(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5220974号
(P5220974)
(45)発行日　平成25年6月26日(2013.6.26)　(24)登録日　平成25年3月15日(2013.3.15)

(51)Int.Cl.　F I
G06F 13/00 (2006.01)　G06F 13/00 354A

請求項の数 1 (全 31 頁)

(21)出願番号	特願2001-530282 (P2001-530282)
(86)(22)出願日	平成12年10月12日 (2000.10.12)
(65)公表番号	特表2003-511777 (P2003-511777A)
(43)公表日	平成15年3月25日 (2003.3.25)
(86)国際出願番号	PCT/EP2000/010277
(87)国際公開番号	WO2001/028179
(87)国際公開日	平成13年4月19日 (2001.4.19)
審査請求日	平成19年10月12日 (2007.10.12)
審判番号	不服2011-22151 (P2011-22151/J1)
審判請求日	平成23年10月13日 (2011.10.13)
(31)優先権主張番号	09/418,558
(32)優先日	平成11年10月14日 (1999.10.14)
(33)優先権主張国	米国(US)

(73)特許権者　505160898
ブルアーク　ユーケー　リミテッド
イギリス国　アールジー１２　１アールビー　バークシアー，　ブラックネル，　クックハム　ロード，　クイーンズゲート　ハウス
(74)代理人　110000279
特許業務法人ウィルフォート国際特許事務所
(72)発明者　バレル、ジェフリー・エス
英国、バークシャー・アールジー４６・４エイチワイ、ロウアー・アーリー、ボウチーフ・クロース　９

最終頁に続く

(54)【発明の名称】ハードウェア実行又はオペレーティングシステム機能の加速のための装置及び方法

(57)【特許請求の範囲】
【請求項１】
ネットワークノードからネットワークを介してネットワークサービス要求を受信し、前記ネットワークサービス要求を扱って、記憶装置に記憶装置アクセス要求を送信する要求処理装置であって、
ソフトウェアプログラムの制御下にありシステムバスに接続されたプロセッサと
前記ネットワークサービス要求を受信するネットワークサブシステムと、
前記ネットワークサブシステムに接続され前記記憶装置アクセス要求を前記記憶装置に送信するサービスサブシステムと
を備え、
前記ネットワークサブシステム及びサービスサブシステムは、前記システムバスから独立した受信高速パスインターフェース及び送信高速パスインターフェースにより接続され、
前記ネットワークサブシステム及び前記サービスサブシステムは、それぞれ、前記ソフトウェアプログラムの直接の制御外で駆動する専用ハードウェアであり、
前記ネットワークサブシステムは、
前記システムバスと前記受信高速パスインターフェースとに接続されており、前記ネットワークサービス要求を前記サービスサブシステム或いは前記プロセッサに予定されているのかを決定し、前記サービスサブシステムに予定されているカプセルから取り出したネットワークサービス要求を、前記受信高速パスインターフェースを介して前記サービスサ

ブシステムに送信する受信モジュールと、
前記システムバスと前記送信高速パスインターフェースとに接続されており、前記ネットワークサービス要求のネットワークサービス応答を、前記サービスサブシステムから前記送信高速パスインターフェースを介して受信し、前記ネットワークサービス応答を前記ネットワークノードに送信する送信モジュールと
を備え、
前記サービスサブシステムは、
前記ネットワークサービス要求を受信し前記記憶装置アクセス要求を発するサービス／ファイルシステムモジュールと、
前記サービス／ファイルシステムモジュールから前記記憶装置アクセス要求を受信し、前記記憶装置アクセス要求を記憶装置に送信する記憶装置モジュールと
を備え、
前記サービス／ファイルシステムモジュールが、
前記システムバスと前記受信高速パスインターフェースとに接続されており、前記受信高速パスインターフェースを介して前記ネットワークサービス要求を受信し、前記記憶装置アクセス要求を発する受信器と、
前記システムバスと前記送信高速パスインターフェースとに接続されており、前記記憶装置アクセス要求の記憶装置アクセス応答を前記記憶装置モジュールから受信し、前記記憶装置アクセス応答に基づくネットワークサービス応答を生成し、前記ネットワークサービス応答を、前記送信高速パスインターフェースを介して前記ネットワークサブシステムの前記送信モジュールに送信する送信器と、
前記システムバス、前記受信高速パスインターフェース及び前記送信高速パスインターフェースから独立しており前記受信器と前記送信器とを結ぶ専用バスと、
前記システムバス、前記受信高速パスインターフェース及び前記送信高速パスインターフェースのいずれにも接続されておらず前記専用バスに接続されており前記受信器及び前記送信器の処理を制御する専用プロセッサと
を備え、
前記受信器が、
前記ネットワークサービス要求で要求されたデータの物理的位置の要求を出し、その要求の応答を受けて、第１の待ち行列に入れ、第１の待ち行列内の応答を基に、アクセス要求を発する受信制御エンジンと、
前記物理的位置の要求を受けて、その要求を第２の待ち行列に入れ、前記第２の待ち行列内の要求を処理することで、前記物理的位置を表す情報を発し、その要求の応答を前記受信制御エンジンに送る物理位置制御エンジンと、
前記受信制御エンジンからの前記アクセス要求と前記物理位置制御エンジンからの前記物理的位置を表す情報とを基に前記記憶装置アクセス要求を生成する生成エンジンと、
前記生成エンジンにより生成された前記記憶装置アクセス要求を発する要求送信インターフェースと
を備え、
前記記憶装置モジュールが、
前記記憶装置アクセス要求を受信し、前記記憶装置アクセス要求を前記記憶装置に適したフォーマットに変換する要求インターフェースと、
前記記憶装置からの応答を取得し、その応答を前記サービス／ファイルシステムモジュールに適したフォーマットに変換する承認インターフェースと、
データの部分への高速読込アクセスを可能にするために前記記憶装置に含まれるデータの一部のローカルコピーを保持するキャッシュ制御部と
を備え、
前記ネットワークノードから前記ネットワークサービス要求を受信したか否かの判断を含んだ第１のループと、記憶装置アクセス要求が完了したか否かを含んだ第２のループとで構成された単一スレッドが実行されるようになっており、

前記ネットワークサブシステムにおいて前記第１のループが実行され、前記サービスサブシステムにおいて前記第２のループが実行される、
ことを特徴とする要求処理装置。
【発明の詳細な説明】
【０００１】
発明の技術分野
本発明は、オペレーティングシステム機能及びハードウェア実行、あるいは、そのような機能の加速に関する。
【０００２】
発明の背景技術
コンピュータのオペレーティングシステムは、コンピュータが外部供給源と通信することを可能にする。オペレーティングシステムは、典型的に、キーボード、ディスプレイ、ディスク記憶装置、ネットワーク設備、プリンタ、モデムなどを含むコンピュータ使用法に結び付けられたアイテムの直接制御を扱う。コンピュータのオペレーティングシステムは、典型的に、中央演算処理装置（ＣＰＵ）にローカル及びネットワークファイルシステム、メモリ、周辺装置ドライバ、及びアプリケーション処理を含む処理を管理することを含むタスクを実行させるように設計される。これらのすべての機能に対する責任をＣＰＵ上に置くことは、特に、オペレーティングシステムが例えば、Windows NT（登録商標）（ワシントン州レドモンドのマイクロソフト社から入手可能）、Unix（登録商標）（カリフォルニア州サンタクルーズのSCO Softwareから、及びマサチューセッツ州ケンブリッジのRed Hat Softwareの「Linux」と呼ばれるバージョンを含む多くの供給源から利用可能）、及びNetWare（ユタ州プロヴォのＮｏｖｅｌ１から利用可能）のように高性能であるとき、重大な処理負担を強要する。負担がアプリケーション結び付けられた以外の処理を実行するＣＰＵ上に置かれると、アプリケーションのパフォーマンスが低下し得る結果となり、アプリケーションを実行するのに利用可能なＣＰＵ時間が益々小さくなる。それに加えて、ＣＰＵの外部装置のスループットは、オペレーティングシステムがこれらの装置を管理する責任をＣＰＵ上に置くとき、ＣＰＵによって強要される制限を受けやすい。さらに、ＣＰＵを含み、オペレーティングシステムを実行し、装置に結び付けられるソフトウエアーハードウェアシステム全体の信頼性は、とりわけ、オペレーティングシステムに依存する。オペレーティングシステムに内在する複雑さに起因して、ソフトウエアーハードウェアシステム全体の安定性を害する不測の状態が発生し得る。
【０００３】
発明の概要
この概要で列挙される本発明のある態様は、これより同日に提出された他の出願の目的である。本発明の一態様では、ネットワークを介してサービス要求を扱うための装置が提供され、ネットワークはプロトコルを利用する。この態様では、該装置は、
ａ．ネットワークプロトコルを用いてネットワークサービス要求を送受信するためのネットワークサブシステムと、
ｂ．前記ネットワークサービス要求を満たすための、前記ネットワークサブシステムに接続されるサービスサブシステムとを含む。
【０００４】
同じく、この態様では、ネットワークサブシステムとサービスサブシステムの少なくとも一つは、ハードウェアで実行され、ネットワークサブシステムとサービスサブシステムの他方は、任意的にハードウェアで加速されてもよい。その代わりに、又はそれに加えて、サービスサブシステムは、ハードウェアで加速されてもよい。
【０００５】
関連する実施の形態では、サービス要求は、長期の電子記憶装置へのデータの読み込み及び書込みの一つを含み、追加的に、ネットワークサブシステムはハードウェアで加速される。同じく、追加的に、長期の記憶装置は、ネットワークを介してコンピュータにアクセス可能なネットワークディスク記憶装置である。その代わりに、長期の記憶装置は、ロ

ーカルコンピュータにアクセス可能であるが、ネットワークを介して他のあらゆるコンピュータにはアクセス不可能なローカルディスク記憶装置である。同じく追加的に、長期の記憶装置は、ネットワークを介して電子メールの供給に関連し、あるいは、それはネットワーク上のウェブページへのアクセスを供給してもよい。

【０００６】

同様に、サービス要求は、記憶装置システムのデータのアクセスを含んでもよく、サービスサブシステムもまた、記憶装置システムのデータの記憶を管理するためのハードウェアで実行されたモジュールを含んでもよい。したがって、一実施の形態では、そのような装置がファイルサーバーであって、記憶装置システムのデータがファイルに配列され、サービス要求は、記憶装置システムのファイルへの要求を含み、サービスサブシステムもまた、記憶装置システムに関連したファイルシステムを管理するためのハードウェアで実行されたモジュールを含む。

【０００７】

もう一つの関連する態様では、プロトコルは、ファイルシステムプロトコルを含み、ファイルシステムプロトコルは、ファイル読み込み及びファイル書込みを含む操作を定義する。装置がウェブサーバーであってもよく、記憶装置システムのデータがウェブページを含んでもよく、サービス要求は、記憶装置システム内のウェブページのための要求を含んでもよい。同様に、プロトコルがＩＰを含んでもよい。さらなる関連する態様では、記憶装置システムが記憶装置プロトコルを有し、サービスサブシステムは、記憶装置システムと相互接続するためのハードウェアで実行されたモジュールを含む。

【０００８】

もう一つの態様では、ネットワークを介してデータを送受信するためのサブシステムであって、ネットワークは層３及び４の少なくとも一つを有するプロトコルを用い、サブシステムは、

ネットワークからカプセル収納データを受信し、プロトコルに従ってそのようなデータをカプセルから出す受信器と、

プロトコルに従ってデータをカプセルに収納し、ネットワークを介してカプセル収納データを送信する送信器とを含む。

【０００９】

受信器及び送信器の少なくとも一つはハードウェアで実行され、その代わりに、又はそれに加えて、受信器及び送信器の少なくとも一つはハードウェアで加速される。さらなる実施の形態では、ネットワークは、TCP/IPプロトコルを用いる。関連する実施の形態では、データは、ネットワークを介してパケットで受信され、各パケットはプロトコルヘッダを有し、サブシステムは、また、受信器によって受信される各パケットのプロトコルヘッダ内に含まれる情報から独特な接続を決定する接続識別子を含む。もう一つの関連する実施の形態では、カプセル収納データは、ネットワーク接続に関連し、サブシステムは、接続の状態を格納する、ネットワーク接続に関連したメモリ領域をさらに含む。

【００１０】

もう一つの関連する態様では、記憶装置アクセス要求を生成し得るネットワークと記憶装置配置を相互接続するためのサービスサブシステムが提供される。この態様のサービスサブシステムは、

ａ．ネットワークサービス要求を受信し、そのようなサービス要求を満たし、そうして、データ記憶装置アクセス要求を発することができるサービスモジュールと、

ｂ．前記サービスモジュールからデータ記憶装置アクセス要求を受信し、そのような記憶装置アクセス要求を満たし、そうして、記憶装置配置アクセス要求を発することができる、該サービスモジュールに接続されたファイルシステムモジュールと、

ｃ．前記ファイルシステムモジュールから記憶装置配置アクセス要求を受信し、そのような記憶装置配置アクセス要求を満たすために、記憶装置配置を制御する、該ファイルシステムモジュールに接続された記憶装置モジュールとを含む。

【００１１】

モジュールの少なくとも一つはハードウェアで実行され、その代わりに、又はそれに加えて、モジュールの少なくとも一つは、ハードウェアで加速される。関連する実施の形態では、サービスモジュールは、

ｉ．ネットワークサービス要求を受信し、そのような要求が適切な否か決定し、もし適切ならば、情報が利用可能か否か応答し、さもなければ、データ記憶装置アクセス要求を発する、受信制御エンジンと、

ii．前記受信制御エンジンからの命令に基づいて、ネットワークサービス応答を生成し、前記データ記憶装置アクセス要求へのデータ記憶装置アクセス応答がある場合、該データ記憶装置アクセス応答を処理する、送信制御エンジンとを含む。

【００１２】

エンジンの少なくとも一つはハードウェアで実行され、その代わりに、又はそれに加えて、エンジンの少なくとも一つはハードウェアで加速される。他の関連する実施の形態では、サービスサブシステムは、コンピュータのマザーボードに直接統合され、又は、コンピュータに接続され得るアダプタカードに統合される。

【００１３】

もう一つの態様では、ネットワークサービス要求を受信し、そのようなサービス要求を満たすサービスモジュールが提供される。そのサービスモジュールは、

ａ．ネットワークサービスを受信し、そのような要求が適切であるか否かを決定し、もし適切ならば、情報が利用可能であるか否かを応答し、さもなければ、データ記憶装置アクセス要求を発する、受信制御エンジンと、

ｂ．前記受信制御エンジンからの命令に基づいて、ネットワークサービス応答を受信し、前記データ記憶装置アクセス要求へのデータ記憶装置アクセス応答がある場合、該データ記憶装置アクセス応答を処理する、送信制御エンジンとを含む。

【００１４】

エンジンの少なくとも一つはハードウェアで実行され、その代わりに、又はそれに加えて、エンジンの少なくとも一つはハードウェアで加速される。関連する実施の形態では、ネットワークサービス要求は、ＣＩＦＳプロトコル、ＳＭＢプロトコル、ＨＴＴＰプロトコル、ＮＦＳプロトコル、ＦＣＰプロトコル、又はＳＭＴＰプロトコルである。さらに関連する実施の形態では、サービスモジュールは、受信器によって受信されたネットワーク要求が該要求を発する権限を有する情報源から発せられたか否かを決定する認証エンジンを含む。まださらなる実施の形態では、認証エンジンは、受信器によって受信されたネットワーク要求が要求された動作を実行する権限を有する情報源から発せられたか否かを決定する。同じく、サービスモジュールは、コンピュータのマザーボードに直接統合され、又は、コンピュータに接続され得るアダプタカードに統合される。

【００１５】

もう一つの態様では、データ記憶装置アクセス要求を受信し、そのようなデータ記憶装置アクセス要求を満たすファイルシステムモジュールが提供される。そのファイルシステムモジュールは、

そのようなデータ記憶装置アクセス要求を受信し、それを翻訳し、そうして、記憶装置アクセス要求を発し得る、受信器と、

データ記憶装置アクセス応答を構築し、それを発する、前記受信器に接続される送信器であって、前記記憶装置アクセス要求への応答に基づいて適切なとき、そのような応答が情報を含む、前記送信器とを含む。

【００１６】

受信器及び送信器の少なくとも一つは、ハードウェアで実行され、その代わりに、又はそれに加えて、受信器及び送信器の少なくとも一つは、ハードウェアで加速される。さらなる実施の形態では、記憶装置アクセス要求は、前記モジュールが接続され得る記憶装置によって用いられるプロトコルに矛盾がない。まださらなる実施の形態では、プロトコルは、ＮＴＦＳ、ＨＰＦＳ、ＦＡＴ、ＦＡＴ１６、又はＦＡＴ３２である。もう一つの関連する実施の形態では、ファイルシステムモジュールは、また、モジュールが接続され得る

記憶装置におけるファイルの物理的位置を画定するテーブルを格納する、前記受信器に接続されるファイルテーブルキャッシュを含む。種々の実施の形態では、プロトコルは、記憶装置内の連続的な物理定位置に置かれるべきファイルを要求しない。他の実施の形態では、ファイルシステムモジュールは、コンピュータのマザーボードに直接統合され、又はコンピュータに接続され得るアダプタカードに統合される。

【００１７】

もう一つの態様では、要求情報源から記憶装置アクセス要求を受信し、そのような記憶装置アクセス要求を満たすために、記憶装置制御部と通信する記憶装置モジュールが提供される。その記憶装置モジュールは、

ａ．そのような記憶装置アクセス要求を受信し、それらを前記記憶装置制御部に適したフォーマットに変換する記憶装置要求インターフェースと、

ｂ．前記記憶装置制御部からの応答を取得し、そのような応答を前記要求情報源に適したフォーマットに変換する記憶装置承認インターフェースとを含む。

【００１８】

記憶装置要求インターフェースと記憶装置承認インターフェースの少なくとも一つはハードウェアで実行され、その代わりに、又はそれに加えて、記憶装置要求インターフェースと記憶装置承認インターフェースの少なくとも一つはハードウェアで加速される。さらなる実施の形態では、記憶装置モジュールは、また、データの部分への高速読込アクセスを可能にするために前記記憶装置に含まれるデータの一部のローカルコピーを保持するキャッシュ制御部を含む。他の関連する実施の形態では、記憶装置要求インターフェースと記憶装置承認インターフェースは、ポートに接続され、そのポートは、光ファイバチャネルを介して、あるいはＳＣＳＩ関連のプロトコルを利用する前記記憶装置制御部との通信を可能にする。さらなる実施の形態では、記憶装置モジュールは、コンピュータのマザーボードに直接統合され、又は、コンピュータに接続され得るアダプタカードに統合される。

【００１９】

もう一つの態様では、記憶装置アクセス要求を置かれ得るラインと記憶装置配置を相互接続するためのシステムが提供される。この態様のシステムは、

ａ．前記記憶装置アクセス要求を処理し、前記記憶装置配置へのアクセスに必要なところを生成し、応答の生成をさせる、前記記憶装置配置に接続されるサービス受信ブロックと、

ｂ．前記記憶装置配置内のファイルの物理的位置を画定するテーブルを格納する、前記受信ブロックに接続されるファイルテーブルキャッシュと、

ｃ．前記応答を送信するための、前記サービス受信ブロックに接続されるサービス送信ブロックとを含む。

【００２０】

サービス受信ブロックとサービス送信ブロックの少なくとも一つは、ハードウェアで実行され、その代わりに、又はそれに加えて、サービス受信ブロックとサービス送信ブロックの少なくとも一つは、ハードウェアで加速される。さらなる実施の形態では、システムは、また、サービス受信ブロックとサービス送信ブロックのそれぞれに接続される応答情報メモリであって、該メモリは前記要求に関連したヘッダに存在する情報を格納し、該情報は前記応答を構築する前記サービス送信ブロックによって使用される、応答情報メモリを含む。もう一つの関連する実施の形態では、記憶装置アクセス要求は、ネットワーク要求である。まだもう一つの実施の形態では、記憶装置アクセス要求は、前記ラインが接続されるローカルプロセッサによって生成される。

【００２１】

もう一つの態様では、多数のクライアントからの記憶装置アクセス要求を扱うための処理が提供される。その処理は、

前記クライアントのいずれかからの記憶装置アクセス要求の受信をテストするステップと、

あらゆる未決定の要求に従って記憶装置へのアクセスの完了をテストするステップとを含む。

【００２２】

この実施の形態では、記憶装置アクセス要求の受信テストと記憶装置へのアクセスの完了テストは、クライアントの数から独立して、多数のスレッドで実行される。さらなる実施の形態では、その処理は、また、要求の受信テストからの肯定的な決定を条件付けられ、該肯定的な決定を生じさせる要求を処理し、そのような要求に従って記憶装置アクセスを始めるステップを含む。関連する実施の形態では、前記処理は、また、未決定の要求に従って記憶装置へのアクセスの完了テストから肯定的な決定を条件付けられ、そのような未決定の要求を発するクライアントに応答を送るステップを含む。まだもう一つの関連する実施の形態では、スレッドの数が３よりも小さい。実際には、全処理は、単一のスレッドで具体化されてもよい。

【００２３】

もう一つの態様では、ネットワークを介してサービス要求を扱うための拡大縮小可能な装置であって、ネットワークはプロトコルを利用する拡大縮小可能な装置が提供される。この態様の装置は、

前記ネットワークプロトコルを用いるネットワークサービス要求を送受信するための第１の複数のネットワークサブシステムと、

前記ネットワークサービス要求を満たすための第２の複数のサービスサブシステムとを含む。

【００２４】

該ネットワークサブシステムと該サービスサブシステムのそれぞれ一つがハードウェアで実行され、あるいはハードウェアで加速されるものである。それに加えて、装置は、第１の複数のネットワークサブシステムのそれぞれを第２の複数のサービスサブシステムのそれぞれに接続するインタコネクトを含む。関連する実施の形態では、インタコネクトはスイッチであり、あるいはインタコネクトはバスである。

【００２５】

関連する態様では、記憶装置アクセス要求を生成され得るネットワークと記憶装置配置を相互接続するための拡大縮小可能なサービスサブシステムが提供される。サービスサブシステムは、

ネットワークサービス要求を受信し、そのようなサービス要求を満たし、そうして、データ記憶装置アクセス要求を発する第１の複数のサービスモジュールと、

データ記憶装置アクセス要求を受信し、そのような記憶装置アクセス要求を満たし、そうして、記憶装置配置アクセス要求を発する第２の複数のファイルシステムモジュールとを含む。

【００２６】

該サービスモジュールと該ファイルシステムモジュールのそれぞれ一つがハードウェアで実行され、あるいはハードウェアで加速されるものである。サービスサブシステムは、また、第１の複数のサービスモジュールのそれぞれを第２の複数のファイルシステムモジュールのそれぞれに接続するインタコネクトを含む。関連する実施の形態では、インタコネクトはスイッチであり、あるいはインタコネクトはバスである。さらなる実施の形態では、拡大縮小可能サービスサブシステムは、記憶装置配置アクセス要求を受信し、そのような記憶装置配置アクセス要求を満たすために該記憶装置配置を制御する第３の複数の記憶装置モジュールを含み、該記憶装置モジュールのそれぞれ一つはハードウェアで実行され、あるいはハードウェアで加速されるものである。同じく、サービスサブシステムは、前記ファイルシステムモジュールのそれぞれを前記記憶装置モジュールのそれぞれと接続する第２のインタコネクトを含む。同様に、関連する実施の形態では、インタコネクト及び第２のインタコネクトのそれぞれは、スイッチであり、あるいはインタコネクト及び第２のインタコネクトのそれぞれは、バスである。

【００２７】

好ましい実施の形態の詳細な記述

本発明の前述の特徴は、添付図面を参照するとともに、次の詳細な説明を参照することによっていっそう容易に理解されるだろう。

【００２８】

本記述と添付の特許請求の範囲の目的のために、次の用語は、文脈が別の方法で必要としないならば、示された意味を有する。

【００２９】

「ハードウェア実行（hardware-implemented）」サブシステムは、主なサブシステム機能がソフトウェアプログラムの直接の制御外で駆動する専用ハードウェアで実行されるサブシステムを意味する。そのようなサブシステムがソフトウェアの制御下にあるプロセッサと相互作用するが、サブシステム自体がソフトウェアによって直接的に制御されていないことに注意されたい。「主な」機能は、最も頻繁に使用されるものである。

【００３０】

「ハードウェア加速（hardware-accelerated）」サブシステムは、主なサブシステム機能が専用プロセッサ又は専用メモリを用いて実行され、それに加えて、（あるいはその代わりに）特別な用途ハードウェア、すなわち、専用プロセッサ及びメモリがＣＰＵに組み込まれるあらゆる中央演算処理装置（ＣＰＵ）及びメモリと性質が異なるものを意味する。

【００３１】

「TCP/IP」は、他の場所間でwww.ietf.orgにおけるインターネット管理委員会のウェブサイト上に定義されるプロトコルである。それは、参照によってここに組み込まれる。「ＩＰ」は、同じ場所で定義されるインターネットプロトコルである。

【００３２】

「ファイル（file）」は、データの論理的結合である。

【００３３】

プロトコル「ヘッダ（header）」は、プロトコルのユーザーと結び付けられたデータの輸送のためにプロトコルによって指定されたフォーマットにおける情報である。

【００３４】

「ＳＣＳＩ関連（SCSI-related）」プロトコルは、SCSI、SCSI-2、SCSI-3、Wide SCSI、Fast SCSI、Fast Wide SCSI、Ultra SCSI、Ultra2 SCSI、Wide Ultra2 SCSI、又はあらゆる類似の若しくは後継のプロトコルを含む。ＳＣＳＩは、「小型コンピュータシステムインターフェース（Small Computer System Interface）」であり、www.ansi.orgにウェブＵＲＬアドレスを持つ米国規格協会（ＡＮＳＩ）に従うコンピュータ周辺機器のパラレル接続の標準である。

【００３５】

「層３及び４（layers 3 and 4）」への言及は、ＩＳＯ標準である開放型システム間相互接続（「ＯＳＩ」）７層モデルの層３及び４を意味する。ＩＳＯ（国際標準化機構）は、www.iso.chにウェブＵＲＬアドレスを有する。

【００３６】

図１は、ネットワークを介してサービス要求を処理するように配置された本発明の一実施の形態の概略表示である。したがって、この実施の形態は、ファイルサーバー又はウェブサーバーが提供される配列を含む。本発明の実施の形態１１は、ネットワークインターフェース１３を介してネットワーク１０に接続される。ネットワーク１０は、例えば、複数のワークステーションへの通信リンクを含んでもよい。ここで、実施の形態１１は、また、記憶装置内部接続１４を介して複数の記憶装置に接続される。実施の形態１１は、ハードウエアで実行され、あるいは、ハードウェアで加速され（又はハードウェア実行及びハードウェア加速を組み合わせて利用され）てもよい。

【００３７】

図２は、図１に示された実施の形態のブロック図である。ネットワークサブシステム２１は、ネットワークサービス要求及び応答を送受信する。ネットワークサブシステム２１

は、ネットワークサービス要求に応ずるサービスサブシステム２２に接続される。ネトワークサブシステム２１、サービスサブシステム２２、あるいは両サブシステムは、ハードウェアで実行されるかハードウェアで加速されるかのいずれかであってよい。

【００３８】

図３は、より詳細にファイルサーバーとして配置された図１の実施の形態のブロック図である。ネットワークサブシステム３１は、ネットワークサービス要求及び応答を送受信する。ネットワークサブシステム３１は、サービスサブシステム３２に接続される。サービスサブシステムは、３つのモジュール、すなわち、サービスモジュール３３、ファイルシステムモジュール３４、及び記憶モジュール３５を含む。サービスモジュール３３は、サービスサブシステム３２を通過したネットワークサービス要求を解析し、適切であるとき、対応する記憶アクセス要求を発する。ネットワークサービス要求は、ＣＩＦＳ、ＳＭＢ、ＮＦＳ、又はＦＣＰのような種々のプロトコルのいずれかで伝達されてもよい。サービスモジュール３３は、ファイルシステムモジュール３４に接続される。もし、ネットワークサービス要求が記憶アクセス要求を含むならば、ファイルシステムモジュール３４は、その要求を記憶媒体によって利用されるファイル記憶プロトコル（例えば、ＨＴＦＳ、ＮＴＦＳ、ＦＡＴ、ＦＡＴ１６、又はＦＡＴ３２）と一致するフォーマットに変換することによって、記憶装置へのアクセスの要求を変換する。記憶モジュール３５は、サービスサブシステムが接続され得る記憶媒体に直接アクセスするためのバス要求と一致する（ＳＣＳＴのような）フォーマットにファイルシステムモジュール３４の出力を変換する。

【００３９】

図４は、図３に類似しており、ウェブサーバーとして配置される図１の実施の形態のブロック図である。ネットワークサーバー４１は、ネットワークサービス要求及び応答を送受信する。ネットワークサブシステム４１は、サービスサブシステム４２に接続される。サービスサブシステムは、３つのモジュール、すなわち、サービスモジュール４３、ファイルシステムモジュール４４、及び記憶モジュール４５を含む。サービスモジュール４３は、サービスサブシステム４２に通過したネットワークサービス要求を解析し、適切であるとき、対応する記憶アクセス要求を発する。ここで、ネットワークサービス要求は、典型的に、ＨＴＴＰプロトコルである。サービスモジュールは、記憶モジュール４５に接続されるファイルシステムモジュール４４に接続される。ファイルシステムモジュール４４及び記憶モジュール４５は、図３に関連して上述された、対応するモジュール３４及び３５に類似する方法で駆動する。

【００４０】

図５は、図２～４の実施の形態のネットワークサブシステム及びサービスサブシステムである。ネットワークサブシステム５１は、ネットワーク受信インターフェース５４からカプセル収納データを受信し、TCP/IP又は他のプロトコルバス５３に従ってそのデータをカプセルから出す。ネットワークサブシステム５１は、また、ネットワークを介してデータにアクセスする（同じく、ＰＣＩバスに接続される）ローカルプロセッサに供給するために、ＰＣＩバス５３に接続される。ネットワークサブシステム５１は、また、サービスサブシステム５２にデータを送信し、送信されたデータは、ＰＣＩバス５３を介してネットワーク受信インターフェース５４又なローカルプロセッサから来てもよい。サービスサブシステム５２は、また、それぞれ図２、３、及び４のサービスサブシステム２２、３２、及び４２に類似の方法で駆動する。

【００４１】

図６は、図５のネットワークサブシステム５１の詳細なブロック図である。図６のネットワークサブシステムは、（受信器６０１、受信バッファメモリ６０３、及び受信制御メモリ６０４を含む）受信器モジュール６１４、並びに、（送信器６０２、送信バッファメモリ６０５、及び送信制御メモリ６０６を含む）送信器モジュール６１３を含む。プロセッサ６１１は、受信器モジュール６１４と送信器モジュール６１３の両方に使用される。受信器６０１は、ネットワーク受信インターフェース６０７からカプセル収納データを受信し、それを翻訳処理する。受信器６０１は、受信制御メモリ６０４及び送信制御メモリ

６０６に含まれる制御情報を用いてそのデータをカプセルから出し、受信バッファメモリ６０５にそのカプセルから出されたデータを格納する。それは、受信バッファメモリ６０５からＰＣＩバス６１３を介してプロセッサ６１１によって取り出されるか、受信高速パス（fast path）インターフェース６０５に出力される。メモリ６１２は、データと命令の貯蔵のためにプロセッサ６１１に用いられる。送信器６０２は、送信後続パスインターフェース６１０かあるいはＰＣＩバス６１３を介してプロセッサ６１１から送信要求を受ける。

【００４２】

送信器６０２は、送信バッファメモリ６０５にそのデータを格納する。送信器６０２は、送信制御メモリ６０６に含まれる制御情報を用いて送信データをカプセル収納し、ネットワーク送信インターフェース６０９を介してネットワーク上のカプセル収納されたデータを送信する。

【００４３】

図７は、図６のネットワークサブシステムの受信モジュール６１４のブロック図である。パケットは、ネットワーク受信インターフェース６０７から受信エンジン７０１によって受信される。受信エンジン７０１は、パケットを解析し、そのパケットがエラーを含むか否か、TCP/IPパケットか、あるいは、TCP/IPパケットでないかを決定する。パケットは、パケット内に含まれるネットワークプロトコルヘッダの検査によってTCP/IPパケットかそうでないかを決定される。もしパケットがエラーを含むならば、それは落とされる。

【００４４】

もし、パケットがTCP/IPパケットでないならば、そのパケットは、受信バッファメモリアービタ７０９を介して受信バッファメモリ６０３に格納される。パケットが受信されたという表示は、プロセッサイベント待ち行列７０２に書き込まれる。プロセッサ７１５は、ＰＣＩバス７０４と受信ＰＣＩインターフェースブロック７０３を用いて受信バッファメモリ６０３からパケットを受信することができる。

【００４５】

もし、パケットがTCP/IPパケットならば、受信エンジン７０１は、パケットのプロトコルヘッダ内に含まれるネットワークアドレスとポートナンバーをこのパケットが属する接続、すなわち接続同定を特有の方法で識別する数に変換しようと試みるために、受信制御メモリ６０４に含まれるハッシュテーブルを用いる。もし、これが新しい接続道程ならば、パケットは、受信バッファメモリアービタ７０８を介して受信バッファメモリ６０３に格納される。パケットが受信されたという表示は、プロセッサイベント待ち行列７０２に書き込まれる。プロセッサ７１３は、ＰＣＩバス７０４と受信ＰＣＩインターフェースブロック７０３を用いて受信バッファメモリ６０３からパケットを受け取る。プロセッサは、TCP/IPプロトコルで指定されるように要求されるならば、新しい接続を確立することができ、あるいは他の適切なアクションをとり得る。

【００４６】

もし、接続同定が既に存在するならば、受信エンジン７０１は、各接続状態についての情報を含むデータのテーブルへのインデックスとしてこの接続同定を用いる。この情報は、「ＴＣＰ制御ブロック」（「ＴＣＢ」）と呼ばれる。各接続のためのＴＣＢは、送信制御メモリ６０６に格納される。受信エンジン７０１は、受信器ＴＣＢアクセスインターフェース７１０を介してこの接続のためのＴＣＢにアクセスする。それは、TCP/IPプロトコルに従ってこのパケットを処理し、受信バッファメモリ６０５のこの接続のための受信されたバイトストリームに結果として生じるバイトを加える。もし、この接続におけるデータがプロセッサ７１３のために予定されているならば、幾らかのバイトが受信されたという表示は、プロセッサイベント待ち行列７０２に書き込まれる。プロセッサは、ＰＣＩバス７０４と受信ＰＣＩインターフェースブロック７０３を用いて受信バッファメモリ６０３からバイトを受け取る。もし、この接続におけるデータが高速パスインターフェース６０８のために予定されているならば、幾らかのバイトが受信されたという表示は、高速パスイベント待ち行列７０５に書き込まれる。受信ＤＭＡエンジン７０６は、受信バッファ

メモリ６０３からバイトを受け取り、高速パスインターフェース６０８にそれらを出力する。

【００４７】

受信エンジン７０１によって受信されたいくつかのパケットは、ＩＰパケットのフラグメントであり得る。もし、これがその場合ならば、フラグメントは、最初に受信バッファメモリ６０３で新たに組み立てられる。完全なＩＰパケットが新たに組み立てられたとき、通常のパケット処理が上述のように適用される。

【００４８】

ＴＣＰプロトコルによれば、接続は、SYN_SENT、SYN_RECEIVED、及びESTABLISHEDを含む多くの異なる状態で存在し得る。ネットワークノードがネットワークサブシステムとの接続を確立することを望むとき、それは、最初にSYNフラグセットを持つTCP/IPパケットを送信する。このパケットは、新しい接続同定を有するので、プロセッサ７１３によって検索される。プロセッサ７１３は、SYN_RECEIVEDへのこの接続のためにＴＣＢ内に接続状態を設定することを含むすべての要求される初期化を実行する。SYN_RECEIVEDからESTABLISHEDへの移行は、TCP/IPプロトコルに従って受信エンジン７０１によって実行される。プロセッサ７１３がネットワークサブシステムを介してネットワークノードへの接続を確立することを望むとき、最初にSYN_SENTへのこの接続のためにＴＣＢ内に接続状態を設定することを含むすべての要求される初期化を実行する。それは、それから、SYNフラグセットを持つTCP/IPパケットを送信する。SYN_SENTからESTABLISHEDへの移行は、TCP/IPプロトコルに従って受信エンジン７０１によって実行される。

【００４９】

もし、プロトコルヘッダ内にSYNフラグ又はFINフラグあるいはRSTフラグセットを有するパケットが受信され、そして、これがプロセッサ７１３によるアクションを要求するならば、受信エンジン７０１は、プロセッサイベント待ち行列７０２へのエントリを書き込むことによって、このイベントをそのプロセッサに通知する。プロセッサ７１３は、TCP/IPプロトコルによって要求されるように適切なアクションをとり得る。

【００５０】

受信されたパケットにTCP/IPプロトコルを適用する結果として、１以上のパケットが今この接続で送信されるべきことが可能になる。例えば、受信データの承認が送信される必要があってもよく、あるいは、受信パケットは、そのようなデータが送信のために適用可能であるならば、より多くのデータがこの接続で送信されることを可能にする増加されたウィンドウサイズを示してもよい。受信エンジン７０１は、それに応じてＴＣＢを変更し、それから、受信器送信待ち行列要求インターフェース７１１を介して図８の送信待ち行列８０２に接続同定を書き込むことによって送信試みを要求することによって、このことを達成する。

【００５１】

受信データは、受信バッファメモリ６０３内の別々のユニット（バッファ）に格納される。バッファ内のすべてのデータがプロセッサ７１３によって検索されるか、高速パスインターフェース６０５に出力されるとすぐに、バッファは、解放、すなわち、新しいデータを格納するために再利用され得る。類似のシステムは、送信バッファメモリ６０５のために駆動するが、しかしながら、送信の場合、そのバッファは、その中のすべてのデータがTCP/IPプロトコルを用いて、送信データを受信するネットワークノードによって完全に承認されるときのみ解放される。送信されたデータが承認されたことをパケットのプロトコルヘッダが示すとき、受信エンジン７０１は、受信器フリー送信バッファ要求インターフェース７１２を介して図８のフリー送信バッファブロック８０５にこのことを示す。

【００５２】

さらに、受信エンジン７０１がTCP/IP自身と同様にTCP/IPをうまく処理する上部層プロトコル（ＵＬＰ）を処理することを可能にする。この場合、完全なＵＬＰプロトコルデータユニット（ＰＤＵ）が受信されたときのみ、イベント待ち行列エントリは、処理又はイベント待ち行列７０２と高速パスイベント待ち行列７０５に書き込まれる。完全なＵＬＰ

のＰＤＵのみがプロセッサ７１３によって受信され、高速パスインターフェース６０８に出力される。ＵＬＰの一例はNetBIOSである。ＵＬＰ処理を可能にすることは、接続毎に基づいてなされてもよい、すなわち、いくつかの接続は、ＵＬＰ処理を可能にし、その他のものはそうでなくてもよい。

【００５３】

図８は、図６のネットワークサブシステムの送信モジュール６１３のブロック図である。TCP/IPを用いてネットワーク上で送信されるべきデータは、送信ＤＭＡエンジン８０７に入力される。このデータは、送信高速パスインターフェース６１０か、ＰＣＩバス７０４と送信ＰＣＩインターフェース８０８を介してプロセッサ７１３からのいずれかからの入力である。それぞれの場合、ＴＣＰ／ＩＰ接続がデータを送信するために用いられるべきか否かを決定する接続同定は、同じく入力される。上述のように、各接続は、接続状態についての情報を含む、組み合わされたＴＣＢを有する。

【００５４】

送信ＤＭＡエンジンは、この接続のために格納されたバイトストリームに入力されたバイトを加えて、送信バッファメモリ６０５内にそのデータを格納する。入力の終わりに、それは、それに応じて接続のためにＴＣＢを変更し、同じく、接続同定を送信待ち行列８０２に書き込む。

【００５５】

送信行列８０２は、３つの情報源、すなわち、受信器送信待ち行列要求インターフェース７１１、タイマー機能ブロック８０６、及び送信ＤＭＡエンジン８０７から接続同定の形式で送信要求を受け取る。要求が受信されるので、それらは待ち行列に置かれる。待ち行列がからであるときはいつも、待ち行列の前における接続同定のための送信要求は、送信エンジン８０１へ通過される。送信エンジン８０１が送信要求の処理を完了すると、この接続同定は、待ち行列の前から取り除かれ、その処理が繰り返す。

【００５６】

送信エンジン８０１は、送信待ち行列８０２から送信行列を受け取る。各要求のために、送信エンジン８０１は、その接続と要求される送信パケットにTCP/IPプロトコルを適用する。これをするために、それは、送信制御メモリアービタ８０５を介して送信制御メモリ６０６内の接続のためにＴＣＢにアクセスし、送信バッファメモリアービタ８０４を介して送信バッファメモリ６０５から接続のための格納されたバイトストリームを検索する。

【００５７】

接続のための格納されたバイトストリームは、送信バッファメモリ６０５内の別々のユニット（バッファ）に格納される。上述のように、各バッファは、TCP/IPプロトコルを用いて、送信データを受信するネットワークノードによってその中のすべてのデータが完全に承認されたときのみ解放され得る。送信されたデータが承認されたことをパケットのプロトコルヘッダが示すとき、受信エンジン７０１は、受信器フリー送信バッファ要求インターフェース７１２を介してフリー送信バッファブロック８０５にこのことを示す。フリー送信バッファブロック８０５は、完全に承認されたすべてのバッファを解放し、これらのバッファは、新しいデータを格納するために再利用され得る。

【００５８】

TCP/IPは、ある条件が満たされるならば一定間隔で実行されるべきある動作を要求する多くのタイマー機能を有する。これらの機能は、タイマー機能ブロック８０６によって実行される。一定間隔で、タイマー機能ブロック８０６は、送信制御メモリアービタ８０３を介して各接続のためのＴＣＢにアクセスする。もし、あらゆる操作が特定の接続のために実行される必要があるならば、その接続のためのＴＣＢは、それに応じて変更され、接続同定は、送信待ち行列８０２に書き込まれる。

【００５９】

そのうえ、送信ＤＭＡエンジン８０７がTCP/IPをうまく処理する上部層プロトコルを処理することができる。この場合、完全なＵＬＰプロトコルデータユニットのみは、プロセ

ッサ７１３か、送信高速パスインターフェース６１０のいずれかから送信ＤＭＡエンジン８０７に入力される。送信ＤＭＡエンジン８０７は、それから、ＰＤＵの前にＵＬＰヘッダを添付し、接続のための格納されたバイトストリームに「前に添付された」ＵＬＰヘッダと入力されたバイトを加える。図２に関連して上述されるように、ＵＬＰの一例はNetBIOSである。ＵＬＰ処理を可能にすることは、接続毎に基づいてなされてもよい。すなわち、いくつかの接続がＵＬＰ処理可能であり、そのたのものがそうでなくてもよい。

【００６０】

もし、プロセッサ７１３が生のパケットを送信すること、すなわち、TCP/IPを用いてデータのハードウェアの自動送信することなく、データを送信することを望むならば、プロセッサ７１３が送信ＤＭＡエンジン８０７にデータを入力するとき、それは、特定の接続同定を用いる。この特定の接続同定は、送信エンジン８０１に、プロセッサ７１３によって送信ＤＭＡエンジン８０７に入力として正確に生のパケットを送信させる。

【００６１】

図９は、ワークステーション又はサーバーのようなネットワークノードで使用するネットワークインターフェースアダプタとして図５のネットワークサブシステムの使用を示すブロック図である。この実施の形態では、ネットワークサブシステム９０１は、コンピュータに接続されるアダプタカード９００に統合される。アダプタカード９００は、ネットワークインターフェース９０４を介してネットワークに接続される。アダプタカード９００はまた、ＰＣＩバス９０７とＰＣＩブリッジ９１２を介してコンピュータのマイクロプロセッサに接続される。ＰＣＩバス９０７はまた、ビデオシステム９１３のような周辺装置にアクセスするためにコンピュータによって用いられてもよい。受信モジュール９０２と送信モジュール９０３は、図６の受信モジュール６１４と送信モジュール６１３と類似の方法で動作する。その代わりに又はそれに加えて、アダプタカード９００は、ネットワーク上のリモートノード又はマイクロプロセッサ１０によって記憶装置配列への高速アクセスを提供するために、単一プロトコル高速受信パイプ９０６と単一プロトコル高速送信パイプ９０５を介して、図２、３、４、又は５のアイテム２２、３２、４３、又は５２のそれぞれのいずれかに相当するサービスモジュールに接続されてもよい。

【００６２】

図１０は、図３に示されるような一実施の形態で使用する図３のＳＭＢサービスモジュール３３とファイルシステムモジュール３４のハードウェアで実行される組み合わせのブロック図である。図１０の実施の形態では、ＳＭＢ要求は、サービス受信ブロック１０１への入力１０５で受信される。最終的に、この実施の形態による処理は、出力１０６上の対応するＳＭＢ応答の送信を結果としてもたらす。この応答の一部はヘッダを含む。出力ヘッダを作るために、入力ヘッダは、ＳＭＢ応答情報メモリ１０３に格納される。ブロック１０１は、ＳＭＢ要求を処理し、応答を生成する。要求の性質に依存して、ブロック１０１は、ファイルテーブルキャッシュ１０４にアクセスし、ディスクアクセス要求を発してもよい。さもなければ、応答が送信ブロック１０２を直接中継される。サービス送信ブロック１０２は、出力１０６上にブロック１０１によって生成される応答を送信する。ディスクアクセス要求がブロック１０１によって発せられ、ディスク応答のライン１０８上で受信された場合、送信ブロック１０２は、ライン１０６上に適切なＳＭＢ応答を発する。受信及び送信の両モジュール１０１及び１０２は、ＰＣＩバス１０９を介してホストシステムと選択的に通信する。供給されるとき、そのような通信は、伝統的なオペレーティングシステムの範囲外に高速で、ハードウェアで実行されるファイルシステムアクセスをホストシステムに与えるように、ホストシステムがネットワークを介する代わりに実施の形態と直接通信することを可能にする。

【００６３】

図１１は、図３に示されるような一実施の形態において使用する図３のＳＭＢサービスモジュール３３とファイルシステムモジュール３４のハードウェアで加速される組み合わせのブロック図である。動作は、同じく番号を付されたブロック及びライン１０５、１０７、１０８、及び１０６に関して図１０に関連して上述されたものに類似している。しか

しながら、専用のファイルシステムプロセッサ１１０は、専用バス１１２を介して動作する専用メモリ１１１に関連して、ブロック１０１及び１０２の処理を制御する。それに加えて、これらのアイテムは、ソフトウェアで変更され得るので、そのような処理の取扱いに柔軟性を提供する。

【００６４】

図１２Ａは、それぞれ図３又は図４のアイテム３３又は４３のようなハードウェアで実行されるサービスモジュールのブロック図である。サービスモジュール１２００は、ネットワークサービス要求を受信し、そのようなサービス要求を実現し、データ記憶アクセス要求を発してもよい。サービスモジュール１２００は、送信器１２０２に接続された受信器１２０１と、受信器１２０１及び送信器１２０２の両方に接続されたデータ記憶アクセスインターフェース１２０３とを含む。受信器１２０１は、ネットワークサービス要求を受信し、それを解釈する。サービス要求を受信し次第、受信器１２０１は、データ記憶アクセスインターフェース１２０３にその要求を送るか、あるいは、ネットワークサービス要求を満たす情報を送信器１２０２に送る。もし、その要求がデータ記憶アクセスインターフェース１２０３に送られるならば、データ記憶アクセスインターフェース１２０３は、データ記憶アクセス要求を構成し、それを発する。データ記憶アクセスインターフェースは、また、データ記憶アクセス要求への応答を受信し、オリジナルのネットワークサービス要求を満たすために必要とされる情報を抽出する。その情報は、それから送信器１２０２に送られる。送信器１２０２は、受信器１２０２又はデータ記憶アクセスインターフェース１２０３からそれに送られる情報を処理し、ネットワークサービス応答を構築し、それを発する。

【００６５】

図１２Ｂは、それぞれ図３又は図４のアイテム３４又は４４のようなハードウェアで実行されるファイルモジュールのブロック図である。ファイルシステムモジュール１２１０は、データ記憶アクセス要求を受信し、そのようなデータサービスアクセス要求を実現し、記憶装置アクセス要求を発してもよい。ファイルシステムモジュール１２１０は、送信器１２１２に接続された受信器１２１１と、受信器１２１１及び送信器１２１２の両方に接続されたデータ記憶装置アクセスインターフェース１２１３とを含む。受信器１２１１は、データ記憶アクセス要求を受信して解釈し、データ記憶装置アクセスインターフェース１２１３へその要求を送るか、あるいは、データ記憶アクセス要求を満たす情報を送信器１２１２に送る。もし、その要求がデータ記憶装置アクセスインターフェース１２１３に送られるならば、データ記憶装置アクセスインターフェース１２１３は、データ記憶装置アクセス要求を構築し、それを発する。データ記憶装置アクセスインターフェース１２１３は、また、データ記憶装置アクセス要求への応答を受信し、オリジナルのデータ記憶アクセス要求を満たすために必要とされる情報を抽出する。その情報は、それから送信器１２１２に送られる。送信器１２１２は、受信器１２１１又はデータ記憶装置アクセスインターフェースモジュール１２１３からそれに送られた情報を処理し、データ記憶アクセス応答を構築し、それを発する。

【００６６】

図１２Ｃは、結合されたサービスモジュール及びファイルモジュールを提供する、図１０のハードウェアで実行されるサービスサブシステムの詳細なブロック図である。図１２Ｃ内の点線１２９は、この実行の機能間の分割を示す。ライン１２９の左側にはサービスモジュール部があり、ライン１２９の右側にはファイルシステムモジュール部がある。（しかしながら、ＳＭＢ受信制御エンジン１２１とＳＭＢ送信制御エンジン１２２を結ぶ二頭の矢がサービスモジュール部とファイルシステムモジュール部のそれぞれのためのエンジン１２１及び１２２問の２方向通信を適切に供給することを理解されるだろう。）

【００６７】

図１２Ｃでは、ＳＭＢフレームは、ネットワーク受信インターフェース１２１ｆを介してネットワークサブシステムから受信され、ＳＭＢフレーム翻訳エンジン１２１ｂに送られる。ここで、フレームは解析され、多くのタスクが実行される。ヘッダの最初のセクシ

ョンは、ＳＭＢ応答の接続基礎毎に適切な情報をメモリ１０３に格納する、ＳＭＢ応答情報制御１２３にコピーされる。完全なフレームは、受信バッファメモリ１２１Ｃ内のバッファに書き込まれ、受信制御メモリ１２１ｄは更新される。ＳＭＢフレームヘッダの適切な部分は、ＳＭＢ受信制御エンジン１２１に送られる。

【００６８】

図１２ＣのＳＭＢ受信制御エンジン１２１は、ヘッダからその情報を解剖し、適切な場合には、批准エンジン１２４からファイルアクセス許可を要求する。ファイルアクセスが要求されるＳＭＢフレームのために、ＳＭＢ受信制御エンジン１２１は、ＳＭＢフレームヘッダからファイルパス情報かファイル同定のいずれかを抽出し、ＭＦＴ制御エンジン１２５に要求されたファイルデータの物理的位置を要求する。

【００６９】

ＭＦＴ制御エンジン１２５は、ＳＭＢ受信制御エンジン１２１からの要求を待ち行列に入れることができ、同様に、ＳＭＢ受信制御エンジン１２１は、ＭＦＴ制御エンジン１２５からの要求を待ち行列に入れることができる。これは、２っのエンジンが互いに非同期に動作することを可能にし、従ってＭＦＴ要求が顕著な間入ってくるＳＭＢフレームが処理されることを可能にする。

【００７０】

ＭＦＴ制御エンジン１２５は、ＳＭＢ受信制御エンジン１２１からの要求を処理する。典型的に、SMB OPENコマンドのために、要求は、必要な物理的ファイル位置情報を得るためのディスクアクセスが必要である。これが必要な場合、ＭＦＴ制御エンジン１２５は、必要な圧縮ＳＣＳＩ要求を生成する圧縮ＳＣＳＩフレーム生成エンジン１２１ａに要求を送る。圧縮ＳＣＳＩプロトコル（「ＣＳＰ」）は、ＳＣＳＩコマンドが図１７Ａと他の図に関連して以下に記述される方法で生成され得るデータフォーマットに関する。圧縮ＳＣＳＩデータがＳＣＳＩから得られないが、むしろＳＣＳＩデータが得られ得る情報源であるので、我々は、時々圧縮されたＳＣＳＩデータを「プロトーＳＣＳＩ」データとして言及する。適切なプロトーＳＣＳＩ応答は、それが処理されるＭＦＴ制御エンジン１２５に送り返され、ＭＦＴキャッシュ１０４は更新され、物理的ファイル情報は、ＳＭＢ受信制御エンジン１２１に送り返さる。

【００７１】

典型的に、最近アクセスされた小さなファイルに関してSMB READ又はWRITEコマンドのために、ファイル情報は、ＭＦＴキャッシュ１０４内に存在する。したがって、ディスクアクセスは要求されない。

【００７２】

ＳＭＢ受信制御エンジン１２１がＭＦＴ要求からその応答を受信し、ファイルデータのためのディスクアクセスが要求されるとき、典型的なREAD又はWRITEコマンドに必要であるように、１以上のプロトーＳＣＳＩ要求は、プロトーＳＣＳＩフレーム生成エンジン１２１ａに送られる。

【００７３】

プロトーＳＣＳＩフレーム生成エンジン１２１ａは、プロトーＳＣＳＩヘッダを構築し、必要な場合、例えば、WRITEコマンドのために、受信バッファメモリ１２１ｃからファイルデータを引くために、ファイルデータＤＭＡエンジン１２１ｅをプログラムする。プロトーＳＣＳＩフレームは、プロトーＳＣＳＩ送信インターフェース１２１ｇを介してプロトーＳＣＳＩモジュールに送られる。ディスクアクセスが要求されない場合、ＳＭＢ応答要求は、直接ＳＭＢ送信制御エンジン１２２に送られる。

【００７４】

プロトーＳＣＳＩフレームは、プロトーＳＣＳＩモジュールから受信され、プロトーＳＣＳＩ受信インターフェース１２２ｆを介してプロトーＳＣＳＩフレーム翻訳エンジン１２２ｂに送られる。ここで、そのフレームは解析され、多くのタスクが実行される。ＭＦＴ応答は、ＭＦＴ制御エンジン１２５に送り返される。他のすべてのフレームは、受信バッファメモリ１２１ｃ内のバッファに書き込まれ、受信制御メモリ１２１ｄは更新される

。プロトーＳＣＳＩフレームヘッダの適切な部分は、ＳＭＢ送信制御エンジン１２２に送られる。

【００７５】

各ＳＭＢ接続は、独特な同定を以前に割り当てられた。すべてのプロトーＳＣＳＩフレームは、この同定を含み、ＳＭＢ送信制御エンジン１２２は、ＳＭＢ受信制御エンジン１２１からの状態情報を要求し、必要な場合これを更新するために、この独特な同定を用いる。ＳＭＢ応答のための必要な情報がプロトーＳＣＳＩモジュールから受信されたとき、ＳＭＢ送信制御エンジン１２２は、ＳＭＢフレーム生成エンジン１２１ａに要求を送る。

【００７６】

ＳＭＢフレーム生成エンジン１２１ａは、ＳＭＢ応答情報メモリ１０３内に含まれるデータ及びＳＭＢ送信バッファメモリ１２２ｃに格納されたファイルデータからＳＭＢ応答：フレームを構築する。それは、結果としてそれをネットワークサブシステムに転送するＳＭＢ送信インターフェース１０６にそのフレームを送る。

【００７７】

図１３は、図１１のハードウェア加速サービスサブシステムの詳細なブロック図である。入力１０５を越えて供給されるＩＰブロックから入ってくるＳＭＢフレームは、ＳＭＢ受信ＦＩＦＯ１３１７を介して、ＳＭＢ受信バッファメモリ１２１Ｃ内のフリーバッファに書き込まれる。ＳＭＢ受信バッファメモリ１２１ｃは、一実施の形態では、２Ｋｂの長さであり、一つのＳＭＢフレームが多くの受信バッファにまたがり得る一連の受信バッファを含む。フレームがＳＭＢ受信バッファメモリ１２１ｃに書き込まれるので、ＳＭＢ受信バッファ記述子は、ＳＭＢ受信制御メモリ１２１ｄで更新される。

【００７８】

３２ビット接続同定及び３２ビットフレームバイトカウントは、フレームの始まりでＩＰブロックからＳＭＢブロックに送られる。これら２つのフィールドは、受信バッファメモリ１２１ｃの受信バッファの始めの２つの位置に書き込まれる。

【００７９】

フレームが格納されるが、ＳＭＢヘッダもまた、ＳＭＢ送信処理による後の使用のためにＳＭＢ応答情報メモリ１０３に書き込まれる。ＩＰブロックによってＳＭＢブロックに送られる独特な接続同定は、ＳＭＢ応答情報メモリ１０３内の適切な情報フィールドへのポインタとして用いられる。このメモリは、１６ワードのブロックで、各独特な接続同定のために一つのブロックとして配置される。１２８ＭｂのＳＤＲＡＭで適合させて、これは２Ｍの接続を可能にする。目下、ＳＭＢフレームの最初の３２バイトは各情報フィールドに書き込まれる。

【００８０】

完全なフレームが受信バッファメモリ１２１ｃに書き込まれたとき、ＳＭＢバッファロケータは、ＳＭＢ受信イベント待ち行列１３１４に書き込まれ、ホストプロセッサ１３０１への割り込みが生成される。ＳＭＢバッファロケータは、バッファポインタと「最後の」ビットを含むＳＭＢフレームに関する情報を含む。バッファポインタは、ＳＭＢフレームの始まりを含む受信バッファメモリ１２１ｃを指し示す。「最後の」ビットは、このバッファがＳＭＢフレームの終わりも含むか否か（すなわち、ＳＭＢフレームが２Ｋｂの長さより小さいか否か、）を示す。

【００８１】

ホストプロセッサ１３０１は、イベント待ち行列１３１４に結び付けられた適切なＳＭＢ受信イベントレジスタを読むことによってＳＭＢ受信待ち行列１３１４内のＳＭＢバッファロケータを読むことができる。ＳＭＢバッファロケータから読み取られたバッファポインタから、ホストプロセッサ１３０１は、受信バッファメモリ１２１ｃ内のＳＭＢフレームの最初のバッファのアドレスを決定することができ、従ってＳＭＢヘッダとフレームの最初の部分を読むことができる。

【００８２】

もし、ＳＭＢフレームが２Ｋｂよりも長く、ＳＭＢフレームの最初の２Ｋｂ以上を読む

必要があるならば、この受信バッファに結び付けられた受信バッファ記述子は、受信制御メモリ１２１ｄから読まれるべきである。この受信バッファ記述子は、ＳＭＢフレームの次のバッファへのポインタを含む。受信バッファがＳＭＢフレームの終わりを含むことを指摘することを示す「最後の」ビットを前のバッファの記述子が含まなければ、この次のバッファは、同様に、それに結び付けられた受信バッファ記述子を有する。

【００８３】

受信されたＳＭＢフレームを呼んだ後、もし、フレーム内に含まれるデータがさらに使用されるできではないならば、受信されたフレームのバッファは、受信バッファ制御メモリ１２１ｄに含まれる受信フリーバッファ待ち行列にポインタを書き込むことによって、結び付けられた受信返却フリーバッファレジスタに書き込むことによって、再び使用するために利用可能にされる。

【００８４】

プロトーＳＣＳＩフレームを送信するために、ホストプロセッサ１３０１は、受信フェッチフリーバッファレジスタから読み込むことによってフリーＳＭＢ受信バッファへのポインタを最初に得る。このアクションは、受信制御メモリ１２１ｄに含まれるフリーバッファ待ち行列からフリーバッファへのポインタを持ってくる。このバッファでは、プロトーＳＣＳＩ要求フレームの始まりが構築され得る。

【００８５】

プロトーＳＣＳＩフレームをプロトーＳＣＳＩエンティティーに移すようにプロトーＳＣＳＩ送信エンティティーに要求するために、ホストプロセッサ１３０１は、プロトーＳＣＳＩ送信イベント待ち行列１３１５に結び付けられた受信プロトーＳＣＳＩイベントレジスタにそれらを書き込むことによって、プロトーＳＣＳＩ送信イベント待ち行列１３１５にバッファロケータとバッファオフセット対を書き込む。

【００８６】

バッファロケータは、プロトーＳＣＳＩフレームのためのデータを含むバッファへのポインタを含む。バッファオフセットは、バッファと長さフィールドの中のデータの始まりに対するオフセットを含む。バッファロケータは、また、さらなるバッファロケータ／バッファオフセット対が、このプロトーＳＣＳＩフレームのためのより多くのデータへのポインタを含むプロトーＳＣＳＩ送信イベント待ち行列１３１５に書き込まれるか否かを示すためのさごのビットを含む。

【００８７】

もし、プロトーＳＣＳＩフレームがもう一つのＳＭＢ受信バッファからのデータを含むべきならば、SMB WRITEコマンドのために典型的であるように、ホストプロセッサ１３０１は、このＳＭＢ受信バッファを記述するもう一つのバッファロケータ／バッファオフセット対をプロトーＳＣＳＩ送信イベント待ち行列１３１５に書き込まなければならない。もし、プロトーＳＣＳＩフレームに含まれるべきデータが一つのＳＭＢ受信バッファ以上にまたがるならば、プロトーＳＣＳＩ送信エンティティーは、データにともにリンクするために、受信制御メモリ１２１ｄに位置される結び付けられたＳＭＢ受信バッファ記述子内のバッファポインタを用いることができる。もし、余分なデータがＳＭＢ受信フレームからのものであるならば、これらの記述子は、ＳＭＢ受信エンティティーによって予め満たされたであろう。

【００８８】

ＳＭＢ受信バッファからのデータが一つ以上のプロトーＳＣＳＩフレームのために用いられ得るので、それらが用いられた後のＳＭＢ受信バッファを自由化することは単純な処理ではない。プロトーＳＣＳＩ送信に含まれない受信ＳＭＢフレームのセクションを含むＳＭＢ受信バッファは、関連付けられた受信戻りフリーバッファレジスタを介して受信制御メモリに含まれるフリーバッファ待ち行列にそれらを書き戻すことによって自由にされ得る。プロトーＳＣＳＩフレームに含まれるべきデータを含むＳＭＢ受信バッファは、それらの中のデータが送信されるまで自由にされないような同一の方法で自由にされ得ない。

【００８９】
それで、ＳＭＢデータを含む種々のプロトーＳＣＳＩフレームに対するバッファロケータ／バッファオフセット対がプロトーＳＣＳＩ送信イベント待ち行列１３１５に書き込まれた後、オリジナルのＳＭＢ受信バッファへのポインタも同じく、プロトーＳＣＳＩに書き込まれる。これらのポインタは、受信制御メモリに含まれるフリーバッファ待ち行列に戻されて自由にされるべきであることを示すために記録される。プロトーＳＣＳＩ送信が次々と処理されるので、ＳＭＢ受信バッファは、それらの中のあらゆるデータが送信された後にのみ自由にされる。

【００９０】
ＩＰブロックから入ってくるプロトーＳＣＳＩフレームは、プロトーＳＣＳＩ受信ＦＩＦＯ１３２７を介してＳＭＢ送信バッファメモリ１２２ｃ内のフリーバッファに書き込まれる。ＳＭＢ送信バッファは２Ｋｂ長であり、一つのプロトーＳＣＳＩフレームは、多くの送信バッファにまたがってもよい。フレームがＳＭＢ送信バッファメモリ１２２ｃに書き込まれるので、ＳＭＢ送信バッファ記述子は、ＳＭＢ送信制御メモリ１２２ｄ内で更新される。

【００９１】
完全なフレームがＳＭＢ送信バッファメモリ１２２ｃに書き込まれるとき、ＳＭＢバッファロケータは、プロトーＳＣＳＩ受信イベント待ち行列１３２４に書き込まれ、ホストプロセッサ１３０１への割り込みが生成される。ＳＭＢバッファロケータは、バッファポインタと「最後の」ビットを含むプロトーＳＣＳＩフレームに関する情報を含む。バッファポインタは、プロトーＳＣＳＩフレームの始まりを含む送信バッファメモリ１２１ｃ内のバッファを指し示す。「最後の」ビットは、このバッファが同じくプロトーＳＣＳＩフレームの終わりを含むか否か（すなわち、フレームが長さで２Ｋｂより小さいか否か）を示す。

【００９２】
ホストプロセッサ１３０１は、イベント待ち行列１３２４に関連した適切なプロトーＳＣＳＩ受信イベント待ち行列１３２４を読むことによって、プロトーＳＣＳＩ受信イベント待ち行列のバッファロケータを読むことができる。バッファロケータから読まれたバッファポインタから、ホストプロセッサ１３０１は、送信バッファメモリ１２２ｃ内のプロトーＳＣＳＩフレームの最初のバッファのアドレスを決定することができ、従ってヘッダとそのフレームの最初の部分を読むことができる。

【００９３】
もし、プロトーＳＣＳＩフレームが２Ｋｂより長く、フレームの最初の２Ｋｂより多くを読む必要があるならば、この送信バッファに関連した送信記述子は、受信制御メモリ１２１ｄから読まれるべきである。バッファがプロトーＳＣＳＩフレームの終わりを含むことを指摘することを示す「最後の」ビットを前のバッファの記述子が含まないならば、記述子は、プロトーＳＣＳＩフレームの次のバッファへのポインタを含む。この次のバッファは、同様に、それに関連した送信記述子を有する。

【００９４】
受信されたプロトーＳＣＳＩフレームを読んだ後、もし、そのフレーム内に含まれるデータが更に用いられるべきでないならば、受信されたフレームのバッファは、それに関連した送信戻りフリーバッファレジスタに書き込むことによって、送信制御メモリ１２２ｄに含まれる送信フリーバッファ待ち行列に戻されるべきである。

【００９５】
ＳＭＢフレームを送信するために、ホストプロセッサは、関連したレジスタから読むことによって、送信制御メモリ１２２ｄに含まれる送信フリーバッファ待ち行列から送信バッファメモリ１２２ｃ内のフリーＳＭＢ送信バッファへのポインタを最初に得る。このバッファでは、ＳＭＢ応答フレームの始まりが構築され得る。

【００９６】
３２ビット接続道程と３２ビットＳＭＢ送信制御フィールドは、バッファ内のＳＭＢフ

レーム前に置かれる。ＳＭＢ送信制御フィールドは、２４ビットフレームバイトカウントと前置ヘッダビットを含む。もし、前置ヘッダビットが設定されるならば、接続同定とＳＭＢ送信制御フィールドがＩＰブロックに送られた後に、応答情報メモリ１０３に格納されたＳＭＢヘッダは、自動的に挿入される。

【００９７】

ＳＭＢ送信エンティティーにＳＭＢフレームをＳＭＢエンティティーに移すことを要求するために、ホストプロセッサ１３０１は、関連した送信ＳＭＢ送信イベントレジスタにそれらを書き込むことによって、バッファロケータとバッファオフセット対をＳＭＢ送信イベント待ち行列１３２５に書き込む。

【００９８】

バッファロケータは、ＳＭＢフレームのためのデータを含むバッファへのポインタを含む。バッファオフセットは、バッファと長さフィールド内のデータの始まりに対するオフセットを含む。バッファロケータは、また、バッファロケータ／バッファオフセット対がさらにこのＳＭＢフレームのためのより多くのデータへのポインタを含んで書き込まれるか否かを示す最後のビットを含む。

【００９９】

もし、ＳＭＢフレームがバッファメモリ１２２ｃ内のもう一つのＳＭＢ送信バッファからのデータを含むべきであるならば、ホストプロセッサは、ＳＭＢ送信イベント待ち行列１３２５にこのＳＭＢ送信バッファを記述するもう一つのバッファロケータ／バッファオフセット対を書き込まなければならない。もし、ＳＭＢフレームに含まれるべきデータが一つ以上のＳＭＢ送信バッファにまたがるならば、ＳＭＢ送信エンティティーは、データをともにリンクするために、関連した送信バッファ記述子内のバッファポインタを用い得る。もし、余分なデータがプロトーＳＣＳＩ受信フレームからのものであるならば、これらの記述子は、プロトーＳＣＳＩ受信エンティティーによって以前に満たされたであろう。

【０１００】

送信バッファメモリ１２２ｃのＳＭＢ送信バッファからのデータが一つ以上のＳＭＢフレームのために用いられ得るので、それらが用いられた後にＳＭＢ送信バッファを自由化することは、単純な処理ではない。ＳＭＢ送信に含まれない受信されたプロトーＳＣＳＩフレームのセクションを含むＳＭＢ送信バッファは、関連した送信戻りフリーバッファレジスタを介して、送信制御メモリに含まれる送信フリーバッファ待ち行列にそれらを書き戻すことによって自由にされ得る。ＳＭＢフレームに含まれるべきデータを含むＳＭＢ送信バッファは、その中のデータが送信されるまでそれらが自由にされ得ないので、同一の方法で自由にされ得ない。

【０１０１】

それで、プロトーＳＣＳＩデータを含む種々のＳＭＢフレームへのバッファロケータ／バッファオフセット対がＳＭＢ送信イベント待ち行列１３２５に書き込まれた後、オリジナルのＳＭＢ送信バッファへのポインタは、また、ＳＭＢ送信イベント待ち行列１３２５に書き込まれる。これらのポインタは、送信フリーバッファ待ち行列に戻って自由化されるべきことを示すことを記録する。ＳＭＢ送信イベント待ち行列１３２５が次々に処理されるので、ＳＭＢ送信バッファは、その中のあらゆるデータが送信された後にのみ自由にされる。

【０１０２】

図１４は、多数のサービス要求を多数のスレッドとして扱うための、ソフトウェアで実行された典型的な先行技術のアプローチを表すフローチャートである。伝統的な多数スレッドの構成では、典型的に各クライアントにサービスを提供するための少なくとも一つのスレッドがある。スレッドは、クライアントがサーバーに接続詞、それから分離するように、始められ、終了させられる。各クライアントは、サービス要求を処理するためのサーバー上のスレッドとディスク要求を処理するためのスレッドとを有してもよい。サービス処理１４００は、ボックス１４０１でクライアント接続要求の存在をテストする繰り返し

ループを含み、もし、テストが肯定的であるならば、処理は、ボックス１４０２、クライアント処理１４３０を始める。クライアント処理１４３０がボックス１４３５でディスクアクセスを要求するとき、それは最初にディスクへアクセスするための適切なディスク処理を要求し、それから、ディスクアクセスが完了するまでボックス１４３６で待機する。ディスク処理１４０２は、それがサービス要求を発するクライアントにボックス１４３７で応答を送信することを可能にするために、クライアント処理１４３０を目覚めさせる。それで、ディスクアクセスを要求する各クライアント要求のための少なくとも二つの処理スイッチがある。これらの多数のスレッド処理をハードウェアで実行することは、問題である。なぜならば、通常、それらはマルチタスク処理オペレーティングシステムによって扱われるからである。

【０１０３】

図１５は、図２のサービスサブシステムと、例えば、図１２及び１３の実施の形態とを関連して使用するために、多数のサービス要求の処理を示すフローチャートである。単純なスレッドアーキテクチャでは、一つのサービス処理１５００が単一のスレッドで多数のクライアントからの要求を扱い、一つのディスク処理１５０２がサービス処理１５００からのすべての要求を扱う。要求をなす各クライアントのために別々の処理を用いる先行技術のアプローチは無視され、その機能は、一つのサービス処理１５００によってここで扱われた。それに加えて、これら２つの処理、サービス処理とディスク処理は、示されるように、同一のスレッド内に含まれてもよく、あるいは、ロードバランシングを容易にするために２つの別々のスレッド間に分担されてもよい。

【０１０４】

図１５の単一スレッドサービス処理は、同時に傑出した多数のクライアントからのディスク要求を有することができる。単一スレッドは、２つのテストを持つメインループを含む。ボックス１５０１の第１のテストは、クライアントからの要求を受信したか否かである。ボックス１５０８の第２のテストは、前に始められたディスクアクセス要求が完了したか否かである。その結果、ディスクアクセスが完了させられるべきボックス１５０８で決定されたので、ボックス１５０７でサービス処理は、クライアントに適切な応答を送り返す。サービス処理１５００がボックス１５０１を介してディスクアクセス要求を扱い、要求をボックス１５０２で処理させ、ボックス１５０４でディスクアクセスの開始を起こすとすぐに、サービス処理は、前のディスクアクセスが完了するまで停止し及び待つことなく、ボックス１５０１を介してもう一つのクライアントからのもう一つの要求を扱うために自由である。ディスクアクセスが完了されたボックス１５０８での決定において、ボックス１５０７のディスク処理は、結果のサービス処理を通知し、クライアントに応答を送信する。従って、サービス及びディスク処理は、クライアントから送られる要求がある限り、絶えず走っている。

【０１０５】

図１６は、ファイル記憶装置を有するコンピュータシステムに関連して、図３に示されるようなファイルシステムモジュールの使用を示すブロック図である。（図１６のものに類似する実行は、ファイル記憶装置を有するコンピュータシステムに関連して、図３に示されるような記憶装置モジュールを提供するために用いられてもよい。）この実施の形態では、ファイルシステムモジュール１６０１は、マイクロプロセッサ１６０５、メモリ１６０６、及び、ここでは従来のディスクドライブ制御部であるディスクサブシステム１６０２を介してアクセスされるディスクドライブ１６１０と同様に、ビデオ１６０９のような周辺機器を含むコンピュータシステムに統合化される。ファイルシステムモジュール１６０１は、また、ＰＣＩバス１６０７上のＰＣＩブリッジ１６０４を介してコンピュータマルチプロセッサ１６０５とコンピュータメモリ１６０６に接続される。ＰＣＩバス１６０７は、また、マイクロプロセッサ１６０５をコンピュータ周辺機器１６０９に接続する。ファイルシステムモジュールの受信エンジン１６１０は、図１０、１１、１２Ｂ、及び１３に関連して上述された方法と類似の方法で、マイクロプロセッサ１６０５からのディスクアクセス要求を処理する。同じく、送信エンジン１６１１は、図１０、１１、１２Ｂ

、及び１３に関連して上述された方法と類似の方法で、そのようなディスクアクセス要求に対する応答を供給する。

【０１０６】

図１７Ａは、図３の記憶装置モジュールにおけるデータフローのブロック図である。図１７Ａと１７Ｂにおけるカリフォルニア州パロアルトのヒューレットパッカード社（Hewlett Packard Co.,Palo Alto, California）から利用可能なタキオンＸＬ光ファイバチャネル制御部が１／Ｏ装置として用いられ得、本発明の実施の形態が他の１／Ｏ装置を等しく用い得ることに気付かれたい。プロトーＳＣＳＩ要求は、プロトーＳＣＳＩ要求プロセッサ１７０２によってプロトーＳＣＳＩ入力を越えて受信される。この要求に関する情報はSEST情報テーブルに格納され、もし、これがWRITE要求ならば、同じくプロトーＳＣＳＩ入力１７００を介して供給されるWRITEデータは、WRITEバッファメモリ１７３６に格納される。

【０１０７】

交換要求（ＥＲＱ）発生器１７１６は、WRITEバッファメモリ１７３６からの情報を取得する。もし、書き込まれるべきすべてのバッファが現在蓄えられ、あるいは書き込まれるべきデータが書き込まれるべきバッファを完全に満たすならば、WRITEは、すぐに実行され得る。書き込まれるべきデータは、キャッシュメモリ１７４０の適切な領域にWRITEバッファメモリ１７３６からコピーされる。ファイバチャネルＩ／Ｏ制御部１７２０は、その制御部１７２０と通信するディスク記憶装置の適切な領域にそのデータを書き込むように構成される。さもなければ、ディスクからのREADは、適切なディスクから要求されるデータを得るために、WIRTEの前になされなくてはならない。

【０１０８】

プロトーＳＣＳＩ承認発生器１７３０は、プロトーＳＣＳＩ要求を生成する責任を負う。プロトーＳＣＳＩ応答を生成することができる３つの情報源、プロセッサ１７３８、ファイバチャネルＩ／Ｏ制御部１７２０、及びキャッシュメモリ１７４０があり、そのそれぞれは、SESTインデックスを供給する。すべての移動のために、ステータス情報とともに、プロトーＳＣＳＩ要求が承認に関連付けられることを可能にする同定は、プロトーＳＣＳＩ承認インターフェース１７３４に戻される。

【０１０９】

図１７Ｂは、図３の記憶装置モジュールの制御フローを示す詳細なブロック図である。プロトーＳＣＳＩ要求がプロトーＳＣＳＩ要求プロセッサ１７０２によってプロトーＳＣＳＩ入力１７００を越えて受信されるとき、それは、独特な識別子（SESTインデックスと呼ばれる）を割り当てられる。この要求に関する情報は、SEST情報テーブルに格納され、もし、これがWRITE要求ならば、プロトーＳＣＳＩ入力１７００に同じく供給されるWRITEデータは、WRITEバッファメモリ１７３６に格納される。SESTインデックスは、それからプロトーＳＣＳＩ要求待ち行列１７０４に書き込まれる。

【０１１０】

キャッシュ制御部１７０６は、プロトーＳＣＳI要求待ち行列１７０４と使用されたバッファ待ち行列１７０８からのエントリを取得する。エントリがプロトーＳＣＳＩ要求待ち行列１７０４から取得されるとき、このSESTインデックスに関する情報は、SEST情報テーブルから読み込まれる。キャッシュ制御部１７０６は、どのディスクブロックがこの転送のために要求されるかを計算し、ディスクブロック番号とアクセスされるべきディスク装置のハッシュルックアップを用いるキャッシュバッファ位置にこれを翻訳する。もし、この転送のために要求される書込みバッファメモリ１７３６のいずれのバッファも現在他の転送によって使用されているならば、SESTインデックスは、他の転送の完了を持つ顕著な要求待ち行列１７１０に入れられる。さもなければ、もし、これがREAD転送であり、要求されたバッファのすべてがキャッシュにあるならば、SESTインデックスは、蓄えられたREAD待ち行列１７１２に入れられる。さもなければ、SESTインデックスは、記憶装置要求待ち行列１７１４に書き込まれる。このアルゴリズムに対する可能な高度化は、バッファが現在蓄えられているとすれば、同一のバッファの多数のＲＥＡＤが進行中であることを

可能にするべきである。

【０１１１】

エントリが使用されたバッファ待ち行列１７０８から取得されるとき、チェックは、いずれの要求が利用可能になるためにこのバッファを待っているか否かについてなされる。これは、最も古い要求で始まる顕著な要求待ち行列を探すことによってなされる。もし、利用可能になるためにこのバッファを待っていた要求が見出されるならば、バッファは、その要求に割り当てられる。もし、その要求がこの転送のために要求されたすべてのバッファを有するならば、SESTインデックスは、記憶装置要求待ち行列１７１４に書き込まれ、この要求は、顕著な要求待ち行列１７１０から取り除かれる。さもなければ、その要求は、顕著な要求待ち行列１７１０に残されたままにされる。

【０１１２】

交換要求（ＥＲＱ）発生器１７１６は、記憶装置要求待ち行列１７１４と部分的なWRITE待ち行列１７１８からのエントリを取得する。SESTインデックスがいずれかの待ち行列から読み出されるとき、このSESTインデックスに関する情報は、SEST情報テーブルから読み出される。もし、それがREAD転送ならば、ファイバチャネル１／Ｏ制御部１７２０は、適切なディスクからのデータを読み込むように構成される。もし、それがWRITE転送であり、書き込まれるべきすべてのバッファが現在蓄えられているか、あるいは、書き込まれるべきデータが書き込まれるべきバッファを完全に満たしているならば、WRITEは、すぐに実行され得る。書き込まれるべきデータは、WRITEバッファメモリ１７３６からキャッシュバッファの適切な領域にコピーされる。ファイバチャネルＩ／Ｏ制御部１７２０は、適切なディスクにデータを書き込むように構成される。さもなければ、図１７Ａに関連して上述したように、我々は、適切なディスクからの要求されたデータのREADを始めるためにWRITEをする前に、ディスクからのREADをする必要がある。

【０１１３】

ＩＭＱプロセッサ１７２２は、入ってくるメッセージ待ち行列１７２４からのメッセージを取得する。これは、ファイバチャネルＩ／Ｏ制御部１７２０が完了した変換、または問題に遭遇した変換の待ち行列である。もし、ファイバチャネル変換に問題があったならば、ＩＭＱプロセッサ１７２２は、適切なエラー回復を可能にするために、プロセッサメッセージ待ち行列１７２６を介してプロセッサにメッセージを送る。もし、変換が許容できたならば、SEST情報は、このSESTインデックスのために読み出される。もし、この変換がWRITE変換の始まりにおけるREAD変換であったならば、SESTインデックスは、部分的なWRITE待ち行列１７１８に書き込まれる。さもなければ、それは、記憶装置承認待ち行列１７２８に書き込まれる。

【０１１４】

図１７Ａに関連して上述されるように、プロトーＳＣＳＩ承認発生器１７３０は、プロトーＳＣＳＩ要求を生成する責任を負う。再び、プロトーＳＣＳＩ応答を生成することができ、それぞれがSESTインデックスを供給する３つの可能な情報源がある。

【０１１５】

プロセッサ承認待ち行列１７３２は、エラーを発生し、エラーが解決されるとすぐにハードウェアに戻ってプロセッサ１７３８によって解決されなければならない要求を送るために、プロセッサ１７３８によって用いられる。記憶装置承認待ち行列１７２８は、通常完了したファイバチャネル要求に送り戻すために用いられる。

【０１１６】

これらの待ち行列のいずれかにエントリがあるとき、SESTインデックスは、読み出される。このインデックスのためのSEST情報はそれから読み込まれる。転送のために、ステータス情報に沿って、プロトーＳＣＳＩ要求が承認を関連付けるのを可能にする同定は、プロトーＳＣＳＩ承認インターフェース１７３４の向こう側に戻される。READのために、読まれたデータは、また、プロトーＳＣＳＩ承認インターフェース１７３４の向こう側へ戻される。

【０１１７】

プロトーＳＣＳＩ転送が完了されるとすぐに、この転送に関連したすべてのバッファのアドレスは、使用されたバッファ待ち行列１７０８に書き込まれる。この転送に用いられるあらゆるWRITEバッファメモリは、また、フリーWRITEバッファメモリのプールに戻される。

【０１１８】

図１８は、ファイル記憶装置を有するコンピュータシステムに関連して、図３に示されるような記憶装置モジュールの使用を示すブロック図である。ここで、記憶装置モジュール１８０１は、マイクロプロセッサ１８０２、メモリ１８０３、ビデオシステム１８０５のような周辺機器、並びに、記憶装置１８０９、１８１０、及び１８１１を含むコンピュータシステムのためのファイバチャネルホストバスアダプタ及びドライバとして行動する。記憶装置モジュール１８０１は、ＰＣＩバス１８０７上でＰＣＩブリッジ１８０４を介してマイクロプロセッサ１８０２とコンピュータメモリ１８０３に接続される。記憶装置モジュール１８０１は、ＰＣＩバスから要求を受信し、図１７Ａ及び１７Ｂに関連して上述された方法でその要求を処理する。記憶装置モジュール１８０１は、記憶装置アクセスインターフェース１８０８を介して記憶装置１８０９、１８１０、及び１８１１にアクセスする。

【０１１９】

図１９は、本発明の実施の形態、特に、複数のネットワークサブシステムとサービスサブシステムが連続するサブシステム及び／又はモジュールのポート間で通信を確立するために拡張スイッチを利用して使用される一実地の形態の拡大縮小可能性を示すブロック図である。余分なネットワーク接続がユニットの帯域幅能力を増加し、より多くの記憶要素をサポートすることを可能にするために、この実施の形態では、拡張スイッチ１９０１、１９０２、１９０３は、多くのモジュールとともに相互接続するために用いられる。拡張スイッチは、拡張スイッチの一面におけるモジュールから他面におけるあらゆるモジュールまでのあらゆる接続の経路を定める。拡張スイッチは、非ブロック化であり、多数の入力を取得し、特定の接続のために最良なルートを決定する情報処理能力を持つ拡張スイッチ制御モジュールによって制御されてもよい。

【０１２０】

図１９の実施の形態では、示された全システムは、ネットワークサブシステム１９０４と類似のサブシステム１９０８及び１９１２を含むコラム１９２１に示される複数のネットワークサブシステムを利用する。それは、ここでは（コラム１９２２の）ファイルアクセスモジュール、（コラム１９２３の）ファイルシステムモジュール、及び（コラム１９２４の）記憶装置モジュールの組み合わせとして実行される複数のサービスサブシステムである。モジュールの各コラム間（及びネットワークサブシステムコラムとファイルアクセスモジュールコラムの間）にスイッチ配置があり、ファイルアクセスプロトコル拡張スイッチ１９０１、記憶装置アクセス拡張スイッチ１９０２、及びプロトーＳＣＳＩプロトコル拡張スイッチ１９０３として実行される。ファイルアクセスプロトコルレベルでは、拡張スイッチ１９０１は、各ファイルアクセスモジュール１９０５の存在する家業負荷を含む基準に依存して、ネットワークサブシステム１９０４から特定のファイルアクセスモジュール１９０５に入ってくるネットワーク接続を動的に割り当てる。

【０１２１】

記憶装置アクセスプロトコルレベルでは、拡張スイッチ１９０２は、ファイルシステムモジュール１９０６の存在する作業負荷を含む基準に依存して、ファイルアクセスモジュール１９０５から特定のファイルシステムモジュール１９０６に入ってくるファイルアクセス接続を動的に割り当てる。

【０１２２】

プロトーＳＣＳＩプロトコルレベルでは、拡張スイッチ１９０３は、記憶要素の物理的位置を含む基準に依存して、特定の記憶装置モジュール１９０７へ入ってくるファイルシステム接続を動的に割り当てる。

【０１２３】

その代わりに、アイテム１９０１、１９０２、及び１９０３は、バスとして実装されてもよい。この場合、入力信号を受けるコラム内の各モジュールは、信号の重複処理を避けるためにコラム内の他のモジュールと通信し、それによって、他の信号を扱うために他のモジュールを自由にする。アイテム１９０１、１９０２、及び１９０３がバスまたはスイッチのいずれとして実現されるとしても、ファイル要求への応答が含められるとき、対応する要求からの適切なヘッダ情報が応答ヘッダの便利なフォーマット化を許すために利用可能であるように、システムを通して信号処理パスを追跡することは本発明の範囲内である。

【図面の簡単な説明】

【図１】　図１は、ファイルサーバーやウェブサーバーのようなネットワークサービスを提供するために配置された本発明の一実施の形態の概略表示である。

【図２】　図２は、図１に示された実施の形態のブロック図である。

【図３】　図３は、ファイルサーバーとして配置される実施の形態のブロック図である。

【図４】　図４は、ウェブサーバーとして配置される実施の形態のブロック図である。

【図５】　図５は、図２～４の実施の形態のネットワークサブシステムである。

【図６】　図６は、図５のネットワークサブシステムのブロック図である。

【図７】　図７は、図６のネットワークサブシステムの受信モジュールのブロック図である。

【図８】　図８は、図６のネットワークサブシステムの送信モジュールのブロック図である。

【図９】　図９は、ワークステーション又はサーバーのようなネットワークノードで使用するネットワークインターフェースアダプタとして、図５のネットワークサブシステムの使用を例証するブロック図である。

【図１０】　図１０は、図３に例証されるような一実施の形態で使用する図３のＳＭＢサービスモジュール３３とファイルシステムモジュール３４のハードウェアで実行される組み合わせのブロック図である。

【図１１】　図１１は、図３に例証されるような一実施の形態で使用する図３のＳＭＢサービスモジュール３３とファイルシステムモジュール３４のハードウェア加速の組み合わせのブロック図である。

【図１２】　図１２Ａは、それぞれ図３又は図４のアイテム３３又は３４めようなハードウェアで実行されるサービルモジュールのブロック図である。図１２Ｂは、それぞれ図３又は図４のアイテム３４又は４４のようなハードウェアで実行されるファイルモジュールのブロック図である。図１２Ｃは、組み合わされたサービスモジュールとファイルモジュールを提供する図１０のハードウェアで実行されるサービスサブシステムの詳細なブロック図である。

【図１３】　図１３は、図１１のハードウェア加速サービスサブシステムの詳細なブロック図である。

【図１４】　図１４は、多数のサービス要求を多数のスレッドとして扱うために、ソフトウェアで実行される典型的な先行技術アプローチを表すフローチャートである。

【図１５】　図１５は、図２のサービスサブシステムと、例えば、図１２及び１３の実施の形態とを関連して使用するために、多数のサービス要求の処理を示すフローチャートである

【図１６】　図１６は、ファイル記憶装置を有するコンピュータシステムに関連して、図３に示されるようなファイルシステムモジュールの使用を示すブロック図である。

【図１７】　図１７Ａは、図３の記憶装置モジュールにおけるデータフローのブロック図である。図１７Ｂは、図３の記憶装置モジュールにおける制御フローのブロック図である。

【図１８】　図１８は、ファイル記憶装置を有するコンピュータシステムに関連して、図３に示されるような記憶装置モジュールの使用を示すブロック図である。

【図１９】　図１９は、本発明の実施の形態、特に、複数のネットワークサブシステムと

サービスサブシステムが連続するサブシステム及び／又はモジュールのポート間で通信するための拡張スイッチを利用して使用される一実地の形態の拡大縮小可能性を示すブロック図である。

【図１】

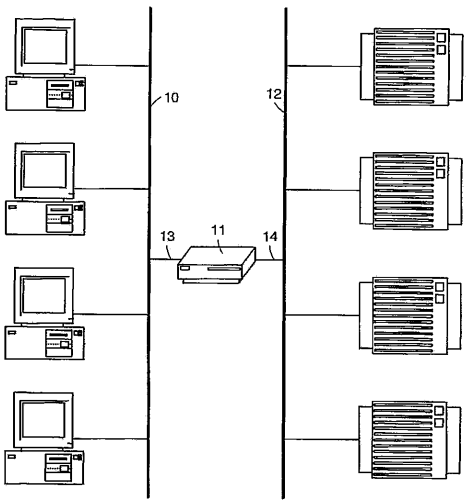


FIG. 1

【図２】

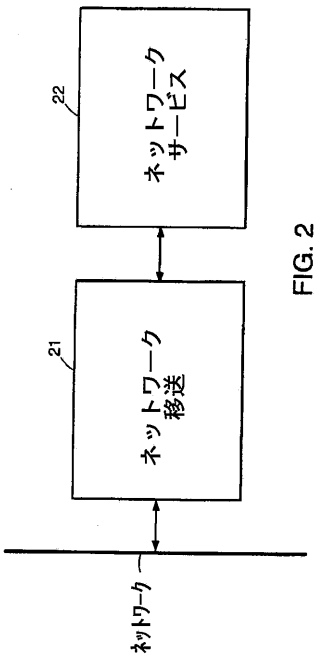


FIG. 2

【図３】

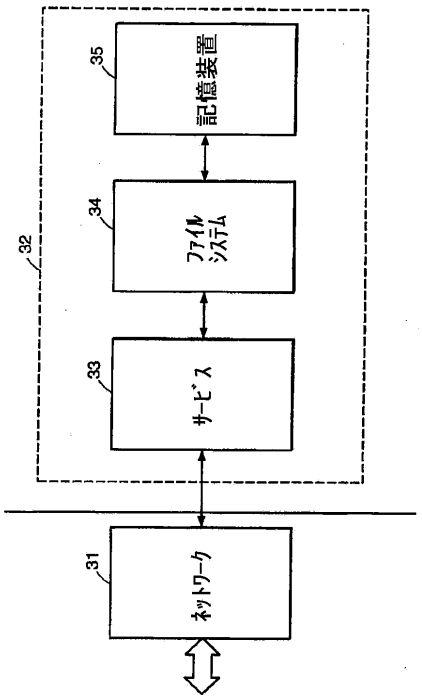


FIG. 3

【図４】

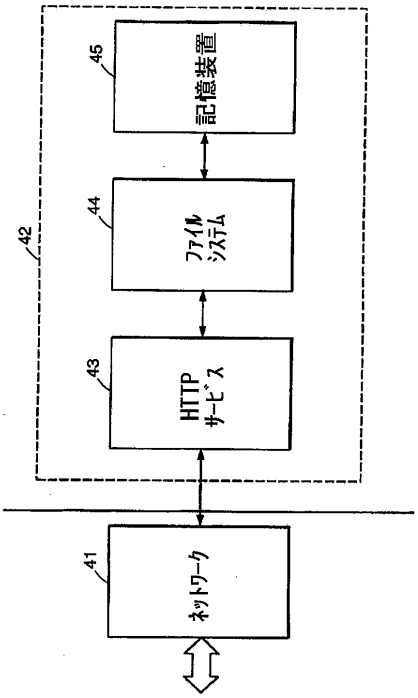


FIG. 4

【図５】

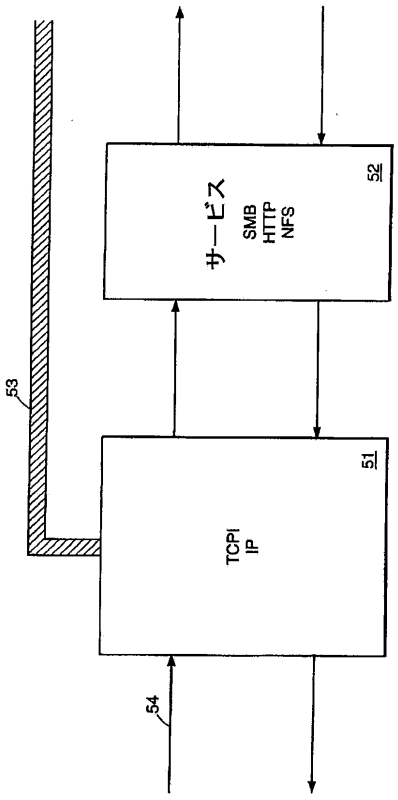


FIG. 5

【図６】

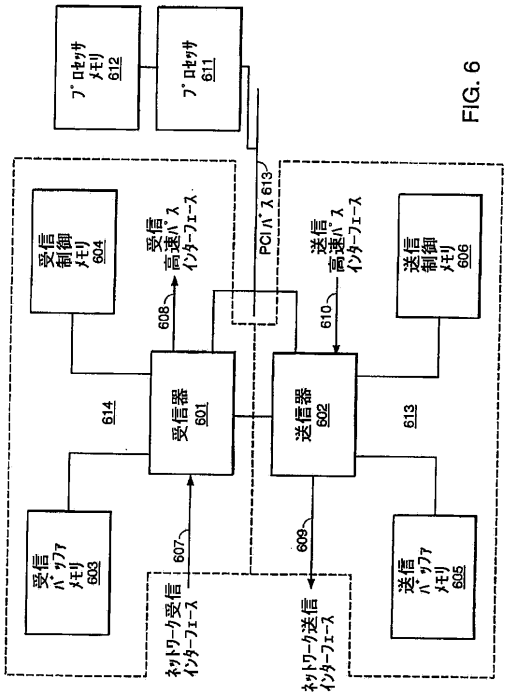


FIG. 6

【図７】

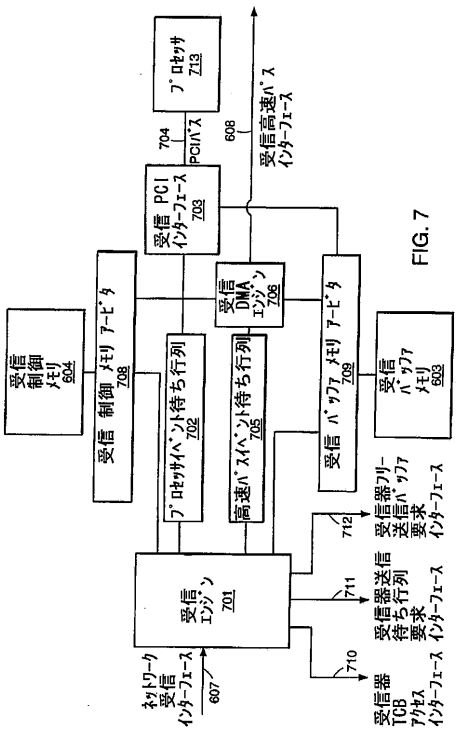


FIG. 7

【図８】

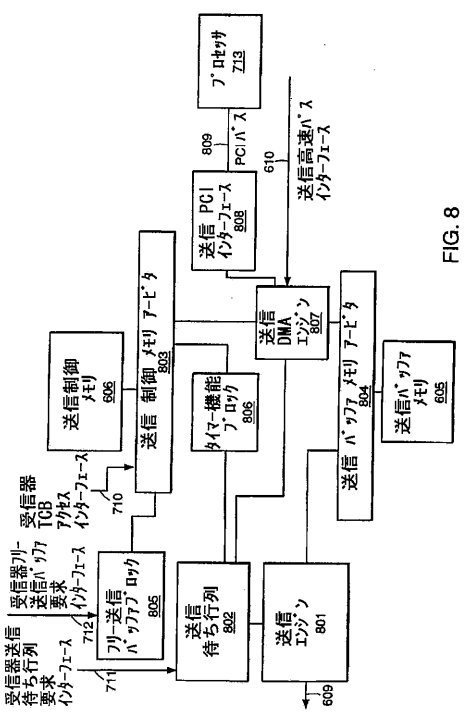


FIG. 8

【図９】

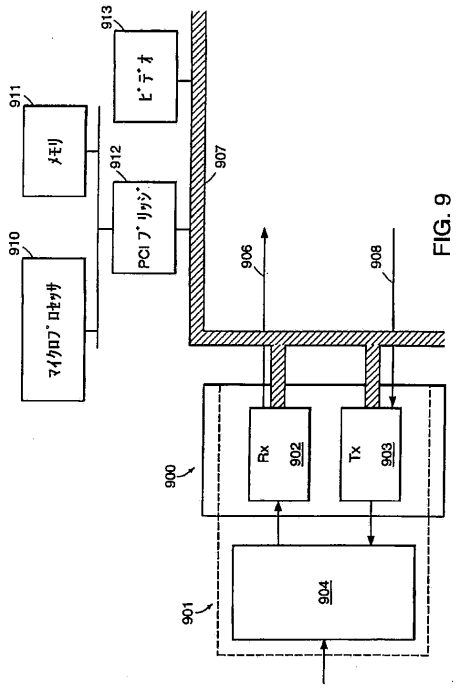


FIG. 9

【図１０】

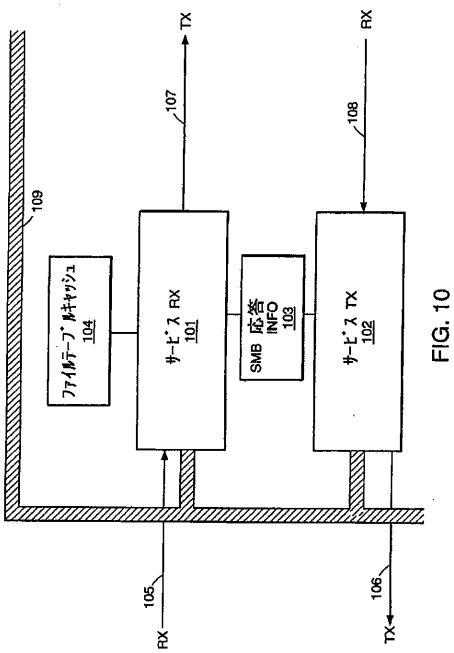


FIG. 10

【図１１】

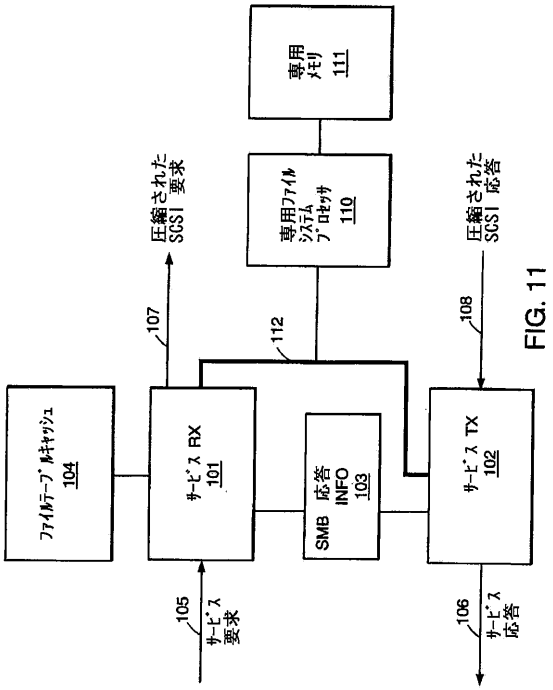


FIG. 11

【図１２Ａ－Ｂ】

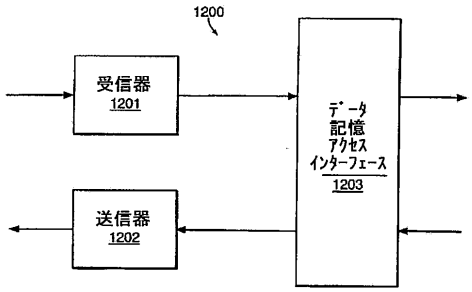


FIG. 12A

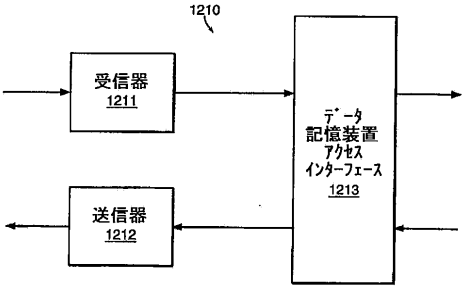


FIG. 12B

【図１２Ｃ】

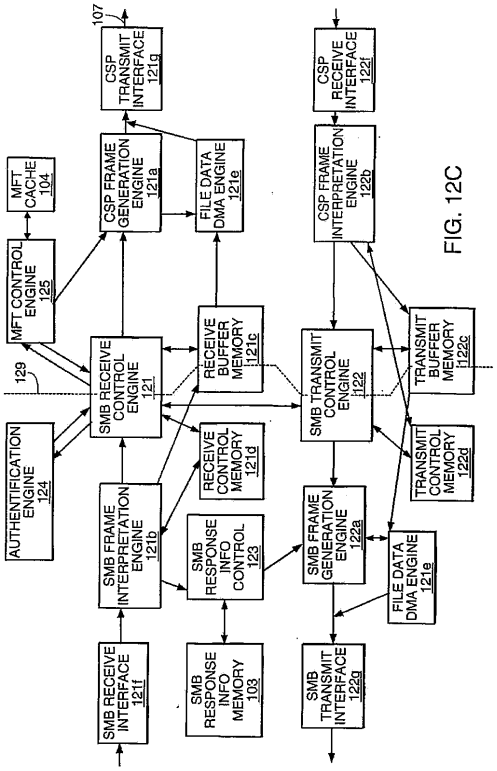


FIG. 12C

【図１３】

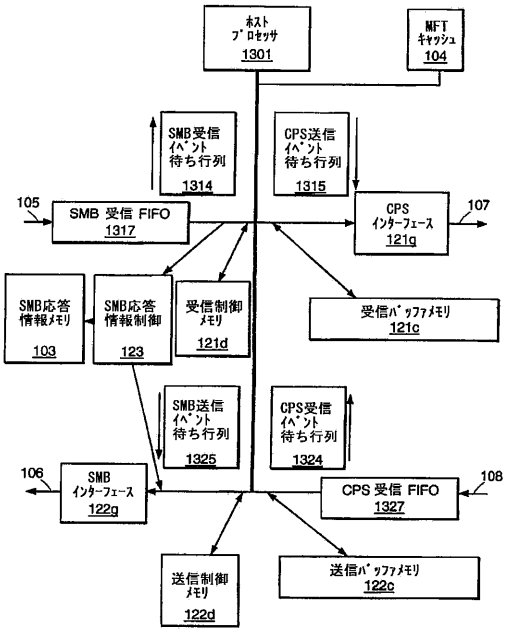


FIG. 13

【図１４】

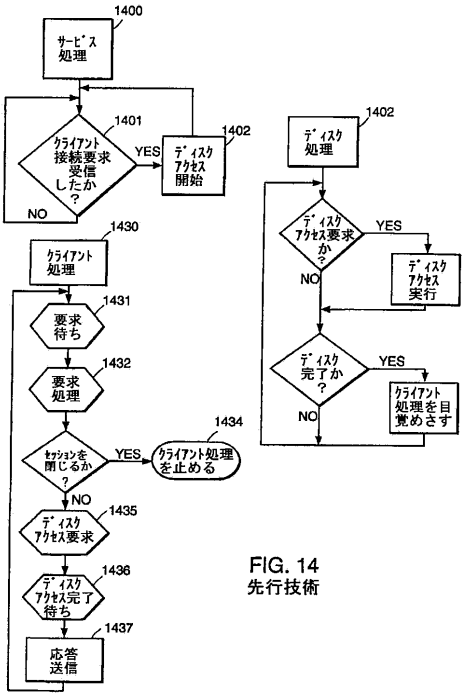


FIG. 14
先行技術

【図１５】

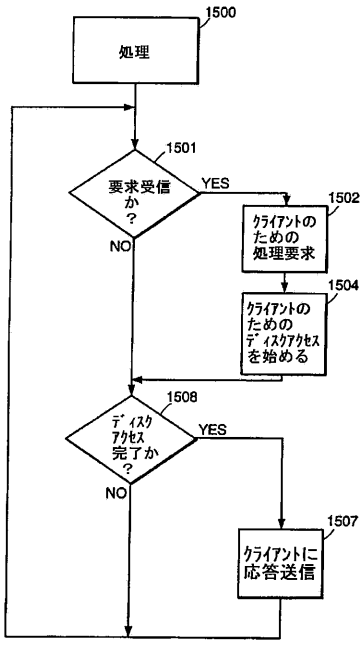
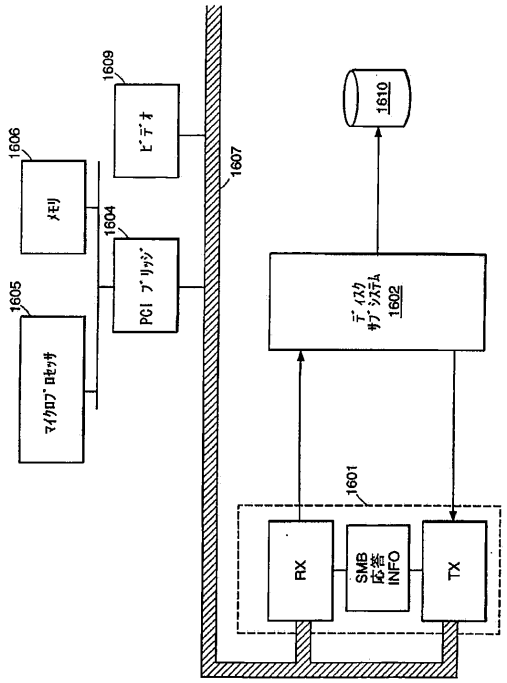


FIG. 15

【図１６】



【図１７Ａ】

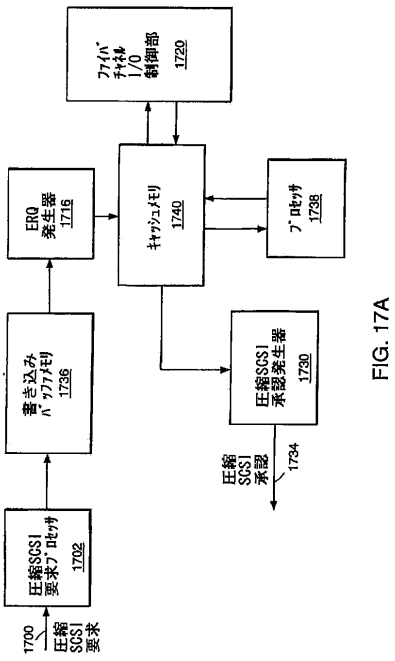


FIG. 17A

【図１７Ｂ】

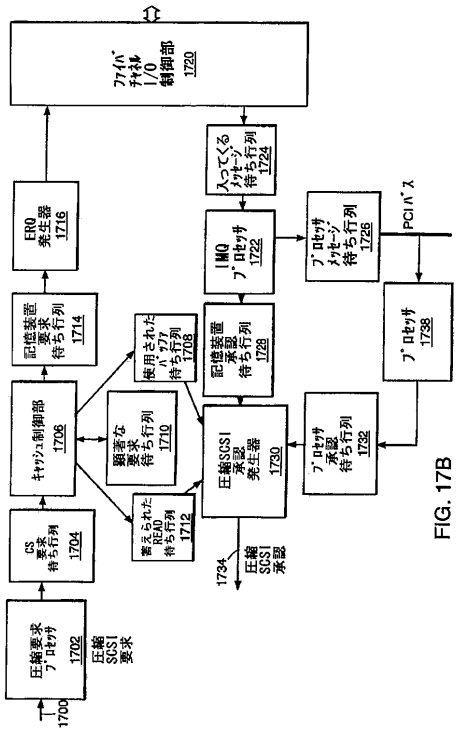


FIG. 17B

【図１８】

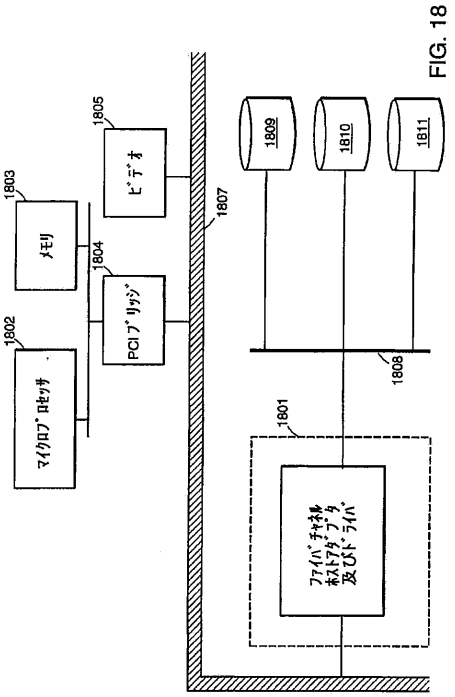


FIG. 18

【図１９】

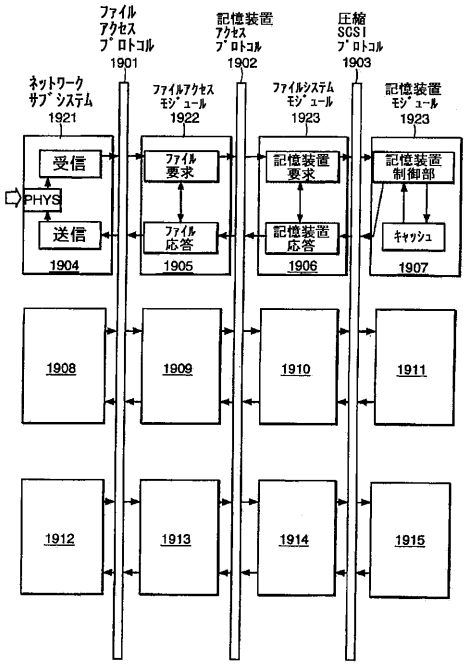


FIG. 19

フロントページの続き

(72)発明者 ウィリス、トレバー
英国、バッキンガムシャー・エイチピー２２・６エイチジー、エイリズベリー、イクフィールド・クロース ９
(72)発明者 ベンハム、サイモン
英国、バークシャー・アールジー１２・２エヌエイ、ブラックネル、ロックバイ・クロース ４
(72)発明者 クーパー、マイケル
英国、バークシャー・アールジー１・３ピーキュー、リーディング、リバプール・ロード １０２
(72)発明者 マイヤー、ジョナサン
英国、サリー・ジーユー１０・３エヌジー、ファーナム、ホワイト・ローズ・レーン ２エイ
(72)発明者 アストン、クリストファー・ジェイ
英国、バッキンガムシャー・エイチピー１３・６ジェイアール、ウィカム、ザ・クレセント ４０
(72)発明者 ウィンフィールド、ジョン
英国、バークシャー・アールジー２・０ビーエヌ、リーディング、エルガー・ロード １２６

合議体
審判長 水野 恵雄
審判官 山田 正文
審判官 稲葉 和生

(56)参考文献 特表平５－５０２５２５（ＪＰ，Ａ）
特開平３－２７３３５０（ＪＰ，Ａ）
特開平９－２６９７０（ＪＰ，Ａ）
特開平１１－２０３３０８（ＪＰ，Ａ）
特開平１０－２３２７８８（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
G06F13/00