

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2013-90072
(P2013-90072A)

(43) 公開日 平成25年5月13日(2013.5.13)

(51) Int.Cl. F I テーマコード(参考)
H04L 12/70 (2013.01) H04L 12/56 B 5K030

審査請求 未請求 請求項の数 7 O L (全 20 頁)

(21) 出願番号	特願2011-227563 (P2011-227563)	(71) 出願人	000005108 株式会社日立製作所 東京都千代田区丸の内一丁目6番6号
(22) 出願日	平成23年10月17日(2011.10.17)	(74) 代理人	100100310 弁理士 井上 学
		(74) 代理人	100098660 弁理士 戸田 裕二
		(74) 代理人	100091720 弁理士 岩崎 重美
		(72) 発明者	東村 邦彦 神奈川県横浜市戸塚区吉田町292番地 株式会社日立製作所横浜研究所内
		(72) 発明者	竹島 由晃 神奈川県横浜市戸塚区吉田町292番地 株式会社日立製作所横浜研究所内 最終頁に続く

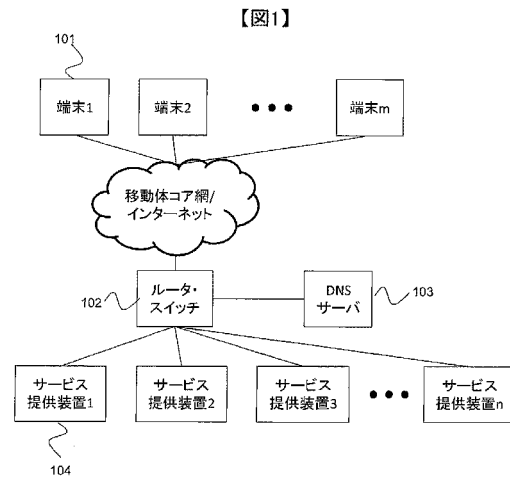
(54) 【発明の名称】 サービス提供システム

(57) 【要約】

【課題】 端末からサービス提供装置群に要求を送信するシステムにおいて、端末からの要求を負荷分散装置で多数のサービス提供装置に分散する場合、負荷分散装置自体が性能上のボトルネックおよび単一障害点となる。また、負荷分散装置を利用せずにサービス提供装置群に負荷を分散させる場合には、端末側にサービス提供装置を適切に選択させるための特殊な機能が必要となる。

【解決手段】 複数のIPアドレスを格納したアドレス群管理テーブルをもち端末からの要求に対して当該アドレス群管理テーブルからアドレスを選択して返答する処理を行う名前解決装置と、アドレス群管理テーブルに格納されているIPアドレスのうち0個以上を重複なく割り当てた複数のサービス提供装置を備え、サービス提供装置の処理能力に応じた数のIPアドレスを割り当てることで負荷の分散を実現する。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

複数のサービス提供装置と一つ以上の名前解決装置を備え、端末からのサービス要求に、いずれかの前記サービス提供装置が応答するサービス提供システムにおいて、

前記名前解決装置は、

k 台のサービス提供装置からなるサービス提供装置グループに対して割り当てられた n 個 (k は n 以上) の IP アドレスからなるアドレス群を管理し、

前記端末からの名前解決要求に対して、管理する前記アドレス群から、いずれかひとつの IP アドレスを選択して応答する名前解決要求処理部と、を備え、

前記複数のサービス提供装置の各々は、

前記名前解決装置が管理する前記 n 個の IP アドレスのうち m (m は 0 以上) 個が、他の前記サービス提供装置と重複することなく当該サービス提供装置に割り当てられており、

前記名前解決要求に対する前記応答に基づいた、前記端末からのサービス要求に対してサービスを提供する端末要求処理部と、

他のサービス提供装置との間で負荷分散を行う負荷分散機能部と、を備え、

監視先サービス提供装置として、他の監視元サービス提供装置から、その負荷を監視されており

前記監視元サービス提供装置の前記負荷分散機能部は、前記監視先サービス提供装置の負荷が高い場合、IP アドレスの割り当てを、前記監視先サービス提供装置から自サービス提供装置へ移動させることを特徴とするサービス提供システム。

【請求項 2】

請求項 1 に記載のサービス提供システムにおいて、

前記監視元サービス提供装置の前記負荷分散機能部は、前記 IP アドレスの割り当てを移動させる処理として、

監視先サービス提供装置の負荷が、自サービス提供装置の負荷より一定値以上高い場合、前記監視先サービス提供装置へ、アドレス引継ぎ要求を送信し、

前記監視先サービス提供装置から、割り当ての移動対象である IP アドレスを含むアドレス引継ぎ応答を受信したら、含まれる前記 IP アドレスを、自装置への割り当て IP アドレスとして登録する

ことを特徴とするサービス提供システム。

【請求項 3】

請求項 2 に記載のサービス提供システムにおいて、

前記監視先サービス提供装置の前記負荷分散機能部は、

前記アドレス引継ぎ要求に応答して、自装置に割り当てられている担当 IP アドレスの一つを選択し、

選択した前記 IP アドレスについて、自装置への割り当ての削除と、前記アドレス引継ぎ応答の送信による前記監視元サービス提供装置への通知と、を行う

ことを特徴とするサービス提供システム。

【請求項 4】

請求項 1 から 3 のいずれかに記載のサービス提供システムにおいて、

前記割り当ての移動対象となる IP アドレスは利用されていないアドレスであって、

前記名前解決装置は、

管理するアドレス群を 2 つ以上の下位の群に分割して管理し、

日時に応じて、前記端末からの名前解決要求に対して選択するアドレスが属する前記下位の群を、日時に応じて選択することを特徴とするサービス提供システム。

【請求項 5】

請求項 1 から 4 のいずれかに記載のサービス提供システムにおいて、

前記監視元サービス提供装置の前記負荷分散機能部は、

IP アドレスの割り当てが監視先サービス提供装置から当該前記監視元サービス提供装置へ移動した前回の日時を管理し、

10

20

30

40

50

現在の日時と、前記前回の日時との差が所定以上である場合に、前記IPアドレスの割り当ての移動を行うことを特徴とするサービス提供システム。

【請求項6】

請求項1から5のいずれか一に記載のサービス提供システムにおいて、
前記監視元サービス提供装置の前記負荷分散機能部は、
他の稼働中の監視先サービス提供装置を監視し、
前記監視先サービス提供装置から自装置へ、割り当てを移動させるIPアドレスの候補を管理し、
監視により、前記監視先サービス提供装置に障害が発生した、と判断した場合、管理する、前記割り当てを移動させるIPアドレスの候補を、自装置への割り当てIPアドレスとして登録することを特徴とするサービス提供システム。

10

【請求項7】

請求項6に記載のサービス提供システムにおいて、
k台の前記サービス提供装置は、稼働中および待機中の2つの群に分けられ、
稼働中の前記監視元サービス提供装置の前記負荷分散機能部は、他の稼働中の監視先サービス提供装置を監視し、
前記監視先サービス提供装置の障害を確認した場合、前記待機中のサービス提供装置を一つ、新たな監視先サービス提供装置として選択し、
障害が発生した前記監視先サービス提供装置に割り当てられていた前記IPアドレスを含むアドレス引継ぎ応答を、新たな監視先サービス提供装置へ送信することを特徴とするサービス提供システム。

20

【発明の詳細な説明】

【技術分野】

【0001】

本明細書で開示される主題は、端末からの要求に対してネットワークを介してサービスを提供するサーバシステムに関する。

【背景技術】

【0002】

Webサービスなど、ネットワーク上のサービス提供装置に端末からアクセスすることにより何らかのサービスを受けるといった形態が一般的になってきている。サービス提供システム側は、サービス利用者数の増加に応じてシステムの増強を行うことが必要である。

30

【0003】

システムの性能の増強には、構成要素の性能を増強させる(スケールアップ)アプローチと、構成要素の数を増やす(スケールアウト)アプローチがある。前者のアプローチでは、構成要素の性能増強の上限に達した時点で性能拡張ができなくなるため、大規模なシステムでは主に後者のアプローチをとることが必要とされる。

【0004】

端末からのアクセスを受けサービスを行うサービス提供システムのスケールアウトを行う場合、従来から負荷分散装置とよばれる特殊アプライアンスを複数のサービス提供装置の前段におくという方法がとられている。この方法では、端末は前段におかれた負荷分散装置に対してアクセスする。端末からのアクセスを受けた負荷分散装置は、配下にあるサービス提供装置の負荷などを勘案して適切なサービス提供装置を選択し、端末からの要求を転送する。システムとして処理能力が不足した場合は、配下のサービス提供装置を増やすことでスケールアウト型の処理能力向上が可能となる。

40

【0005】

また、特許文献1に開示されている方法では、端末がサービス提供装置にアクセスする前に名前解決を行うことを利用し、端末からの名前解決要求への返答時に前記負荷分散装置と同様に適切なサービス提供装置を選択し、振り分け先のサービス提供装置へのアクセスに必要な情報を返すというアプローチをとっている。

50

【 0 0 0 6 】

また、非特許文献 1 では、スケールアウト型のサービス提供装置においてストレージサービスを提供する際に、各サービス提供装置の能力に応じて担当するデータ量を調整するために、各装置に1つ以上の仮想ノードを割り当て、その割り当てを変化させることで各サービス提供装置が担当するデータ格納量を調整する方法が開示されている。

【 先行技術文献 】

【 特許文献 】

【 0 0 0 7 】

【 特許文献 1 】 米国特許第7284051号明細書

【 非特許文献 】

【 0 0 0 8 】

【 非特許文献 1 】 Giuseppe DeCandia等, “ Dynamo: Amazon's Highly Available Key-Value Store ”, in the Proceedings of the 21st ACM Symposium on Operating Systems Principles, Stevenson, WA, October 2007.

【 発明の概要 】

【 発明が解決しようとする課題 】

【 0 0 0 9 】

前述した、負荷分散装置を用いたスケールアウトが有効な範囲は、1つのサービス提供装置での処理量が負荷分散装置での振り分けにかかる処理量より非常に大きい場合に限られる。すなわち、サービス提供装置での処理が軽量なものであるか、サービス提供装置単体の性能が大きい場合などは、負荷分散装置での振り分け処理自体がボトルネックとなり、システム全体としての性能の上限を規定してしまう。また、負荷分散装置で障害が発生した場合、その配下すべてのサービス提供装置へのアクセスが不能となるため、二重化等負荷分散装置での障害対応に大きなコストがかかる。

【 0 0 1 0 】

特許文献 1 に開示されている名前解決を利用した負荷分散では、名前解決を行った後のサービス提供装置へのアクセスは、名前解決装置に障害が発生した場合であっても可能である。しかし、名前解決を適切に行うために、端末とサービス提供装置間の経路の情報などを収集しその状態を保持しておく必要がある。このため、この情報を保持している部分に障害が発生した場合には、以後の名前解決は困難になりシステム全体へのアクセスも困難になる。

【 0 0 1 1 】

また、非特許文献 1 に開示されている負荷分散方法では、仮想ノード数によって各サービス提供装置が担当するデータ格納量を調整可能であるが、端末側はどこにデータが格納されているかを意識せずにサービス提供装置のいずれかにアクセスするため、実際にはデータにアクセスするため再度の処理転送が発生する。このため、最初に挙げた負荷分散方法による負荷分散と同様、処理の転送がボトルネックとなりうる。これを回避するために、端末側にデータ格納先を判別する処理部を埋め込む方法も考えられるが、すべての端末に格納先判別のプログラムを配付し、サービス提供装置群の状況に応じて必要な情報を全端末に通知する必要があるなどの問題が残る。

【 課題を解決するための手段 】

【 0 0 1 2 】

本明細書で開示されるのは、負荷分散のための機能 (Feature) がボトルネックや障害点となりうる点、および負荷分散のために端末側に負荷分散のための機能を備えることを要求する、という従来の課題を解決する方法と、当該方法を実行するサービス提供システムである。

【 0 0 1 3 】

開示されるサービス提供システムは、端末からサービス提供装置にアクセスする際に名前解決装置に対してサービス提供装置の識別子の名前解決を要求し、名前解決の結果得られたIPアドレスに対してサービス要求を行うシステムを前提とするものであり、さらに、

10

20

30

40

50

以下の特徴をもつ。

【0014】

複数のIPアドレスを格納したアドレス群管理テーブルをもち端末からの要求に対して当該アドレス群管理テーブルからアドレスを選択して返答する処理を行う名前解決装置と、アドレス群管理テーブルに格納されているIPアドレスのうち0個以上を重複なく割り当てた複数のサービス提供装置を備え、サービス提供装置の処理能力に応じた数のIPアドレスを各サービス提供装置に割り当てることで負荷の分散を実現する。

【0015】

端末からの要求を受け付ける各サービス提供装置は、監視先となる他のサービス提供装置の負荷を計測し、監視先の負荷が監視元である自サービス提供装置の負荷より、大きい場合、負荷が大きい監視先装置に割り当てられているIPアドレスを負荷の小さい自装置に移動させることにより負荷の差を小さくする処理を行うことにより負荷の変動への対応を行う。

10

【0016】

また、名前解決装置において、アドレス群管理テーブル内のアドレスを2つ以上の群に分割して管理し、日時から1つあるいは複数の群を選択する関数によって、端末からの要求に対して返答対象となるアドレスを日時ごとに切り替える機能を持つことにより、前記負荷譲渡に必要な未使用アドレスを高い確率で確保可能とする。

【0017】

開示されるサービス提供システムのより具体的な態様は、複数のサービス提供装置と一つ以上名前解決装置を備え、端末からのサービス要求に、いずれかのサービス提供装置が応答するサービス提供システムであって、

20

名前解決装置は、

k台のサービス提供装置からなるサービス提供装置グループに対して割り当てられたn個(kはn以上)のIPアドレスからなるアドレス群を管理し、

端末からの名前解決要求に対して、管理するアドレス群から、いずれかひとつのIPアドレスを選択して応答する名前解決要求処理部と、を備え、

複数のサービス提供装置の各々は、

名前解決装置が管理するn個のIPアドレスのうちm(mは0以上)個が、他のサービス提供装置と重複することなく当該サービス提供装置に割り当てられており、

30

名前解決要求に対する応答に基づいた、端末からのサービス要求に対してサービスを提供する端末要求処理部と、

他のサービス提供装置との間で負荷分散を行う負荷分散機能部と、を備え、

監視先サービス提供装置として、他の監視元サービス提供装置から、その負荷を監視されており、

監視元サービス提供装置の負荷分散機能部は、監視先サービス提供装置の負荷が高い場合、IPアドレスの割り当てを、監視先サービス提供装置から自サービス提供装置へ移動させることを特徴とする。

【0018】

開示されるサービス提供システムにおいて、さらに、

40

監視元サービス提供装置の負荷分散機能部が、IPアドレスの割り当てを移動させる処理として、

監視先サービス提供装置の負荷が、自サービス提供装置の負荷より一定値以上高い場合、監視先サービス提供装置へ、アドレス引継ぎ要求を送信し、

監視先サービス提供装置から、割り当ての移動対象であるIPアドレスを含むアドレス引継ぎ応答を受信したら、含まれるIPアドレスを、自装置への割り当てIPアドレスとして登録するように構成しても良い。

【0019】

上記態様により、端末からサービス提供装置の間に、負荷分散装置のような単一障害点およびボトルネックとなる装置をおかずにサービス提供装置群に対して負荷を分散させる

50

ことが可能となり、障害耐性が向上する。

【0020】

また、名前解決の際に必要な名前解決装置は、各サービス提供装置の負荷情報という動的なデータを保持せず、IPアドレス群という静的なデータを状態として保持するため、IPアドレス群のデータを安定ストレージに格納するなどの手法によって、容易に障害から復旧が可能となる。

【0021】

また、複数の名前解決装置を配備した場合であっても、静的なデータをそれぞれの名前解決装置に与えるだけでよいため、共有データの動的な更新に対する一貫性の保持などの複雑な手順が不要となる。

【0022】

また、処理能力に応じたIPアドレス数を割り当てることで各サービス提供装置にかかる負荷を調整することにより、端末が標準的にもつIPでの通信機構を用いて適切なサーバへ要求を誘導できるとともに、IPアドレスの譲渡によってあるIPアドレスをもったサービス提供装置が変更になった場合も、IPネットワークが標準的にもつルーティング機能によって、新たな要求が適切にリダイレクトされる。

【0023】

また、名前解決装置において返すアドレス群を時系列で変動させることにより、端末とサービス提供装置の間でセッションが確立していないサービス提供装置側アドレスを確保できるようになる。これによって、サービス提供装置においてIPアドレスを譲渡する際に、未使用のアドレスを選択することが可能となる。TCP/IP上のプロトコルなど、サービス提供装置に何らかのセッション管理が必要な場合は、セッション確立中に当該セッションが使用しているIPアドレスを譲渡するには複雑な処理が必要であり、未使用のIPアドレスを譲渡することが望ましい。本方式により、未使用のIPアドレスを確保できる可能性を向上させ、負荷均等化にかかる処理コストを削減することが可能となる。

【発明の効果】

【0024】

開示に従えば、負荷分散機能を備えながら、負荷分散機能がボトルネックや障害点となりにくく、かつ、負荷分散機能を端末側に備える必要の無い、サービス提供システムが実現可能になる。

【図面の簡単な説明】

【0025】

【図1】サービス提供システムの全体構成を例示する図

【図2】DNSサーバの機能構成を例示する図

【図3】アドレス群管理テーブルの構成を例示する図

【図4】DNSサーバの処理フローチャートを例示する

【図5】サービス提供装置の機能構成を例示する図

【図6】サービス提供装置リストの構成を例示する図

【図7】担当IPアドレスリストの構成を例示する図

【図8】監視先サービス提供装置リストの構成を例示する図

【図9】監視元サービス提供装置リストの構成を例示する図

【図10】サービス提供装置の監視と被監視の関係を構築する処理を例示するフローチャート

【図11】サービス提供装置の負荷分散機能部の処理を例示するフローチャート

【図12】サービス提供装置の負荷分散機能部の処理を例示するフローチャート

【図13】サービス提供システムのシーケンス構成を例示する図

【図14】サービス提供システムのシーケンス構成を例示する図

【図15】N+M構成の場合のサービス提供装置リスト構成を例示する図

【図16】日時に応じて返答するIP群を切り替える場合のDNSサーバの処理を例示するフローチャート

10

20

30

40

50

【図17】DNSサーバが複数存在する場合のサービス提供システムの全体構成を例示する図

【図18】DNSサーバと、サービス提供装置のハードウェア構成を例示する図

【発明を実施するための形態】

【0026】

(実施例1)

図1は、実施例1の全体構成図である。

【0027】

サービス提供装置104およびDNSサーバ103は、外部ネットワークを構成するルータ・スイッチ102に接続されており、ルータ・スイッチ102は移動体コア網/インターネットなどを介して端末101と接続されている。なお、ここで挙げた接続トポロジは一例であり、端末101、DNSサーバ102、サービス提供装置104がそれぞれ互いにIP到達性のある接続トポロジであれば本実施例を適用可能である。

10

【0028】

図2に本実施例の名前解決装置の一例であるDNSサーバ103の機能構成を示す。DNSサーバ103は、名前解決の一例として、ドメイン名をIPアドレスに変換する。DNSサーバ103はルータ・スイッチ102とパケット送受信部204を介して接続している。端末101からの名前解決要求は名前解決要求処理部(以下、DNS要求処理部という)202で処理される。また、アドレス群管理テーブル201へのアドレスの登録や削除などの処理は管理機能部203にて行う。

20

【0029】

上記DNSサーバ103と、機能構成を後述するサービス提供装置104のハードウェア構成の例を図18に示す。

【0030】

これらの装置は、主記憶装置1802、中央処理装置1803、ハードディスクドライブなどの二次記憶装置1805、ディスプレイやキーボードなどの入出力装置1807、ルータ・スイッチ102と接続するための通信装置1806、およびそれらの装置間を接続するバスなどの内部通信線1804を備えた一般的なコンピュータ1801により実現できる。

【0031】

たとえばDNSサーバ内のアドレス群管理テーブル201は、主記憶装置1802に置かれるとともに、永続的にデータを記憶するために二次記憶装置1805に置かれる。

30

【0032】

また、以下で説明する各処理は、二次記憶装置1805に格納されたプログラムが、実行時に主記憶装置1802にコピーされたうえで中央処理装置1803によって解釈・実行されることにより実現する。また、送受信処理は、通信装置1806を介して行う。

【0033】

図3は、DNSサーバ103がもつアドレス群管理テーブル201の一例である。ここで示したアドレス群管理テーブル201では、解決すべき名前(本実施例ではドメイン名)に対するアドレスとして、複数のエントリにIPv6アドレスをサブネット単位で登録している。図3では、アドレス群をIPv6アドレスで表現しているが、IPv4アドレスでアドレス群を表現してもよい。アドレス群管理テーブル201の各エントリには、全てのサービス提供装置104に割り当てたアドレスを登録する。

40

【0034】

次に、このアドレス群管理テーブル201を用いて、DNSサーバ103がどのように端末101からの要求に応えるかを示す。

【0035】

図4に、DNSサーバ103が端末101からの要求を受けた場合の処理フローを示す。ステップ401にて、端末101が要求してきた名前解決要求がどの名前に対するものであるかを調べる。たとえば、図3に示した形のアドレス群管理テーブルを持っていた場合、www.example.comの名前解決を要求している場合はステップ402に進む。

50

【 0 0 3 6 】

ステップ402では、アドレス群管理テーブルに含まれるアドレスから1つを選択して返答する。この時、アドレス群から返答するアドレスを選択する方法は、各アドレスを返答する確率がほぼ等しいのであれば方法は問わない。例としては、先頭から順番に1つずつ選択する方法や、全アドレス群の中から一様乱数を用いて選択する、という方法がある。

【 0 0 3 7 】

ステップ401でいいえ、すなわちアドレス群管理テーブル201として管理していない名前に対する要求の場合は、他のDNSサーバに要求を転送し、転送先の他のDNSサーバへの問い合わせによって得られた結果を返答する（ステップ403）。これは、サービス提供装置104と関係のない名前に対する名前解決要求が来た場合の対処である。本サービス提供システムと関係のない名前への要求をサポートしない場合、単に名前解決のエラーを返してもよい。

10

【 0 0 3 8 】

次に、サービス提供装置104に関して説明する。

【 0 0 3 9 】

図5にサービス提供装置104の構成図を示す。サービス提供装置104は、パケット送受信部502を介してルータ・スイッチ102と接続している。サービス提供装置104が端末101に対して提供するサービスの処理は端末要求処理部501で処理する。また、負荷分散に関しては負荷分散機能部503で処理する。負荷分散機能部503は4つのテーブルを持ち、それぞれサービス提供装置リスト601、監視先サービス提供装置リスト801、監視元サービス提供装置リスト901、担当IPアドレスリスト701とよぶ。

20

【 0 0 4 0 】

図6にサービス提供装置リスト601の一例を示す。

【 0 0 4 1 】

サービス提供装置リスト601は、負荷を分散しあうn台のサービス提供装置104の識別子を管理する。サービス提供装置リスト601の値は、サービス提供装置104の起時に管理者によって設定されるとともに、サービス提供装置104がシステムに追加、あるいはシステムから削除された場合に負荷分散機能部によって更新される。

【 0 0 4 2 】

図7に担当IPアドレスリスト701の一例を示す。

30

【 0 0 4 3 】

担当IPアドレスリスト701には、本リストをもつ自サービス提供装置104に割り当てられたm個のIPアドレスが格納されている。担当IPアドレスリスト701は、サービス提供装置104の起動時に管理者によって設定されるとともに、他のノードとのIPアドレスの移動、すなわち譲渡と譲受、の際に負荷分散機能部によって更新される。

【 0 0 4 4 】

図8に監視先サービス提供装置リスト801の一例を示す。

【 0 0 4 5 】

監視先サービス提供装置リスト801は、このリストを持つサービス提供装置104が負荷の監視を行う対象である他のサービス提供装置104の情報を示す。

40

【 0 0 4 6 】

ここでは、当該サービス提供装置104は、識別子 "serv2" をもつサービス提供装置104を監視対象としており、監視対象サービス提供装置が高負荷になった場合や、障害が発生した場合に譲り受けるIPアドレスの候補リストと、前回の負荷調整によって、当該IPアドレスの割り当てが、当該監視対象サービス提供装置104へ移動した日時を負荷調整日時として記録している。負荷調整日時は、負荷によるアドレス割り当ての移動が短期間に連続して発生しないようにするために参照されるものである。

【 0 0 4 7 】

本実施例では、IPアドレスの割り当ての移動と負荷の移動は同期しないため、負荷に差がある状態ですぐにIPアドレスの割り当ての移動を行うと、負荷の移動に必要な量以上の

50

IPアドレスが移動してしまうため、負荷調整日時の参照により、短期間でのIPアドレスの移動を制限する。

【 0 0 4 8 】

図9に監視元サービス提供装置リスト901の一例を示す。

【 0 0 4 9 】

監視元サービス提供装置リスト901は、本リストをもつサービス提供装置104を監視している他のサービス提供装置104を記録している。

【 0 0 5 0 】

図9では、自サービス提供装置104は、"serv4"という識別子をもつ他のサービス提供装置104により監視されており、自サービス提供装置104が高負荷になった場合や、障害が発生した際に、"serv4"という識別子をもつサービス提供装置104に割り当てを移動すべきアドレスの候補を引継IPリストに格納している。

【 0 0 5 1 】

当該サービス提供装置104が、単一のサービス提供装置104から監視されている場合は、引継IPリストの内容は、当該サービス提供装置104の担当IPアドレスリスト701の内容と同一になる。複数のサービス提供装置から監視されている場合は、担当IPアドレスリスト701に含まれるアドレスを分担することになる。

【 0 0 5 2 】

また、当該サービス提供装置104が、IPアドレスの移動（他から譲り受け、または、他への譲り渡し）によって、担当IPアドレスリスト701を更新した場合、併せて、この引継IPリストも更新する。

【 0 0 5 3 】

初期設定時にすべてのサービス提供装置104間に監視と被監視の関係を構築する方法について以下に述べる。図10に負荷分散機能部503が実行する構築手順を示す。

【 0 0 5 4 】

ステップ1001では、サービス提供装置リスト601に含まれるすべてのサービス提供装置104の識別子に対して何らかの形の順序付けを行う。これは、すべての識別子に対して順序関係が定義できればいいため、ハッシュ値を計算して数の大小で決める、辞書順で決める、などの方法が考えられる。一例としてここではハッシュ値を用いて計算するものとしている。

【 0 0 5 5 】

そして、ステップ1002において、順序付けが可能となったサービス提供装置のリストを昇順でソートする。次にステップ1003において、サービス提供装置リスト601の中から自装置の識別子を探す。

【 0 0 5 6 】

最後にステップ1004で自装置のエントリの次にあるサービス提供装置104を監視先、前にあるサービス提供装置104を監視元とする。さらに、サービス提供装置リスト601はリング構造をしていると考え、リストの末尾の装置の監視先はリストの先頭の装置、リストの先頭の装置の監視元はリストの末尾の装置、とする。この処理によって、すべてのサービス提供装置104が一つのリング状のトポロジで仮想的に接続されたものと扱うことができる。

【 0 0 5 7 】

次に、一つのサービス提供装置104の負荷分散機能部503の処理フローを図11に示す。

【 0 0 5 8 】

ステップ1101として、サービス提供装置104の起動時、初期設定としてサービス提供装置リスト601と担当IPアドレスリスト701を設定する。担当IPアドレスリスト701の設定の一例としては、サービス提供装置104の性能が初期状態では均一であると仮定して、すべてのサービス提供装置104に同数のアドレスを割り当てる方法や、各サービス提供装置104に性能差によって重みづけを行い、性能が高いサービス提供装置104に多くのアドレスを割り当てる、などの方法がある。

10

20

30

40

50

【 0 0 5 9 】

次にステップ1102では、図10で示した処理手順に従って監視先サービス提供装置を決め、装置情報取得要求メッセージを送信する。そしてステップ1103でメッセージの受信待ちを行う。以下では、受信したメッセージの内容に応じて処理を行う。

【 0 0 6 0 】

装置情報取得要求メッセージを受信した場合、監視元サービス提供装置リスト901に送信元サービス提供装置のエントリを追加し、監視元全体に対して引継IPアドレスを再構成する。そして、全ての監視元サービス提供装置に対して、監視元サービス提供装置リスト901を含めた装置情報メッセージを送信する(ステップ1104)。

【 0 0 6 1 】

装置情報メッセージを受信した場合、監視先サービス提供装置リスト801を、受信したメッセージに含まれる装置情報で更新する(ステップ1105)。

【 0 0 6 2 】

負荷情報要求メッセージを受信した場合、現在の装置の負荷情報を負荷情報メッセージとして送信する(ステップ1106)。負荷情報としては、CPU負荷、単位時間当たりの送受信トラフィック量、I/O負荷、メモリ使用量などがある。ディスクへのアクセスがボトルネックとなる処理ではI/O負荷を負荷情報として使用するなど、サービス提供装置104で端末に提供するサービスの種類に応じて選択する。

【 0 0 6 3 】

監視元サービス提供装置リスト901に記載されている他のサービス提供装置104からアドレス引継ぎ要求メッセージを受信した場合、担当IPアドレスリスト701から現在利用されていないアドレスを一つ選択する。高負荷時など、すべてのアドレスが利用されている場合は、利用中のアドレスからランダムに一つ選択する。

【 0 0 6 4 】

そして、前記担当IPアドレスリスト701から選択した前記IPアドレスを削除して、自装置から当該IPアドレスの割り当てを外し、そのIPアドレスを含むアドレス引継ぎ応答メッセージを送信元装置へ通知する(ステップ1107)。また、監視元サービス提供装置リスト901の引継IPアドレスリストから当該IPアドレスを削除する。

【 0 0 6 5 】

上記から判るように、本明細書において、アドレス引継ぎ要求とは、アドレス割り当て先を、要求送信先のサービス提供装置から、要求送信元のサービス提供装置へ移動することの要求、を意味する。

【 0 0 6 6 】

自装置からアドレス引継ぎ要求メッセージを送信し、送信相手である他のサービス提供装置104からアドレス引継ぎ応答メッセージを受信した場合、担当IPアドレスリスト701に受信したメッセージに含まれるアドレスを登録し、自装置に当該アドレスを割り当てる。そして、自装置に新たにアドレスが割り当てられたことをシステム内の他装置に知らせるため、gratuitous ARPをパケット送受信部から送信する。また、監視元サービス提供装置リスト901内の引継IPアドレスリストを更新し、自装置を監視している監視元サービス提供装置に、新たに登録されたIPアドレス情報を含む装置情報メッセージを送信する(ステップ1108)。

【 0 0 6 7 】

サービス提供装置追加要求メッセージ、または削除要求メッセージを受信した場合、サービス提供装置リスト601の更新を行う。そして図10で示したフローにしたがって監視元と監視先関係の再計算を行い、自装置の監視元・監視先の変化があった場合、更新対象の情報を監視先サービス提供装置リスト801および監視元サービス提供装置リスト901の更新を行う。そして、新たに監視先となったサービス提供装置104に対して装置情報取得要求メッセージを送信する(ステップ1109)。

【 0 0 6 8 】

サービス提供装置の加入要求メッセージを受信した場合、サービス提供装置リスト601

10

20

30

40

50

に送信元サービス提供装置を追加する。そして、加入要求メッセージ送信元のサービス提供装置104には、送信元サービス提供装置を除くすべてのサービス提供装置104の追加要求メッセージを、また他のすべてのサービス提供装置に送信元のサービス提供装置の追加要求メッセージを送信する(ステップ1110)。

【0069】

アドレス追加要求メッセージを受信した場合、担当IPアドレスリスト701に受信したメッセージに含まれるアドレスを追加し、自装置に当該アドレスを割り当てる。また、監視元サービス提供装置リスト901の引継IPアドレスリストに当該アドレスを追加し、変更があった監視元サービス提供装置104に対して、装置情報を含む装置情報メッセージを送信する(ステップ1111)。

10

【0070】

アドレス削除要求メッセージを受信した場合、担当IPアドレスリスト701から該当するアドレスを削除する。そして、監視元サービス提供装置リスト901の引継IPアドレスリストから当該アドレスを削除し、変更があった監視元サービス提供装置104に対して装置情報を含む装置情報メッセージを送信する(ステップ1112)。

【0071】

サービス提供装置104の負荷分散機能部503は、図11に示した手順と並行して、他のサービス提供装置104の稼働情報を監視するための処理を行う。この処理の手順を図12に示す。

【0072】

監視が開始されると、サービス提供装置104は、繰り返し(例えば定期的に)、負荷情報要求メッセージを監視対象である監視先サービス提供装置104に送信する(ステップ1201)。負荷情報メッセージを一定時間内に受信した場合は、監視先のサービス提供装置104は障害が発生していないものと認識する。

20

【0073】

受信までに、監視先サービス提供装置リスト801の負荷調整日時から一定期間が経過している場合は、自装置との負荷差を計算する(ステップ1202)。そして、監視先サービス提供装置の負荷が、自装置の負荷より一定値以上高い場合は、アドレス引継ぎ要求を監視先サービス提供装置104に対して送信する(1203)。負荷の差が一定値未満、あるいは前回の負荷調整から一定期間が経過していない場合は、一定期間の休止ののち再度負荷情報要求の送信を行う。

30

【0074】

一定時間内に負荷情報メッセージを受信できないことが続いた場合、監視先サービス提供装置に障害が発生したと判断して、負荷を分散しあう他のn台のサービス提供装置104に対して、サービス提供装置リスト601から当該サービス提供装置の削除を要求する削除要求メッセージを送信する。

【0075】

そして監視先サービス提供装置リスト801内の当該装置のエントリから、当該サービス提供装置から譲り受ける引継ぎIPアドレスリストを抽出し、担当IPアドレスリスト701に追加する(ステップ1205)。

40

【0076】

次にサービス提供装置リスト601から当該装置を削除し、図10で示した手順で監視先サービス提供装置104を決定する。次に、監視先サービス提供装置104に対して装置情報取得要求メッセージを送信する(ステップ1206)。そして、例えば一定期間休止したのち、またステップ1201からの手順を繰り返す。

【0077】

図13および図14に、ここまで述べた手順で生じる一連のシーケンスの例を示す。

【0078】

まず、初期設定および端末からの要求処理について示す。

【0079】

50

図13のステップ1301では、管理者が各サービス提供装置104に対してサービス提供装置リスト601および担当IPアドレスリスト701の設定を行い、図10のフローにしたがって監視元と監視先の関係を構築している。

【0080】

そして、監視元となる各サービス提供装置104は、監視先のサービス提供装置104に対して装置情報取得要求を送る(ステップ1302)。

【0081】

装置情報取得要求を受け取ったサービス提供装置104は、図11のステップ1104にしたがって、装置情報メッセージを返答する(ステップ1303)。ここまでで、各サービス提供装置104の初期設定が完了する。

10

【0082】

端末101は、サービスを受けるためにDNSサーバ103に対して名前解決要求を送る(ステップ1304)。

【0083】

DNSサーバ103は図4のフローにしたがって、端末101にアドレスを返答する(ステップ1305)。

【0084】

端末101は返答に示されるIPアドレスに対して処理要求を送り(ステップ1306)、当該アドレスを担当とするサービス提供装置104-3が要求に対して返答する(ステップ1307)。

20

【0085】

次に、サービス提供装置104-1と104-2の間で負荷が不均衡となった場合の負荷均等化の処理について示す。

【0086】

サービス提供装置104-1は、図12に示すフローのステップ1201にて、繰り返し、例えば定期的に監視先であるサービス提供装置104-2に負荷情報要求を送信する(ステップ1308)。

【0087】

負荷情報要求を受け取ったサービス提供装置104-2は、図11のステップ1106にしたがって現在の装置の負荷情報を返答する(ステップ1309)。

30

【0088】

ここで、負荷情報を受信したサービス提供装置104-1は、ステップ1202によって負荷差が一定以上であると判定したものとし、ステップ1203にしたがって、アドレス引継ぎ要求をサービス提供装置104-2に送信する(ステップ1310)。

【0089】

アドレス引継ぎ要求を受け取ったサービス提供装置104-2は、ステップ1107に示すように、引継ぎアドレスを選び、アドレス引継ぎ応答をサービス提供装置104-1に返答する(ステップ1311)。

【0090】

アドレス引継ぎ応答を受信したサービス提供装置104-1は、自装置の担当アドレスリスト701の更新を行い、サービス提供装置104-1の監視元であるサービス提供装置104-4に対して装置情報を送信する。サービス提供装置104-4は、ステップ1105にしたがって監視先サービス提供装置リスト801を更新する(ステップ1312)。

40

【0091】

次に、監視先のサービス提供装置104-2に障害が発生した場合について示す。

【0092】

サービス提供装置104-1は、図12にフローにしたがって、定期的に負荷情報要求を監視先サービス提供装置104-2に送信している(ステップ1313)。

【0093】

負荷情報の受信が、あらかじめ定めたR回以上タイムアウトすると、ステップ1204にし

50

たがって監視先装置は障害が発生したと判定する。そして、サービス提供装置リスト601に従って、全てのサービス提供装置104に対して、サービス提供装置104-2を削除する要求を送信する(ステップ1314)。

【0094】

削除要求を受け取ったサービス提供装置104-3および104-4は、ステップ1109にしたがってそれぞれの装置のサービス提供装置リスト601を更新する。そして、改めて図10のフローにしたがって監視元と監視先の関係を構築し直し、監視先装置サービス提供装置リスト801および監視元サービス提供装置リスト901を更新する。

【0095】

ここで、監視先のサービス提供装置104-1は監視先がサービス提供装置104-3に更新されたので、新たに監視先となったサービス提供装置104-3に対して情報取得要求を送信する(ステップ1315)。

10

【0096】

情報取得要求を受け取ったサービス提供装置104-3は、ステップ1104にしたがって装置情報を返答する(ステップ1316)。

【0097】

次に、図14にて、サービス提供装置104-2が障害から復旧してシステムに改めて追加された場合を示す。

【0098】

管理者は、サービス提供装置104-2のサービス提供装置リスト601にサービス提供装置104-2の情報を入れる。サービス提供装置104-2は、サービス提供中のサービス提供装置104を任意に選び(ここではサービス提供装置104-4)、加入要求を送信する(ステップ1402)。

20

【0099】

加入要求を受信したサービス提供装置104-4は、図11のステップ1110にしたがって、送信元のサービス提供装置104-2にサービス提供装置リスト601の内容を含んだ追加要求を送信する。また、送信元以外のサービス提供装置104-1および104-3に、サービス提供装置104-2に追加要求を送信する(ステップ1403)。

【0100】

追加要求を受けとった各サービス提供装置104はステップ1109にしたがってサービス提供装置リスト601を更新し、図10のフローにしたがって監視元と監視先の関係を再構築し、監視先に変更があった場合は監視先サービス提供装置に情報取得要求を送信する(ステップ1404)。

30

【0101】

情報取得要求を受け取ったサービス提供装置104は、ステップ1104にしたがって装置情報を返答する(ステップ1405)。

【0102】

なお、DNSサーバ103自体に障害がおこった場合、何らかの外部サーバがそれを検知し、当該DNSサーバ103を再起動するか、別のDNSサーバ103を起動する必要がある。このとき、DNSサーバがサービス提供装置群の負荷情報や稼働情報などを管理していると、その情報を何らかの形で稼働中に安定なストレージなどに保存し、再起動時に復元しなくてはならない。これは障害耐性を向上させるためにかかるコストが大幅に増えてしまう。これにたいし、本実施例では、DNSサーバ自体は、サービス提供装置104群の動作状況に関係なく不変な情報であるアドレス群管理テーブル201を状態として持ち、かつ、サービス提供装置群の負荷情報や稼働情報を持たないため、DNSサーバ103は常に同一の設定で起動できればよい。このため、DNSサーバ103の障害への対応コストは非常に低く抑えることが可能となる。

40

【0103】

(実施例2)

実施例1では、すべてのサービス提供装置104が対等な関係で管理され、正常なすべてのサービス提供装置104が端末101からの要求を受け入れるという構成の例であった。

50

【0104】

実施例2では、N+M台のサービス提供装置104のうち、N台のサービス提供装置104が端末101の要求を処理し、M台のサービス提供装置104は待機状態にある、という構成での実施例を示す。なお、構成と処理手順には共通な点が多いため、ここではその差分を説明する。

【0105】

図15に、本実施例でのサービス提供装置リスト1501を示す。実施例1のサービス提供装置リスト601と比べ、各サービス提供装置104が稼働状態(ACT)であるか待機状態(SBY)であるかを示すカラムを追加している。

【0106】

このように設定した状態で、稼働状態にあるサービス提供装置104間で実施例1にある監視を実施する。そして、監視先サービス提供装置104に障害が発生した場合、サービス提供装置リスト1501から状態がSBYとなっているサービス提供装置104を一つ選択し、そのサーバの状態をACTに変更の上、すべてのサービス提供装置104にその情報をメッセージで送信する。また、障害がおこったサービス提供装置104の引継IPアドレスをアドレス引継ぎ応答メッセージとして新規サービス提供装置に送信する。これによって、N+M構成においても同様の方法で負荷分散が実現可能である。

【0107】

(実施例3)

実施例1および2では、IPアドレスの引継ぎ時に利用されていないアドレスを選択していた。しかし、一つのセッションが長時間に及ぶ場合などでは、すべてのアドレスが使用中となり安全に引き継ぐことができるアドレスが確保できない場合が生じる可能性がある。

【0108】

本実施例では、DNSサーバ103が返答するアドレス群を複数の下位の群に分け、日時に応じて返答対象となる群を切り替えることで、IPアドレス引継ぎ時に利用中でないIPアドレスを確保できるようにする方法について示す。

【0109】

図16に本実施例でのDNSサーバ103の動作を示す。実施例1、2での相当する手順である図4と比べ、テップ1602が異なっている。

【0110】

ステップ1602では、あらかじめ定められた、現在日時からアドレス群番号への関数を用いてアドレス群を1つ以上選択し、その群の中からアドレスを選択して返答をおこなう。

【0111】

この手順により、ある日時においては配布されないアドレス群が生じる。端末101とサービス提供装置104のセッションの継続が一定時間以上は続かないと仮定すると、アドレスが配布され終わってから一定の時間経過したのちはそのアドレスを利用している端末101が存在する確率は十分低くなる。これによって、アドレス引継ぎ時に、利用していないアドレスを確保することが可能となる。

【0112】

(実施例4)

実施例1,2,3では、DNSサーバ103は一つとして説明した。しかし、本実施例のDNSサーバ103は静的なアドレス群管理テーブル201と日時からアドレス群番号への関数を状態として持ち、負荷の変化や稼働しているサービス提供装置104の状況など、動的に変化する情報をもたない。

【0113】

このため、複数のDNSサーバ103をシステム内におき、端末側でそれらのDNSサーバを適宜切り替えて利用する場合であってもDNSサーバ103間での状態同期は不要である。図17に実施例4の全体構成図を示す。この構成により、DNSサーバ103の性能スケーラビリティと障害耐性を向上させることが可能である。

【符号の説明】

【0114】

10

20

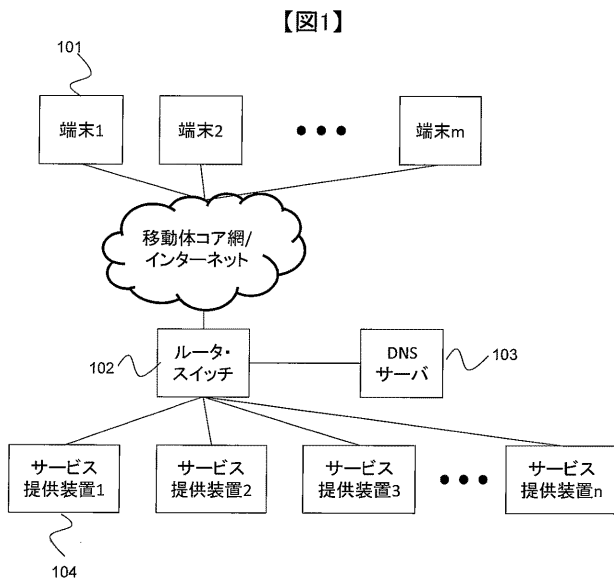
30

40

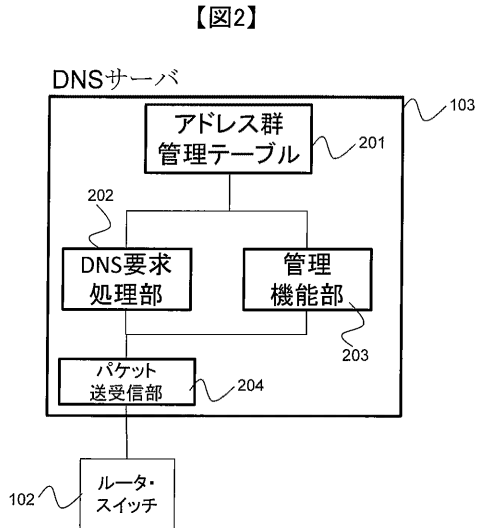
50

101：端末、102：ルータ・スイッチ、103：DNSサーバ、104：サービス提供装置、201：アドレス群管理テーブル、202：DNS要求処理部、203：管理機能部、204：パケット送受信部、501：端末要求処理部、502：パケット送受信部、503：負荷分散機能部、601：サービス提供装置リスト、701：担当IPアドレスリスト、801：監視先サービス提供装置リスト、901：監視元サービス提供装置リスト、1501：サービス提供装置リスト。

【図1】



【図2】



【 図 3 】

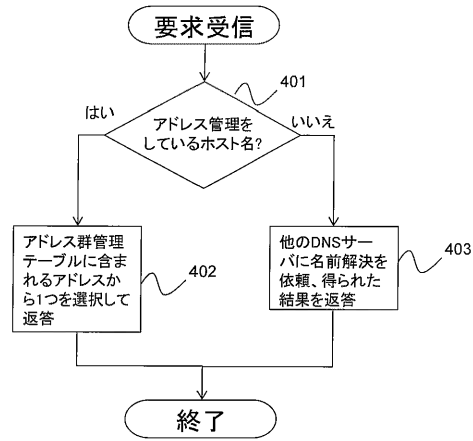
【図3】

アドレス群管理テーブル

www.example.com	
#	アドレス群
1	2001:db8::0000:0000/96
2	2001:db8::0001:0000/96
...	
15	2001:db8::000f:0000/96

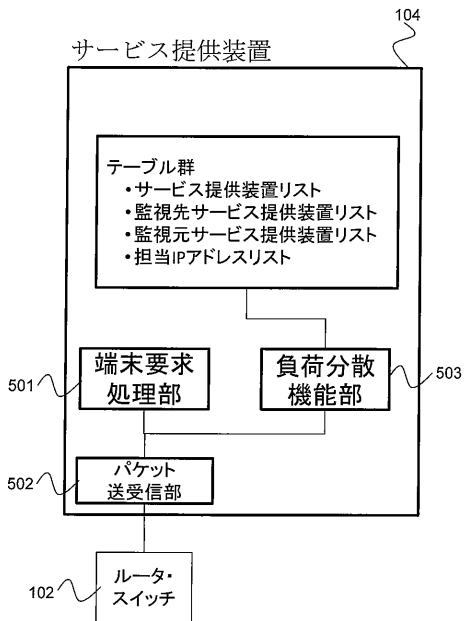
【 図 4 】

【図4】



【 図 5 】

【図5】



【 図 6 】

【図6】

サービス提供装置リスト

#	サービス提供装置識別子
1	serv1
2	serv2
...	
n	servn

【 図 7 】

【図7】

担当IPアドレスリスト

#	IPアドレス
1	2001:db8::0000:0001
2	2001:db8::0000:0005
...	
m	2001:db8::000f:ffff

【図8】

【図8】

監視先サービス提供装置リスト

#	サービス提供装置識別子	引継IPリスト	負荷調整日時
1	serv2	2001:db8::0000:0002, 2001:db8::0000:0006, ...	2011/1/1 10:00

【図9】

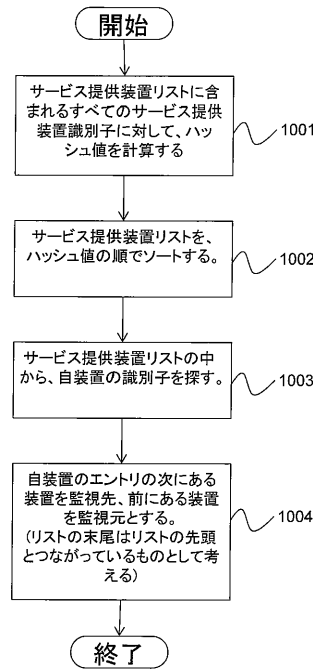
【図9】

監視元サービス提供装置リスト

#	サービス提供装置識別子	引継IPリスト
1	serv4	2001:db8::0000:0004, 2001:db8::0000:0008, ...
2	serv2	2001:db8::0000:0014, 2001:db8::0000:0018, ...

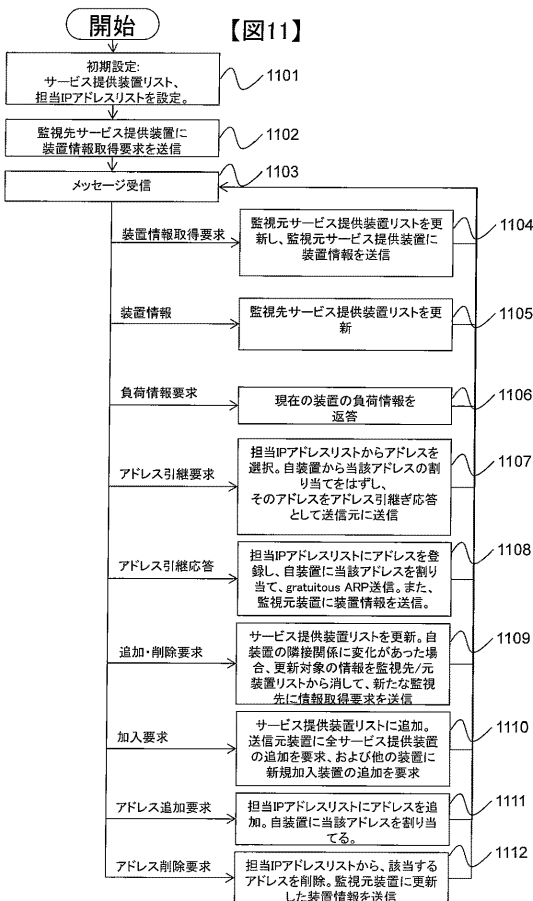
【図10】

【図10】



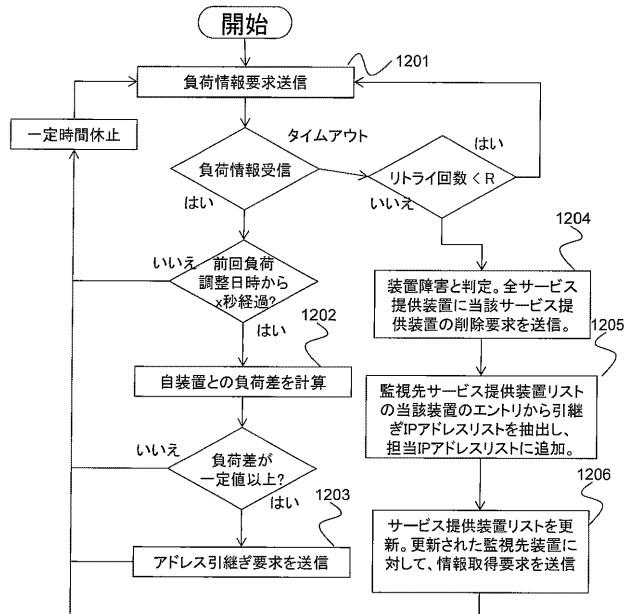
【図11】

【図11】

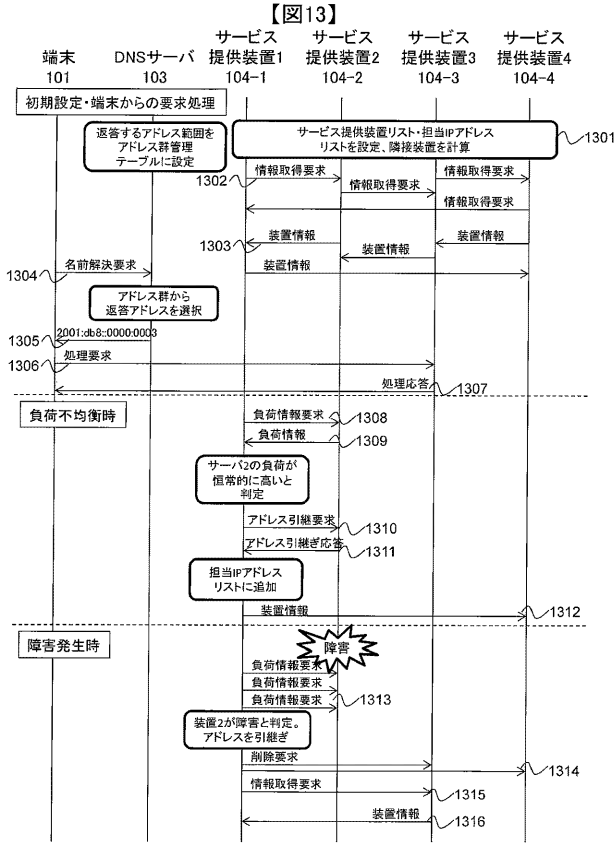


【図12】

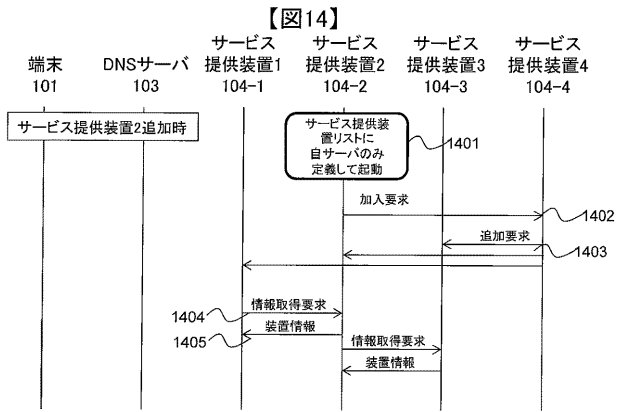
【図12】



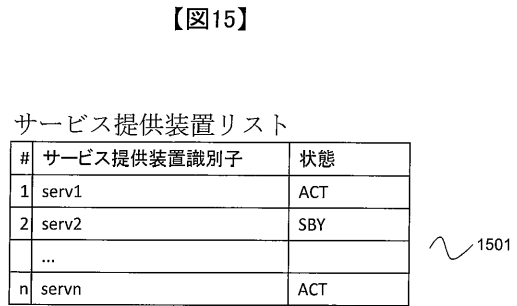
【図13】



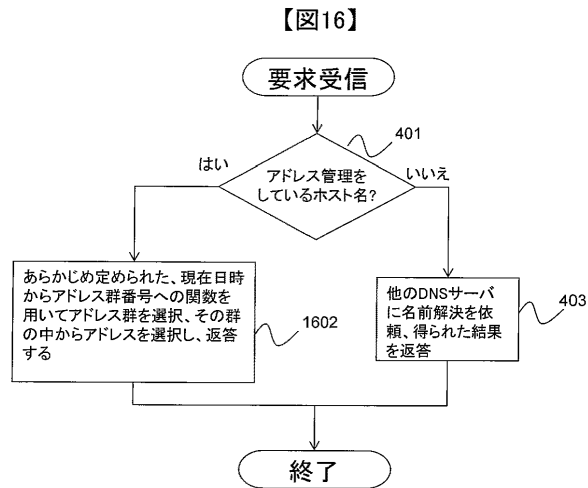
【図14】



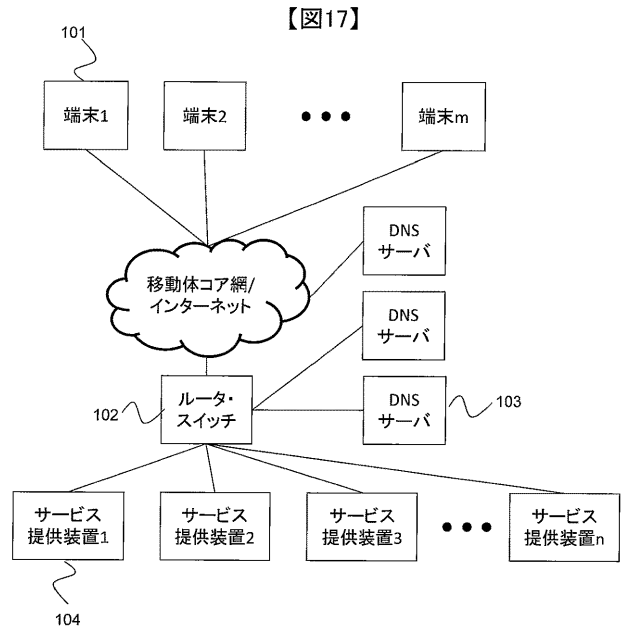
【図15】



【図16】

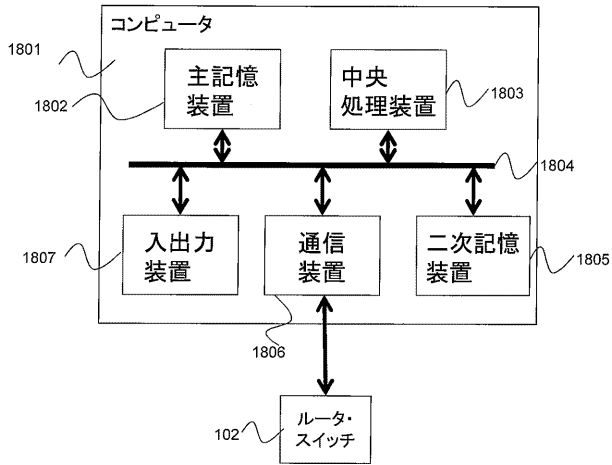


【図17】



【図18】

【図18】



フロントページの続き

(72)発明者 原口 直規

神奈川県川崎市幸区鹿島田 8 9 0 番地 株式会社日立製作所ネットワークソリューション事業部内
Fターム(参考) 5K030 GA19 HA08 LE03 MB02 MD07