(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4660362号
(P4660362)
(45)発行日　平成23年3月30日(2011.3.30)　(24)登録日　平成23年1月7日(2011.1.7)

(51)Int.Cl.　F I
G06F 13/14 (2006.01)　G06F 13/14 310H

請求項の数 10　(全 29 頁)

(21)出願番号　特願2005-340088(P2005-340088)
(22)出願日　平成17年11月25日(2005.11.25)
(65)公開番号　特開2007-148621(P2007-148621A)
(43)公開日　平成19年6月14日(2007.6.14)
審査請求日　平成20年9月19日(2008.9.19)

(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　100075513
弁理士　後藤　政喜
(74)代理人　100084537
弁理士　松田　嘉夫
(74)代理人　100114236
弁理士　藤井　正弘
(72)発明者　對馬　雄次
東京都国分寺市東恋ヶ窪一丁目２８０番地
株式会社日立製作所　中央研究所内
(72)発明者　森木　俊臣
東京都国分寺市東恋ヶ窪一丁目２８０番地
株式会社日立製作所　中央研究所内
最終頁に続く

(54)【発明の名称】計算機システム

(57)【特許請求の範囲】
【請求項１】
複数のサーバと、
該複数のサーバとＩ／Ｏカードを接続するスイッチと、を備え、
前記サーバが、前記Ｉ／Ｏカードから書き込み可能なメモリ空間を備えた計算機システムにおいて、
前記スイッチとＩ／Ｏカードの間に配置されて、前記複数のサーバと前記Ｉ／Ｏカードとの間のアクセス要求信号及び応答信号を操作するＩ／Ｏカード共有部と、
前記Ｉ／ＯカードとＩ／Ｏカード共有部にアクセス可能であって、前記Ｉ／Ｏカード共有部から書き込み可能なメモリと、前記Ｉ／Ｏカード共有部からの割り込みを受け付けるプロセッサとを備えて、前記複数のサーバに対する前記Ｉ／Ｏカードの割り当てを管理するＩ／Ｏプロセッサと、
を備え、
前記スイッチは、前記サーバからＩ／Ｏカードへの指令と前記メモリ空間を指定するベースアドレスとを含んだアクセス要求信号に、当該アクセス要求信号の宛先と要求元の経路情報を前記アクセス要求信号のヘッダ情報に付加して前記Ｉ／Ｏカード共有部へ送信し、前記Ｉ／Ｏカードからサーバへの応答と前記メモリ空間を指定するベースアドレスとを含んだ応答信号を、当該応答信号のヘッダ情報に含まれる宛先のサーバに転送するヘッダ処理部を有し、
前記Ｉ／Ｏカード共有部は、

前記スイッチから受信したアクセス要求信号に含まれる指令が、予め設定した第１の指令のときには前記Ｉ／Ｏカード共有部のメモリに当該アクセス要求信号を書き込む要求信号書込部と、
前記スイッチから受信した前記アクセス要求信号に含まれる指令が、予め設定した第２の指令のときには前記Ｉ／Ｏプロセッサへ割り込みをかける割り込み発生部と、
前記Ｉ／Ｏカードからサーバへの応答信号を受信したときには、当該応答信号に含まれるベースアドレスから要求元の経路情報を抽出して、当該抽出した要求元の経路情報を応答信号のヘッダ情報の宛先に設定するヘッダ修正部と、
前記応答信号のベースアドレスに埋め込まれた要求元の経路情報を削除するベースアドレス修正部と、
当該応答信号を前記スイッチへ送信する送信部と、を有し、
前記Ｉ／Ｏプロセッサは、
前記割り込みがあったときに、前記メモリに書き込まれたアクセス要求信号に含まれるベースアドレスの所定の上位ビットに、前記アクセス要求信号のヘッダ情報から前記要求元の経路情報を設定するアドレス変換部と、
前記割り込みに基づいて前記要求元のヘッダ情報を上位ビットに設定した前記ベースアドレスを前記Ｉ／Ｏカードの応答先として設定する応答アドレス設定部と、
前記割り込みに基づいて前記アクセス要求信号をＩ／Ｏカードへ送信し、前記Ｉ／Ｏカードの動作を起動するＩ／Ｏカード起動部と、を有し、
前記Ｉ／Ｏカードは、前記アクセス要求信号に応じた処理の応答信号を、前記Ｉ／Ｏカード共有部に返信することを特徴とする計算機システム。

【請求項２】
前記Ｉ／Ｏカードは、
前記アクセス要求信号に含まれる指令を設定するコマンドレジスタと、
前記アクセス要求信号に対して応答するサーバのメモリ空間を指定するアドレスレジスタと、を有し、
前記要求信号書込部は、前記アクセス要求信号に含まれる指令が前記コマンドレジスタを除くＩ／Ｏカードへの書き込み処理を含む第１の指令のときに、前記スイッチから受信したアクセス要求信号を前記Ｉ／Ｏカード共有部のメモリに書き込み、
前記割り込み発生部は、前記スイッチから受信したアクセス要求信号に含まれる指令が前記コマンドレジスタへの書き込み処理を含む第２の指令のときに、前記Ｉ／Ｏプロセッサへ割り込みをかけ、
前記応答アドレス設定部は、前記割り込みに基づいて、前記要求元のヘッダ情報を上位ビットに設定した前記ベースアドレスを前記Ｉ／Ｏカードのアドレスレジスタに書き込むことを特徴とする請求項１に記載の計算機システム。

【請求項３】
前記要求信号書込部は、前記アクセス要求信号に含まれる指令がアドレスレジスタへの書き込みを要求するＤＭＡ初期化処理を含むときに、前記スイッチから受信したアクセス要求信号を前記Ｉ／Ｏカード共有部のメモリに書き込み、
前記割り込み発生部は、前記スイッチから受信したアクセス要求信号にコマンドレジスタへの書き込みを要求するＤＭＡ開始指令が含まれるときに、前記Ｉ／Ｏプロセッサへ割り込みをかけることを特徴とする請求項２に記載の計算機システム。

【請求項４】
前記Ｉ／Ｏプロセッサは、
前記各サーバが前記Ｉ／Ｏカードを利用可能か否かを識別し、かつ前記Ｉ／Ｏカードが共有されているか否かを識別する属性を予め設定したＩ／Ｏカード共有設定管理部を有し、
前記Ｉ／Ｏカード共有部は、
前記Ｉ／Ｏカード共有設定管理部を参照して、前記スイッチから受信したアクセス要求信号の宛先のＩ／Ｏカードと要求元のサーバの前記属性が利用可能かつ占有のときには、

前記サーバからの要求信号をそのままＩ／Ｏカードに転送し、当該Ｉ／Ｏカードから前記サーバへの応答信号をそのまま転送することを特徴とする請求項１に記載の計算機システム。

【請求項５】

前記Ｉ／Ｏプロセッサは、

前記各サーバが前記Ｉ／Ｏカードを利用可能か否かを識別し、かつ前記Ｉ／Ｏカードが共有されているか否かを識別する属性を予め設定したＩ／Ｏカード共有設定管理部を有し、

前記Ｉ／Ｏカード共有部は、

前記Ｉ／Ｏカード共有設定管理部を参照して、前記スイッチから受信したアクセス要求信号の宛先に設定されているＩ／Ｏカードと要求元のサーバの前記属性が利用不可のときには、前記サーバからの要求信号に対してマスタアボートまたは全ビットが１のデータの何れか一方を返信することを特徴とする請求項１に記載の計算機システム。

【請求項６】

前記Ｉ／Ｏプロセッサは、

前記各サーバが前記Ｉ／Ｏカードを利用可能か否かを識別し、かつ前記Ｉ／Ｏカードが共有されているか否かを識別する属性を予め設定したＩ／Ｏカード共有設定管理部と、

前記Ｉ／Ｏプロセッサに接続された入力部を介して前記Ｉ／Ｏカード共有設定部の属性を設定する属性設定部と、

を有することを特徴とする請求項１に記載の計算機システム。

【請求項７】

前記Ｉ／Ｏカード共有部は、

前記要求信号書込部または割り込み発生部で用いるアドレス情報を、前記スイッチから受信したアクセス要求信号に含まれるヘッダ情報と予め設定した情報とを比較して出力するアドレス情報判定部を有し、

当該アドレス情報判定部は、前記サーバの起動時に実行されるＩ／Ｏカードへの書き込み指令に基づいて前記情報を設定することを特徴する請求項１に記載の計算機システム。

【請求項８】

前記Ｉ／Ｏカードは、

前記アクセス要求信号に含まれる指令を設定するコマンドレジスタと、

前記アクセス要求信号に対して応答するサーバのメモリ空間を指定するアドレスレジスタと、

前記サーバのメモリに格納されたデータ構造体を読み込んでコマンドチェイン処理を行うコマンドチェイン処理部と、

を有し、

前記第１及び第２の指令が一つの起動要求で構成されて、当該起動要求を前記Ｉ／Ｏカードのコマンドレジスタに書き込むことで行われ、

前記要求信号書込部は、前記要求信号をＩ／Ｏプロセッサのメモリへ書き込み、また、前記サーバのメモリに設定されたデータ構造体をＩ／Ｏプロセッサのメモリへ書き込み、

前記Ｉ／Ｏカードは、当該Ｉ／Ｏカードの起動後に、前記Ｉ／Ｏプロセッサのメモリに設定されたデータ構造体を読み込んで処理を行うことを特徴とする請求項２に記載の計算機システム。

【請求項９】

複数のサーバと、

該複数のサーバとＩ／Ｏカードを接続するスイッチと、を備え、

前記サーバが、前記Ｉ／Ｏカードから書き込み可能なメモリ空間を備えた計算機システムにおいて、

前記スイッチとＩ／Ｏカードの間に配置されて、前記複数のサーバと前記Ｉ／Ｏカードとの間のアクセス要求信号及び応答信号を操作するＩ／Ｏカード共有部と、

前記Ｉ／ＯカードとＩ／Ｏカード共有部にアクセス可能であって、前記Ｉ／Ｏカード共

有部から書き込み可能なメモリと、前記Ｉ／Ｏカード共有部からの割り込みを受け付けるプロセッサとを備えて、前記複数のサーバに対する前記Ｉ／Ｏカードの割り当てを管理するＩ／Ｏプロセッサと、
を備え、
前記スイッチは、
前記サーバからＩ／Ｏカードへの指令と前記メモリ空間を指定するベースアドレスとを含んだアクセス要求信号に、当該アクセス要求信号の宛先と要求元の経路情報を前記アクセス要求信号のヘッダ情報に付加して前記Ｉ／Ｏカード共有部へ送信し、前記Ｉ／Ｏカードからサーバへの応答と前記メモリ空間を指定するベースアドレスとを含んだ応答信号を、当該応答信号のヘッダ情報に含まれる宛先のサーバに転送するヘッダ処理部と、前記Ｉ／Ｏカード共有部を有し、
前記Ｉ／Ｏカード共有部は、
前記ヘッダ処理部から受信したアクセス要求信号に含まれる指令が、予め設定した第１の指令のときには前記Ｉ／Ｏカード共有部のメモリに当該アクセス要求信号を書き込む要求信号書込部と、
前記ヘッダ処理部から受信した前記アクセス要求信号に含まれる指令が、予め設定した第２の指令のときには前記Ｉ／Ｏプロセッサへ割り込みをかける割り込み発生部と、
前記Ｉ／Ｏカードからサーバへの応答信号を受信したときには、当該応答信号に含まれるベースアドレスから要求元の経路情報を抽出して、当該抽出した要求元の経路情報を応答信号のヘッダ情報の宛先に設定するヘッダ修正部と、
前記応答信号のベースアドレスに埋め込まれた要求元の経路情報を削除するベースアドレス修正部と、
当該応答信号を前記スイッチへ送信する送信部と、を有し、
前記Ｉ／Ｏプロセッサは、
前記割り込みがあったときに、前記メモリに書き込まれたアクセス要求信号に含まれるベースアドレスの所定の上位ビットに、前記アクセス要求信号のヘッダ情報から前記要求元の経路情報を設定するアドレス変換部と、
前記割り込みに基づいて前記要求元のヘッダ情報を上位ビットに設定した前記ベースアドレスを前記Ｉ／Ｏカードの応答先として設定する応答アドレス設定部と、
前記割り込みに基づいて前記アクセス要求信号をＩ／Ｏカードへ送信し、前記Ｉ／Ｏカードの動作を起動するＩ／Ｏカード起動部と、を有し、
前記Ｉ／Ｏカードは、前記アクセス要求信号に応じた処理の応答信号を、前記Ｉ／Ｏカード共有部に返信することを特徴とする計算機システム。
【請求項１０】
前記サーバとスイッチの間と、前記スイッチとＩ／Ｏカード共有部及び前記Ｉ／Ｏカード共有部とＩ／Ｏカードの間を、ＰＣＩまたはＰＣＩ－ＥＸＰＲＥＳＳで接続したことを特徴とする請求項１または請求項９に記載の計算機システム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、Ｉ／Ｏデバイスの共有を行う計算機システムに関し、特に、サーバ統合を実現可能なブレードサーバシステムに関する。
【背景技術】
【０００２】
近年、サーバ台数の増加と共に運用に関する複雑さが増加し、運用コストの増大が問題化している。この運用コストを低減する技術として複数サーバを１台にまとめるサーバコンソリデーション（サーバ統合）が注目を集めている。
【０００３】
サーバ統合を実現する技術として、バックプレーンに複数のサーバブレード及び複数のＩ／Ｏブレードを装着可能なブレードサーバシステムが知られており、統合するサーバに

応じてサーバブレードを装着し、一つの筐体内に複数のサーバを統合している。つまり、一つの筐体に複数のサーバを集約した高性能なサーバを提供する。

【０００４】

サーバ統合では、複数の筐体で独立して稼動していたサーバを、一つの高性能なブレードサーバシステムに集約し、複数のＯＳとアプリケーションを稼動させることになる。ここで、複数のサーバを一つのブレードサーバシステムに統合する際には、独立して稼動していたサーバのＩ／Ｏカード（またはＩ／Ｏデバイス）も統合する必要がある。各サーバが統合前に使用していたＩ／Ｏカードをそのまま利用しようとすると、ブレードサーバシステムのバックプレーン（またはＩ／Ｏスロット）が不足する。

【０００５】

ブレードサーバシステムの筐体にＩ／Ｏ拡張用の筐体を付加して、必要な数のＩ／Ｏカードを装着することも可能ではあるが、筐体のサイズが大形化するためより大きな設置スペースが必要なってしまい、サーバ統合の効果が薄れてしまう。

【０００６】

このため、ブレードサーバシステムを用いたサーバ統合では、サーバブレード間でＩ／Ｏカードを共有し、Ｉ／Ｏカードの数を削減する必要が生じる。

【０００７】

Ｉ／Ｏカードを複数のサーバで共有する技術としては、一つのコンピュータを任意論理区画に分割する仮想計算機が知られている。これは、ホストＯＳ上で複数のゲストＯＳを稼動させ、各ゲストＯＳを論理区画として提供し、ゲストＯホストＯＳのアプリケーションとして稼動させ、ゲストＯＳからのＩ／Ｏ要求はホストＯＳが一元的にＩ／Ｏ要求を処理することで、Ｉ／Ｏデバイスの共有を行っている（例えば、特許文献１）。

【０００８】

あるいは、ＡＳ（Advanced Switching）等に代表されるＩ／Ｏスイッチを用いて複数のサーバからＩ／Ｏカードを共有する技術も知られている。これは、ＰＣＩ－ＥＸＰＲＥＳＳ規格のＩ／Ｏカードとサーバ間を専用のプロトコルで通信を行うスイッチであり、一つのＩ／Ｏカードを複数のサーバで切り換えて使用し、共有することができる。

【特許文献１】米国特許第６，４９６、８４７号

【発明の開示】

【発明が解決しようとする課題】

【０００９】

しかしながら、上記特許文献１のような従来例では、ゲストＯＳとＩ／Ｏカードとの間のＩ／Ｏアクセスは、常にホストＯＳが中継することになる。このため、ゲストＯＳとＩ／Ｏカードの間でＤＭＡ転送を行う場合などでは、ホストＯＳのメモリ領域からゲストＯＳのメモリ領域へ転送する処理が必要となり、このホストＯＳの処理がオーバーヘッドとなってＩ／Ｏアクセスの性能（転送速度やレスポンス）が低下する、という問題があった。

【００１０】

また、上記従来例のＡＳでは、ＰＣＩ－ＥＸＰＲＥＳＳを拡張した規格（専用のプロトコル）に対応する専用のＩ／Ｏカードが必要となり、従来のＰＣＩやＰＣＩ－ｘ、ＰＣＩ－ＥＸＰＲＥＳＳといった汎用のインターフェースで構成されたＩ／Ｏカードをそのまま利用することはできない。したがって、ＡＳを利用してＩ／Ｏカードを共有するには、従来のＩ／Ｏカードを置き換えるために多額の費用が必要になるという問題がある。さらに、ＡＳに対応したＩ／Ｏカードでは、専用のプロトコルを使用するため、デバイスドライバなども新規に導入する必要が生じる。このため、サーバ統合の際には、従来のＯＳに組み込んでいたデバイスドライバを変更する作業が必要となり、上記費用の問題に加えてサーバ統合に要する労力が増大する、という問題があった。

【００１１】

そこで本発明は、上記問題点に鑑みてなされたもので、Ｉ／Ｏアクセスの性能低下を防いで複数のサーバ間でＩ／Ｏカードの共有を実現することを目的とし、さらに、共有する

Ｉ／Ｏカードを汎用のインターフェースで構成可能にすることで、サーバ統合を行う際のコストを抑制することを目的とする。

【課題を解決するための手段】

【００１２】

本発明は、複数のサーバと、該複数のサーバとＩ／Ｏカードを接続するスイッチと、を備え、前記サーバが、前記Ｉ／Ｏカードから書き込み可能なメモリ空間を備えた計算機システムにおいて、前記スイッチとＩ／Ｏカードの間に配置されて、前記複数のサーバと前記Ｉ／Ｏカードとの間のアクセス要求信号及び応答信号を操作するＩ／Ｏカード共有部と、前記Ｉ／ＯカードとＩ／Ｏカード共有部にアクセス可能であって、前記Ｉ／Ｏカード共有部から書き込み可能なメモリと、前記Ｉ／Ｏカード共有部からの割り込みを受け付けるプロセッサとを備えて、前記複数のサーバに対する前記Ｉ／Ｏカードの割り当てを管理するＩ／Ｏプロセッサと、を備え、前記スイッチは、前記サーバからＩ／Ｏカードへの指令と前記メモリ空間を指定するベースアドレスとを含んだアクセス要求信号に、当該アクセス要求信号の宛先と要求元の経路情報を前記アクセス要求信号のヘッダ情報に付加して前記Ｉ／Ｏカード共有部へ送信し、前記Ｉ／Ｏカードからサーバへの応答と前記メモリ空間を指定するベースアドレスとを含んだ応答信号を、当該応答信号のヘッダ情報に含まれる宛先のサーバに転送するヘッダ処理部を有し、前記Ｉ／Ｏカード共有部は、前記スイッチから受信したアクセス要求信号に含まれる指令が、予め設定した第１の指令のときには前記Ｉ／Ｏカード共有部のメモリに当該アクセス要求信号を書き込む要求信号書込部と、前記スイッチから受信した前記アクセス要求信号に含まれる指令が、予め設定した第２の指令のときには前記Ｉ／Ｏプロセッサへ割り込みをかける割り込み発生部と、前記Ｉ／Ｏカードからサーバへの応答信号を受信したときには、当該応答信号に含まれるベースアドレスから要求元の経路情報を抽出して、当該抽出した要求元の経路情報を応答信号のヘッダ情報の宛先に設定するヘッダ修正部と、前記応答信号のベースアドレスに埋め込まれた要求元の経路情報を削除するベースアドレス修正部と、当該応答信号を前記スイッチへ送信する送信部と、を有し、前記Ｉ／Ｏプロセッサは、前記割り込みがあったときに、前記メモリに書き込まれたアクセス要求信号に含まれるベースアドレスの所定の上位ビットに、前記アクセス要求信号のヘッダ情報から前記要求元の経路情報を設定するアドレス変換部と、前記割り込みに基づいて前記要求元のヘッダ情報を上位ビットに設定した前記ベースアドレスを前記Ｉ／Ｏカードの応答先として設定する応答アドレス設定部と、前記割り込みに基づいて前記アクセス要求信号をＩ／Ｏカードへ送信し、前記Ｉ／Ｏカードの動作を起動するＩ／Ｏカード起動部と、を有し、前記Ｉ／Ｏカードは、前記アクセス要求信号に応じた処理の応答信号を、前記Ｉ／Ｏカード共有部に返信する

【発明の効果】

【００１３】

したがって、本発明は、Ｉ／Ｏカードからサーバへ向かう応答信号を中継するＩ／Ｏカード共有部は、ベースアドレスに埋め込まれた要求元のサーバの識別子をヘッダ情報の宛先に付け替えることで、汎用バスのＩ／Ｏカードを共有することができる。そして、Ｉ／Ｏカードの起動後は、Ｉ／Ｏプロセッサは応答信号の転送へ介入せず、Ｉ／Ｏカード共有部のハードウェアがアドレス付け替えを行うため、前記従来例のようなソフトウェア処理によるオーバーヘッドを防いで、Ｉ／Ｏアクセスの性能低下を防ぎながら複数のサーバ間でＩ／Ｏカードの共有を実現することが可能となる。

【発明を実施するための最良の形態】

【００１４】

以下、本発明の一実施形態を添付図面に基づいて説明する。

【００１５】

図１は、第１の実施形態を示すブレードサーバシステムのブロック図である。

【００１６】

ブレードサーバシステムは、複数のサーバブレード１０－１～ｎと、各種Ｉ／Ｏインターフェースを備えたＩ／Ｏカード５０１、５０２と、サーバブレード１０－１～ｎとＩ／

Ｏカードを接続するスイッチ２００と、Ｉ／Ｏカード５０１、５０２を複数のサーバブレード１０－１～ｎで共有するためのＩ／Ｏカード共有機構４００と、Ｉ／Ｏカードの共有を管理するＩ／Ｏプロセッサブレード６００とを備える。そして、これらの各ブレードとスイッチ２００及びＩ／Ｏカード共有機構４００は一つの筐体（図示省略）内に収納される。

【００１７】

サーバブレード１０－１～ｎは、それぞれＣＰＵ１０１とメモリ１０２がチップセット（あるいはＩ／Ｏブリッジ）１０３を介して接続され、また、チップセット１０３は汎用バス１１－１～ｎを介してスイッチ２００に接続される。ここで、汎用バス１１－１～ｎとしては、例えば、ＰＣＩ－ＥＸＰＲＥＳＳ（図中ＰＣＩ－ｅｘ）を適用した場合を示す。

【００１８】

ＣＰＵ１０１はメモリ１０２にロードしたＯＳやアプリケーションを実行することでサーバ＃１～ｎを提供する。そして、ＣＰＵ１０１は、チップセット１０３からスイッチ２００及びＩ／Ｏカード共有機構４００を介してＩ／Ｏカード５０１、５０２にアクセスを行う。

【００１９】

スイッチ２００は、サーバブレード１０－１～ｎとＩ／Ｏカード５０１、５０２との間で送受信されるパケットのヘッダ情報を付加し、このヘッダ情報に基づいてパケットを転送するヘッダ処理部２１０を有する。

【００２０】

ヘッダ処理部２１０は、サーバブレード１０－１～ｎからＩ／Ｏカード５０１、５０２へ向けて送信されたパケット（アクセス要求信号）にヘッダ情報を付加して、このヘッダ情報に含まれる宛先のノード（Ｉ／Ｏカード）に転送する。このヘッダ情報は、サーバブレード１０－１～ｎのアドレス情報（識別子）を要求元に設定し、Ｉ／Ｏカードのアドレス情報を宛先に設定する。また、スイッチ２００のヘッダ処理部２１０は、Ｉ／Ｏカードからサーバブレード１０－１～ｎへ向けたパケット（応答信号）を、このパケットに含まれる宛先（サーバ識別子）のサーバブレード１０－１～ｎに転送する。ここで、本実施形態のパケットは、汎用バスとしてＰＣＩ－ＥＸＰＲＥＳＳを採用しているので、ＰＣＩトランザクション（ＰＣＩ－Ｔｘ）である。

【００２１】

スイッチ２００とＩ／Ｏカード５０１、５０２の間には、汎用バス３０１、３１１、３１２を介してＩ／Ｏカード５０１、５０２を複数のサーバブレード１０－１～ｎで共有するためのＩ／Ｏカード共有機構４００が接続される。また、Ｉ／Ｏカード共有機構４００は汎用バス４０１を介してＩ／Ｏプロセッサブレード６００に接続されており、Ｉ／Ｏプロセッサブレード６００は後述するように、Ｉ／Ｏカードの共有に関するアドレス変換や共有状態などの管理を実行する。なお、Ｉ／Ｏプロセッサブレード６００には、コンソール５が接続されており、管理者等の入力によりＩ／Ｏカード５０１、５０２の共有状態を設定する。

【００２２】

Ｉ／Ｏカード５０１、５０２は、ＳＣＳＩ（またはＳＡＳ）やＦＣ（Fibre Channel）あるいはEthernet（登録商標）等のインターフェースを備えている。Ｉ／Ｏカード５０１、５０２には、サーバブレード１０－１～ｎのメモリ１０２へ直接アクセスするＤＭＡ（Direct Memory Access）コントローラ５１３をそれぞれ備えている。そして、Ｉ／Ｏカード５０１、５０２は、ＤＭＡコントローラ５１３でＤＭＡを行うサーバブレード１０－１～ｎ上のメモリ１０２のＭＭＩＯ（Memory Mapped I/O）のベースアドレスを指定するベースアドレスレジスタ５１１と、Ｉ／Ｏカード５０１、５０２に対する指令を指定するコマンドレジスタ５１２を備えている。ＤＭＡコントローラ５１３は、ベースアドレスレジスタ５１１に書き込まれたメモリ１０２のアドレスに対して、コマンドレジスタ５１２に書き込まれたコマンドに対応する動作を実行する。なお、Ｉ／Ｏカード５０１、５０２は

、ＰＣＩ規格に準拠した図示しないレジスタ（コンフィグレーションレジスタ、レイテンシタイマレジスタ等）を有するものである。

【００２３】

次に、Ｉ／Ｏプロセッサブレード６００は、ＣＰＵ６０２とメモリ６０３がチップセット（あるいはＩ／Ｏブリッジ）６０１を介して接続され、また、チップセット６０１は汎用バス４０１を介してＩ／Ｏカード共有機構４００に接続される。そして、Ｉ／Ｏプロセッサブレード６００では、後述するように、所定の制御プログラムが実行され、サーバブレード１０－１～ｎからのＩ／Ｏアクセスに対してアドレス変換などの処理を実行する。

【００２４】

＜Ｉ／Ｏカード共有機構＞

次に、本発明のＩ／Ｏカード共有機構４００の詳細について、図２のブロック図を参照しながら以下に詳述する。

【００２５】

Ｉ／Ｏカード共有機構４００は、サーバブレード１０－１～ｎとＩ／Ｏカード５０１、５０２の間に位置して、サーバブレード１０－１～ｎからのＩ／Ｏアクセスパケットのアドレスの変換を行うことで、ひとつのＩ／Ｏカードを複数のサーバブレード１０－１～ｎで共有することを実現する。ここで、スイッチ２００と汎用バス及びＩ／Ｏカード５０１、５０２はＰＣＩ－ＥＸＰＲＥＳＳに準拠しており、以下、Ｉ／ＯアクセスパケットをＰＣＩトランザクションという。

【００２６】

Ｉ／Ｏカード共有機構４００の主な機能は、

１）サーバブレード１０－１～ｎからＩ／Ｏカード５０１、５０２へのＰＣＩトランザクションを、Ｉ／Ｏプロセッサブレード６００のメモリ６０３へ書き込む機能

２）Ｉ／Ｏカード５０１、５０２のコマンドレジスタ５１２への書き込み要求に基づいて、Ｉ／Ｏプロセッサブレード６００のＣＰＵ６０２へ割り込みを要求する機能

３）Ｉ／Ｏカード５０１、５０２からサーバブレード１０－１～ｎへのＤＭＡによるＰＣＩトランザクションの宛先を変換する機能

の３つである。

【００２７】

図２において、Ｉ／Ｏカード共有機構４００は、後述するアドレス情報テーブル４１１を格納する連想メモリ４１０と、サーバブレード１０－１～ｎからのＰＣＩトランザクションからヘッダ情報を分離するヘッダ情報抽出部４０６と、ヘッダ情報を除くＰＣＩトランザクションの本体を解析してＩ／Ｏプロセッサブレード６００に指令を送る第１のトランザクションデコーダ４０２（図中Ｔｘデコーダ１）と、Ｉ／Ｏプロセッサブレード６００からの信号を解析してＩ／Ｏカード５０１、５０２に指令を送る第２のトランザクションデコーダ４０３（図中Ｔｘデコーダ２）と、Ｉ／Ｏカード５０１、５０２からのＰＣＩトランザクションを解析して、ＤＭＡ転送であれば宛先アドレスを修正（変換）する第３のトランザクションデコーダ４０４（図中Ｔｘデコーダ３）と、トランザクションデコーダ４０２からの指令によりＩ／Ｏプロセッサブレード６００のＣＰＵ６０２へ割り込みをかける割り込み発生部４０７と、トランザクションデコーダ４０２からの指令によりＩ／Ｏプロセッサブレード６００のメモリ６０３へ書き込みを行うメモリ書込部４０８と、第１のトランザクションデコーダ４０２の指令に基づいてＩ／Ｏプロセッサブレード６００へ出力する信号を選択する信号選択部４１２と、第３のトランザクションデコーダ４０４の指令に基づいてサーバ＃１～＃ｎ側（スイッチ２００）へ出力する信号を選択する信号選択部４１３とを主体にして構成されている。

【００２８】

ここで、ＰＣＩトランザクションは、図３で示すように、コマンドやオーダーまたはサーバアドレス（ＭＭＩＯベースアドレス）４６２などのデータを格納するＰＣＩトランザクション本体４６１と、経路情報を格納するヘッダ情報４５１から構成される。そして、ヘッダ情報４５１は、宛先４５２を先頭にして、ＰＣＩトランザクションを発行した要求

元４５３が格納される。例えば、サーバブレード１０－１からＩ／Ｏカード５０１へのＰＣＩトランザクションは、宛先４５２にＩ／Ｏカード５０１への経路情報（アドレスなど）が設定され、要求元４５３にサーバブレード１０－１の経路情報が設定され、ＰＣＩトランザクション本体４６１には、コマンドやオーダーに加えてサーバブレード１０－１のＩ／ＯレジスタのアドレスMMIOベースアドレス４６２に設定される。

【００２９】

図２において、３０１－１はスイッチ２００（サーバブレード１０－１～ｎ側）とＩ／Ｏカード共有機構４００を接続する汎用バス３０１上で、スイッチ２００からＩ／Ｏカード共有機構４００に向かう下りのＰＣＩトランザクションを示し、３０１－２は汎用バス３０１上でＩ／Ｏカード共有機構４００からスイッチ２００（サーバブレード１０－１～ｎ側）に向かう上りのＰＣＩトランザクションを示す。同様に、４０１－２はＩ／Ｏカード共有機構４００とＩ／Ｏプロセッサブレード６００を接続する汎用バス４０１上で、Ｉ／Ｏカード共有機構４００からＩ／Ｏプロセッサブレード６００へ向かう下りの指令信号を示し、４０１－１は、汎用バス４０１上でＩ／Ｏプロセッサブレード６００からＩ／Ｏカード共有機構４００へ向かう上りの指令信号（またはＰＣＩトランザクション）を示す。さらに、３１１－１、３１２－１はＩ／Ｏカード共有機構４００とＩ／Ｏカード５０１、５０２を接続する汎用バス３１１、３１２上でＩ／Ｏカード５０１、５０２からＩ／Ｏカード共有機構４００へ向かう上りのＰＣＩトランザクションを示し、３１１－２、３１２－２は汎用バス３１１、３１２上でＩ／Ｏカード共有機構４００からＩ／Ｏカード５０１、５０２へ向かう下りのＰＣＩトランザクションを示す。

【００３０】

Ｉ／Ｏカード共有機構４００は、スイッチ２００（サーバ側）から下りのＰＣＩトランザクション３０１－１を受信すると、ヘッダ情報抽出部４０６がＰＣＩトランザクションを図３に示したヘッダ情報４５１とＰＣＩトランザクション本体４６１に分離する。また、ヘッダ情報抽出部４０６はＰＣＩトランザクションに設定されたMMIOベースアドレスからオフセットを抽出する。そして、ヘッダ情報抽出部４０６は、ヘッダ情報４５１とオフセットを連想メモリ４１０へ入力し、ＰＣＩトランザクション本体４６１を第１のトランザクションデコーダ４０２へ入力する。

【００３１】

連想メモリ４１０は、ＣＡＭ（Contents Addressable Memory）等で構成され、Ｉ／Ｏプロセッサブレード６００により設定されたアドレス情報テーブル４１１を保持する。このアドレス情報テーブル４１１には、後述するように、Ｉ／Ｏカード共有機構４００に接続されたＩ／Ｏカード５０１、５０２に対する各サーバ＃１～＃ｎのアクセス許可情報（割り当て情報）が格納されている。

【００３２】

そして、連想メモリ４１０は、検索を行うアドレス（ヘッダ情報４５１）を検索キー（ビット列）として入力し、入力した検索キーに対応するアドレスを予め設定したテーブル（アドレス情報テーブル４１１）から出力するものである。後述するように、連想メモリ４１０は、入力されたヘッダ情報４５１から該当するMMIOのベースアドレスとＩ／Ｏプロセッサブレード６００上のメモリ６０３のアドレスを出力する。

【００３３】

第１のトランザクションデコーダ４０２は、ヘッダ情報抽出部４０６から受信したＰＣＩトランザクション本体４６１と、連想メモリ４１０から受信したMMIOベースアドレスとから、ＰＣＩトランザクション本体４６１の指令を解析してＩ／Ｏプロセッサブレード６００へ出力する下りの指令信号４０１－２を選択する。また、トランザクションデコーダ４０２は、ＰＣＩトランザクション本体４６１の指令が所定の指令でない場合には、Ｉ／Ｏカード共有機構４００が受信したＰＣＩトランザクションをそのままＩ／Ｏカード５０１または５０２へ下りのＰＣＩトランザクションとして転送する。

【００３４】

ここで、複数のサーバブレード１０－１～ｎ（サーバ＃１～＃ｎ）間でひとつのＩ／Ｏ

カードを共有する際の、メモリ空間について説明する。以下の例では、３つのサーバ＃１～＃３がひとつのＩ／Ｏカード５０１を共有する場合を示す。

【００３５】

ＭＭＩＯのアドレス空間は、ひとつのＩ／Ｏカードについて各サーバ＃１～＃３が自身のメモリ１０２にそれぞれ設定したＩ／Ｏ領域である。例えば、図４で示すように、サーバ＃１～＃３（サーバブレード１０－１～３）には、使用（または共有）するＩ／Ｏカード（ここでは、Ｉ／Ｏカード５０１）毎にそれぞれＭＭＩＯベースアドレスが０ｘＡ、０ｘＢ、０ｘＣ、メモリ空間のサイズを示すオフセットが０ｘＸ、０ｘＹ、０ｘＺのＭＭＩ／Ｏ領域が設定されている。なお、これらのＭＭＩＯベースアドレスとオフセットは、各サーバブレード１０－１～ｎで起動したＢＩＯＳまたはＯＳが決定するものである。

【００３６】

これら各サーバ＃１～＃３のＭＭＩＯに対応して、プロセッサブレード６００のメモリ６０３には、図５で示すように、各サーバ＃１～＃３で共有するＩ／Ｏカード５０１のメモリ空間６０３１～６０３２が後述するように設定される。図５では、サーバ＃１のＭＭＩＯベースアドレス０ｘＡに対応してプロセッサブレード６００のメモリ６０３にメモリ空間６０３１（０ｘＰ）が設定される。なお、Ｉ／Ｏプロセッサブレード６００は、メモリ６０３に設定したＩ／Ｏカード共有設定テーブル（図７参照）に基づいて、共有対象のＩ／Ｏカードを共有するサーバのＭＭＩＯについてのみ、メモリ６０３上にメモリ空間を設定する。同様にＩ／Ｏカード５０１を共するサーバ＃２のＭＭＩＯベースアドレス０ｘＢに対応してメモリ空間６０３２（０ｘＱ）が設定され、サーバ＃３のＭＭＩＯベースアドレス０ｘＣに対応してメモリ空間６０３２（０ｘＲ）が設定される。

【００３７】

そして、上記サーバ＃１～＃３のＭＭＩＯベースアドレス＝０ｘＡ、０ｘＢ、０ｘＣのＩ／Ｏ領域がＩ／Ｏカード５０１を共有するため、連想メモリ４１０のアドレス情報テーブル４１１は、図６で示すようにＩ／Ｏプロセッサブレード６００により設定される。

【００３８】

図６のアドレス情報テーブル４１１には、Ｉ／Ｏカード共有機構４００がサーバ＃１～＃３から受信したＰＣＩトランザクションのヘッダ情報４５１と比較を行うヘッダ４１１１と、ヘッダ情報４５１とアドレス情報テーブル４１１のヘッダ４１１１が一致したときに出力する各サーバのＭＭＩＯベースアドレス４１１２と、ヘッダ情報４５１とアドレス情報テーブル４１１のヘッダ４１１１が一致したときに出力するＩ／Ｏプロセッサブレード６００のメモリ空間のアドレス（図中ＩｏＰ　ＡＤＤＲ）４１１３と、Ｉ／Ｏプロセッサブレード６００のＣＰＵ６０２への割り込みを行う際に、連想メモリ４１０へ入力されたオフセットと比較を行うためのオフセット４１１４が予め設定される。

【００３９】

上述のようにサーバ＃１～＃３でＩ／Ｏカード５０１を共有する場合、図６のアドレス情報テーブル４１１のヘッダ４１１１には、Ｉ／Ｏカード５０１とサーバ＃１～＃３のアドレス情報が設定され、ＭＭＩＯベースアドレス４１１２には、図４に示した各サーバ＃１～＃３のＭＭＩＯベースアドレスが設定され、メモリ空間アドレス４１１３には図５で示したように各サーバ＃１～＃３のＭＭＩＯベースアドレスに対応したメモリ空間６０３１～６０３３のアドレス情報が設定される。また、オフセット４１１４は、サーバ＃１～＃３のメモリ空間のサイズを求めるためにＭＭＩＯベースアドレスからの差分が設定される。

【００４０】

ここで、ヘッダ４１１１には、ヘッダ情報４５１の宛先４５２がＩ／Ｏカード５０１のアドレス情報を示す「Ｉｏ１」と、Ｉ／Ｏアクセスを要求した要求元４５３のサーバ＃１～＃３のアドレス情報を示す「ＳＶ１」～「ＳＶ３」が一対の比較用アドレスとしてヘッダ４１１１に設定される。

【００４１】

例えば、サーバ＃１（サーバブレード１０－１）からＩ／Ｏカード５０１へのＰＣＩト

ランザクションには、ヘッダ情報４５１の宛先４５２に「Ｉｏ１」が設定され、要求元４５３にサーバ＃１のアドレス情報である「ＳＶ１」が設定されている。ヘッダ情報抽出部４０６で抽出されたヘッダ情報４５１を、連想メモリ４１０のアドレス情報テーブル４１１へ入力すると、サーバ＃１のＭＭＩＯベースアドレス０ｘＡと、サーバ＃１がＩ／Ｏカード５０１を共有するためのＩ／Ｏプロセッサブレード６００のメモリ空間６０３１のアドレス＝０ｘＰが連想メモリ４１０から出力される。

【００４２】

連想メモリ４１０から出力されたＭＭＩＯベースアドレスは第１のトランザクションデコーダ４０２へ入力され、メモリ空間６０３１のアドレスは信号選択部４１２へ入力される。

【００４３】

以上のようなメモリ空間により、各トランザクションデコーダ４０２～４０４の動作について以下に説明する。

【００４４】

第１のトランザクションデコーダ４０２は、ヘッダ情報抽出部４０６から受信したＰＣＩトランザクション本体４６１よりＩ／Ｏカード５０１、５０２に対する指令を抽出し、この指令の内容に応じて、Ｉ／Ｏカード共有機構４００がＩ／Ｏプロセッサブレード６００またはＩ／Ｏカード５０１、５０２へ出力する信号を以下のように決定する。

【００４５】

Ａ）ＰＣＩトランザクション本体４６１から抽出した指令が、Ｉ／Ｏカード５０１、５０２のコマンドレジスタ５１２への書き込み指令（Ｉ／Ｏカードの動作を起動させる指令、例えば、ＤＭＡ転送開始コマンド）であれば、トランザクションデコーダ４０２は割り込み発生部４０７に割り込みの出力を指令し、信号選択部４１２へ割り込み発生部４０７の出力を選択して汎用バス４０１の下りの指令信号４０１－２としてＩ／Ｏプロセッサブレード６００へ出力するよう指令する。この割り込み信号には、連想メモリ４１０が出力したＩ／Ｏプロセッサブレード６００のアドレス４１１３を含める。また、割り込みを行う場合には、連想メモリ４１０へ入力したヘッダ情報４５１とオフセットが、アドレス情報テーブル４１１のヘッダ４１１１とオフセット４１１４に一致している必要がある。

【００４６】

Ｂ）ＰＣＩトランザクション本体４６１から抽出した指令が、Ｉ／Ｏカード５０１、５０２のコマンドレジスタ５１２以外のレジスタ（例えば、ベースアドレスレジスタ５１１等）への書き込み指令（例えば、ＤＭＡの初期化要求等でＩ／Ｏカードの動作を起動しない指令）であれば、トランザクションデコーダ４０２はメモリ書込部４０８に対してＩ／Ｏプロセッサブレード６００の所定のメモリ空間へ、ＰＣＩトランザクション（ヘッダ情報４５１とＰＣＩトランザクション本体４６１）を書き込むように指令し、信号選択部４１２へメモリ書込部４０８の出力を選択して汎用バス４０１の下りの指令信号４０１－２としてＩ／Ｏプロセッサブレード６００へ出力するよう指令する。なおメモリ書込部４０８は、連想メモリ４１０から出力されたＩ／Ｏプロセッサブレード６００のアドレス４１１３に対して、ヘッダ情報４５１とＰＣＩトランザクション本体４６１を書き込む。

【００４７】

Ｃ）ＰＣＩトランザクション本体４６１から抽出した指令が、Ｉ／Ｏカード５０１、５０２のレジスタへの書き込み要求以外の指令であれば、トランザクションデコーダ４０２は信号線４２０から受信したＰＣＩトランザクションを、Ｉ／Ｏカード共有機構４００とＩ／Ｏカード５０１、５０２を接続する汎用バス３１１、３１２から下りのＰＣＩトランザクション３１１－２、３１２－２としてそのまま出力する。この場合、トランザクションデコーダ４０２は、ＰＣＩトランザクションのヘッダ情報４５１の宛先４５２に基づいて、該当するＩ／Ｏカード５０１または５０２が接続された汎用バス３１１または３１２を選択して出力する。

【００４８】

なお、上記Ａ）～Ｃ）において、トランザクションデコーダ４０２は後述するＩ／Ｏカ

ード共有設定４０５を参照して、要求元のサーバに要求先のＩ／Ｏカードが割り当てられていない場合にはアクセスを禁止する。また、トランザクションデコーダ４０２は後述するＩ／Ｏカード共有設定４０５を参照して、割り当て状態（属性）が占有を示す「Ｄｅｄｉｃａｔｅ」の場合には、連想メモリ４１０によるアドレスのデコードを行わず、上記Ｃ）の機能によりＰＣＩトランザクションをそのまま宛先のＩ／Ｏカードに転送し、サーバ＃１～＃ｎとＩ／Ｏカード５０１、５０２が直接Ｉ／Ｏ処理を行う通常のアクセスを行う。

【００４９】

以上のように、Ｉ／Ｏカード共有機構４００のトランザクションデコーダ４０２は、サーバ＃１～＃ｎとＩ／Ｏカード５０１、５０２の間に介在して、Ｉ／Ｏカード５０１、５０２のレジスタへの書き込みをＩ／Ｏプロセッサブレード６００への操作（メモリ空間への書き込み処理、割り込み処理）に変換し、共有機能のないＩ／Ｏカード５０１、５０３を共有可能にするのである。例えば、サーバ＃１～＃ｎがＩ／Ｏカード５０１、５０２へＤＭＡ転送要求（ＤＭＡ初期化要求）を行うと、上記Ｂ）の機能により、Ｉ／Ｏカード共有機構４００がＤＭＡ転送を行うＭＭＩＯベースアドレスをＩ／Ｏプロセッサブレード６００のメモリ空間に書き込む。次に、サーバ＃１～＃ｎがＤＭＡ転送の開始を指令すると、上記Ａ）の機能によりＩ／Ｏプロセッサブレード６００に割り込みをかけ、後述するように、Ｉ／Ｏプロセッサブレード６００のＣＰＵ６０２がサーバ＃１～＃ｎに代わってＤＭＡを実行させるＩ／Ｏカード５０１または５０２のコマンドレジスタ５１２、ベースアドレスレジスタ５１１に書き込みを行う。そして、Ｉ／Ｏカード５０１または５０２は、サーバ＃１～＃ｎの代理でＩ／Ｏアクセス要求を行ったＩ／Ｏプロセッサブレード６００の指令により、要求元のサーバ＃１～＃ｎにＤＭＡ転送を行う。なお、Ｉ／Ｏプロセッサブレード６００の詳細な動作については後述する。また、サーバブレード１０－１～ｎから起動時の初期化のためにＩ／Ｏカードのコンフィグレーションレジスタ（図示省略）参照要求があったときには、トランザクションデコーダ４０２は、Ｉ／Ｏプロセッサブレード６００のＣＰＵ６０２へ割り込みをかけ、さらに、ＰＣＩトランザクションをメモリ６０３へ書き込む。

【００５０】

次に、Ｉ／Ｏカード共有機構４００の第２のトランザクションデコーダ４０３の主な機能は、汎用バス４０１を介してＩ／Ｏプロセッサブレード６００から受信した上りの指令信号４０１－１を、共有されているＩ／Ｏカード５０１または５０２に対してのみ出力するようにフィルタリングを行うことである。

【００５１】

このため、Ｉ／Ｏカード共有機構４００は、Ｉ／Ｏプロセッサブレード６００のメモリ６０３上に設定したＩ／Ｏカード共有設定テーブル６１０（図７参照）を参照する領域としてＩ／Ｏカード共有設定４０５を格納するレジスタ４３０を備える。

【００５２】

ここで、Ｉ／Ｏカード共有設定テーブル６１０は、図７で示すように、Ｉ／Ｏカード毎に、どのサーバが割り当てられているか（利用可能か）を示す属性のテーブルであり、コンソール５などから管理者が設定するものである。このテーブルは、Ｉ／Ｏカードのアドレス情報（デバイス番号など）などで構成される識別子６１１と、Ｉ／Ｏカードの機能などを示す種別６１２と、各サーバ＃１～＃３の割り当て状態６１３～６１５から構成される。

【００５３】

図７はサーバ＃１～＃３とＩ／Ｏカード５０１、５０２の関係（属性）を示し、識別子６１１のＩＯカード１はＩ／Ｏカード５０１を示し、種別６１２がＳＣＳＩカードであり、属性は「Ｓｈａｒｅ」でサーバ＃１、＃２、＃３がＩ／Ｏカード１を共有していることを示している。なお、稼動するサーバの数に応じて割り当て状態６１３～６１５の数も増減する。

【００５４】

また、識別子６１１のＩＯカード２はＩ／Ｏカード５０２を示し、種別６１２がＮＩＣカードであり、サーバ＃２のみに割り当てられて占有（非共有）されている属性「Ｄｅｄｉｃａｔｅ」を示し、このＩ／Ｏカード２が他のサーバ＃１、＃３に共有されていないことを示している。なお、稼動するサーバの数に応じて割り当て状態６１３～６１５の数も増減する。また、Ｉ／Ｏカード２は、サーバ＃１、＃３について割り当てられていない（利用可能ではない）ため、Ｉ／Ｏカード２に対してこれらのサーバからのアクセスは禁止される。

【００５５】

トランザクションデコーダ４０３は、Ｉ／Ｏプロセッサブレード６００からＰＣＩトランザクションを受信すると、ＰＣＩトランザクション本体４６１からサーバアドレス（ＭＭＩＯベースアドレス）を抽出し、アドレス情報テーブル４１１のＭＭＩＯベースアドレス４１１２と比較し、一致するＭＭＩＯベースアドレスがあればそのエントリのヘッダ４１１１から宛先と要求元を取得する。次に、Ｉ／Ｏカード共有設定４０５の識別子と取得した宛先を比較し、一致するエントリで取得した要求元と同一のサーバを検索する。該当するサーバに宛先のＩ／Ｏカードが割り当てられていれば、トランザクションデコーダ４０３が受信したＰＣＩトランザクションは正当なものであるので、宛先のＩ／Ｏカードに出力する。一方、該当するサーバに宛先のＩ／Ｏカードが割り当てられていなければ、不等なＩ／Ｏアクセス要求であるので、ＰＣＩトランザクションを破棄する。なお、Ｉ／Ｏカード共有機構４００は、ＰＣＩトランザクションを破棄してから要求元のサーバにエラーの通知を行っても良い。

【００５６】

次に、Ｉ／Ｏカード共有機構４００の第３のトランザクションデコーダ４０４の主な機能は、汎用バス３１１、３１２を介してＩ／Ｏカード５０１、５０２から受信した上りのＰＣＩトランザクション３１１－１、３１２－１に対して、Ｉ／Ｏアクセスの要求元のサーバ＃１～＃ｎへ返信するためにヘッダ情報４５１とＰＣＩトランザクション本体４６１のサーバアドレス４６２を変換することである。

【００５７】

また、トランザクションデコーダ４０４は、Ｉ／Ｏカード側から受信したＰＣＩトランザクション３１１－１、３１２－１が、アドレス変換が必要なトランザクション（ＤＭＡ等）であるか、アドレス変換が不要なトランザクション（例えば、割り込みなどのイベント）であるかを判定し、信号選択部４１３でトランザクションデコーダ４０４の出力または上りのＰＣＩトランザクション３１１－１、３１２－１の何れかを選択する。

【００５８】

トランザクションデコーダ４０４は、Ｉ／Ｏカード側から受信したＰＣＩトランザクション３１１－１、３１２－１がＤＭＡ転送などであれば、アドレス変換が必要であると判定してトランザクションデコーダ４０４の出力を選択するよう信号選択部４１３に指令する。一方、アドレス変換が不要なＰＣＩトランザクションであれば、受信したＰＣＩトランザクション３１１－１、３１２－１をそのまま出力するように信号選択部４１３へ指令する。

【００５９】

ここで、アドレス変換の有無の判定は、後述するように、図３で示したＰＣＩトランザクション本体４６１のＭＭＩＯベースアドレス４６２の未使用領域に設定された所定の上位ビットにサーバ＃１～＃ｎの識別子（アドレス情報など）が含まれているか否かに応じて、トランザクションデコーダ４０４が決定する。つまり、トランザクションデコーダ４０４は、ＰＣＩトランザクション本体４６１のＭＭＩＯベースアドレス４６２の上位ビットにサーバ＃１～＃ｎの識別子（以下、サーバ識別子）が含まれていれば、アドレス変換が必要なＰＣＩトランザクションであると判定し、サーバ識別子が含まれていなければアドレス変換が不要なＰＣＩトランザクションであると判定する。

【００６０】

＜Ｉ／Ｏプロセッサブレード＞

次に、Ｉ／Ｏプロセッサブレード６００の機能について以下に説明する。図８は、Ｉ／Ｏプロセッサブレード６００を主体とする機能ブロック図である。

【００６１】

図８において、Ｉ／Ｏプロセッサブレード６００のメモリ６０３には、上記図７で示したＩ／Ｏカード共有設定テーブル６１０と、複数のサーバが共有するＩ／Ｏカードのメモリ空間６０３１～６０３２（図中６０３ｘ）が格納され、さらに、Ｉ／Ｏカード共有機構４００からの割り込み（図中ＩＮＴ）により起動する割り込み処理部６２０が図示しないＲＯＭなどからロードされている。また、サーバブレード１０－１～ｎの起動時には、Ｉ／Ｏカード共有機構４００からの割り込みによって、初期化処理部６３０が図示しないＲＯＭなどからメモリ６０３へロードされる。

【００６２】

Ｉ／Ｏカード共有設定テーブル６１０は、上記したようにＩ／Ｏプロセッサブレード６００に接続されたコンソール５から、管理者などにより適宜設定されるもので、Ｉ／Ｏカードとサーバ＃１～＃ｎの割り当てを規定する。メモリ６０３のメモリ空間６０３ｘは、後述するように、サーバ＃１～＃ｎの起動時にＣＰＵ６０２により設定される。そして、Ｉ／Ｏカード共有機構４００は、受信したスイッチ２００からのＰＣＩトランザクションが、上記Ｂ）に該当するとき、例えば、Ｉ／Ｏカードのベースアドレスレジスタ５１１への書き込みコマンド（ＤＭＡ初期化要求）を含むとき、このＰＣＩトランザクション本体４６１とヘッダ情報４５１を、アクセスを行うＩ／Ｏカードと要求元のサーバに対応するメモリ空間６０３ｘに書き込む。

【００６３】

その後、Ｉ／Ｏカード共有機構４００は、受信したスイッチ２００からのＰＣＩトランザクションが、上記Ａ）に該当するとき（コマンドレジスタ５１２への書き込み）、Ｉ／Ｏプロセッサブレード６００のＣＰＵ６０２に割り込みをかけて、割り込み処理部６２０を起動する。

【００６４】

割り込み処理部６２０は、Ｉ／Ｏカード共有機構４００からの割り込み指令に含まれるメモリ空間６０３ｘのアドレスに基づいて、メモリ空間６０３ｘに予め書き込まれたヘッダ情報４５１とＰＣＩトランザクション本体４６１を読み込む。ＰＣＩトランザクション本体４６１に含まれるコマンドがＤＭＡ転送の場合、ヘッダ情報４５１とＰＣＩトランザクション本体４６１に含まれるＭＭＩＯベースアドレス４６２を後述するように一時的に変換し、起動するＩ／Ｏカードのアドレスレジスタ５１１に変換したＭＭＩＯベースアドレスを書き込む。

【００６５】

次に、割り込み処理部６２０は、ヘッダ情報４５１の宛先４５２となっているＩ／Ｏカードのコマンドレジスタ５１２へ、割り込みを発生させたＰＣＩトランザクション本体４６１に含まれる指令（例えば、ＤＭＡ転送開始）を書き込んでＩ／Ｏカードの動作を起動する。

【００６６】

次に、割り込み処理部６２０がＤＭＡ転送の場合に行う上記アドレス変換について以下に説明する。

【００６７】

図３で示すように、ＰＣＩ－ＥＸＰＲＥＳＳまたはＰＣＩにおいてはＰＣＩトランザクションのＭＭＩＯのアドレス空間として６４ビット（図中０～６３ｂｉｔ）が定義されている。また、サーバ＃１～＃ｎ（サーバブレード１０－１～ｎ）のＣＰＵ１０１も６４ビットのアドレッシングが可能なものが普及しつつある。しかし、６４ビットのアドレス空間を使い切るようなメモリ１０２をサーバブレード１０－１～ｎに実装するのは現実的ではでなく、現状では数十ＧＢを搭載可能なメモリ空間の上限とするのが一般的である。このため、ＣＰＵ１０１等のアドレスバスも、図３の使用領域のように、６４ビット未満の所定値、例えば、５２ビット（０～５１ビット）等に設定されている。

【００６８】
したがって、ＭＭＩＯのアドレス空間としては６４ビットが規定されてはいるものの、サーバーブレード１０－１～ｎの実装上は、アドレス空間の上位ビットは未使用となっている。
【００６９】
上述のように、下位５２ビットをアクセス可能なアドレス空間とした場合、ＰＣＩトランザクション本体４６１に格納されるＭＭＩＯベースアドレス４６２の上位ビットのうち、図３の５２～６３ｂｉｔは未使用領域４６３となる。ブレードサーバシステムでは、実装可能なサーバブレードの数を数十程度とすると、図３の未使用領域は１２ビットあるので、この未使用領域の１２ビットのうち少なくとも６ビットなどを使用すれば、筐体内の全てのサーバを識別することができる。
【００７０】
一方、ＰＣＩ－ＥＸＰＲＥＳＳに準拠したＩ／Ｏカードは、Ｉ／Ｏカード側で複数のサーバ＃１～＃ｎを識別することはできず、一旦ＤＭＡ転送が開始されると、初期化時のＭＭＩＯベースアドレス４６２しか認識できないので、複数のサーバ＃１～＃ｎにサーバ＃１～＃ｎにＤＭＡ転送を行うことはできない。
【００７１】
そこで、本発明では、ＤＭＡ転送の際に、ＰＣＩトランザクションのＭＭＩＯベースアドレス４６２の未使用の上位ビットを、サーバ識別子（アドレス情報）を格納する領域として利用し、Ｉ／Ｏプロセッサブレード６００の割り込み処理部６２０がＭＭＩＯベースアドレス４６２の上位ビットにサーバ識別子となる要求元４５３を埋め込んでアドレス変換を行う。
【００７２】
そして、割り込み処理部６２０がＩ／Ｏカードのベースアドレスレジスタ５１１にアドレス変換を行ったＭＭＩＯベースアドレス４６２を書き込み、コマンドレジスタ５１２へＤＭＡ転送開始を書き込んでＩ／ＯカードのＤＭＡを起動する。
【００７３】
Ｉ／ＯカードのＤＭＡ転送開始後には、Ｉ／Ｏカード共有機構４００のトランザクションデコーダ４０４が、Ｉ／ＯカードからのＰＣＩトランザクションを受信すると、ＭＭＩＯベースアドレス４６２の上位ビットの未使用領域４６３に埋め込まれたサーバ識別子を抽出し、ＰＣＩトランザクションのヘッダ情報４５１の宛先４５２に書き込む。そして、トランザクションデコーダ４０４は、ＭＭＩＯベースアドレス４６２のサーバ識別子の領域に「０」を書き込んで埋め込まれた要求元の経路情報を削除してから、このＰＣＩトランザクションをスイッチ２００へ送信する。スイッチ２００は、ＰＣＩトランザクションのヘッダ情報に基づいて宛先に指定されたサーバ、つまりＤＭＡ転送を要求したサーバ＃１～＃ｎへＰＣＩトランザクションを転送する。
【００７４】
すなわち、Ｉ／Ｏプロセッサブレード６００の割り込み処理部６２０が、ＭＭＩＯの未使用領域４６３に要求元４５３のアドレスをサーバ識別子として埋め込んだものをＩ／Ｏカードのアドレスレジスタ５１１に書き込んで、Ｉ／ＯカードのＤＭＡ転送を起動し、Ｉ／Ｏカードから出力されるＤＭＡ転送はＩ／Ｏカード共有機構４００によって、ＰＣＩトランザクション内のＭＭＩＯベースアドレス４６２の未使用領域４６３からサーバ識別子を抽出して、ヘッダ情報４５１の宛先４５２に設定する。これにより、Ｉ／Ｏカード自体には複数のサーバ＃１～＃ｎを識別する機能がなくても、Ｉ／Ｏカード共有機構４００とＩ／Ｏプロセッサブレード６００によりＩ／Ｏカードの共有が可能になるのである。
【００７５】
＜Ｉ／Ｏカード共有処理＞
次に、ＰＣＩトランザクションを主体にしたＩ／Ｏカード共有機構４００とＩ／Ｏプロセッサブレード６００によるＩ／Ｏカード共有の処理を図９に示す。
【００７６】

図９において、Ｓ１ではサーバ＃１～＃ｎが、Ｉ／Ｏアクセスを行うＩ／Ｏカードに対して、ＰＣＩトランザクションにＤＭＡ初期化コマンドなどを設定して送信する。サーバ＃１～＃ｎは、ＰＣＩトランザクションのヘッダ情報４５１の要求元４５３に自分のアドレス情報（サーバ識別子）を設定し、ＭＭＩＯベースアドレス４６２にこのＩ／Ｏカードにサーバが割り当てたＭＭＩＯベースアドレスを設定する。

【００７７】

Ｓ２では、Ｉ／Ｏカードとサーバ間に介在するＩ／Ｏカード共有機構４００が、このＰＣＩトランザクションを受信すると、Ｉ／Ｏカードのアドレスレジスタ５１１への書き込み指令を含むＤＭＡ初期化コマンドがあるので、このＰＣＩトランザクションをＩ／Ｏプロセッサブレード６００のメモリ空間６０３ｘに書き込む。この時点では、Ｉ／Ｏカードへのアクセスは行われない。

【００７８】

Ｓ３では、サーバ＃１～＃ｎがＩ／Ｏカードに対してＤＭＡ転送開始のＰＣＩトランザクションを送信すると、Ｉ／Ｏカード共有機構４００はＩ／Ｏカードの動作を起動する指令を含むためＩ／Ｏプロセッサブレード６００に割り込みをかけて、割り込み処理部６２０を起動する。

【００７９】

割り込み処理部６２０は、メモリ空間６０３ｘに書き込まれたＰＣＩトランザクションからＭＭＩＯベースアドレス４６２を読み込んで、Ｉ／Ｏカードのアドレスレジスタ５１１に書き込む。このとき、割り込み処理部６２０は、ＰＣＩトランザクション内のＭＭＩＯベースアドレス４６２の未使用領域４６３に、要求元のサーバを示すヘッダ情報４５１の要求元４５３を埋め込んでおく。そして、割り込み処理部６２０がＩ／Ｏカードのコマンドレジスタ５１２へＤＭＡ転送開始を書き込んで、Ｉ／Ｏカードの動作を起動する。

【００８０】

Ｓ４では、Ｉ／Ｏカードがアドレスレジスタ５１１に設定されたＭＭＩＯベースアドレス４６２に対してＤＭＡ転送（書き込みまたは読み込み）を行う。

【００８１】

Ｉ／ＯカードからのＤＭＡによるＰＣＩトランザクションには、ＭＭＩＯベースアドレス４６２の上位ビットに設定された未使用領域４６３にサーバ識別子が埋め込まれている。

【００８２】

Ｓ５では、Ｉ／Ｏカードとサーバ＃１～＃ｎの間に介在するＩ／Ｏカード共有機構４００が、ＰＣＩトランザクションを受信すると上記図２のトランザクションデコーダ４０４でＤＭＡによるＰＣＩトランザクションであるか否かを判定する。

【００８３】

このトランザクションデコーダ４０４で行われるＩ／ＯカードからのＰＣＩトランザクションがＤＭＡであるか否かの判定は、上記ＭＭＩＯベースアドレス４６２の未使用領域４６３の全ビットが０でなければ、サーバ識別子が埋め込まれていると判定して、ＤＭＡによるＰＣＩトランザクションであると判定することができる。

【００８４】

そして、ＤＭＡのＰＣＩトランザクションの場合、トランザクションデコーダ４０４は、ＭＭＩＯベースアドレス４６２の未使用領域４６３の内容をヘッダ情報４５１の宛先４５２に設定し、スイッチ２００で識別可能なサーバ＃１～＃ｎのアドレス情報に変換する。この後、トランザクションデコーダ４０４は未使用領域４６３の内容を消去するため、未使用領域４６３の全ビットに０をセットしてからこのＰＣＩトランザクションを送信する。

【００８５】

この宛先４５２に基づいてスイッチ２００は、ＤＭＡのＰＣＩトランザクションを、宛先４５２に設定されたＤＭＡの要求元のサーバに転送し、サーバ＃１～＃ｎに設定したＭＭＩＯに対して所定のアクセスが行われる。

【００８６】
このように、Ｉ／Ｏカードのアドレスレジスタ５１１に設定されたＭＭＩＯベースアドレスの所定の上位ビットに、ＤＭＡを要求したサーバ識別子（要求元４５３）が設定されているので、ＤＭＡを繰り返して行ってもＩ／Ｏカード共有機構４００が各ＰＣＩトランザクションの宛先４５２をＤＭＡの要求元であるサーバのアドレス情報に付け替えるので、汎用のＩ／Ｏカードを用いながらも複数のサーバブレード１０－１～ｎで共有することが可能となるのである。
【００８７】
上記処理を時系列的に示したものが、図１０のタイムチャートである。まず、Ｓ１１ではサーバ＃１～＃ｎが、Ｉ／ＯアクセスをＩ／Ｏカードに対して、ＰＣＩトランザクションにＤＭＡ初期化コマンドなどを設定して送信する。
【００８８】
Ｓ１２では、Ｉ／Ｏカード共有機構４００が、Ｉ／Ｏカードのコマンドレジスタ５１２以外のレジスタに対する書き込み要求であるので、Ｉ／Ｏプロセッサブレード６００のメモリ空間にＰＣＩトランザクションの内容を書き込む。
【００８９】
次に、Ｓ１３では、サーバ＃１～＃ｎがＩ／ＯアクセスをＩ／Ｏカードに対して、ＰＣＩトランザクションにＤＭＡ転送開始コマンドなどコマンドレジスタ５１２への書き込み指令を設定して送信する。
【００９０】
Ｓ１４では、Ｉ／Ｏカード共有機構４００が、Ｉ／Ｏカードのコマンドレジスタ５１２に対する書き込み指令であるので、Ｉ／Ｏプロセッサブレード６００のＣＰＵ６０２へ割り込みを要求する。
【００９１】
Ｓ１５では、Ｉ／Ｏプロセッサブレード６００のＣＰＵ６０２が割り込み処理部６２０を起動して、アドレスレジスタ５１１等コマンドレジスタ５１２以外の内容を読み出す。

【００９２】
Ｓ１６では、割り込み処理部６２０が読み込んだＰＣＩトランザクションの内容がＤＭＡであれば、ヘッダ情報４５１の要求元４５３をＭＭＩＯベースアドレス４６２の未使用領域４６３にセットする。そして、このアドレス変換を行ったＭＭＩＯベースアドレス４６２をＩ／Ｏカードのアドレスレジスタ５１１に書き込み、コマンドレジスタ５１２にＤＭＡ転送開始コマンドを書き込んでＩ／Ｏカードの動作を起動する。
【００９３】
Ｓ１７では、Ｉ／ＯカードのＤＭＡコントローラ５１３が、アドレスレジスタ５１１のメモリ空間にＤＭＡアクセスを行う。
【００９４】
Ｓ18では、Ｉ／ＯカードからのＰＣＩトランザクションを受信し、ＤＭＡであるかをトランザクションデコーダ４０４が上述のように判定する。そしてトランザクションデコーダ４０４は、ＤＭＡのＰＣＩトランザクションであれば、ＰＣＩトランザクション本体４６１のＭＭＩＯベースアドレス４６２の未使用領域４６３のサーバ識別子を、ヘッダ情報４５１の宛先４５２に設定してアドレス情報の変換（復元処理）を実施する。そして、トランザクションデコーダ４０４が、スイッチ２００を介してＤＭＡを要求したサーバにＰＣＩトランザクションを送信する。
【００９５】
以上の手順により、ＭＭＩＯベースアドレス４６２の未使用領域４６３にＤＭＡを要求したサーバの識別子を埋め込んでＩ／ＯカードのＤＭＡコントローラ５１３に処理を実行させることで、一つのＩ／Ｏカードを複数のサーバ＃１～＃ｎで共有することが可能となる。
【００９６】

＜アドレス情報の設定処理＞

次に、図１１はサーバブレード１０－１～ｎを起動したときに実行される、アドレス情報テーブル４１１の設定処理の一例を示すタイムチャートである。

【００９７】

連想メモリ４１０に格納されるアドレス情報テーブル４１１は、サーバブレード１０－１～ｎが起動するたびに図１１の処理により更新される。なお、以下では、サーバブレード１０－１を起動した場合について説明する。

【００９８】

まず、Ｓ２０でサーバブレード１０－１の電源をＯＮにする。Ｓ２１では、電源の投入によりＣＰＵ１０１が図示しないＢＩＯＳを起動して、Ｉ／Ｏカード（デバイス）の初期化を行うために各Ｉ／Ｏカードのコンフィグレーションレジスタに対して読み込みを要求する。

【００９９】

Ｓ２２では、Ｉ／Ｏカード共有機構４００のトランザクションデコーダ４０２は、Ｉ／Ｏカードのコンフィグレーションレジスタの読み込み要求を受けたので、上述したように、Ｉ／Ｏプロセッサブレード６００のＣＰＵ６０２へ割り込みをかけ、さらに、このコンフィグレーションレジスタの読み込み要求のＰＣＩトランザクションをメモリ６０３へ書き込む。このとき、コンフィグレーションレジスタの読み込み要求を出したサーバブレード１０－１ではＭＭＩＯが未設定であるため、トランザクションデコーダ４０２は予め設定したアドレスに書き込みを行う。

【０１００】

Ｓ２３では、Ｉ／Ｏプロセッサブレード６００のＣＰＵ６０２が、トランザクションデコーダ４０２の割り込みによって図８に示した初期化処理部６３０を起動する。初期化処理部６３０は、所定のアドレスに書き込まれたコンフィグレーションレジスタの読み込み要求から、Ｉ／Ｏカード共有設定テーブル６１０を読み込んで当該サーバブレードに割り当てられたＩ／Ｏカードを確認する。

【０１０１】

Ｓ２４では、Ｉ／Ｏカードが当該サーバブレード１０－１に割り当てられていなければ（アクセス禁止）、初期化処理部６３０はＩ／Ｏカード共有機構４００を介して割り当てがないことを通知する（マスターアボート扱い）。一方、Ｉ／Ｏカードが当該サーバブレードに割り当てられていれば、初期化処理部６３０はこのＩ／Ｏカードのコンフィグレーションレジスタの内容を読み込んで、サーバブレード１０－１に応答する。なお、Ｓ２４の処理は、Ｉ／Ｏカード共有６１０に設定されている全てのＩ／Ｏカードについて順次実行する。

【０１０２】

Ｓ２５では、Ｉ／Ｏプロセッサブレード６００からＩ／Ｏカードのコンフィグレーションレジスタの内容を受信すると、サーバブレード１０－１は取得したＩ／Ｏカードの情報からＭＭＩＯ空間やＩＯ空間を設定し、ＭＭＩＯベースアドレスの設定などを行う。そして、サーバブレード１０－１は、ＭＭＩＯベースアドレスをＩ／Ｏカードに通知する。なお、この通知はＩ／Ｏカード毎に実行される。

【０１０３】

Ｓ２６では、Ｉ／Ｏカード共有機構４００がサーバブレード１０－１からのＭＭＩＯベースアドレスを通知するＰＣＩトランザクションを受信する。そして、Ｉ／Ｏカード共有機構４００は、ＭＭＩＯベースアドレスの設定通知であるので、Ｉ／Ｏプロセッサブレード６００のＣＰＵ６０２へ割り込みをかけ、さらに、ＭＭＩＯベースアドレスを通知するＰＣＩトランザクションをメモリ６０３へ書き込む。なお、この処理は上記Ｓ２２と同様に行われる。

【０１０４】

Ｓ２７では、Ｉ／Ｏプロセッサブレード６００のＣＰＵ６０２が、トランザクションデコーダ４０２の割り込みによって初期化処理部６３０を起動する。初期化処理部６３０は

、所定のアドレスに書き込まれたサーバブレード１０－１のＭＭＩＯベースアドレス設定通知から、メモリ６０３にサーバブレード１０－１の当該Ｉ／Ｏカードに対応するメモリ空間６０３１を割り当てる。そして、初期化処理部６３０は、サーバブレード１０－１のＭＭＩＯベースアドレス及びオフセットと、Ｉ／Ｏプロセッサブレード６００のメモリ空間６０３１のアドレスと、割り当てたＩ／Ｏカードのアドレス情報と、サーバブレード１０－１のアドレス情報をＩ／Ｏカード共有機構４００に通知し、連想メモリ４１０のアドレス情報テーブル４１１にこれらのアドレスを反映させる。なお、Ｓ２７の処理は、サーバブレード１０－１が利用するＩ／Ｏカード毎に繰り返して実行すればよい。

【０１０５】

以上の処理により、複数のサーバブレード１０－１～ｎで共有されるＩ／Ｏカード５０１、５０２へ起動したサーバブレード１０－１～ｎが直接アクセスすることなく、Ｉ／Ｏプロセッサブレード６００が代理でコンフィグレーションレジスタの内容を取得し、メモリ空間６０３ｘの設定などを行う。ブレードサーバシステムでは、他のサーバブレードが稼動しているときに、新たなサーバブレードを設置し起動することができる。このような場合に、新たなサーバブレードの起動時にＩ／Ｏプロセッサブレード６００が、Ｉ／Ｏカードに代わって応答をすることで、他の稼働中のサーバブレードのＩ／Ｏアクセスに影響を当てることなく新規のサーバブレードを起動することが可能となる。

【０１０６】

なお、上記ではサーバブレードに搭載したＢＩＯＳにより起動する例を示したが、図示はしないが、ＥＦＩ（Extensible Firmware Interface）により起動する場合でも上記と同様に処理することが可能である。

【０１０７】

＜Ｉ／Ｏカード共有設定テーブル＞

図７に示したＩ／Ｏカード共有設定テーブル６１０について、以下に説明する。

【０１０８】

上述したように、Ｉ／Ｏカード共有設定テーブル６１０は、Ｉ／Ｏカード毎に、どのサーバに割り当てられているかを示すテーブルであり、コンソール５などから管理者が適宜設定するものである。なお、各サーバ＃１～＃ｎとＩ／Ｏカードの共有、占有、アクセス禁止の属性を、図７に示すインターフェースで、コンソール５のディスプレイへ表示し、図示しないマウスやキーボードなどのインターフェースで設定することができる。

【０１０９】

Ｉ／Ｏカード共有設定テーブル６１０の識別子６１１は、Ｉ／Ｏカードの追加、変更の度に、管理者が適宜設定する。また、Ｉ／Ｏカードの機能などを示す種別６１２は、Ｉ／Ｏカードのコンフィグレーションレジスタのクラスコードおよびサブクラスコードを読み込んで、本テーブルに設定することができる。

【０１１０】

各サーバ＃１～＃３の割り当て状態６１３～６１５は、管理者がサーバ＃１～＃３やＩ／Ｏカードの特性や性能等から共有の有無、割り当ての有無を適宜設定する。図７において、複数のサーバに割り当てて共有する場合には「Ｓｈａｒｅ」を設定し、割り当てない場合には「Ｎｏ　Ａｓｓｉｇｎｅｄ」を設定し、一つのサーバで占有する場合には「Ｄｅｄｉｃａｔｅ」を設定して他のサーバからこのＩ／Ｏカードへのアクセスを禁止する。なお、Ｉ／Ｏカードの割り当が「Ｎｏ　Ａｓｓｉｇｎｅｄ」に設定されたサーバは、当該Ｉ／Ｏカードへアクセスするとアクセスが拒否される。この拒否されたアクセスが読み込みの場合には、Ｉ／Ｏカード共有機構４００は全ビットが１のデータを応答し、上記アクセスが書き込みの場合には、マスターアボートを通知する。

【０１１１】

このＩ／Ｏカード共有設定テーブル６１０は、Ｉ／Ｏカード共有機構４００のＩ／Ｏカード共有設定４０５に反映される。そして、Ｉ／Ｏプロセッサブレード６００からＩ／ＯカードへのＰＣＩトランザクションを管理する第２のトランザクションデコーダ４０３は、Ｉ／Ｏカード共有設定４０５に基づいて、正当なトランザクションのみを許可し、不正

なトランザクション、すなわち、Ｉ／Ｏカード共有設定テーブル６１０で割り当てられていないサーバによるアクセスを禁止する。

【０１１２】

このＩ／Ｏカード共有設定テーブル６１０により、全てのＩ／Ｏカードを共有することもできるが、スループットを確保するためなどで一つのＩ／Ｏカードあるサーバに占有させ、他のＩ／Ｏカードを共有するように設定することができる。これにより、Ｉ／Ｏカードの共有と占有を混在させることが可能となって、ブレードサーバシステムのＩ／Ｏデバイスの構成に柔軟性を持たせることが可能となって、Ｉ／Ｏデバイスのリソースを有効に利用できるのである。

【０１１３】

＜第２実施形態＞

図１２は、第２の実施形態を示すブレードサーバシステムのブロック図である。このブレードサーバシステムでは、前記第１実施形態のＩ／Ｏカードにコマンドチェイン制御部５１４を付加し、サーバ＃１～＃ｎのメモリ０２に設定されたデータ構造体を順次読み込んでＩ／Ｏ処理を行うＩ／Ｏカード１５０１、１５０２で構成したもので、その他の構成は前記第１実施形態と同様である。

【０１１４】

Ｉ／Ｏカード１５０１、１５０２は、サーバブレード１０－１～ｎのメモリ１０２に設定されたデータ構造体を順次読み込んで、各データ構造体の記述に従ってＩ／Ｏ動作を行うもので、いわゆるコマンドチェイン処理を行うものである。

【０１１５】

サーバブレード１０－１～ｎで稼動するサーバ＃１～＃ｎは、Ｉ／Ｏカードを使用する際に、図１３で示すように、各サーバ＃１～＃ｎのメモリ空間の所定のアドレスにデータ構造体１０２０～１０２２を設定しておく。

【０１１６】

例えば、サーバ＃１のメモリ空間には、３つのデータ構造体１０２０（ＣＣＷ１～３）がアドレス０ｘＤに設定され、サーバ＃２のメモリ空間には、２つのデータ構造体１０２１（ＣＣＷ１１、１２）がアドレス０ｘＥに設定され、サーバ＃３のメモリ空間には、４つのデータ構造体１０２２（ＣＣＷ２１～２４）がアドレス０ｘＦに設定される。なお、これらのデータ構造体は、各サーバ＃１～＃３のＯＳまたはアプリケーションが適宜設定するものである。

【０１１７】

先頭及び中間のデータ構造体には、次のデータ構造体があることを示すフラグが設定されている。例えば、データ構造体１０２０のＣＣＷ１とＣＣＷ２には、次のデータ構造体があることを示すフラグが設定され、ＣＣＷ３にはこのフラグが設定されず、最後のデータ構造体であることを示す。

【０１１８】

そして、各サーバ＃１～＃３は、使用するＩ／Ｏカード１５０１または１５０２に対して動作を起動させる指令を送信し、メモリ空間に設定したデータ構造体１０２０～１０２２のアドレスを通知する。

【０１１９】

Ｉ／Ｏカード１５０１、１５０２は、サーバ＃１～＃ｎから起動の指令とともにメモリ空間のアドレスを受信すると、指定されたメモリ空間のデータ構造体を読み込んで、データ構造体１０２０～１０２２に記述されたＩ／Ｏアクセスを実行する。

【０１２０】

このようなコマンドチェイン処理を行うＩ／Ｏカード１５０１、１５０２を備えたブレードサーバシステムに、前記第１実施形態と同様のＩ／Ｏカード共有機構４００と、Ｉ／Ｏプロセッサブレード６００を適用する例を以下に示す。

【０１２１】

Ｉ／Ｏカード共有機構４００とＩ／Ｏプロセッサブレード６００によるＩ／Ｏカードの

共有処理のタイムチャートを図１４に示す。なお、以下では、サーバ＃１（サーバブレード１０－１）がＩ／Ｏカード１５０１を使用する場合について説明する。

【０１２２】

まず、サーバ＃１が、Ｉ／Ｏアクセスを行うＩ／Ｏカード１５０１に対して動作の起動を指令するＰＣＩトランザクションを送信する（Ｓ３１）。このＰＣＩトランザクションは、サーバ＃１がＩ／Ｏカード１５０１に割り当てたＭＭＩＯベースアドレスをＭＭＩＯベースアドレス４６２に設定し、動作の起動を指令するコマンドとデータ構造体１０２０のアドレス０ｘＤをＰＣＩトランザクション本体４６１内に設定する。

【０１２３】

Ｉ／Ｏカード共有機構４００は、サーバ＃１からのＰＣＩトランザクションを受信するとＰＣＩトランザクションの指令を解析し、動作の起動の指令はＩ／Ｏカード１５０１のコマンドレジスタ５１２への書き込み指令であるので、Ｉ／Ｏプロセッサブレード６００のＣＰＵ６０２へ割り込みをかける（Ｓ３２）。同時に、Ｉ／Ｏカード共有機構４００は、Ｉ／Ｏプロセッサブレード６００の所定のメモリ空間６０３１（図５参照）にＰＣＩトランザクションの内容を書き込む（Ｓ３３）。

【０１２４】

割り込みにより起動したＣＰＵ６０２は、前記第１実施形態の割り込み処理部６２０を起動する。そして割り込み処理部６２０が、メモリ空間に書き込まれたＰＣＩトランザクションのデータ構造体１０２０のアドレス０ｘＤを読み込んで、ＭＭＩＯベースアドレス４６２の要求元４５３を取得する。次に、割り込み処理部６２０は、要求元４５３のサーバ＃１のメモリ１０２に対して取得したアドレスからデータ構造体１０２０を読み込み、Ｉ／Ｏプロセッサブレード６００の所定のメモリ空間にデータ構造体１０２０をコピーする（Ｓ３４）。なお、このメモリ空間は図１５で示すように、サーバ＃１～＃ｎ毎に予め設定したデータ構造体を格納する領域であり、この例では、サーバ＃１～＃３のデータ構造体用メモリ空間として０ｘＳ、０ｘＴ、０ｘＵが設定され、サーバ＃１のデータ構造体１０２０は、図示のようにアドレス０ｘＳから格納される。

【０１２５】

次に割り込み処理部６２０は、ＤＭＡ等で使用するＭＭＩＯの処理を実施する。メモリ空間６０３１に書き込んだＰＣＩトランザクションから、ヘッダ情報４５１の要求元４５３をＭＭＩＯベースアドレス４６２の未使用領域４６３にセットしたものを、目的のＩ／Ｏカード１５０１のアドレスレジスタ５１１に書き込んで、アドレス変換を実施しておく（Ｓ３５）。

【０１２６】

この後、割り込み処理部６２０は、受信したＰＣＩトランザクションのＩ／Ｏカードの動作の起動指令に基づいて、Ｉ／Ｏカード１５０１のコマンドレジスタに動作の起動指令を書き込み、コマンドチェイン制御部５１４にデータ構造体１０２０のアドレス０ｘＳを通知する（Ｓ３６）。

【０１２７】

Ｉ／Ｏカード１５０１は、割り込み処理部６２０の起動指令に基づいて起動し、受信したＩ／Ｏプロセッサブレード６００のメモリ６０３のアドレス０ｘＳからひとつのデータ構造体（ＣＣＷ１）１０２０を読み込む（Ｓ３７）。なお、Ｉ／Ｏカード共有機構４００は、Ｉ／Ｏカード１５０１からＩ／Ｏプロセッサブレード６００への通信については、そのまま転送するものとする。

【０１２８】

Ｉ／Ｏカード１５０１はこのデータ構造体１０２０の記述に応じてＩ／Ｏ操作を行う。例えば、読み込んだデータ構造体１０２０がＤＭＡの場合、Ｉ／Ｏカード１５０１はアドレスレジスタ５１１にセットされたＭＭＩＯベースアドレスに対してＤＭＡ転送を実施する（Ｓ３８）。Ｉ／ＯカードからのＤＭＡによるＰＣＩトランザクションには、ＭＭＩＯベースアドレス４６２の上位ビットに設定された未使用領域４６３にサーバ識別子が埋め込まれている。

【０１２９】
Ｉ／Ｏカード１５０１とサーバ＃１の間に介在するＩ／Ｏカード共有機構４００が、ＰＣＩトランザクションを受信すると上記図２のトランザクションデコーダ４０４でＤＭＡによるＰＣＩトランザクションであるか否かを判定する。
【０１３０】
このトランザクションデコーダ４０４で行われるＩ／ＯカードからのＰＣＩトランザクションがＤＭＡであるか否かの判定は、上記ＭＭＩＯベースアドレス４６２の未使用領域４６３の全ビットが０でなければ、サーバ識別子が埋め込まれていると判定して、ＤＭＡによるＰＣＩトランザクションであると判定することができる。
【０１３１】
そして、ＤＭＡのＰＣＩトランザクションの場合、トランザクションデコーダ４０４は、ＭＭＩＯベースアドレス４６２の未使用領域４６３の内容をヘッダ情報４５１の宛先４５２に設定し、スイッチ２００で識別可能なサーバ＃１～＃ｎのアドレス情報に変換する。この後、トランザクションデコーダ４０４は未使用領域４６３の内容を消去するため、未使用領域４６３の全ビットに０をセットしてからこのＰＣＩトランザクションを送信する（Ｓ３９）。
【０１３２】
この宛先４５２に基づいてスイッチ２００は、ＤＭＡのＰＣＩトランザクションを、宛先４５２に設定されたＤＭＡの要求元のサーバに転送し、サーバ＃１～＃ｎに設定したＭＭＩＯに対して所定のアクセスが行われる。
【０１３３】
データ構造体（ＣＣＷ１）１０２０で指定されたＩ／Ｏ操作が完了すると、Ｉ／Ｏカード１５０１は、次のデータ構造体（ＣＣＷ２）を指定されたＩ／Ｏプロセッサブレード６００のメモリ６０３のアドレス０ｘＳから読み込んで、上記と同様に実行する。
【０１３４】
このように、コマンドチェイン処理を行うＩ／Ｏカード１５０１、１５０２の場合では、Ｉ／Ｏプロセッサブレード６００のメモリ６０３のメモリ空間に、各サーバ＃１～＃３のデータ構造体１０２０～１０２２をコピーしておき、サーバ＃１～＃３に代わってＩ／Ｏプロセッサブレード６００がＩ／Ｏカード１５０１、１５０２からの読み込み要求に応答する。
【０１３５】
したがって、Ｉ／Ｏカード１５０１、１５０２の起動時に、Ｉ／Ｏアクセスを要求したサーバのデータ構造体１０２０～１０２２を格納したメモリ６０３上のアドレスをＩ／Ｏカードへ通知することで、コマンドチェイン処理を行うＩ／Ｏカードを複数のサーバ＃１～＃３で共有することが可能となる。
【０１３６】
＜第３実施形態＞
図１６は、第３の実施形態を示すブレードサーバシステムのブロック図を示す。本実施形態は、前記第１または第２実施形態のスイッチ２００に、Ｉ／Ｏカード共有機構４００を組み込んで一体にしたものである。
【０１３７】
スイッチ２００Ａは、Ｉ／Ｏカード共有機構４００を含んで構成され、前記第１または第２実施形態で述べたように動作する。Ｉ／Ｏカード共有機構４００をスイッチ２００Ａに包含することで、ブレードサーバシステムに搭載するスロットの数を低減でき、筐体をコンパクトに構成することが可能となる。
【０１３８】
＜まとめ＞
以上のように、本発明によれば、複数のサーバブレード１０－１～ｎで一つのＩ／ＯカードをＤＭＡ転送を行う場合は、ＤＭＡ転送開始時にはＩ／Ｏプロセッサブレード６００がＩ／Ｏカードのアドレスレジスタ５１１にＤＭＡを要求したサーバの識別子を

ＭＭＩＯベースアドレス内に埋め込んでおき、Ｉ／Ｏカードからサーバへ向かうＰＣＩトランザクションを中継するＩ／Ｏカード共有機構４００は、ＭＭＩＯベースアドレスに埋め込まれたサーバ識別子をヘッダ情報４５１の宛先４５２に付け替えることで、汎用バスのＩ／Ｏカードを共有することができる。

【０１３９】

そして、ＤＭＡ転送開始後は、Ｉ／Ｏプロセッサブレード６００はＰＣＩトランザクションの転送へ介入せず、Ｉ／Ｏカード共有機構４００のハードウェアがアドレス変換を行うため、前記従来例のようなソフトウェア処理によるオーバーヘッドを防いで、Ｉ／Ｏアクセスの性能低下を防ぎながら複数のサーバ間でＩ／Ｏカードの共有を実現することが可能となる。

【０１４０】

さらに、複数のサーバブレード１０－１～ｎで共有するＩ／Ｏカードは、汎用のバスインターフェースで構成することができるので、上述のようなサーバ統合を行う際には、従来使用していたＩ／Ｏカードをそのまま利用することができるので、サーバ統合の際のコストを抑制することが可能となる。

【０１４１】

また、サーバ統合の際には、複数のサーバブレード１０－１～ｎで一つのＩ／Ｏカードを共有できるので、Ｉ／Ｏカードの数を低減することができ、前記従来例のようなＩ／Ｏカードの増大を防いでコンパクトな筐体を用いることができる。

【０１４２】

また、Ｉ／Ｏカード共有設定テーブル６１０により、共有するＩ／Ｏカードと、ひとつのサーバブレードのみに占有させるＩ／Ｏカードを混在させることができ、ブレードサーバシステムの柔軟な構成が可能となる。

【０１４３】

なお、上記各実施形態では、汎用バスとしてＰＣＩ－ＥＸＰＲＥＳＳを適用した例を示したが、ＰＣＩやＰＣＩ－Ｘ等の汎用バスを採用してもよい。

【０１４４】

また、上記各実施形態では、サーバブレード１０－１～ｎとサーバ＃１～＃ｎが１対１で対応した例を示したが、サーバ＃１～＃ｎが仮想計算機の論理区画で構成されるときは、サーバ識別子を論理区画番号とすればよい。

【０１４５】

また、上記各実施形態では、Ｉ／Ｏカード共有機構４００に直接Ｉ／Ｏプロセッサブレード６００を接続した例を示したが、Ｉ／Ｏプロセッサブレード６００をスイッチ２００に接続しても良い。さらに、Ｉ／Ｏプロセッサブレード６００に代わって、サーバブレードがＩ／Ｏプロセッサブレード６００の処理を実行するようにしても良い。

【産業上の利用可能性】

【０１４６】

以上のように、本発明は複数のサーバとＩ／ＯカードまたはＩ／Ｏデバイスを汎用バスで接続する計算機システムに適用することができ、特に、サーバ統合を行うブレードサーバシステムに適用することで、コストの低減と性能の確保実現できる。

【図面の簡単な説明】

【０１４７】

【図１】第１の実施形態を示すブレードサーバシステムのブロック図。

【図２】Ｉ／Ｏカード共有機構のブロック図。

【図３】ＰＣＩトランザクションの説明図。

【図４】サーバのメモリ空間を示す説明図。

【図５】Ｉ／Ｏプロセッサブレードのメモリ空間を示す説明図。

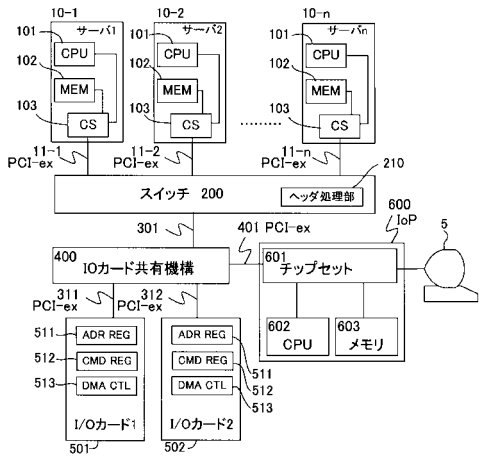
【図６】アドレス情報テーブルの一例を示す説明図。

【図７】Ｉ／Ｏカード共有設定テーブルの一例を示す説明図。

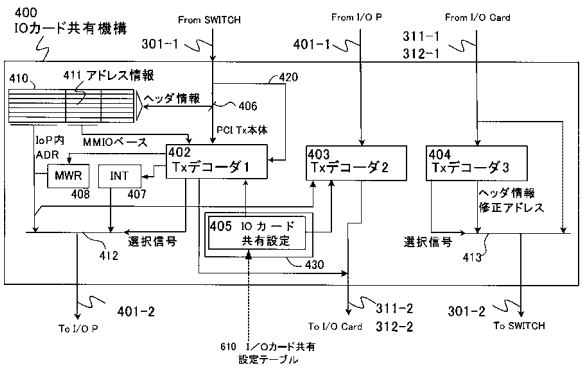
【図８】Ｉ／Ｏカード共有機構とＩ／Ｏプロセッサブレードの関係を示すブロック図。

【図９】Ｉ／Ｏカード共有処理の流れを示す説明図。
【図１０】Ｉ／Ｏカード共有処理の流れを示すタイムチャート。
【図１１】サーバブレードを起動したときに実行される、アドレス情報の設定処理の一例を示すタイムチャートである。
【図１２】第２の実施形態を示すブレードサーバシステムのブロック図。
【図１３】第２の実施形態を示し、サーバのメモリ空間を示す説明図。
【図１４】第２の実施形態を示し、Ｉ／Ｏカード共有処理の流れを示すタイムチャート。
【図１５】第２の実施形態を示し、Ｉ／Ｏプロセッサブレードのメモリ空間を示す説明図。
【図１６】第３の実施形態を示すブレードサーバシステムのブロック図。
【符号の説明】
【０１４８】
１０－１～１０－ｎ　サーバブレード
２００　スイッチ
４００　Ｉ／Ｏカード共有機構
５０１、５０２　Ｉ／Ｏカード
６００　Ｉ／Ｏプロセッサブレード
４０５　Ｉ／Ｏカード共有設定
４１１　アドレス情報テーブル
６１０　Ｉ／Ｏカード共有設定テーブル

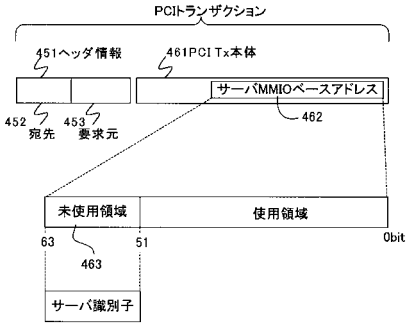
【図１】



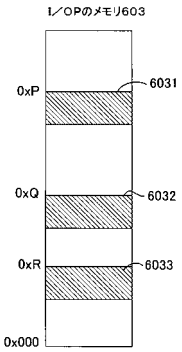
【図２】



【図３】



【図５】



【図６】

411 アドレス情報テーブル

HEADER	MMIo ADDR	IoP ADDR	OFFSET
Io1:SV1	0xA	0xP	0xX
Io1:SV2	0xB	0xQ	0xY
Io1:SV3	0xC	0xR	0xZ

4111 4112 4113 4114

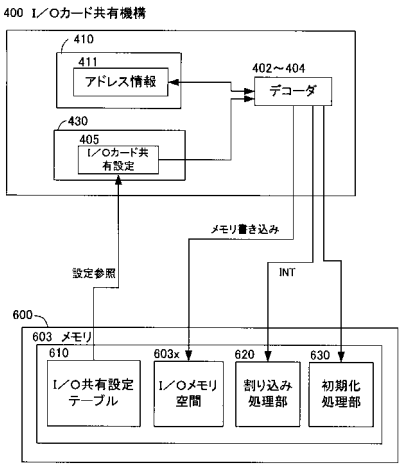
【図７】

610 I/Oカード共有設定テーブル

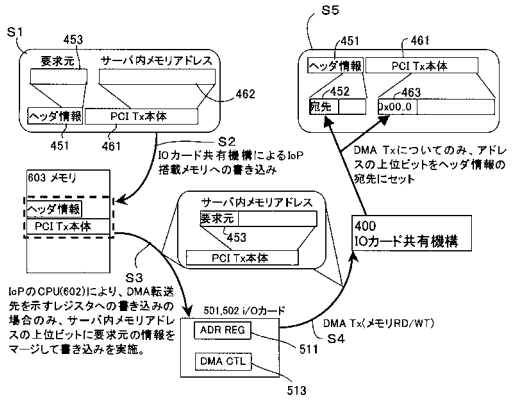
	種別	サーバ1	サーバ2	サーバ3	
IOカード1	SCSIカード	Share	Share	Share	.....
IOカード2	NICカード	No Assigned	Dedicate	No Assigned	.....

611 612 613 614 615

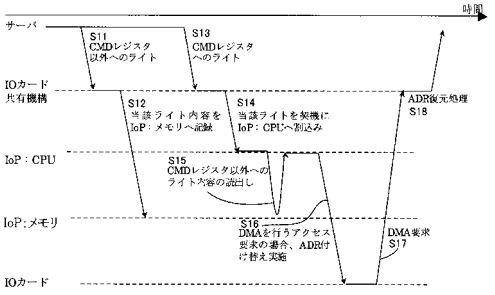
【図８】



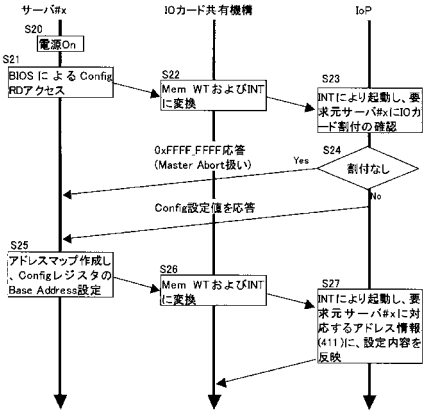
【図９】



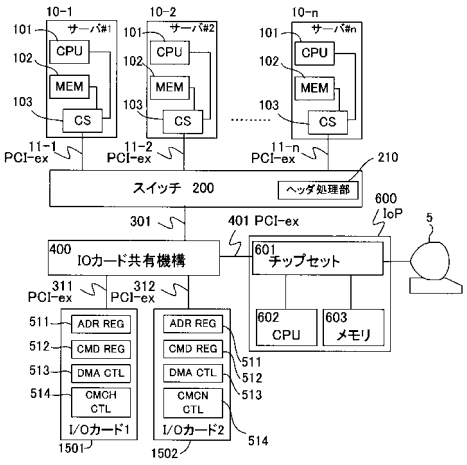
【図１０】



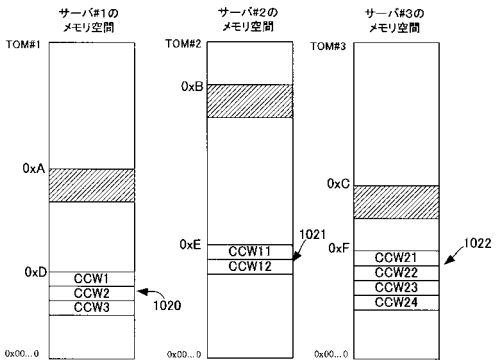
【図１１】



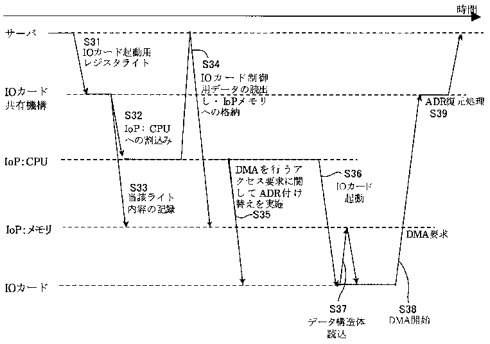
【図１２】



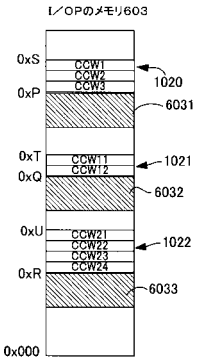
【図１３】



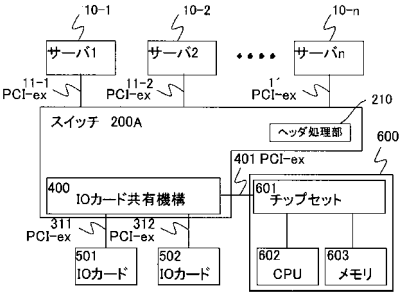
【図１４】



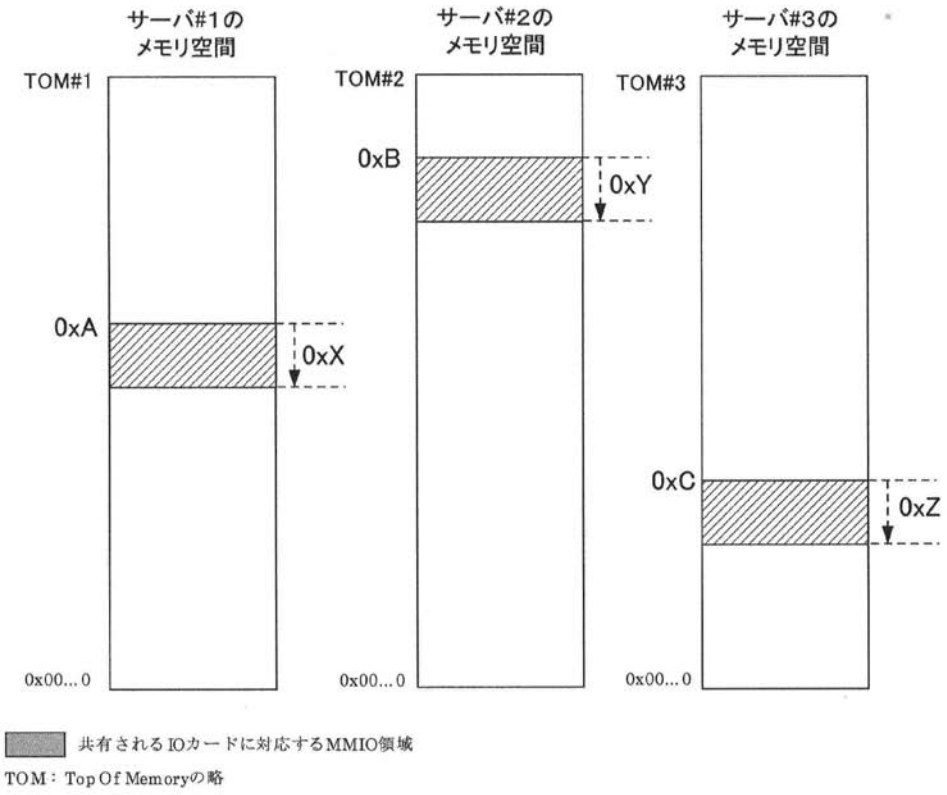
【図１５】



【図１６】



【図４】


サーバ#1の
メモリ空間
サーバ#2の
メモリ空間
サーバ#3の
メモリ空間
TOM#1
TOM#2
TOM#3
0xB
0xY
0xA
0xX
0xC
0xZ
0x00... 0
0x00... 0
0x00... 0
共有されるIOカードに対応するMMIO領域
TOM：Top Of Memoryの略

フロントページの続き

(72)発明者　保田　淑子
東京都国分寺市東恋ヶ窪一丁目２８０番地　株式会社日立製作所　中央研究所内

審査官　坂東　博司

(56)参考文献　特表２００７－５０２５７９（ＪＰ，Ａ）
特開２００４－５４９４９（ＪＰ，Ａ）
特開２００５－２０９１９７（ＪＰ，Ａ）
特表２００７－５０７０４５（ＪＰ，Ａ）
特表２００６－５０６７３６（ＪＰ，Ａ）
米国特許出願公開第２００６／２５３６１９（ＵＳ，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１３／１４