(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5244717号
(P5244717)

(45)発行日　平成25年7月24日(2013.7.24)　(24)登録日　平成25年4月12日(2013.4.12)

(51)Int.Cl.　F I
H04L 12/713 (2013.01)　H04L 12/56　G
G06F 13/00 (2006.01)　G06F 13/00　357Z

請求項の数 4　(全 17 頁)

(21)出願番号　特願2009-157717(P2009-157717)
(22)出願日　平成21年7月2日(2009.7.2)
(65)公開番号　特開2011-15196(P2011-15196A)
(43)公開日　平成23年1月20日(2011.1.20)
審査請求日　平成24年2月17日(2012.2.17)

(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　100080001
弁理士　筒井　大和
(72)発明者　江頭　ちひろ
神奈川県秦野市堀山下１番地　株式会社日立製作所　エンタープライズサーバ事業部内
(72)発明者　小国　哲
神奈川県秦野市堀山下１番地　株式会社日立製作所　エンタープライズサーバ事業部内

最終頁に続く

(54)【発明の名称】負荷割当制御方法および負荷分散システム

(57)【特許請求の範囲】
【請求項１】
複数のクライアントからの複数の要求を、複数のサーバに割り当てる負荷分散システムにおける負荷割当制御方法であって、
前記負荷分散システムにより、前記要求による処理状況に対する前記複数のサーバのそれぞれのサービスレベルを保つための閾値を保持し、前記複数のサーバへの前記要求の割り当ての際、前記複数のサーバの優先度に従い、前記優先度の高いサーバから、前記閾値に達するまで前記要求を割り当て、前記要求に対する前記複数のサーバのそれぞれの処理状況に応じて、前記複数のサーバの電源制御を行い、前記複数のサーバへの前記要求の割り当ての際、前記複数のサーバの全ての処理状況が、前記閾値に達していた場合、前記要求を前記複数のサーバに均等に分配することを特徴とする負荷割当制御方法。
【請求項２】
複数のクライアントからの複数の要求を、複数のサーバに割り当てる負荷分散システムであって、
前記複数のサーバのそれぞれの、優先度、コネクション数、レスポンス時間、割り当て時刻、および状態を格納する情報テーブルと、
前記要求による処理状況に対する前記複数のサーバのそれぞれのサービスレベルを保つための閾値を保持する閾値保持部と、
前記複数のサーバの稼動状況を監視管理するサーバ管理部と、
前記複数のサーバへの前記要求の割り当ての際、前記複数のサーバの優先度に従い、前

記優先度の高いサーバから、前記閾値に達するまで前記要求を割り当てる転送先判定部と、

前記要求に対する前記複数のサーバのそれぞれの処理状況に応じて、前記複数のサーバの電源制御を行う電源制御部とを備え、

前記転送先判定部は、前記複数のサーバへの前記要求の割り当ての際、前記複数のサーバの全ての処理状況が、前記閾値に達していた場合、前記要求を前記複数のサーバに均等に分配することを特徴とする負荷分散システム。

【請求項３】

複数のクライアントからの複数の要求を、複数のサーバに割り当てる負荷分散システムであって、

複数の負荷分散装置と、前記複数の負荷分散装置を管理する管理サーバとを備え、

前記管理サーバは、前記複数のサーバのそれぞれの、優先度、コネクション数、レスポンス時間、割り当て時刻、および状態を格納する情報テーブルと、前記要求による処理状況に対する前記複数のサーバのそれぞれのサービスレベルを保つための閾値を保持する閾値保持部と、前記複数のサーバの稼動状況を監視管理するサーバ管理部と、前記複数のサーバへの前記要求の割り当ての際、前記複数のサーバの前記優先度に従い、前記優先度の高いサーバから、前記閾値に達するまで前記要求を割り当てる転送先判定部と、前記複数のサーバのそれぞれの前記要求に対する処理状況に応じて、前記複数のサーバの電源制御を行う電源制御部と、前記転送先判定部での前記要求の割り当てを前記複数の負荷分散装置に指示する負荷分散装置通信部とを有することを特徴とする負荷分散システム。

【請求項４】

請求項３に記載の負荷分散システムにおいて、

前記電源制御部は、前記複数のサーバの電源制御により、特定の負荷分散装置に接続されているサーバの全てを停止可能な場合、前記特定の負荷分散装置の電源を停止させることを特徴とする負荷分散システム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、情報処理システムなどにおける負荷分散システムに関し、特に、負荷集中機能をもたせ、積極的に負荷量を適度に集中させることで情報処理システム内のサーバなどの稼動台数を抑制することで省電力を実現する方法に関するものである。

【背景技術】

【０００２】

昨今、情報処理システムにおいてクライアントＰＣ（パーソナルコンピュータ）からインターネットあるいはＬＡＮ（ローカルエリアネットワーク）を介してサーバに接続し、様々な情報を入手したり提供したり、情報の検索あるいは物品の販売など様々なサービスを享受できるようになっている。このようなシステムは一般にＷｅｂサーバシステムといわれている。

【０００３】

これらのシステムにおいて、クライアント（クライアントは通常はパーソナルコンピュータであることが多い）からの大量のアクセスへの対応や耐障害性を維持することを目的として、クライアントからのアクセス要求を複数サーバに分配することで大量のアクセスにも対応でき、また、その内の一部が故障した場合にも、他のサーバに処理を振り分けることでシステム全体としてサービスを停止させない、といった処理をさせることがよく行われている。このような処理を行う装置が負荷分散装置と呼ばれる。

【０００４】

この負荷分散装置において負荷分配は負荷分散アルゴリズムを用いて、接続クライアント台数や処理量などの負荷（以下、負荷と呼ぶ）がシステムを構成するサーバの間で均等になるように負荷を分配する機能を提供している。

【０００５】

一方で、地球温暖化などの環境問題への取り組みに関心が集まる中、サーバシステムの市場においても省電力化が求められるようになってきており、省電力を実現する機能を提供する技術も盛んに開発されるようになってきている。

【０００６】

上記のような従来の負荷分散装置における負荷分配方式の場合、各サーバの性能が高くなくても、複数サーバで多数のクライアントからのアクセスを、均等に分散させて処理できるという利点がある。しかし、一般にこのようなＷｅｂサーバシステムではクライアントからのアクセスによる負荷が、例えば１日の内で時間帯によって大きく変動したり、あるいは、時期によって大きく変動したりすることが知られており、それでも一定のサービス品質（応答時間、同時アクセス数の上限確保等）を維持するために最大負荷の時に備えて多数のサーバを稼動させる必要があるが、一方で負荷量が少ない場合などには不必要に多数のサーバを稼動させるということになっており、省電力化のために台数を適正に増減させる技術が必要となってきている。

【０００７】

上記課題の解決策として、特開２００３－２８１００８号公報（特許文献１）に開示された従来技術がある。

【０００８】

特許文献１に開示されたものは、稼動時間が最短であるサーバに対して優先的にクライアントからの要求を分配するというものであり、なおかつクライアントからのデータ要求量とサーバシステムのデータ供給可能量とを比較し、サーバシステム全体として必要最低限のサーバだけを稼動させることを目的としたものである。

【０００９】

特許文献１を含む従来技術では、負荷分配の方法に従来の負荷分散方式を用いて、負荷を複数サーバに分散し、その時点でのサーバシステムの負荷量を計測しながらサーバの台数を増減させていく。システム稼動中において、負荷が存在しないサーバがある場合にはサーバを停止し、稼動中サーバシステムの処理容量が不足してきたら、新たにサーバを起動する方法をとるというものである。

【先行技術文献】

【特許文献】

【００１０】

【特許文献１】特開２００３－２８１００８号公報

【発明の概要】

【発明が解決しようとする課題】

【００１１】

しかしながら、特許文献１に開示された方法の場合、従来の負荷分散装置に比べ電力消費を抑えることが可能ではあるが、負荷分散を行う中で、もし未使用サーバが存在すれば電源を落とし、割り当てサーバの数を少なくしていき、自然と負荷集中が起こることを待っている制御（以後、静的な制御と呼ぶ）しかできていない。

【００１２】

また、この静的な制御を行う上で、ある時点で生起した負荷に対して、その時点以前に各サーバからクライアントに要求データを返信した時のサーバシステムの負荷量を用いて負荷の割当先サーバを決めるため、実際の負荷数とサーバ台数の設定にタイムラグが生じる。そのため、場合によって台数が不足、もしくは過剰になることがあり、不足する場合を恐れると常に過剰な台数のサーバを稼動させる必要がある。その結果、省電力効果については最適な制御でない状態が発生する。

【００１３】

また、クライアントからのアクセスの未来の増減は予測できない中にあって、ある時点での負荷量を見て台数を制限するといった静的な負荷集中を行う手段を提供しているが、ある特定サーバに積極的に負荷を割り当て、負荷集中の状態を自ら作り出す（以後、動的な制御と呼ぶ）ことができない。その結果、省電力化のためのサーバ稼動台数の増減を最

適制御できていない。
【００１４】
また、従来技術において、単純なる動的な負荷集中を行うと、１台のサーバシステムが提供するサービスレベル品質を一定以上に保持することができない。
【００１５】
そこで、本発明の目的は、動的な制御を行い、かつユーザが求めるサービスレベルの品質を守った上で最大の省電力効果を得ることができる負荷分散システムを提供することにある。
【００１６】
本発明の前記ならびにその他の目的と新規な特徴は、本明細書の記述および添付図面から明らかになるであろう。
【課題を解決するための手段】
【００１７】
本願において開示される発明のうち、代表的なものの概要を簡単に説明すれば、次の通りである。
【００１８】
すなわち、代表的なものの概要は、負荷分散システムにより、クライアントからの要求に対する複数のサーバのそれぞれのサービスレベルを保つための閾値を保持し、複数のサーバへの要求の割り当ての際、複数のサーバの優先度に従い、優先度の高いサーバから、閾値に達するまで要求を割り当て、複数のサーバのそれぞれの要求に対する処理状況に応じて、複数のサーバの電源制御を行うものである。
【発明の効果】
【００１９】
本願において開示される発明のうち、代表的なものによって得られる効果を簡単に説明すれば以下の通りである。
【００２０】
すなわち、代表的なものによって得られる効果は、動的な制御を行い、かつユーザが求めるサービスレベルの品質を守った上で最大の省電力効果を得ることができる。
【図面の簡単な説明】
【００２１】
【図１】本発明の実施の形態１に係る負荷分散システムが適用されるネットワークシステムの構成を示す構成図である。
【図２】本発明の実施の形態１に係る負荷分散システムの構成を示す構成図である。
【図３】本発明の実施の形態１に係る負荷分散システムの情報テーブルおよび割り当てテーブルの一例を示す図である。
【図４】本発明の実施の形態１に係る負荷分散システムのアルゴリズムの一例を示す図である。
【図５】本発明の実施の形態１に係る負荷分散システムの負荷集中モードの基本的な概念を説明するための説明図である。
【図６】本発明の実施の形態１に係る負荷分散システムの転送先判定部における負荷集中モードの処理を示すフローチャートである。
【図７】本発明の実施の形態１に係る負荷分散システムの転送先判定部における負荷集中モード（ラウンドロビンモード）の処理を示すフローチャートである。
【図８】本発明の実施の形態１に係る負荷分散システムの負荷集中モード（ラウンドロビンモード）の様子を説明するための情報テーブルと割り当てテーブルの一例を示す図である。
【図９】本発明の実施の形態１に係る負荷分散システムの転送先判定部における負荷集中モード（ラウンドロビンモード）の処理を示すフローチャートである。
【図１０】本発明の実施の形態１に係る負荷分散システムの負荷集中モード（ラウンドロビンモード）の情報テーブルの一例を示す図である。

【図１１】本発明の実施の形態１に係る負荷分散システムによる効果を説明するための説明図である。
【図１２】本発明の実施の形態２に係る負荷分散システムの構成を示す構成図である。
【発明を実施するための形態】
【００２２】
以下、本発明の実施の形態を図面に基づいて詳細に説明する。なお、実施の形態を説明するための全図において、同一の部材には原則として同一の符号を付し、その繰り返しの説明は省略する。
【００２３】
（実施の形態１）
図１により、本発明の実施の形態１に係る負荷分散システムが適用されるネットワークシステムの構成について説明する。図１は本発明の実施の形態１に係る負荷分散システムが適用されるネットワークシステムの構成を示す構成図である。
【００２４】
図１において、ネットワークシステムは、複数のサーバから構成されるサーバシステム（１３０－１～－４）が、負荷分散システム１２０を介して外部ネットワーク１１０に接続されている。
【００２５】
また、この外部ネットワーク１１０に接続されているクライアント１００とのデータパケットの送受信には必ず負荷分散システム１２０が仲介している。
【００２６】
なお、図１に示す例では、クライアント台数と、サーバ台数を４としているが、特に４台には限定されないし、異なる台数であっても問題ない。
【００２７】
次に、図２～図４により、本発明の実施の形態１に係る負荷分散システムの構成ついて説明する。図２は本発明の実施の形態１に係る負荷分散システムの構成を示す構成図、図３は本発明の実施の形態１に係る負荷分散システムの情報テーブルおよび割り当てテーブルの一例を示す図、図４は本発明の実施の形態１に係る負荷分散システムのアルゴリズムの一例を示す図である。
【００２８】
図２において、負荷分散システム１２０は、処理要求受付部１２１と、サーバシステムの管理を行うための手段としてサーバ管理部１２２と、要求転送部１２３と、電源制御部１２４と、転送先判定部１２５とサーバシステム全てのサーバの稼動状況や優先度などを保持する情報テーブル１２６と負荷分散システムが負荷を割り当て可能なサーバのアドレスを保持する割り当てテーブル１２７とサービスレベルアグリーメントを守るための閾値（以後アセスメントレベルと呼ぶ）を保持する閾値保持部１２８とを備えている。
【００２９】
情報テーブル１２６は何らかのイベントが発生すると情報が読み出され、また、サーバの状態が変更されるもしくはされた場合にはその都度、情報テーブル１２６の内容も変更される。
【００３０】
図３（ａ）に情報テーブル１２６の様子、図３（ｂ）に割り当てテーブル１２７の様子を示す。
【００３１】
アセスメントレベルは任意に設定可能で、設定内容としては接続コネクション数やＣＰＵビジー率、ＩＯビジー率、レスポンス時間などがあげられる。
【００３２】
処理要求受付部１２１は、クライアント１００からのデータ要求パケットを受け付ける。受け付けた要求パケットは転送先判定部１２５へ送られ、転送先判定部１２５において転送先サーバを選択し、要求転送部１２３へどのサーバへ要求を送ればよいかを指示する

。
【００３３】
そして、要求転送部１２３は転送先判定部１２５より指示されたサーバの割り当てテーブル１２７のアドレスに要求を無事転送した後、情報テーブル１２６を更新する。
【００３４】
転送先サーバの選択方法としては、負荷分散アルゴリズムを用いて従来の負荷分散方式により転送先サーバを選択する負荷分散モードと、負荷分散アルゴリズムを負荷集中型に変換することで特定のサーバに負荷を集中させる負荷集中モードがあり、某かのアルゴリズムによって制御可能である。
【００３５】
図４に、既存の負荷分散装置の負荷分散アルゴリズムを参考に一般化したアルゴリズムの一覧を示す。
【００３６】
サーバ管理部１２２は、情報テーブル１２６と割り当てテーブル１２７とサーバシステム（１３０－１～４）を監視管理する。
【００３７】
例えば、所定期間内にサーバシステム（１３０－１～４）の各稼動中サーバの処理量を調査し、未使用サーバが存在し、さらに電源停止条件を満たしていると判断したならば、電源制御部に対象サーバの電源停止を指示し、情報テーブル１２６の内容を“停止中”と更新する。
【００３８】
逆に、上記所定期間内にサーバシステム（１３０－１～４）の全ての稼動中サーバの負荷量が供給可能量を超えており、電源起動条件を満たしていると判断したならば、サーバシステム（１３０－１～４）の内、必要数の待機中サーバを優先度が高い順に起動するよう電源制御部に指示する。
【００３９】
本実施の形態による負荷集中の場合、優先度の高いサーバから順に負荷を片寄せしていくため、次に電源を停止させるサーバは、稼動中サーバの中で最も優先度の低いサーバとなるため、明示的にこのサーバの負荷量を見て、未使用の状態で、さらに電源停止条件を満たしていると判断したならば、対象サーバを割り当て可能対象から外し、電源制御部に対象サーバの電源停止を指示した後、対象サーバの電源停止を確認した上で情報テーブル１２６の内容を“停止中”と更新する、という方法も可能である。
【００４０】
そして、情報テーブル１２６が持つ対象サーバの状態を“起動処理中”と更新する。起動処理中サーバの正常な起動を確認したら、割り当てテーブル１２７に対象サーバを追加し割り当て可能となった時点で、情報テーブルの内容を“稼動中”と更新する。
【００４１】
また、サーバ管理部１２２は、サーバシステム（１３０－１～４）の待機中サーバの優先度の変更を行うなどして、電源の起動と停止の動作が特定のサーバに集中することを防ぐことも可能とする。
【００４２】
次に、図５～図１１により、本発明の実施の形態１に係る負荷分散システムの転送先判定部の負荷集中モードの処理について説明する。図５は本発明の実施の形態１に係る負荷分散システムの負荷集中モードの基本的な概念を説明するための説明図、図６は本発明の実施の形態１に係る負荷分散システムの転送先判定部における負荷集中モードの処理を示すフローチャート、図７および図９は本発明の実施の形態１に係る負荷分散システムの転送先判定部における負荷集中モード（ラウンドロビンモード）の処理を示すフローチャート、図８は本発明の実施の形態１に係る負荷分散システムの負荷集中モード（ラウンドロビンモード）の様子を説明するための情報テーブルと割り当てテーブルの一例を示す図、図１０は本発明の実施の形態１に係る負荷分散システムの負荷集中モード（ラウンドロビ

ンモード）の情報テーブルの一例を示す図、図１１は本発明の実施の形態１に係る負荷分散システムによる効果を説明するための説明図である。

【００４３】

まず、負荷集中モードにおいて、図４に示した負荷分散アルゴリズム各々を負荷集中型に変換する方法について、負荷集中モードの基本的な概念を図５を用いて説明した後に、図６～図９を用いて説明する。

【００４４】

本来の負荷分散と本実施の形態の負荷集中との相違点を次に示す。

【００４５】

本来の負荷分散は、各サーバに負荷が均等に割り振られる。

【００４６】

例えば、サーバが３台あり、各々＃０、＃１、＃２とし、何らかのサーバで処理する負荷が６個順にもたらされるとすると、これらの負荷が全て等しいとするならば、３台のサーバに順に負荷１～３を割り振り、さらに負荷４～６を割り振ることで均等にすることができる。負荷１～６が全てサーバに割り振られ、処理中である様子を図５（ａ）に示す。

【００４７】

これを本実施の形態の負荷集中モードでは、サーバに優先度を設け、優先度の高いサーバに集中的に負荷を割り振る。

【００４８】

その際、負荷が特定のサーバに集中すると、そのレスポンス時間が長くなるなどしてサービスの質が落ちるため、サービスレベルアグリーメントで規定するサービスを維持できるアセスメントレベルを設定し、サーバの処理中の負荷量がこのアセスメントレベルに達するまではある特定のサーバに負荷を割り振り、アセスメントレベルに達した後は、次に優先度の高いサーバに負荷の割り振りを行う。

【００４９】

例えば、アセスメントレベルを１台あたり４個までの負荷が許容範囲であることとし、＃０、＃１、＃２の順に優先度が高いサーバが３台あるとする。何らかのサーバで処理する負荷が６個順にもたらされるとすると、サーバ＃０には負荷１～４が、サーバ＃１には負荷５～６が割り振られ、サーバ＃２には何も割り振られていない状態となる（図５（ｂ））。

【００５０】

この状態であればサーバ＃２は電源停止をただちに行うことができる。

【００５１】

しかし、負荷がさらに増大していき、サーバシステムがアセスメントレベルを守れなくなった場合（これは本来、システム設計上あってはならないことではあるが、予め想定しておく必要はある）、既にこの段階でシステム設計上想定していた負荷量を超えており、できるだけ負荷を分散させた方が賢明である。

【００５２】

例えば、先ほどの例に続けて負荷がかかった場合、やがて負荷量がアセスメントレベルを超えることとなる。ここから先は、負荷の増大に対してアセスメントレベルを守ることができないので、負荷分散モードに変更させることが必要となる。この状況を図５（ｃ）に示す。すなわち、図５（ｃ）の負荷１～１２まではサーバ＃０～２に４個づつ割り振られているが、その後の負荷１３～１５はサーバ＃０～２に均等に１個づつを順に割り振っていくことになる。

【００５３】

そして、その後負荷が減ってきてアセスメントレベルを守れるようになれば、また本来の負荷集中モードに戻るものとする。

【００５４】

この場合のアセスメントレベルは、コネクション数やレスポンス時間などユーザが望むもので任意に設定可能である。

【００５５】
この負荷集中モードの基本的概念を基に、負荷分散アルゴリズムでの負荷集中の方法を最小コネクションモード、最速モード、ラウンドロビンモードを例にあげて説明する。
【００５６】
まず、最小コネクションモードの場合については、上記負荷集中モードの基本的概念の負荷量をコネクション数と置き換えて考える。
【００５７】
最速モードの場合には、各サーバの最新のレスポンス時間（あるいは最新の複数個のリクエストに対するレスポンス時間の平均）を見て、特定のサーバについてレスポンス時間がアセスメントレベル以下である内は負荷を割り振り、超えたら別のサーバへ割り振ると考える。
【００５８】
最速モードと最小コネクションモードの違いとしては、アセスメントレベルがコネクション数なのか、レスポンス時間なのかという違いである。
【００５９】
この２つのモードについて図６に示すフローチャートを用いて説明する。
【００６０】
前提条件として、新たにクライアントから接続要求があった場合、要求受付部によって受付処理が行われたリクエストは転送先判定待ち行列にスタックされるものとする。
【００６１】
この転送先判定待ち行列にデータが存在した場合（ステップ７００）、閾値保持部１２８よりアセスメントレベルの読み込みと情報テーブル１２６にアクセスを行い読み書き可能な状態とし、テーブル情報を読み込む（ステップ７０１）。
【００６２】
そして、モードが負荷集中モードで、状態が稼動中の全てのサーバにおいて、負荷量がアセスメントレベルに達している場合には、負荷集中モードを負荷分散モードに切り替え（ステップ７０２、７０４、７１２）、モードが負荷分散モードなら、負荷分散アルゴリズムにより負荷が均等に割り振られるようにサーバを決定し（ステップ７０２、７０３）、負荷分散処理後、負荷集中モードへと戻す（ステップ７０６）。ここで用いる負荷分散アルゴリズムについては公知の技術であるため、ここでは詳細には述べないこととする。
【００６３】
稼動中のサーバにおいて、負荷量がアセスメントレベルに達してないサーバが存在する場合には、優先度が最も高いサーバを“調査サーバ”とする（ステップ７０４、７０５）。
【００６４】
そして、“調査サーバ”の負荷量とアセスメントレベルとを比較し（ステップ７１３）、ステップ７１３で、アセスメントレベルに達している場合には、優先度が次に高いサーバを“調査サーバ”として負荷量がアセスメントレベルに達しているかを確認していく（ステップ７１４）。
【００６５】
ステップ７１３で、“調査サーバ”の負荷量がアセスメントレベルに達していない場合には、そのサーバを割り当てサーバに決定し（ステップ７０７）、情報テーブル１２６のサーバのコネクション数欄を更新する（ステップ７０８）。
【００６６】
その際のアセスメントレベルとサーバの負荷量の比較とは、例えばアセスメントレベルがコネクション数の場合にはサーバの接続コネクション数の値であり、レスポンス時間の場合には、レスポンス時間の値との比較である。
【００６７】
そして、リクエスト情報を１つ転送先判定待ち行列から読み出し、転送先サーバの情報を付加して転送待ち行列にスタックする（ステップ７０９、７１０）。

【００６８】
そして、判定待ち行列にデータがあるかを判断し（ステップ７１１）、ステップ７１１で判定待ち行列にデータがあれば、ステップ７０４に戻り、ステップ７１１で判定待ち行列にデータが無ければステップ７００に戻る。
【００６９】
要求転送部１２３は、この転送待ち行列に積まれているデータを取り出して処理を行う。ただし、上記では、アセスメントレベルに達していないサーバとは、新に１つ負荷を割り振った場合にアセスメントレベルを達するか達しないかの状態にはなるが、アセスメントレベルを超える状態（つまりサービスレベルアグリーメントを満足できない状態）にはならないと仮定している。
【００７０】
この仮定を満足させるにはアセスメントレベルを以下のように設定しておく必要がある。すなわち、サーバに割り振った負荷量に対してサービスレベルがどのように変化するかを予め実測値や理論値を元に算出できるようにしておき（しかもこれは負荷量増大に対しサービスレベルが単純に低下する単調減少となっている必要があり、そうでないならば、そもそも負荷量として設定している計量単位を見直すべきである）、サーバに新に１個の負荷を割り振った場合にサービスレベルアグリーメントを満足できる、最大の負荷量をアセスメントレベルと設定する必要がある。
【００７１】
もし、例えば１個の負荷に対し負荷量の増減にばらつきがあるならば、アセスメントレベルは、あるサーバに新に最大の負荷量をもつ１個の負荷が割り振られてもサービスレベルアグリーメントを満足する最大の負荷量とすべきであるし、もっと工夫するならば、図６に示す処理の中で、アセスメントレベルをそのときに割り振ろうとしている負荷の内容から負荷量の増分を予測し、その予測値に基づいてアセスメントレベルに達しているのかどうかを判断してもよい。
【００７２】
通常、負荷分散装置は処理要求の内容を解析する処理を持っている場合が多いのでこの種のインプリメントは比較的容易に可能である。また、さらに工夫するならば、サーバ毎にこのアセスメントレベルを別にしてもよい。本発明の場合、アセスメントレベルに達しているかどうかを判定するステップ（図６のステップ７１３）より前のステップ（図６のステップ７０５）で調査サーバ（すなわち、割り振り先のサーバ）が決まっているため、このような工夫を行うことができる。
【００７３】
次に、図７および図９のフローチャートにより、ラウンドロビンモードについて説明する。
【００７４】
本来のラウンドロビンモードは、負荷を複数台のサーバに順番に均等に割り振ることで負荷分散を達成することができるが、この考え方で負荷集中に変換させることは困難である。ラウンドロビンモードの場合には、次の２通りの考え方を用いることとする。
【００７５】
１つ目の考え方としては、ラウンドロビンモードを単純に、負荷が発生した場合、前回割り当てたサーバの次に優先度が高いサーバに割り当てると考える、とした場合、割当先が１つしか存在しないという場合であれば、そのサーバにのみ負荷が割り当てられる。
【００７６】
これによって、本実施の形態では、負荷集中は以下のように考えて実現する。
【００７７】
「複数台の稼動中サーバが存在する中で、割当先はある特定の１つのサーバのみとする。負荷量がアセスメントレベルに達するまでは割当先はその特定サーバとし、アセスメントレベルに達したら、割当先を優先度が次に高いサーバとする。」
つまり、以下のように処理を行う。

【００７８】
（１）稼動中の全てのサーバにおいて、負荷量がアセスメントレベルに達しているならば、負荷分散モードに切り替え（ステップ８００、８０１、８０２、８０４、８１３）、負荷分散アルゴリズムに従って、つまり最後に割り当てたサーバの次に優先度の高いサーバに負荷を割り当てる（ステップ８００、８０１、８０２、８０３、８０６）。
【００７９】
（２）稼動中サーバに、アセスメントレベルに達していないサーバがある場合には、優先度が最も高いサーバを“調査サーバ”と指定する（ステップ８０４、８０５）。
【００８０】
（３）調査サーバの負荷量がアセスメントレベルに達しているか確認する（ステップ８０７）。
【００８１】
（４）ステップ８０７で調査サーバの負荷量がアセスメントレベルに達していなければ、調査サーバを割り当てサーバに決定し、割り当てテーブルと情報テーブルを変更し、割り当てテーブルに存在するサーバ情報の削除と調査サーバの割り当てテーブルへの追加を行い（ステップ８０８）、情報テーブル１２６の“調査サーバ”の負荷量を追加する（ステップ８０９）。
【００８２】
（５）転送待ち行列に、転送先サーバの情報を付加したデータをスタックし（ステップ８１０）、判定待ち行列にデータがあるかを判断し（ステップ８１１）、ステップ８１１で判定待ち行列にデータがあれば、ステップ８０４に戻り、ステップ８１１で判定待ち行列にデータが無ければステップ８００に戻る。
【００８３】
（６）ステップ８０７で、調査サーバの負荷量がアセスメントレベルに達している場合には、次に優先度が高いサーバを調査サーバとしてステップ８０７に戻る（ステップ８１２）。
【００８４】
例えば、アセスメントレベルがコネクション数３とし、情報テーブルと割り当てテーブルが図８（ａ）に示す状態の時点で、負荷が１個生起したとする。優先度順にコネクション数を確認すると、コネクション数がアセスメントレベルに達していない稼動中サーバの内、最も優先度が高いのはサーバ番号が２のサーバとなるため、割当先と決定する。そのため、割り当てテーブルの既存のサーバの情報を削除し、新たに対象サーバが追加され、情報テーブルと割り当てテーブルは、図８（ａ）に示す状態から、図８（ｂ）に示す状態へ変更される。
【００８５】
また、２つ目の考え方としては、負荷分散におけるランドロビンモードについて、ある一定期間内に割り振る負荷の数を、対象とするサーバ間で同数になるよう割り振るものと解釈し、これによって負荷集中は以下のように考えて実現する。
【００８６】
「ある一定期間内に割り振る負荷の数を、対象とするサーバの中で特定のサーバに集中して割り振る。ただし、割り振る負荷の数はアセスメントレベル以下となるようにする。」
つまり、以下のように処理を行う。
【００８７】
（１）稼動中の全てのサーバにおいて、負荷量がアセスメントレベルに達しているならば、負荷分散モードに切り替え（ステップ８１４、８１５、８１６、８１７、８１８）、負荷分散アルゴリズムに従って、つまり最後に割り当てたサーバの次に優先度の高いサーバに負荷を割り当てる（ステップ８１４、８１５、８１７、８１９、８２２）。
【００８８】
（２）稼動中サーバに、アセスメントレベルに達していないサーバがある場合には、優

先度が最も高いサーバを“調査サーバ”と指定する（ステップ８１８、８２０）。

【００８９】

（３）現時刻から時刻－Ｔの間に調査サーバに割り当てたコネクション数をカウントする（ステップ８２１）。

【００９０】

（４）調査サーバに割り当てたコネクション数がアセスメントレベルに達しているか確認する（ステップ８２３）。

【００９１】

（５）ステップ８２３で調査サーバに割り当てたコネクション数がアセスメントレベルに達していないなら、そのサーバに割り当てることを決定し、割り当てテーブルを変更し、割り当てテーブルに割り当てた時刻と、調査サーバの情報を追加する（ステップ８２４）。

【００９２】

（６）情報テーブル内の、現時点から時刻Ｔ以前のデータを削除し（ステップ８２５）、判定待ち行列にデータがあるかを判断し（ステップ８２７）、ステップ８２７で判定待ち行列にデータがあれば、ステップ８１８に戻り、ステップ８２７で判定待ち行列にデータが無ければステップ８１４に戻る。

【００９３】

（７）ステップ８２３で、調査サーバに割り当てたコネクション数がアセスメントレベルに達している場合には、次に優先度が高いサーバを調査サーバとしてステップ８２１に戻る（ステップ８２６）。

【００９４】

情報テーブルには、例えば、図１０に示すように過去に割り当てたサーバとその時刻が記録されており、一定時間（例えば１秒とか１０秒とか１分とか１時間とか、アプリケーションに応じてユーザが選択する）内の割り当てた個数を算出できるようになっており、調査サーバについて、その個数がアセスメントレベル以下であるかどうかをステップ８２３の判定で行うようになっている。

【００９５】

次に、本実施の形態の負荷分散システムによる効果について説明する。

【００９６】

まず、比較例として、従来技術による省電力制御での動作を図１１（ａ）に示す。

【００９７】

サーバシステムに稼動サーバが３台存在し、サーバ１台あたり３個の負荷まで処理できるとする（すなわちアセスメントレベルが負荷量３としていることを意味する）。従来技術の場合、各サーバの負荷が均等になるように割り当てられるため、例えば、時刻Ｔ１で生起した７個の負荷に対して、各サーバが処理中である場合、このＴ１の時点での負荷の割り当ては、各サーバに負荷が均等になるよう割り当てられるため、＃１から順に３個、２個、２個となる。

【００９８】

また、時刻Ｔ２、Ｔ３の時点で生起する負荷が６個、５個であった場合、サーバ２台で対応可能な負荷量が６個のため、単純に考えると２台で対応可能のためこの時点でサーバを２台に削減できることになる。

【００９９】

しかし、負荷分散の場合、一旦ここで３台のサーバに負荷が均等となるよう割り当てられるため、例えば、時刻Ｔ２に生起した負荷は２個づつ割り当てられるため、最速でも２個の処理が終了した時点での電源停止となる。

【０１００】

これに対して、本実施の形態のように、アセスメントレベルに達するまでは特定サーバに負荷を割り当て、達したならば、次のサーバに割り当てるという負荷集中を行うと、図１１（ｂ）に示すように、まず、時刻Ｔ１で発生した７個の負荷はサーバ＃１からアセス

メントレベルに達するまで負荷を割り当てていくため、割り当てられた負荷の個数は順に、３個、３個、１個となる。

【０１０１】

その後、Ｔ２で発生した負荷６個はサーバ２台で対応可能であるため、サーバ＃３には既に負荷が割り当てられることはないので、時刻Ｔ２以前で電源をオフにすることができる（場合により△の時点までさかのぼることも可能である）。

【０１０２】

なお、図１１に示す簡略なモデルから、本実施の形態の省電力制御としての効果の大きさは負荷の処理時間の長さに依存していることがわかる。例えば、従来技術では、サーバを電源停止するためには、電源停止するサーバでの負荷の処理が終了しないと電源停止できないが、従来技術では、電源停止させるサーバへの負荷の割り当てが均等に割り当てられているため、本実施の形態と比べて、その負荷の個数が多くなるため、負荷の処理時間が長くなれば電源停止するまでの時間が長くなってしまう。

【０１０３】

従って、例えば、インターネットショッピングやネットでのストリーミングデータを配信する場合のように一旦接続すると相当長い間継続する場合には、本実施の形態での負荷集中処理は効果的である。

【０１０４】

（実施の形態２）

実施の形態１では、負荷分散システム１２０に、サーバ管理部１２２、転送先判定部１２５、電源制御部１２４、情報テーブル１２６、割り当てテーブル１２７、閾値保持部１２８の機能を設けているが、実施の形態２では、複数の負荷分散装置を管理する管理サーバに設けるようにしたものである。

【０１０５】

図１２により、本発明の実施の形態２に係る負荷分散システムの構成ついて説明する。図１２は本発明の実施の形態２に係る負荷分散システムの構成を示す構成図である。

【０１０６】

図１２において、負荷分散システムは、４台のサーバから構成されるサーバシステムを接続した負荷分散装置１６６が２台と、負荷分散装置１６６を管理する管理サーバ１７７から構成されている。

【０１０７】

負荷分散装置１６６は、処理要求受付部１２１と、管理サーバ１７７との通信を行う管理サーバ通信部１６２と、管理サーバ１７７の指示によって、負荷をサーバに転送する要求転送部１２３と割当先サーバの情報を保持する割り当てテーブル１２７とを備える。

【０１０８】

管理サーバ１７７は、各負荷分散装置１６６との通信を行う負荷分散装置通信部１７１と負荷分散装置１６６で受け付けた処理の配分先を判定する転送先判定部１２５と、全てのサーバを監視・管理するサーバ管理部１２２と、計算機の電源制御を行う電源制御部１２４と、各サーバと負荷分散装置１６６の情報を保持する情報テーブル１２６と、アセスメントレベルを保持する閾値保持部１２８を備える。

【０１０９】

本実施の形態の負荷分散システムとしての全体の動作は、実施の形態１と同様である。

【０１１０】

このように、本実施の形態では、管理サーバ１７７上に、サーバ管理部１２２、転送先判定部１２５、電源制御部１２４、情報テーブル１２６、割り当てテーブル１２７、閾値保持部１２８の機能を負荷分散装置１６６と独立してもつことで、複数の負荷分散装置１６６と、複数の負荷分散装置１６６に接続されたサーバシステムを制御可能となり、より省電力効果が得られる。

【０１１１】

なお、図１２に示す例では、管理サーバ１７７による制御対象の負荷分散装置１６６を

２台としているが、２台である必要はなく、複数台の制御が可能である。

【０１１２】

この場合、複数の負荷分散装置１６６をまたいで負荷集中が可能となる。つまり、図１２に示す例では、１台の負荷分散装置１６６に接続されているサーバシステムのみで処理可能な場合、負荷を割り当てないサーバシステム側については、負荷分散装置１６６ごと電源を停止することが可能となる。

【０１１３】

以上、本発明者によってなされた発明を実施の形態に基づき具体的に説明したが、本発明は前記実施の形態に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【０１１４】

例えば、実施の形態１、２において、クライアントからの要求の割当先であるサーバシステムは、物理サーバに限定されず、仮想サーバであってもよい。

【産業上の利用可能性】

【０１１５】

本発明は、情報処理システムなどにおける負荷分散システムに関し、負荷分散と共に、システム全体の省電力化が必要なシステムなどに広く適用可能である。

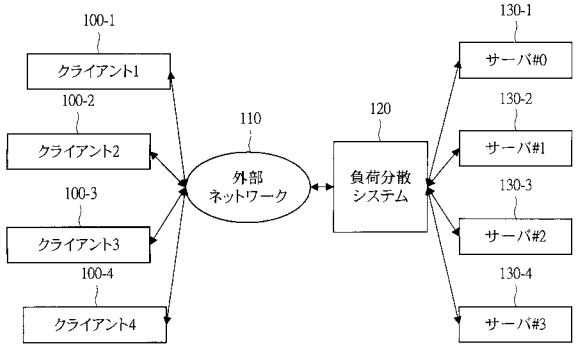
【符号の説明】

【０１１６】

１００（１００－１～１００－４）…クライアント、１１０…外部ネットワーク、１２０…負荷分散システム、１２１…処理要求受付部、１２２…サーバ管理部、１２３…要求転送部、１２４…電源制御部、１２５…転送先判定部、１２６…情報テーブル、１２７…割り当てテーブル、１２８…閾値保持部、１３０（１３０－１～１３０－４）…サーバシステム、１６２…管理サーバ通信部、１６６…負荷分散装置、１７１…負荷分散装置通信部、１７７…管理サーバ。

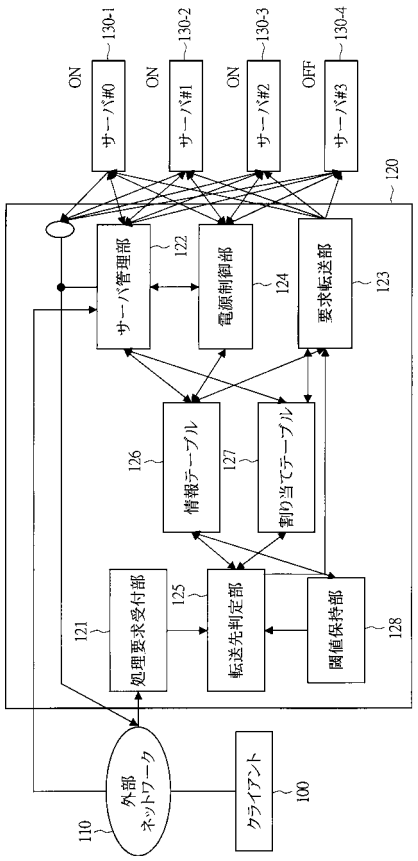
【図１】

図 1



【図２】

図 2



【図３】

図 3

(a)

126

301 サーバ番号	302 優先度	303 状態	304 コネクション数	305 レスポンス時間	306 アドレス
#0	1	稼動中	3	0.0531	xxxx.xxxx.xxxx.xxxx
#1	2	稼動中	3	0.0455	xxxx.xxxx.xxxx.xxxx
#2	3	稼動中	2	0.0155	xxxx.xxxx.xxxx.xxxx
#3	4	停止中	0	0.0000	xxxx.xxxx.xxxx.xxxx

(b)

127

307 サーバ番号	308 アドレス	309 コネクション数	310 レスポンス時間
#0	xxxx.xxxx.xxxx.xxxx	3	0.0531
#1	xxxx.xxxx.xxxx.xxxx	3	0.0455
#2	xxxx.xxxx.xxxx.xxxx	3	0.0155

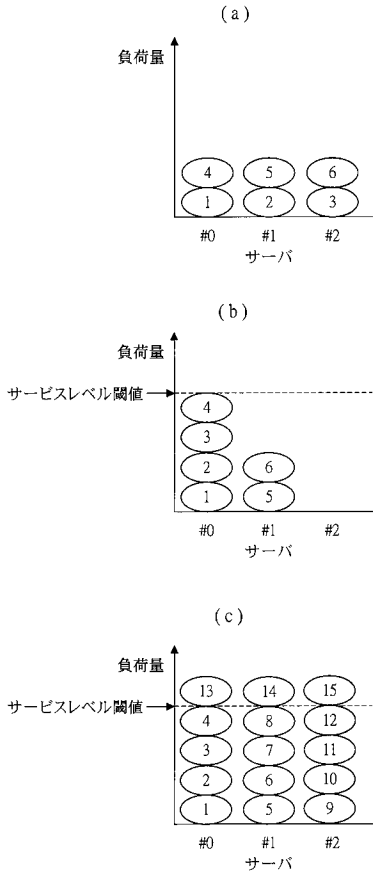
【図４】

図 4

#	401 アルゴリズム名称	402 負荷分散アルゴリズムの説明
1a	最少コネクションモード	リクエストを、各サーバに接続しているTCPコネクションが最も少ないものにリクエストを割り振る。
1b	重み付き最少コネクションモード	最少コネクションモードにサーバごとに割り当ての重み付けをつけて負荷分散を実施。
2a	ラウンドロビンモード	リクエストを複数台のサーバに順番に均等に割り振る。
2b	重み付きラウンドロビンモード	ラウンドロビンにサーバごとに割り当ての重み付けをつけて負荷分散を実施。
3	最速モード	リクエストを最も応答時間が早いサーバへ優先的に割り振る。

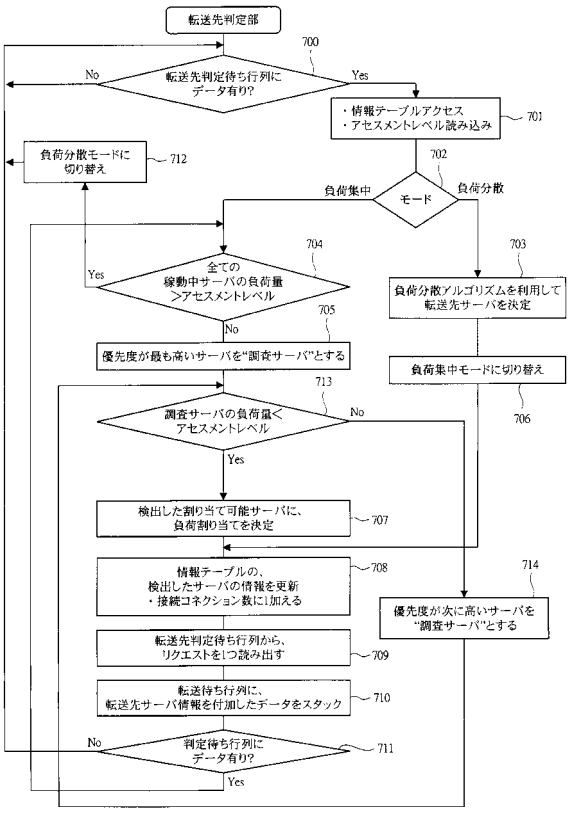
【図５】

図 5



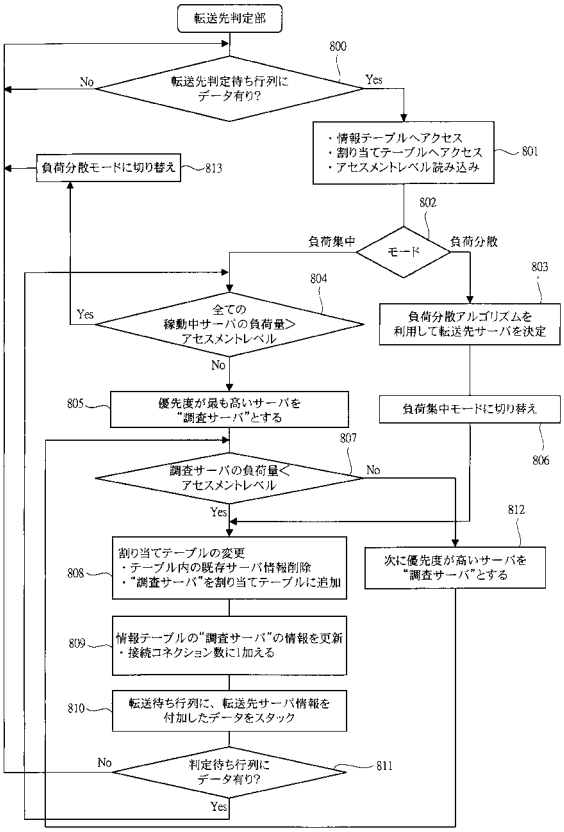
【図６】

図 6



【図７】

図 7



【図８】

図 8

(a)

210

サーバ番号	優先度	状態	コネクション数	レスポンス時間	アドレス
#0	1	稼動中	3	0.0531	xxxx.xxxx.xxxx.xxxx
#1	2	稼動中	3	0.0455	xxxx.xxxx.xxxx.xxxx
#2	3	稼動中	1	0.0155	xxxx.xxxx.xxxx.xxxx
#3	4	起動処理	0	0.0000	xxxx.xxxx.xxxx.xxxx
#4	5	停止中	0	0.0000	xxxx.xxxx.xxxx.xxxx

211

サーバ番号	アドレス	コネクション数	レスポンス時間
#1	xxxx.xxxx.xxxx.xxxx	3	0.0531

(b)

212

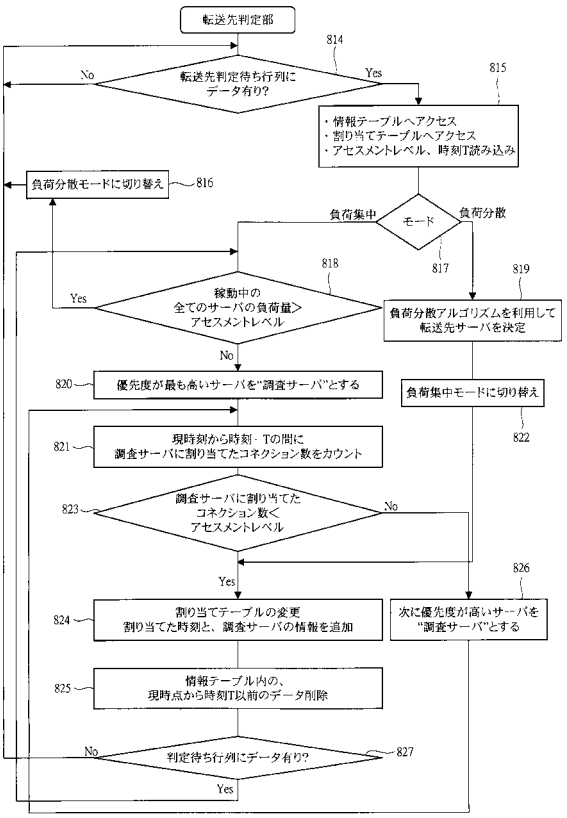
サーバ番号	優先度	状態	コネクション数	レスポンス時間	アドレス
#0	1	稼動中	3	0.0531	xxxx.xxxx.xxxx.xxxx
#1	2	稼動中	3	0.0455	xxxx.xxxx.xxxx.xxxx
#2	3	稼動中	2	0.0255	xxxx.xxxx.xxxx.xxxx
#3	4	起動処理	0	0.0000	xxxx.xxxx.xxxx.xxxx
#4	5	停止中	0	0.0000	xxxx.xxxx.xxxx.xxxx

213

サーバ番号	アドレス	コネクション数	レスポンス時間
#2	xxxx.xxxx.xxxx.xxxx	2	0.0255

【図９】

図 9



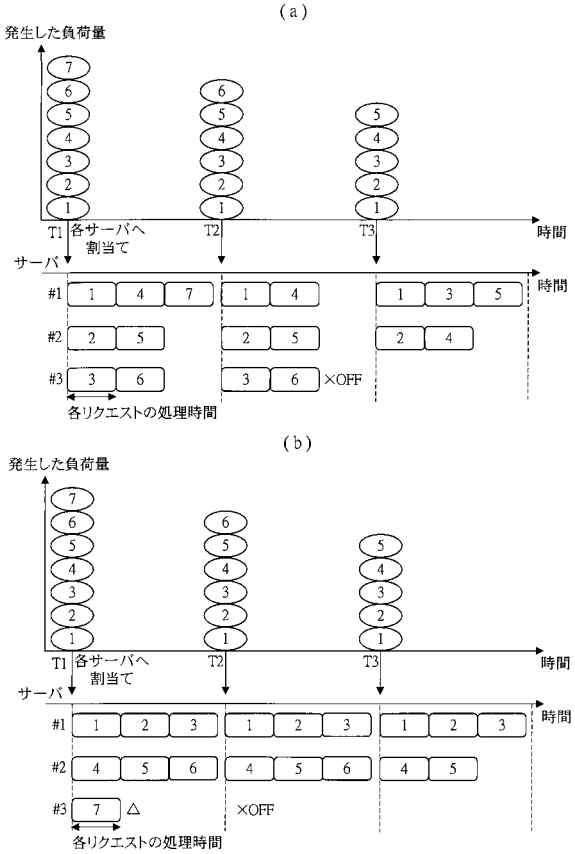
【図１０】

図 10

サーバ番号	時刻	状態	優先度	コネクション数	レスポンス時間	アドレス
#0	15:00	稼動中	1	1		xxxx.xxxx.xxxx.xxxx
#0	17:00	稼動中	1	1		xxxx.xxxx.xxxx.xxxx
#0	17:03	稼動中	1	1		xxxx.xxxx.xxxx.xxxx
#1	18:00	稼動中	2	1		xxxx.xxxx.xxxx.xxxx
#1	18:22	稼動中	2	1		xxxx.xxxx.xxxx.xxxx
#1	18:30	稼動中	2	1		xxxx.xxxx.xxxx.xxxx
#1	18:45	稼動中	2	1		xxxx.xxxx.xxxx.xxxx
#2	18:50	稼動中	2	1		xxxx.xxxx.xxxx.xxxx
#2	20:35	稼動中	2	1		xxxx.xxxx.xxxx.xxxx
#3	21:22	停止中	2	0		xxxx.xxxx.xxxx.xxxx
#4	21:30	停止中	2	0		xxxx.xxxx.xxxx.xxxx

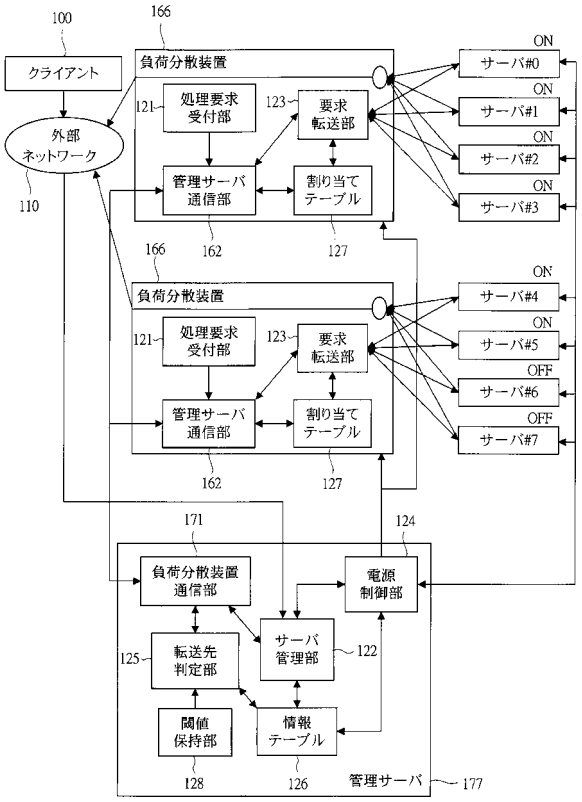
【図１１】

図 11



【図１２】

図 12



フロントページの続き

(72)発明者　亀田　泰弘
神奈川県秦野市堀山下１番地　株式会社日立製作所　エンタープライズサーバ事業部内
(72)発明者　藤田　博文
神奈川県秦野市堀山下１番地　株式会社日立製作所　エンタープライズサーバ事業部内

審査官　浦口　幸宏

(56)参考文献　特開２００８－２２５６４２（ＪＰ，Ａ）
特開２０１０－１６５１９３（ＪＰ，Ａ）
特開２００６－２２７９６３（ＪＰ，Ａ）
特開２００８－２６９２４９（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１／２６－　１／３２
Ｇ０６Ｆ　９／４６－　９／５４
Ｇ０６Ｆ　１３／００
Ｈ０４Ｌ　１２／００－１２／２６
Ｈ０４Ｌ　１２／５０－１２／９５５