第 2 のコンピュータと前記第 3 のコンピュータとを接続する通信路、前記第 3 のコンピュータと前記第 4 のコンピュータとを接続する通信路、および、前記第 4 のコンピュータと前記第 1 のコンピュータとを接続する通信路によって構成され

50

前記第2のリングネットワークは、前記第2のコンピュータと前記第4のコンピュータとを接続する通信路、前記第4のコンピュータと前記第1のコンピュータとを接続する通信路、前記第1のコンピュータと前記第3のコンピュータとを接続する通信路、および、前記第3のコンピュータと前記第2のコンピュータとを接続する通信路によって構成され

前記第1のリングネットワークのブロッキングは、前記第1のリングネットワークの、前記第4のコンピュータと前記第1のコンピュータとを接続する通信路部分に設けられ、前記第2のリングネットワークのブロッキングは、前記第2のリングネットワークの、前記第3のコンピュータと前記第2のコンピュータとを接続する通信路部分に設けられること

10

を特徴とする2重化リングネットワーク。

【請求項10】

前記第1のリングネットワークの、前記第4のコンピュータと前記第1のコンピュータとを接続する通信路部分、および、前記第2のリングネットワークの、前記第3のコンピュータと前記第2のコンピュータとを接続する通信路部分、の少なくとも一方の通信路部分に、前記4台のコンピュータとは別のコンピュータまたはネットワークスイッチからなる1台以上のノード装置が接続されること

を特徴とする請求項9に記載の2重化リングネットワーク。

【発明の詳細な説明】

20

【技術分野】

【0001】

本発明は、所定の処理を相互に同期化しながら実行する4重化コンピュータシステム、および、その4重化コンピュータシステムを構成するコンピュータを相互に接続する2重化リングネットワークに関する。

【背景技術】

【0002】

コンピュータシステムの高信頼化を図る方法として、3台以上のコンピュータに同じ処理を実行させ、それぞれのコンピュータの処理によって得られた結果について、その多数決をとる方法が知られている。このような多数決の機能を有する多重化コンピュータシステムにおいて問題となるのは、多数決をとるタイミングをいかにして同期化するか、つまり、処理の同期化である。

30

【0003】

特許文献1には、クロック同期バスを有し、そのクロック同期バスを用いて、それぞれのコンピュータの動作をクロックレベルで同期化する3重化コンピュータシステムの例が開示されている。その3重化コンピュータシステムにおいては、それぞれのコンピュータは、同期化されたクロックで動作し、適宜、その処理結果を多数決回路へ出力している。従って、多数決回路は、それぞれのコンピュータの処理の結果について、いつでも多数決をとることができる。すなわち、特許文献1に開示された3重化コンピュータシステムにおいては、処理の同期化の問題は生じないように見える。

40

【0004】

しかしながら、特許文献1に開示された3重化コンピュータシステムにおいては、それぞれのコンピュータの動作がクロックレベルで同期化されるため、1のコンピュータが処理結果を出力するタイミングと、他の2つのコンピュータが処理結果を出力するタイミングと、が1クロックでもずれた場合には、その1のコンピュータは誤動作している（障害がある）とみなされる。従って、1のコンピュータ内で、例えば、メモリのエラー訂正などによる処理のクロックずれが生じた場合にも、誤動作しているとみなされる。つまり、特許文献1に開示された3重化コンピュータシステムにおける同期化法は、同期化するタイミングが厳格なため、メモリのエラー訂正などさえもできなくなり、かえって融通のきかないものとなっている。

50

【 0 0 0 5 】

また、特許文献 1 に開示された 3 重化コンピュータシステムに限らず、一般の 3 重化コンピュータシステムの場合には、いずれか 1 のコンピュータに障害が生じた場合には、多数決による処理結果の信頼性を確保することができなくなる。同様に、1 のコンピュータを保守するために停止した場合にも、多数決による処理結果の信頼性を確保することができない。

【特許文献 1】特開平 9 - 1 6 5 3 5 号公報

【発明の開示】

【発明が解決しようとする課題】

【 0 0 0 6 】

そこで、本発明においては、互いに非同期で動作する汎用のパソコンなどを用いて、4 重化コンピュータシステムを構築することを想定する。4 重化コンピュータシステムの場合には、1 のコンピュータが障害または保守のために停止した場合でも、残りの 3 台のコンピュータの処理結果により多数決をとることができる。また、非同期で動作するコンピュータ同士での処理の同期化は、プログラムの処理で行うことになるので、処理の同期ずれを如何様にも一致させることができる。つまり、処理の同期化の融通性を高めることができる。

【 0 0 0 7 】

その一方で、このような 4 重化コンピュータシステムには、新たな課題が発生する。第 1 に、処理の同期化をプログラムで行うため、その所要時間が長くなりがちになる。従って、処理の同期化に要する時間を、如何にして短縮するかという課題が生じる。

【 0 0 0 8 】

また、第 2 に、4 重化コンピュータシステムを構成する 4 台のコンピュータのうち、1 台のコンピュータに障害が生じた場合に、どのような通信路の構成にすれば、残り 3 台のコンピュータ間で同期化を行うための同期化データを相互に送受信することができ、多数決を行うことが可能となるか、つまり、より信頼性の高い 4 重化コンピュータシステムをどのように構築するかという課題が生じる。

【 0 0 0 9 】

本発明は、以上の課題を解決するためになされたものであり、その発明の目的は、4 重化コンピュータシステムにおいて、処理の同期化の所要時間を短縮することができ、さらに、その 4 重化コンピュータシステムを構成する 4 台のコンピュータのうち、1 台のコンピュータに障害が生じた場合であっても、残りの 3 台のコンピュータ間で継続して処理の同期化および多数決が可能であるような高信頼の 4 重化コンピュータシステムおよび 2 重化リングネットワークを提供することにある。

【課題を解決するための手段】

【 0 0 1 0 】

本発明は、前記発明の目的を達成するために、4 重化コンピュータシステムを構成する 4 台のコンピュータを、ストアアンドフォワード型の 2 重のリングネットワークによって相互に接続する。そして、その 2 重のリングネットワークを、互いに独立に動作する第 1 のリングネットワークと第 2 のリングネットワークとによって構成する。このとき、前記 4 台のコンピュータを 4 辺形の各頂点に配置したとしたとき、その各頂点に配置された各コンピュータを、前記 4 辺形の各辺に沿った通信路によって相互に接続して、第 1 のリングネットワークを構成する。また、前記 4 辺形の各頂点に配置された各コンピュータを、前記 4 辺形の 4 辺のうち互いに対向する 2 辺に沿って形成された通信路、および、前記 4 辺形の 2 つの対角線に沿って形成された通信路によって相互に接続して、第 2 のリングネットワークを構成する。そして、第 2 のリングネットワークのブロッキングを、前記第 2 のリングネットワークの、前記 4 辺形の前記互いに対向する 2 辺に沿って形成された通信路のいずれか一方の通信路部分に設け、第 1 のリングネットワークのブロッキングを、前記第 2 のリングネットワークのブロッキングが設けられた通信路が沿う前記 4 辺形の辺に対向する辺に沿って形成された、前記第 1 のリングネットワークの通信路部分に設ける

10

20

30

40

50

〇

【 0 0 1 1 】

以上のようにストアアンドフォワード型の2重のリングネットワークで4台のコンピュータを相互に接続した場合には、詳細は実施形態で説明するように、それぞれのコンピュータは、2重のリングネットワークの少なくとも一方を介して、他の3台のコンピュータすべてが隣接することになる。そのため、4台のコンピュータ間で処理の同期化を行うための同期化データを他の3台のコンピュータすべてから直接に、つまり、途中で、さらに他のコンピュータを介在させることなく受信することができる。そのため、それぞれのコンピュータは、処理の同期化の所要時間を短縮化することができる。

【 0 0 1 2 】

また、以上のように構成したストアアンドフォワード型の2重のリングネットワークの場合には、4台のコンピュータのうち、1台のコンピュータに障害が生じた場合であっても、2重のリングネットワークの少なくとも一方のネットワークを介して、3台のコンピュータ間の接続が保たれる。従って、その3台のコンピュータ間で同期化データの送受信を行うことができるので、その3台のコンピュータの処理結果について、多数決をとることができる。

【 0 0 1 3 】

ここで、ストアアンドフォワード型のリングネットワークとは、コンピュータ、ルータ、スイッチングハブなど、いわゆるノード装置を通信路で相互に接続し、その通信路全体がリング状を呈するネットワークであって、そのネットワークの各ノード装置が、送信されてきたパケットを受信、いったんバッファリング（一時記憶）した後、そのパケットを隣接するノード装置へ送信するように構成されたネットワークをいう。

【 発明の効果 】

【 0 0 1 4 】

本発明の4重化コンピュータシステムによれば、処理の同期化の所要時間を短縮することができる。さらには、その4重化コンピュータシステムを構成する4台のコンピュータのうち、1台のコンピュータに障害が生じた場合であっても、残りの3台のコンピュータ間で継続して処理の同期化および多数決が可能であるような高信頼性を実現することができる。

【 発明を実施するための最良の形態 】

【 0 0 1 5 】

以下、本発明の実施形態について、図面を参照して詳細に説明する。

【 0 0 1 6 】

図1は、本発明の実施形態に係る4重化コンピュータシステムおよび2重化リングネットワークの構成の例を示した図である。

【 0 0 1 7 】

図1(a)に示すように、4重化コンピュータシステム7は、4台のコンピュータ1によって構成され、それぞれのコンピュータ1は、CPU(Central Processing Unit)11、メインメモリ12および2つのNIC(Network Interface Card)13を含んで構成される。また、これら4台のコンピュータ1は、2重のリングネットワークで相互に接続され、図1(a)では、その第1のリングネットワークは、ネットワーク(#1)2aとして太い実線で描かれ、第2のリングネットワークは、ネットワーク(#2)2bとして太い破線で描かれている。

【 0 0 1 8 】

さらに、図1(a)に示すように、各コンピュータ1に含まれる2つのNIC13のそれぞれは、2つの送受信ポート131を備えている。送受信ポート131は、図1(a)では、NIC13を表す長方形ブロックの上辺または下辺に小さな正方形の端子として描かれている。

【 0 0 1 9 】

図

10

20

30

40

50

また、各コンピュータ1の一方(図1(a)では、左側)のNIC13の2つの送受信ポート131には、太い実線で描かれたネットワーク(#1)2aが接続され、他方(図1(a)では、右側)のNIC13の2つの送受信ポート131には、太い破線で描かれたネットワーク(#2)2bが接続される。

【0020】

図1(b)、(c)は、図1(a)における4台のコンピュータの接続関係を取り出して示した略図である。図1(b)に示すように、第1のリングネットワークであるネットワーク(#1)2aは、4台のコンピュータ1を(#1) - (#2) - (#3) - (#4) - (#1)の順でリング状に接続する。また、図1(c)に示すように、第2のリングネットワークであるネットワーク(#2)2bは、4台のコンピュータ1を(#2) - (#4) - (#1) - (#3) - (#2)の順でリング状に接続する。

10

【0021】

また、図1(a)に示すように、ネットワーク(#1)2aの、コンピュータ(#4) - (#1)間をつなぐ通信路の両端には、ブロッキング3が設けられ、また、同様に、ネットワーク(#2)2bの、コンピュータ(#3) - (#2)間をつなぐ通信路の両端には、ブロッキング3が設けられる。なお、本明細書では、リングネットワークのうち、2つのコンピュータ1をつなぐ部分を取り出していう場合には、通信路と呼ぶ。

【0022】

ここで、ブロッキング3とは、ブロッキング3に接する送受信ポート131の packets 送受信機能を停止させることを意味するマークであり、一般のリングネットワークでも、 packets がそのリングネットワークを複数周回するのを防止するために設けられる。このようなブロッキング3は、そのブロッキング3に接する送受信ポート131を備えたNIC13が、その送受信ポート131の packets 送受信機能を停止させることによって実現される。このように、ブロッキング3を設定するために、 packets 送受信機能を停止させた送受信ポート131をブロッキングポートという。

20

【0023】

次に、本実施形態に係る4重化コンピュータシステム7で用いられる2重のリングネットワーク(ネットワーク(#1)2aおよびネットワーク(#2)2b)における4つのコンピュータ1の接続構成の特徴について説明する。

【0024】

本実施形態では、図1(a)~(c)に示したように、4重化コンピュータシステム7を構成する4つのコンピュータ1を、4辺形の各頂点の配置した場合を想定する。その場合、第1のリングネットワークであるネットワーク(#1)2aは、図1(b)に示すように、4辺形の各辺に沿った通信路によって、それぞれのコンピュータ1を接続するように構成される。また、第2のリングネットワークであるネットワーク(#2)2bは図1(c)に示すように、その4辺形の4辺のうち対向する2辺に沿った通信路、および、前記4辺形の2つの対角線に沿った通信路により、それぞれのコンピュータ1を接続するように構成される。

30

【0025】

以上の構成において、ネットワーク(#1)2aは、4台のコンピュータ1を、まさにリング状(0字状)に接続しているが、ネットワーク(#2)2bは、4台のコンピュータ1を、横の8字状に接続している。

40

【0026】

また、この場合、ネットワーク(#2)2bのブロッキング3は、ネットワーク(#2)2bを構成する通信路部分であって、4辺形の前記対向する2辺に沿って設けられた通信路のうち、一方の通信路の通信路部分に設けられる。また、ネットワーク(#1)2aのブロッキング3は、ネットワーク(#1)2aを構成する通信路部分であって、ネットワーク(#2)2bが設けられた通信路部分に沿った辺に対向する辺に沿った通信路の通信路部分に設けられる。

【0027】

50

すなわち、図1(a)に示すように、ネットワーク(#2)2bのブロッキング3が、コンピュータ(#2)-(#3)をつなぐ辺に沿った通信路部分に設けられた場合には、ネットワーク(#1)2aのブロッキング3は、その辺に対向する辺に沿ったコンピュータ(#4)-(#1)をつなぐ通信路部分に設けられる。

【0028】

なお、図1においては、ネットワーク(#2)2bは、4台のコンピュータ1を、横の8字状に接続しているが、縦の8字状に接続するようにしてもよい。その場合には、ネットワーク(#2)2bを構成する通信路のうち、4台のコンピュータ1を頂点とする4辺形の辺に沿った通信路は、前記4辺形の上辺または下辺に沿った通信路になるので、その通信路の一方(例えば、コンピュータ(#3)-(#4)をつなぐ通信路)にブロッキング3が設けられる。また、ネットワーク(#1)2aのブロッキング3は、ネットワーク(#2)2bにブロッキングが設けられた通信路に対向する通信路(例えば、コンピュータ(#1)-(#2)をつなぐ通信路)に設けられる。

10

【0029】

以上、図1を用いて、4台のコンピュータ1を4辺形の各頂点に配置したことを想定して、その2重のリングネットワークの接続構成の特徴について説明したが、これは、4台のコンピュータ1の物理的な配置位置を限定するものではない。以上に説明した2重のリングネットワークの接続構成の特徴は、トポロジカルな接続関係を表したものであって、4台のコンピュータ1の物理的配置は、どのような配置であってもよい。

【0030】

20

続いて、4重化コンピュータシステム7において、以上のようにネットワーク(#1)2aおよびネットワーク(#2)2bを構成したことについての効果を説明する必要があるが、その前に、NIC13の内部構成およびコンピュータ1における処理の同期化の方法について説明する。

【0031】

図2は、本発明の実施形態に係る4重化コンピュータシステム7で用いられるコンピュータ1およびコンピュータ1に含まれるNIC13の構成の例を示した図である。図2に示すように、また、図1でもその構成の一部を示したように、コンピュータ1は、CPU11、メインメモリ12、2つのNIC13などが内部バス14を介して相互に接続されて構成される。このとき、コンピュータ1は、ハードディスク装置、表示装置、キーボード、各種のセンサ装置、制御装置などを、さらに含んでいてもよい。

30

【0032】

また、NIC13は、送受信ポート131、ポート接続スイッチ132、パケットバッファメモリ133、送受信制御部134などを含んで構成される。

【0033】

送受信ポート131は、以下、図示を省略するが、電氣的な送受信回路と、ネットワークケーブルを接続するためのコネクタジャックと、により構成される。そして、そのコネクタジャックには、例えば、イーサネット(登録商標)仕様の通信ケーブルが接続され、その通信ケーブルによって、2つのコンピュータ1をつなぐ通信路21a, 21b(図2参照)が形成される。

40

【0034】

なお、図2の通信路21aは、図1のネットワーク(#1)2aの一部を構成し、通信路21bは、ネットワーク(#2)2bの一部を構成する。従って、図2に示した左側のNIC13aは、ネットワーク(#1)2aにおける通信を制御し、また、右側のNIC13bは、ネットワーク(#2)2bにおける通信を制御する。

【0035】

図2において、ポート接続スイッチ132は、いわゆるL2スイッチに相当し、単なる通路スイッチとしての機能だけでなく、パケットの宛先アドレス判定や属性の判定などの機能、送受信ポート131をブロッキングポートとして動作させる機能などを備える。また、パケットバッファメモリ133は、送受信ポート131を介して受信されたパケット

50

を一時記憶するメモリであり、例えば、D R A M (Dynamic Random Access Memory)などで構成される。

【 0 0 3 6 】

送受信制御部 1 3 4 は、ポート接続スイッチ 1 3 2 やパケットバッファメモリ 1 3 3 を、適宜、制御することにより、パケットの送受信に関し、様々な制御を行う。例えば、受信したパケットが自身のコンピュータ 1 宛であったような場合には、送受信制御部 1 3 4 は、その旨を C P U 1 1 に通知し、その受信したパケットからヘッダ部などを取り除いた受信データをメインメモリ 1 2 へ転送する制御を行う。また、C P U 1 1 が送信データを出力する場合には、送受信制御部 1 3 4 は、その送信データにヘッダ部などを、適宜、付加して送信パケットを構成し、ポート接続スイッチ 1 3 2 を介して左右両方の送受信ポート 1 3 1 からその送信パケットを送出する制御を行う。さらに、送受信制御部 1 3 4 は、他のコンピュータ 1 の障害などに伴うブロッキング 3 の設定変更処理やその設定変更処理時の送受信機能の停止などを制御する機能も有している。

10

【 0 0 3 7 】

図 2 において、メインメモリ 1 2 は、同期化データ領域 1 2 1 を有している。ここで、同期化データ領域 1 2 1 は、自身のコンピュータ 1、つまり、C P U 1 1 によって生成された同期化データ (例えば、同期化データ # 1) と、他の 3 台のコンピュータ 1 から送信されてきた同期化データ (例えば、同期化データ # 2 ~ # 4) と、から構成される。これらの同期化データ # 1 ~ # 4 の扱いについては、次に、図 3 を参照して説明する。

【 0 0 3 8 】

20

図 3 は、各コンピュータ 1 における処理の同期化のフローの例を示した図である。ここで、処理の同期化とは、同じ処理のプログラムを実行する複数のコンピュータ間で、プログラムの実行進度の一致化を行うことをいう。本実施形態では、4 重化コンピュータシステム 7 を構成する 4 台のコンピュータ 1 (図 1 参照) の間で処理の同期化を行うが、これら 4 台のコンピュータ 1 は、互いに非同期のクロックで動作しているものとする。そして、その処理の同期化を、所定の 1 まとまりの処理 (以下、タスクという) を終了するたびにを行うものとする。

【 0 0 3 9 】

図 3 に示すように、C P U 1 1 は、あるタスク A を終了すると (ステップ S 0)、2 つの N I C 1 3 a, 1 3 b のそれぞれに含まれる送受信制御部 1 3 4 に対し、例えば、タスク A の識別番号などを含んでなる同期化データの送出手続きを指示する (ステップ S 1)。

30

【 0 0 4 0 】

N I C 1 3 a, 1 3 b のそれぞれに含まれる送受信制御部 1 3 4 は、C P U 1 1 から同期化データの送出手続きを受けると、指示された同期化データを、所定の形式のパケットにし、ポート接続スイッチ 1 3 2 および 2 つの送受信ポート 1 3 1 を介して、通信路 2 1 a, 2 1 b へ送出手続きする (ステップ S 1 1)。従って、各コンピュータ 1 からは、同じ同期化データがネットワーク (# 1) 2 a およびネットワーク (# 2) 2 b のそれぞれに送出手続きされる。

【 0 0 4 1 】

なお、同期化データを含んだパケットは、ブロードキャストパケットとして送出手続きされるが、そのパケットには、そのパケットが同期化データを含むパケットであることを識別する情報が含まれているものとする。

40

【 0 0 4 2 】

N I C 1 3 a, 1 3 b のそれぞれの送受信制御部 1 3 4 は、他方で、他のコンピュータ 1 から送信される同期化データの受信を待ち、他のコンピュータ 1 から送信された同期化データを受信すると (ステップ S 1 2)、その受信した同期化データを送信元のコンピュータ 1 に対応する同期化データ領域に格納する (ステップ S 1 3)。例えば、受信した同期化データの送信元がコンピュータ (# 2) であれば、同期化データ領域 1 2 1 の同期化データ # 2 の部分に格納する。

【 0 0 4 3 】

50

なお、送受信制御部 134 が受信した同期化データを同期化データ領域 121 に格納する動作は、例えば、DMA (Direct Memory Access) によるものとする。

【0044】

前記したように、各コンピュータ 1 からは、同じ同期化データが通信路 21a, 21b の両方に送られるので、それを受信するコンピュータ 1 においては、NIC 13a, 13b のそれぞれの送受信制御部 134 が、送信元が同じ同期化データをそれぞれ独立に受信し、それぞれのパケットバッファメモリ 133 に格納する。そして、それぞれの送受信制御部 134 は、同期化データが先着した場合に、その同期化データを同期化データ領域 121 へ転送し、後着した場合には、その同期化データを廃棄する。

【0045】

この動作を実現するには、例えば、同期化データ領域 121 の各同期化データ (ただし、自身のコンピュータ 1 に対応する同期化データを除く) に対し、着信済みを表すフラグを設けておき、送受信制御部 134 が着信済みフラグを読み取って、その着信済みフラグがセットされていない場合には、ステップ S13 を実行し、着信済みフラグがセットされている場合には、ステップ S13 をスキップするようにすればよい。

【0046】

以上のようにして他のコンピュータ 1 から同期化データを受信する一方で、CPU 11 は、同期化データ領域 121 に格納された同期化データの検証を行う (ステップ S2)。ここで、同期化データの検証とは、他のコンピュータ 1 から送信されてきた同期化データが有効であるか否かを判定する処理である。例えば、他のコンピュータ 1 から送信されてきた同期化データと、自身のコンピュータ 1 の処理によって生成した同期化データと、が同じ、または、あらかじめ定められた所定の関係を満たす場合には、その同期化データは有効であるとする。なお、所定の関係とは、例えば、同期化データ同士の剰余コードやハッシュが同じになる関係であったり、同じ数列関係であったりする場合をいう。

【0047】

そして、その検証の結果、他のすべてのコンピュータ 1 から有効な同期化データを受信していた場合には (ステップ S3 で Yes)、次のタスクを起動する (ステップ S4)。このとき、次のタスクがタスク B であった場合には、CPU 11 は、タスク B の処理を開始する (ステップ S5)。また、他のすべてのコンピュータ 1 から、未だ、有効な同期化データを受信していない場合には (ステップ S3 で No)、再度、同期化データの検証を行う (ステップ S2)。

【0048】

なお、ここでいう「すべてのコンピュータ 1」とは、4重化コンピュータシステム 7 を構成する 4 台のコンピュータ 1 がすべて正常に稼働している場合には、自身を除く他のコンピュータ 3 台を指し、4 台のコンピュータのうち、障害のため、いずれか 1 台が稼働していない場合には、自身を除く稼働中の他のコンピュータ 2 台を指す。

【0049】

続いて、図 4 ~ 図 6 を参照して、以上のように構成した 4 重化コンピュータシステム 7 が奏する効果について説明する。

【0050】

図 4 は、図 1 に示した 4 重化コンピュータシステム 7 における 4 台のコンピュータ 1 の接続構成の例を示した図であり、(a) は、ネットワーク (#1) 2a の接続順に従ってコンピュータ 1 の配列を示した図、(b) は、ネットワーク (#2) 2b の接続順に従ってコンピュータ 1 の配列を示した図である。すなわち、図 4 (a) と図 4 (b) とでは、コンピュータ 1 の表示上の配列が異なるだけであって、いずれも図 1 に示した 4 重化コンピュータシステム 7 におけるコンピュータ 1 の接続構成を表したものである。なお、図 4 (a), (b) において、ブロッキング 3 の位置も、図 1 に示した 4 台のコンピュータ 1 の接続構成におけるブロッキング 3 の位置に一致させている。

【0051】

以上のように、図 4 (a), (b) は、いずれも同じものを表したものであるが、図 4

10

20

30

40

50

(a) からは、4台のコンピュータ1がネットワーク(# 1) 2 a によってコンピュータ(# 1) - (# 2) - (# 3) - (# 4) の順序でループ状に接続されていることが容易に分かり、図4(b) からは、4台のコンピュータ1がネットワーク(# 2) 2 b によってコンピュータ(# 2) - (# 4) - (# 1) - (# 3) の順序でループ状に接続されていることが容易に分かる。

【 0 0 5 2 】

図5は、図1に示した4重化コンピュータシステム7を構成するそれぞれのコンピュータ1が、2重のリングネットワーク(ネットワーク(# 1) 2 a およびネットワーク(# 2) 2 b) を介して相互に隣接していることを示した図である。

【 0 0 5 3 】

図5では、図4での説明に基づき、コンピュータ1の接続関係をさらに簡略化して表示している。すなわち、図5(a) ~ (d) の各図において、その上段の図は、ネットワーク(# 1) 2 a の接続順にコンピュータ1の配列を表した図、その下段の図は、ネットワーク(# 2) 2 b の接続順にコンピュータ1の配列を表した図である。従って、その上段および下段のコンピュータ1の配列において、同じ#番号が付されたコンピュータ1は、実体として同一のコンピュータ1を表している。

【 0 0 5 4 】

図5(a) において、印を付したコンピュータ(# 1) は、基準となるコンピュータ1を表し、また、上に開いた波括弧({) は、基準となるコンピュータ1に対し、隣接するコンピュータ1の範囲を示した図である。すなわち、図5(a) は、コンピュータ(# 1) がネットワーク(# 1) 2 a を介してコンピュータ(# 2) に隣接し、また、ネットワーク(# 2) 2 b を介してコンピュータ(# 3) および(# 4) に隣接していることを表している。

【 0 0 5 5 】

同様に、図5(b) は、コンピュータ(# 2) がネットワーク(# 1) 2 a を介してコンピュータ(# 1) および(# 3) に隣接し、また、ネットワーク(# 2) 2 b を介してコンピュータ(# 4) に隣接していることを表している。また、図5(c) は、コンピュータ(# 3) がネットワーク(# 1) 2 a を介してコンピュータ(# 2) および(# 4) に隣接し、また、ネットワーク(# 2) 2 b を介してコンピュータ(# 1) に隣接していることを表している。また、図5(d) は、コンピュータ(# 4) がネットワーク(# 1) 2 a を介してコンピュータ(# 3) に隣接し、また、ネットワーク(# 2) 2 b を介してコンピュータ(# 1) および(# 2) に隣接していることを表している。

【 0 0 5 6 】

以上のように、本実施形態においては、4台のそれぞれのコンピュータ1は、ネットワーク(# 1) 2 a またはネットワーク(# 2) 2 b を介して他の3台のコンピュータ1のいずれにも隣接していることになる。

【 0 0 5 7 】

本実施形態では、前記したように、処理の同期化に際しての同期化データは、ネットワーク(# 1) 2 a およびネットワーク(# 2) 2 b の両方に送出され、また、それぞれのコンピュータ1は、先に受信した同期化データを利用して、処理の同期化を行う。すなわち、それぞれのコンピュータ1は、隣接する他の3台のコンピュータ1から同期化データを直接に受信することができる。

【 0 0 5 8 】

ここで、比較例の4重化コンピュータシステムを想定する。比較例の4重化コンピュータシステムでは、その4台のコンピュータを接続する2重のリンクネットワークは、単純な2重構成であるとする。すなわち、その第1のネットワークおよび第2のネットワークのいずれにおいても、その4台のコンピュータを接続する接続関係は同じであるとする。従って、その第1のネットワークおよび第2のネットワークは、いずれも、例えば、図5(a) 上段の図のように、コンピュータを(# 1) - (# 2) - (# 3) - (# 4) の順序で接続したリングネットワークとして表される。

10

20

30

40

50

【 0 0 5 9 】

この比較例の場合には、コンピュータ（# 1）は、コンピュータ（# 2）の同期化データを直接に受信することができるが、コンピュータ（# 3）の同期化データは、コンピュータ（# 2）を介して受信することになる。その場合、コンピュータ（# 3）から送信された同期化データは、コンピュータ（# 2）のバケットバッファメモリ 1 3 3 にいったん記憶された上で、コンピュータ（# 1）へ送信される。その結果、コンピュータ（# 3）から送信された同期化データは、コンピュータ（# 2）を経由する分だけ遅くコンピュータ（# 1）へ到着することになる。また、コンピュータ（# 4）から送信された同期化データは、コンピュータ（# 3）およびコンピュータ（# 2）を経由するので、さらに遅くコンピュータ（# 1）に到着する。

10

【 0 0 6 0 】

従って、この比較例の場合には、本発明の実施形態の場合に比べ、あるコンピュータが他の 3 台のコンピュータのすべての同期化データを受信するのに、長い時間を要していることになり、その分、処理の同期化に要する時間も長くなっていることになる。言い換えれば、本発明の実施形態により、処理の同期化に要する時間が短縮されたことになる。

【 0 0 6 1 】

図 6 は、図 1 の 4 重化コンピュータシステム 7 を構成する 4 台のコンピュータ 1 のうち 1 台のコンピュータ 1 に障害が発生したとき、他の 3 台のコンピュータ 1 が、2 重のリングネットワーク（ネットワーク（# 1）2 a およびネットワーク（# 2）2 b）の一方のネットワークを介して相互に接続されている状態にあることを示した図である。

20

【 0 0 6 2 】

図 6（a）において、×印を付したコンピュータ（# 1）は、障害が発生したコンピュータ 1 を表し、上に開いた波括弧（{）は、3 台のコンピュータ 1 が継続して接続された状態にあることを表している。なお、ここでいうコンピュータ（# 1）の障害に程度は、重度のものであり、CPU 1 1 からすべての送受信ポート 1 3 1 まで一切の機能が失われたような障害であるとする。

【 0 0 6 3 】

図 6（a）に示すように、コンピュータ（# 1）に障害が発生した場合、ネットワーク（# 1）2 a では、コンピュータ（# 2） - （# 3） - （# 4）の接続が維持されるが、ネットワーク（# 2）2 b では、その接続が障害のコンピュータ（# 1）によって切断される。しかしながら、コンピュータ（# 2）、（# 3）、（# 4）は、ネットワーク（# 1）2 a による接続を介して、同期化データを相互に送受信することができるので、処理の同期化を行うことができる。

30

【 0 0 6 4 】

同様に、図 6（b）は、コンピュータ（# 2）に障害が発生した場合であっても、ネットワーク（# 2）2 b により、コンピュータ（# 4） - （# 1） - （# 3）の接続が維持されていることを表している。また、図 6（c）は、コンピュータ（# 3）に障害が発生した場合であっても、ネットワーク（# 2）2 b により、コンピュータ（# 2） - （# 4） - （# 1）の接続が維持されていることを表している。また、図 6（d）は、コンピュータ（# 4）に障害が発生した場合であっても、ネットワーク（# 1）2 a により、コンピュータ（# 1） - （# 2） - （# 3）の接続が維持されていることを表している。

40

【 0 0 6 5 】

従って、いずれの場合にも、1 台のコンピュータ 1 に障害が発生しても、残りの 3 台のコンピュータ 1 の接続は、ネットワーク（# 1）2 a またはネットワーク（# 2）2 b により維持されたままの状態になるので、それら 3 台のコンピュータ 1 の間では、同期化データを相互に送受信することができるので、処理の同期化を行うことができる。

【 0 0 6 6 】

以上のように、本実施形態の 4 重化コンピュータシステム 7 においては、そのいずれかの 1 台のコンピュータ 1 に障害が発生した場合であっても、残りの 3 台の接続が維持されるため、その 3 台のコンピュータ 1 により処理の同期化を行うことができ、さらに、処理

50

の同期化をした時点で、処理結果の多数決をとることができる。

【0067】

ちなみに、前記した比較例の4重化コンピュータシステムにおいて、コンピュータ(#2)または(#3)に障害が発生した場合(図6(b),(c)の各上段の図を参照)には、残りの3台のコンピュータの接続が維持されない。従って、その残りの3台のコンピュータの間で、処理の同期化ができなくなり、また、処理結果の多数決をとることもできなくなる。

【0068】

なお、よく知られているように、リングネットワークにおいては、ブロッキングの位置を変更することによって、障害の発生したコンピュータを除外したネットワークを構成することができる。前記した比較例の場合にも、それが可能であり、障害の発生したコンピュータを除外するようにブロッキングの位置を変更すると、再び、残された3台のコンピュータにより処理の同期化ができ、処理結果の多数決をとることができるようになる。

【0069】

しかしながら、その場合には、ブロッキングの位置を変更するとき、わずかな時間ではあるが、そのネットワークにおける通信が不能となる。それに対し、本実施形態の場合には、2重のリングネットワークのうち、いずれか一方のリングネットワークでは、3台のコンピュータの接続が維持されたままの状態にあるので、他方のリングネットワークでブロッキングの位置を変更しても通信が不能になることはない。

【0070】

従って、以上の点において、本実施形態の4重化コンピュータシステム7は、比較例の4重化コンピュータシステムよりも高い信頼性を有しているといえる。

【0071】

図7は、図1に示した4重化コンピュータシステム7における2重化リングネットワークに対し、他のコンピュータやネットワークスイッチを接続する場合の例を示した図である。図7において、他のコンピュータは、コンピュータ(#5)1a、コンピュータ(#6)1bと表し、ネットワークスイッチは、LANSW(#1)4a、LANSW(#2)4bと表している。なお、ここでいうネットワークスイッチは、いわゆるスイッチングハブやルータを指す。

【0072】

図7に示すように、コンピュータ(#5)1aおよびLANSW(#1)4aは、ネットワーク(#1)1aの、コンピュータ(#4)およびコンピュータ(#1)をつなぐ通信路部分に接続される。この通信路部分は、図4(a)では、ブロッキング3により、通信可能なネットワークから除外されていた部分である。ただし、図7の場合には、コンピュータ(#5)1aおよびLANSW(#1)4aは、4重化コンピュータシステム7を構成するコンピュータ(#1)~(#4)との間で通信する必要があるので、ブロッキング3は、例えば、コンピュータ(#5)1aとLANSW(#1)4aとの間に設けられる(図示省略)。

【0073】

このように、コンピュータ(#5)1aおよびLANSW(#1)4aの少なくとも一方を、この通信路部分に接続する限りは、図5および図6を用いて説明した4重化コンピュータシステム7の効果は、喪失されない。

【0074】

同様に、コンピュータ(#6)1bおよびLANSW(#2)4bは、ネットワーク(#2)1bの、コンピュータ(#3)およびコンピュータ(#4)をつなぐ通信路部分に接続される。この通信路部分は、図4(a)では、ブロッキング3により、通信可能なネットワークから除外されていた部分である。ただし、図7の場合には、コンピュータ(#6)1bおよびLANSW(#2)4bは、4重化コンピュータシステム7を構成するコンピュータ(#1)~(#4)との間で通信する必要があるので、ブロッキング3は、例えば、コンピュータ(#6)1bとLANSW(#2)4bとの間に設けられる(図示省

10

20

30

40

50

略)。

【0075】

このように、コンピュータ(#6)1bおよびLANSW(#2)4bの少なくとも一方を、この通信路部分に接続する限りは、図5および図6を用いて説明した4重化コンピュータシステム7の効果は、喪失されない。

【図面の簡単な説明】

【0076】

【図1】本発明の実施形態に係る4重化コンピュータシステムおよび2重化リングネットワークの構成の例を示した図。

【図2】本発明の実施形態に係る4重化コンピュータシステムで用いられるコンピュータおよびそのコンピュータに含まれるNICの構成の例を示した図。

【図3】各コンピュータにおける処理の同期化のフローの例を示した図。

【図4】4重化コンピュータシステムにおける4台のコンピュータの接続構成の例を示した図。

【図5】4重化コンピュータシステムを構成するそれぞれのコンピュータが2重のリングネットワークを介して相互に隣接していることを示した図。

【図6】4重化コンピュータシステムを構成する4台のコンピュータのうち1台のコンピュータに障害が発生したとき、他の3台のコンピュータが、2重のリングネットワークの一方のネットワークを介して相互に接続されている状態にあることを示した図。

【図7】4重化コンピュータシステムにおける2重化リングネットワークに対し、他のコンピュータやネットワークスイッチを接続する場合の例を示した図。

【符号の説明】

【0077】

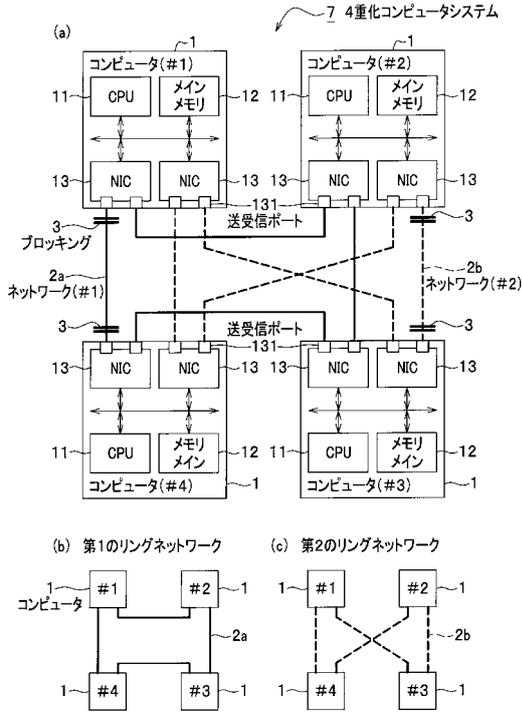
- 1 コンピュータ
- 3 ブロッキング
- 2 a ネットワーク(#1)：第1のリングネットワーク
- 2 b ネットワーク(#2)：第2のリングネットワーク
- 7 4重化コンピュータシステム
- 1 1 CPU
- 1 2 メインメモリ
- 1 3 NIC
- 1 4 内部バス
- 1 2 1 同期化データ領域
- 1 3 1 送受信ポート
- 2 1 a , 2 1 b 通信路
- 1 3 2 ポート接続スイッチ
- 1 3 3 パケットバッファメモリ
- 1 3 4 送受信制御部

10

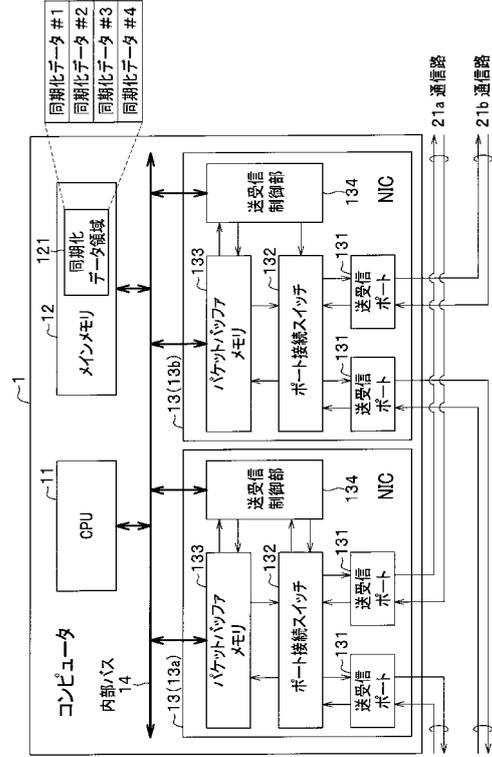
20

30

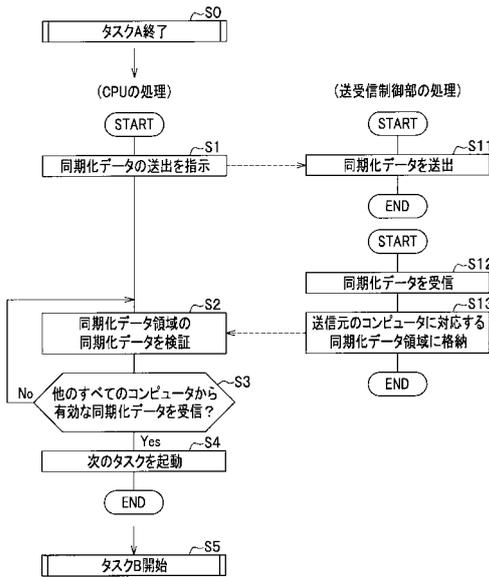
【図1】



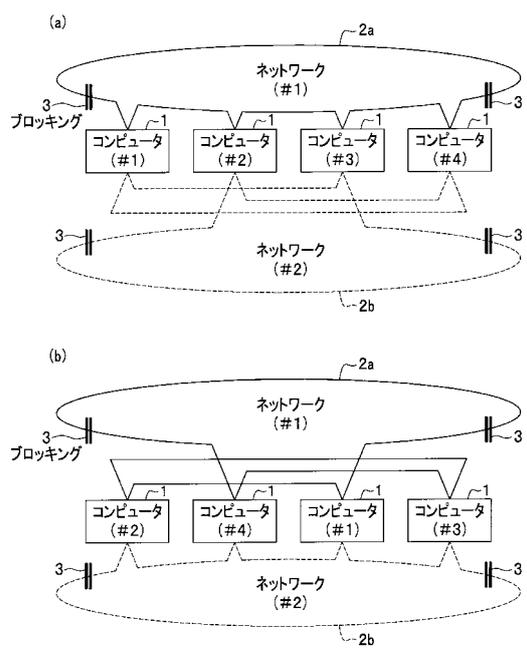
【図2】



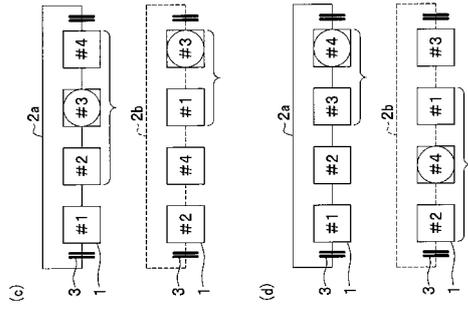
【図3】



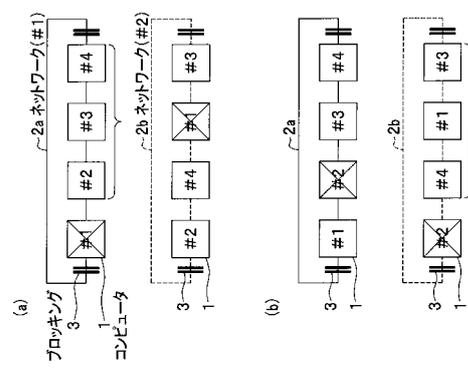
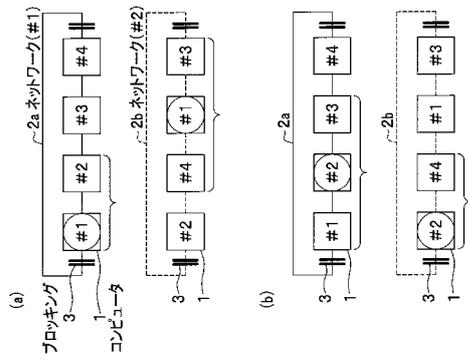
【図4】



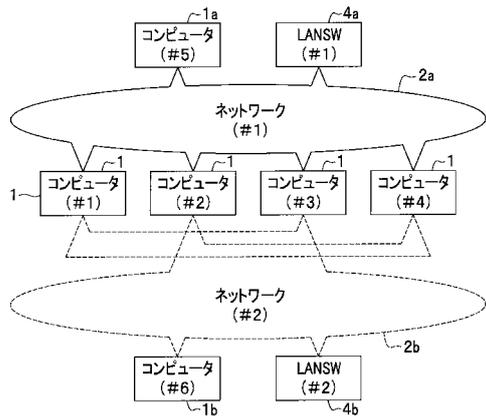
【図 5】



【図 6】



【図 7】



フロントページの続き

(72)発明者 益子 英昭

茨城県日立市大みか町五丁目2番1号 株式会社日立製作所 情報制御システム事業部内

審査官 鈴木 修治

(56)参考文献 特開平09-282296(JP,A)
特開平02-283142(JP,A)
特開平09-128348(JP,A)
特表平11-511634(JP,A)
特開2006-157952(JP,A)
特開2008-118696(JP,A)
特開平02-083744(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 11/18

H04L 12/42