

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5871908号
(P5871908)

(45) 発行日 平成28年3月1日(2016.3.1)

(24) 登録日 平成28年1月22日(2016.1.22)

(51) Int.Cl.

F 1

H04L 12/70 (2013.01)
H04L 12/721 (2013.01)H04L 12/70
H04L 12/721100 Z
Z

請求項の数 15 (全 17 頁)

(21) 出願番号 特願2013-508398 (P2013-508398)
 (86) (22) 出願日 平成23年5月5日 (2011.5.5)
 (65) 公表番号 特表2013-529013 (P2013-529013A)
 (43) 公表日 平成25年7月11日 (2013.7.11)
 (86) 國際出願番号 PCT/EP2011/002238
 (87) 國際公開番号 WO2011/138033
 (87) 國際公開日 平成23年11月10日 (2011.11.10)
 審査請求日 平成26年2月7日 (2014.2.7)
 (31) 優先権主張番号 10004798.4
 (32) 優先日 平成22年5月6日 (2010.5.6)
 (33) 優先権主張国 歐州特許庁 (EP)

(73) 特許権者 597149146
 ドイツチェ テレコム アーゲー
 ドイツ国. 53113 ボン, フリードリ
 ヒ - エベルト - アレエ 140
 (74) 代理人 100094112
 弁理士 岡部 譲
 (74) 代理人 100106183
 弁理士 吉澤 弘司
 (74) 代理人 100128657
 弁理士 三山 勝巳
 (74) 代理人 100160967
 弁理士 ▲濱▼口 岳久
 (74) 代理人 100170601
 弁理士 川崎 孝

最終頁に続く

(54) 【発明の名称】ネットワーク内部のデータ通信を制御するための方法およびシステム

(57) 【特許請求の範囲】

【請求項 1】

データ通信の目的で相互接続された複数のネットワーク要素を有するネットワーク内部のデータ通信を制御するための方法であって、

- ネットワーク・プロバイダに関する中央ネットワーク要素(100)が、ネットワーク情報(110)を取り出すことによって、および/または前記ネットワーク内部のデータ通信をモニタリングすることによって、前記ネットワークの表現(130)を生成し、前記ネットワークの前記表現は前記ネットワークのノードに関連する情報および/または前記ネットワークのノードのペア間の通信経路に関連する情報を含み、

- 前記ネットワーク内の変更が、ネットワーク情報(110)を連続的に取り出すことによって、および/または前記ネットワーク内部のデータ通信をモニタリングすることによって検出され、前記ネットワークの前記表現(130)が前記検出された変更に応じて更新され、

- クライアント・ネットワーク要素(180)から、要求が前記中央ネットワーク要素(100)に送信され、前記要求が、送信元ネットワーク要素および少なくとも2つの宛先ネットワーク要素を識別する情報を含み、

- 各識別された宛先ネットワーク要素に対して、前記送信元ネットワーク要素と前記宛先ネットワーク要素のそれぞれとの間の通信経路のランキング値が、前記ネットワークの前記表現(130)に基づいて前記中央ネットワーク要素(100)によって決定され、

10

20

- 前記識別された宛先ネットワーク要素の順序付けられたリストが、前記それぞれのランク値に基づいて生成され、
- 前記順序付けられたリストが、前記中央ネットワーク要素（100）から前記クライアント・ネットワーク要素（180）に送信され、
- 前記順序付けられたリストに基づいて、前記宛先ネットワーク要素のうちの少なくとも1つが、前記送信元ネットワーク要素とのデータ通信に選択される、方法。

【請求項2】

前記クライアント・ネットワーク要素が、前記送信元ネットワーク要素から要求を受信したことに応答して前記中央ネットワーク要素（100）に送信されるべき要求を生成し、前記クライアント・ネットワーク要素が、前記中央ネットワーク要素（100）によって提供された前記順序付けられたリストに応じて前記送信元ネットワーク要素に応答を送信する、請求項1に記載の方法。10

【請求項3】

前記クライアント・ネットワーク要素（180）によって送信された前記要求が、事前に定められた統一フォーマットの要求に変換される、請求項1または2のいずれか1項に記載の方法。

【請求項4】

許可されたクライアント・ネットワーク要素から受信され、許可された送信元ネットワーク要素を識別する要求のみが、前記中央ネットワーク要素（100）によって処理され、前記クライアント・ネットワーク要素および／または前記送信元ネットワーク要素の許可が、前記クライアント・ネットワーク要素の識別情報および／または前記要求に含まれる前記送信元ネットワーク要素の識別情報によりチェックされる、請求項1乃至3のいずれか1項に記載の方法。20

【請求項5】

前記ランク値が、前記クライアント・ネットワーク要素によって送信された前記要求に含まれる情報に基づいて事前に定められた1組のランク付け機能から選択された事前に定められたランク付け機能によって決定される、請求項1乃至4のいずれか1項に記載の方法。

【請求項6】

少なくとも1つの事前に定められたランク付け機能が、前記ネットワークの検出された変更に応じて動的に適合される、請求項5に記載の方法。30

【請求項7】

複数の要求が解析され、要求が個々のクライアント・ネットワーク要素から受信される頻度ならびに／または個々の送信元ネットワーク要素および／もしくは個々の宛先ネットワーク要素が前記要求のそれぞれで識別される頻度に関する統計情報が決定され、前記統計情報が、前記ランク値を決定するために使用される、請求項1乃至6のいずれか1項に記載の方法。

【請求項8】

データ通信の目的で相互接続された複数のネットワーク要素を有するネットワーク内部のデータ通信を制御するためのシステムであって、前記システムがネットワーク・プロバイダに関連する少なくとも1つの中央ネットワーク要素（100）内に配置され、40

- ネットワーク情報（110）を取り出すおよび／または前記ネットワーク内部でのデータ通信をモニタリングするための取り出しサブシステム（112）と、
- 前記取り出しサブシステムから受け取ったデータに基づいて前記ネットワークの表現を生成および更新するためのネットワーク・マップ・ジェネレータサブシステム（120）であって、前記ネットワークの前記表現は前記ネットワークのノードに関連する情報および／または前記ネットワークのノードのペア間の通信経路に関連する情報を含む、ネットワーク・マップ・ジェネレータサブシステム（120）と、
- 前記ネットワークの前記表現（130）を保存するためのネットワーク・マップ・データベースと、50

- クライアント・ネットワーク要素(180)からの要求を受信する、処理する、およびそれに応答するためのクエリ・マネージャ・サブシステム(140)であって、前記要求が、送信元ネットワーク要素および少なくとも2つの宛先ネットワーク要素を識別する情報を含む、クエリ・マネージャ・サブシステム(140)と、

- 前記ネットワークの前記保存された表現(130)に基づいて、クライアント・ネットワーク要素(180)からの要求で識別された各宛先ネットワーク要素に対して前記送信元ネットワーク要素(180)と前記宛先ネットワーク要素のそれぞれとの間の通信経路のランキング値を決定するように適合された経路ランク付けサブシステム(150)と

を備え、

10

- 前記クエリ・マネージャ・サブシステム(140)が、要求に対する応答を生成し、前記要求が受信された前記クライアント・ネットワーク要素(180)に前記応答を送信するように適合され、前記応答が、前記要求で識別された前記宛先ネットワーク要素の順序付けられたリストを含み、前記リストが、前記経路ランク付けサブシステム(150)によって提供されるそれぞれのランキング値に基づいて決定される、システム。

【請求項9】

前記クライアント・ネットワーク要素(180)から受信された要求を事前に定められた統一フォーマットの要求に変換するための要求トランスレータ・サブシステム(170)をさらに備える、請求項8に記載のシステム。

【請求項10】

20

前記経路ランク付けサブシステム(150)が、前記クライアント・ネットワーク要素(180)によって送信された前記要求に含まれる情報に基づいて事前に定められた1組のランク付け機能からの前記経路ランク付けサブシステム(150)によって選択された事前に定められたランク付け機能によって前記ランキング値を決定するように適合された、請求項8または9のいずれか1項に記載のシステム。

【請求項11】

少なくとも1つの事前に定められたランク付け機能が、前記ネットワークの検出された変更に応じて動的に適合される、請求項10に記載のシステム。

【請求項12】

30

複数の要求を解析し、要求が個々のクライアント・ネットワーク要素から受信される頻度ならびに／または個々の送信元ネットワーク要素および／もしくは個々の宛先ネットワーク要素が前記要求のそれぞれで識別される頻度に関する統計情報を決定するように適合された頻繁ヒッター検出サブシステムをさらに備え、前記統計情報が、前記ランキング値を決定するための前記経路ランク付けサブシステム(150)に提供される、請求項8乃至11のいずれか1項に記載のシステム。

【請求項13】

前記ネットワークの前記表現(130)のバックアップを事前に定められた時刻に行い、前回のバックアップ以降の前記ネットワーク内の検出された変更を保存し、障害が発生した場合に前記前回のバックアップおよび前記保存された変更に基づいて前記ネットワークの前記表現(130)を取り出すように適合されたバックアップ・サブシステムを備える、請求項8乃至12のいずれか1項に記載のシステム。

40

【請求項14】

データ通信を制御するためのシステムと通信するための、請求項8乃至13のいずれか1項に記載のネットワーク要素であって、

- 要求を生成し、前記要求が、送信元ネットワーク要素および少なくとも2つの宛先ネットワーク要素を識別する情報ならびに前記システムが1組の事前に定められたランク付け機能からランク付け機能を選択することを可能にするための情報を含み、

- 前記要求を前記システムに送信し、

- 前記システムから受信した応答を処理する

ように適合されたネットワーク要素。

50

【請求項 15】

- 前記送信元ネットワーク要素から受信した要求に応答して前記要求を生成して前記システムに送信し、
 - 前記システムから受信した前記応答に応じて前記送信元ネットワーク要素に応答を送信する

ように適合された、請求項 14 に記載のネットワーク要素。

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、一般に通信ネットワークに関し、特にネットワーク内部のデータ通信を制御するための方法およびシステムに関し、より詳細には、ネットワーク内の送信元・宛先ペアの選択に関する具体的な推奨事項 (recommendation) が提供される。 10

【背景技術】**【0002】**

現代の通信インフラストラクチャでは、IPベースのコンピュータ・ネットワークが主要な役割を果たしている。これらのネットワークの展開は、企業、公共企業体、および個人などの異なる種類の参加者が高度で複雑なサービスおよび通信システムに依存しているので、指数関数的速度で進んでいる。

【0003】

現在、サーバまたはピアの選択方法は、主に、たとえば ping、trace-route、または利用可能な帯域幅推定ツールを使用することによってユーザにより開始される能動的測定 (active measurement) に基づいているが、能動的測定は、不正確である、かつ/またはエンド・ユーザに負荷を与えることが多い。さらなる選択方法は、地理情報を利用すること、仮想協調システムの作成に利用可能なランドマーク (landmark) を利用すること、ランダムな送信元選択、負荷分散に基づいた送信元選択、または経済的利益もしくは請負契約に基づいた送信元選択に基づく、コンテンツ・プロバイダまたはコンテンツ配信業者による推奨事項およびリダイレクションに従うことを含む。上述のすべての場合において、情報はネットワーク・プロバイダによってサポートされず、上記のスキームが成功するには、他のピア、サービス・プロバイダ、およびネットワークの正確な図を持たずユーザの目的と矛盾する目的を有することがある他のサードパーティによる、既に利用可能なインフラストラクチャに依存することが必要である。 20

【0004】

インターネット・サービス・プロバイダ (ISP) などのネットワーク・プロバイダは、トラヒック・エンジニアリング (traffic engineering) に関する目標の達成ならびにユーザ・エクスペリエンスおよびアプリケーション効率の向上に関心を抱いている。このような目標を達成するための既知の例示的な技法が負荷分散である。この技法では、作業負荷は、ネットワーク内の 2 つ以上のコンピュータまたはネットワーク・リンクにわたって分散し、それによって、たとえばリソース利用率、スループット、または応答時間を向上させる。負荷分散は、一般に、人気のあるウェブ・サイト、インターネット・リレー・チャット・ネットワーク、高帯域幅ファイル転送プロトコル・サイト、および DNS サーバ (DNS: ドメイン・ネーム・システム) に用いられ、典型的には、クライアントからの要求を複数のバックエンド・サーバに転送するロード・バランサが利用される。 40

【0005】

複数の IP アドレスが单一のドメイン名と関連付けられることも一般的であり、DNS 要求への応答は、同一のサービスをホストする数台のサーバの IP アドレスのリストを含む。この点における負荷分散の方法がラウンド・ロビン DNS であり、リスト内の IP アドレス・シーケンスは各 DNS 応答により変えられる。しかし、ラウンド・ロビン DNS には、DNS サーバにクエリが行われるたびにアドレスの順序を交替させるだけにすぎな 50

いという欠点があり、したがって、サーバ間で基本的に均一な負荷分散が達成されるが、最適でないことが多い。

【発明の概要】

【発明が解決しようとする課題】

【0006】

したがって、本発明の目的は、通信ネットワーク特にコンピュータ・ネットワーク内部のデータ通信を制御するための新しい改良された方法を示すことである。本発明のさらなる目的は、負荷の重いリンクにおける輻輳の減少、ダウンしているリンクの回避、または輸送コストの削減など、ネットワーク・プロバイダがネットワーク内のトラヒック・エンジニアリングを実行できる新しい改良された方法を示すことである。

10

【課題を解決するための手段】

【0007】

本発明の目的の解決策は、それぞれの添付の独立請求項の主題のそれぞれによって達成される。有利なおよび／または好ましい実施形態または改良点は、それぞれの添付の従属請求項の主題である。

【0008】

したがって、ネットワーク内部のデータ通信を制御するための本発明の方法は、ネットワーク・プロバイダに関連する中央ネットワーク要素が、ネットワーク情報を取り出す(retrieve)ことによって、および／またはネットワーク内部のデータ通信をモニタリングすることによって、ネットワークの表現を生成するステップを含む。ネットワーク内の変更は、ネットワーク情報を連続的に取り出すことによって、および／またはネットワーク内部のデータ通信をモニタリングすることによって検出され、したがって、ネットワークの表現(representation)は、検出された変更に応じて更新される。ネットワークは、データ通信の目的で相互接続された複数のネットワーク要素を有する任意のネットワーク特にインターネットであつてよい。中央ネットワーク要素は、典型的には、ネットワーク・プロバイダによって運用されるサーバまたはサーバのグループである。ネットワークの表現は、好ましくは、ネットワークのノードに関連する情報および／またはネットワークのノードのペア間の通信経路に関連する情報を含む。好ましくは、通信経路の性能特性が計算され、表現が高速アクセスのために保守される。有利には、中央ネットワーク要素で収集されるネットワーク情報は、生のネットワーク情報、たとえば物理的情報、モニタリング情報、ポリシー情報、および／またはメタ情報を含む。有利には、この情報が処理され、詳細な注釈付き(annotation)ネットワーク表現が構築され、保守される。

20

【0009】

この方法は、クライアント・ネットワーク要素から中央ネットワーク要素に要求を送信するステップをさらに含み、前記要求が、送信元ネットワーク要素および少なくとも2つの宛先ネットワーク要素を識別する情報を含む。各識別された宛先ネットワーク要素に対して、送信元ネットワーク要素とそれぞれの宛先ネットワーク要素との間の通信経路のランキング値は、ネットワークの表現に基づいて中央ネットワーク要素によって決定され、識別された宛先ネットワーク要素の順序付けられたリストが、それぞれのランディング値に基づいて生成され、前記順序付けられたリストは、中央ネットワーク要素からクライアント・ネットワーク要素に送信され、順序付けられたリストに基づいて、宛先ネットワーク要素のうちの少なくとも1つが、送信元ネットワーク要素とのデータ通信に選択される。

30

【0010】

クライアント・ネットワーク要素、送信元ネットワーク要素、および宛先ネットワーク要素という用語については、以下では簡単に、クライアント、送信元、および宛先という用語もそれぞれ使用される。用語は、本質的に交換可能なように使用され、それぞれのハードウェア・ユニットまたはソフトウェア・ユニットまたはその識別情報を定義する。ユーザという用語は、以下では、典型的にはクライアント・ネットワーク要素を指す。送信元および宛先という用語は、ネットワークにおけるデータ通信の2つのエンドポイントを

40

50

定義するために使用され、データ通信の方向は一方向または他の方向に限定されない。しかし、典型的には、データ通信は送信元によって開始される。

【0011】

本発明の基本概念は、ネットワーク・プロバイダ内部でのネットワーク情報の収集、処理、および保守であり、保守される情報は、送信元 - 宛先ペアのランキング値を推定するために使用され、したがって、少なくとも 2 つの宛先候補を含む要求は、ランク付けされたリストの形をとる推奨事項により応答され、ネットワークのトラヒック・エンジニアリング、ユーザ・エクスペリエンス、およびアプリケーション効率を向上させる。ランギング値は、好ましくは、それぞれの送信元 - 宛先ペアの近接度の尺度とすることができます。送信元 - 宛先ペアの送信元および / または宛先は、ネットワーク・プロバイダの管理権限に属しても属さなくてもよい。ランギング値は、有利には、遅延、帯域幅、エラー送出、ネットワーク信頼性、輻輳などのネットワーク性能特性、およびアプリケーションの種類、価格設定、相互接続契約 (peering agreement)、地理的ロケーション、法的問題、地理的適用範囲、クライアントによって発動される制限、サービス負荷利用率などの他の特性に応じて計算される。それに基づいてランギング値が計算されるパラメータは、上述のパラメータに限定されず、他の任意の適切なパラメータを含むこともできる。10

【0012】

送信元および宛先は、IP アドレスなど、ネットワーク・プロトコルで使用される識別子であってよく、IP アドレスは、たとえばネットワーク・アドレス変換 (NAT) が用いられているとき、またはルータ IP のみが分かっているとき、実 IP、サブネットワーク、またはアクセス IP とすることができます。ユーザは、たとえばコンテンツ配信システム (CDN)、ピア・ツー・ピア・システム、ストリーミング・システム、キャッシュのグループ、ならびにサーバ・プール、直接ダウンロード・プロバイダ (direct download provider)、ドメイン・ネーム・サーバ (DNS) などのインターネット・インフラストラクチャを含む、クライアント - サーバ・モデルに基づくアプリケーション、ならびに他のネットワーク・プロバイダによって許可されたサーバである。20

【0013】

方法の好ましい一実施形態では、クライアント・ネットワーク要素は、送信元ネットワーク要素から要求を受信したことに応答して中央ネットワーク要素に送信されるべき要求を生成し、クライアント・ネットワーク要素は、中央ネットワーク要素によって提供された順序付けられたリストに応じて送信元ネットワーク要素に応答を送信する。これは、たとえば、クライアント・ネットワーク要素が DNS サーバであるときに当てはまる。このようなクライアント・ネットワーク要素からの要求は、以下では、プロキシ要求とも呼ばれる。この実施形態の特別な利点は、送信元ネットワーク要素が中央ネットワーク要素で実行された順序変更に気付き、その結果、たとえば、本発明を実施しなければ受信するであろう IP アドレスと異なるウェブ・サイトの IP アドレスが送信元ネットワーク要素によって受信されるので、送信元ネットワーク要素の変更が必要でないことである。30

【0014】

あるいは、クライアント・ネットワーク要素は、同時に送信元ネットワーク要素であってもよく、その結果、要求が最初にクライアント・ネットワーク要素によって生成される。

【0015】

要求が、異なるサービスに関連する異なるタイプのネットワーク要素によって送信できるので、要求は、典型的には、異なるフォーマットを有することができる。したがって、クライアント・ネットワーク要素によって送信された要求は、好ましくは、事前に定められた統一フォーマットの要求に変換される。統一フォーマットの要求は、典型的には、少なくとも送信元ネットワーク要素の識別情報と宛先ネットワーク要素すなわち宛先候補の識別情報とを含む。4050

【0016】

最も好ましい実施形態では、ランキング値は、クライアント・ネットワーク要素によって送信された要求に含まれる情報に基づいて事前に定められた1組のランク付け機能から選択された事前に定められたランク付け機能によって決定される。

【0017】

この情報は、たとえばコード値とることができ、異なるコード値は、有利には、トランジット・エンジニアリングによる最適化の異なる変形と関連付けられ、この変形は、送信元ネットワーク要素によって、またはクライアント・ネットワーク要素によって選択される。要求を変換するとき、コード値は当然のことながら、統一フォーマットの要求にも取り入れられる。コード値は、関連するランク付け機能にマッピングされ、次いで、中央ネットワーク要素に保存されたマッピング・テーブルに応じた宛先ネットワーク要素の順序付けられたリストの決定に使用され、マッピングは、さらなるパラメータに応じて実行されることができる。このように、一定のランク付け機能は、有利には、ネットワーク・プロバイダにのみ知られている。10

【0018】

1組のランク付け機能およびそれらに関連するコード値は好ましくは、事前に定められ、中央ネットワーク要素に保存される。柔軟性をさらに向上させるために、少なくとも1つの事前に定められたランク付け機能は、ネットワークの検出された変更に応じて動的に適合されてもよい。

【0019】

ランク付け機能の非常に単純な例は、送信元ネットワーク要素とそれぞれの宛先ネットワーク要素との間の通信経路のホップカウント率および最小帯域幅である。しかし、好ましくは、事前に定められたランク付け機能は、さまざまなパラメータに応じて大きく変化することができる。しかし、特に有利には、ランク付け機能は、ネットワーク・プロバイダのみによってアクセス可能な情報に基づいて定められる。20

【0020】

好ましくは、複数の要求が解析され、要求が個々のクライアント・ネットワーク要素から受信される頻度ならびに／または個々の送信元ネットワーク要素および／もしくは個々の宛先ネットワーク要素がそれぞれの要求で識別される頻度に関する統計情報が決定され、この統計情報は、ランキング値を決定するために使用される。30

【0021】

さらに、方法は、好ましくは承認制御を含み、承認制御は、たとえば少なくとも1つのアクセス制御リストによって実現されることができる。この好ましい実施形態では、許可されたクライアント・ネットワーク要素から受信され、許可された送信元ネットワーク要素を識別する要求のみが、中央ネットワーク要素によって処理され、クライアント・ネットワーク要素および／または送信元ネットワーク要素の許可が、クライアント・ネットワーク要素の識別情報および／または要求に含まれる送信元ネットワーク要素の識別情報によりチェックされる。

【0022】

アクセス制御リストは、ポジティブ・リストとして、またはネガティブ・リストとして提供されることができる。たとえば、その要求が受け入れられるサブネットを含む第1のポジティブ・リストが提供されることがある。第2のポジティブ・リストは、たとえば、そのプロキシ要求が受け入れられるIPアドレスのリストを作成することができる。プロキシ要求は、たとえば、送信元ネットワーク要素としてのDNSサーバによって送信される。上述の統計情報に、短期間に多数の要求が受信されたことが示される場合、それぞれの送信元ネットワーク要素および／またはクライアント・ネットワーク要素のIPアドレスがネガティブ・リストに取り入れられてもよく、その結果、このような要求はブロックされる。40

【0023】

方法の好ましい一実施形態では、ネットワークの表現のバックアップが事前に定められ50

た時刻に作成され、前回のバックアップ以降のネットワーク内の検出された変更が保存され、障害が発生した場合に前回のバックアップおよび保存された変更に基づいてネットワークの表現が取り出される。

【0024】

データ通信の目的で相互接続された複数のネットワーク要素を有するネットワーク内部のデータ通信を制御するための本発明のシステムは、ネットワーク・プロバイダに関連する少なくとも1つの中央ネットワーク要素内に配置される。この方法は、ネットワーク情報を取り出すおよび／またはネットワーク内部でのデータ通信をモニタリングするための取り出しサブシステムと、この取り出しサブシステムから受け取ったデータに基づいてネットワークの表現を生成および更新するためのネットワーク・マップ・ジェネレータサブシステムと、ネットワークの表現を保存するためのネットワーク・マップ・データベースと、クライアント・ネットワーク要素からの要求を受信する、処理する、およびそれに応答するためのクエリ・マネージャ・サブシステムであって、前記要求が、送信元ネットワーク要素および少なくとも2つの宛先ネットワーク要素を識別する情報を含む、クエリ・マネージャ・サブシステムと、ネットワークの保存された表現に基づいて、クライアント・ネットワーク要素からの要求で識別された各宛先ネットワーク要素に対して送信元ネットワーク要素とそれぞれの宛先ネットワーク要素との間の通信経路のランキング値を決定するように適合された経路ランク付けサブシステムとを備え、クエリ・マネージャ・サブシステムは、要求に対する応答を生成し、要求が受信されたクライアント・ネットワーク要素に応答を送信するように適合され、前記応答は、要求で識別された前記宛先ネットワーク要素の順序付けられたリストを含み、前記リストは、経路ランク付けサブシステムによって提供されるそれぞれのランキング値に基づいて決定される。10

【0025】

サービスのモードでは、本発明のシステムは、好ましくは、生のネットワーク情報を収集および処理し、ネットワークの注釈付きマップを構築および保守し、ネットワーク内のマッピングまたは割り当て可能な任意の送信元・宛先ペアの評価に使用される経路特性を事前に推定し、ネットワークの前述の保守された図への高速アクセスのためのインターフェースを提供する。サービスのモードでは、ユーザは要求を送出でき、システムは、好ましくは、要求を承認または拒否する。要求が承認される場合、要求はさらに処理され、保守されたネットワーク図からの情報が取り出され、送信元と宛先の両方の集約された統計情報が計算され、あらゆる送信元・宛先ペア候補が評価される。この評価は、ランク付けされたリストとして、たとえば降順で、ユーザに送信される。20

【0026】

システムのユーザによる要求が、宛先候補たとえばコンテンツをダウンロードまたはアップロードするサーバまたはピアのリストと共に到着すると、システムは、ネットワーク・マップ・データベースに保存されたネットワークの表現を利用して送信元・宛先経路特性を推定し、事前に定められたまたは動的に定められたランク付け機能に基づいて、宛先の順序付けられたリストにより要求に応答して、トラヒック・エンジニアリングに関する目標を達成し、ユーザ・エクスペリエンスおよびアプリケーション効率を向上させる。30

【0027】

したがって、経路ランク付けサブシステムは、好ましくは、クライアント・ネットワーク要素によって送信された要求に含まれる情報に基づいて事前に定められた1組のランク付け機能からの経路ランク付けサブシステムによって選択された事前に定められたランク付け機能によってランキング値を決定するように適合され、有利には、少なくとも1つの事前に定められたランク付け機能は、ネットワークの検出された変更に応じて動的に適合される。40

【0028】

好ましい実施形態では、システムは、クライアント・ネットワーク要素から受信された要求を事前に定められた統一フォーマットの要求に変換するための要求トランスレータ・サブシステムをさらに備える。50

【0029】

さらに好ましい実施形態では、システムは、複数の要求を解析し、要求が個々のクライアント・ネットワーク要素から受信される頻度ならびに / または個々の送信元ネットワーク要素および / もしくは個々の宛先ネットワーク要素がそれぞれの要求で識別される頻度に関する統計情報を決定するように適合された頻繁ヒッター (frequent hit tester) 検出サブシステムを備え、この統計情報は、ランキング値を決定するための経路ランク付けサブシステムに提供される。

【0030】

有利には、ネットワークの表現のバックアップを事前に定められた時刻に行い、前回のバックアップ以降のネットワーク内の検出された変更を保存し、障害が発生した場合に前回のバックアップおよび保存された変更に基づいてネットワークの表現を取り出すように適合されたバックアップ・サブシステムも提供される。 10

【0031】

前述のようにシステムと通信するための本発明のネットワーク要素は、要求を生成し、前記要求が、送信元ネットワーク要素および少なくとも 2 つの宛先ネットワーク要素を識別する情報ならびに前記システムが 1 組の事前に定められたランク付け機能からランク付け機能を選択することを可能にするための情報を含み、前記要求をシステムに送信し、システムから受信した応答を処理するように適合される。好ましくは、ネットワーク要素は、送信元ネットワーク要素から受信した要求に応答して要求を生成してシステムに送信し、システムから受信した応答に応じて送信元ネットワーク要素に応答を送信するように適合される。 20

【0032】

本発明のシステムの機能の少なくとも一部は、好ましくは、ソフトウェア構成要素によって提供されることができる。したがって、少なくとも 1 つのコンピュータで実行されるときに上述の方法を実行するように適合された電子的に読み取り可能な制御命令を含むデジタル記憶媒体も本発明の範囲に含まれる。

【図面の簡単な説明】**【0033】**

【図 1】本発明のシステムの好ましい一実施形態の構成要素の概略図である。

【図 2】図 1 に示されるシステムの取り出しサブシステムの概略図である。 30

【図 3】図 1 に示されるシステムのネットワーク・マップ・ジェネレータサブシステムの機能の概略流れ図である。

【図 4】図 1 に示されるシステムの要求トランスレータ・サブシステムの機能の概略流れ図である。

【図 5】図 1 に示されるシステムのクエリ・マネージャ・サブシステムの機能の概略流れ図である。

【図 6】図 1 に示されるシステムのバックアップ・サブシステムの機能の概略流れ図である。

【図 7 a】図 1 に示されるシステムを用いなくても達成される、5 つのノードからなるネットワークのためのトラヒック・エンジニアリングおよびトラヒック制御の一例である。 40

【図 7 b】図 1 に示されるシステムを用いる場合に達成される、5 つのノードからなるネットワークのためのトラヒック・エンジニアリングおよびトラヒック制御の一例である。

【図 8 a】図 1 に示されるシステムを用いた場合と用いない場合に達成される例示的な平均ダウンロード速度を示すグラフである。

【図 8 b】図 1 に示されるシステムを用いた場合と用いない場合に達成される例示的な平均ホップカウントを示すグラフである。

【発明を実施するための形態】**【0034】**

以下では、本発明の好ましいが例示的な実施形態について、図を参照してより詳細に説明する。 50

【0035】

図1は、本発明のシステム100の好ましい一実施形態の構成要素の構成の概略図を示し、前記構成要素は、システムのデータ収集用サブシステム、データ処理用サブシステム、ネットワーク・マップの構築および保守用サブシステム、ならびにクエリ処理サブシステム、ランク付けサブシステム、および応答サブシステムを含む。情報取り出しサブシステム112は、物理的情報、モニタリング情報、ポリシー情報、およびメタ情報を含みうる生のネットワーク情報110をネットワーク・プロバイダにおいて収集する役割を果たす。次に、情報取り出しサブシステム112は、この情報をネットワーク・マップ・ジェネレータ120に送信する。ネットワーク・マップ・ジェネレータ120は、ネットワークで発生した変更のレベルを評価し、それに応じて注釈付きネットワーク・マップ・データベース130を構築および更新する役割を果たす。注釈付きネットワーク・マップ・データベース・サブシステム130は、ネットワークの最も正確な図を保守し、ネットワーク内のマップできるすべての使用可能な送信元および宛先で測定される種々の特性を事前に推定し、処理の結果を保存し、クエリ・マネージャ・サブシステム140とのインターフェースを利用する可能にする役割を果たす。障害復旧の目的で、ネットワーク・マップ・データベース130のバックアップおよび前回のバックアップ以降の変更に関する情報もネットワーク・マップ・ジェネレータ120から定期的な間隔で受け取るバックアップ・サブシステム125が設けられ、その結果、障害が発生した場合に、ネットワーク・マップ・データベース130の内容を復旧することができる。上述の部品は、システム100のデータ管理および保守部分の一部である。要求の処理は、システム100の第2の部分の一部である。クライアント・ネットワーク要素180によって送出される要求は、要求トランスレータ170で最初に処理されることができる。次に、この要求は、事前に定められた承認基準に基づいて承認または捨てられる。要求が承認された場合、すべての送信元・宛先ペアに関連するすべての情報を取り出す役割を果たすクエリが、クエリ・マネージャ140に送出される。あるいは、事前に定められた統一フォーマットの要求も、クライアント・ネットワーク要素180からクエリ・マネージャ140に直接送信されることができる。集約された統計情報が、送信元および宛先の使用率(popularity)を得るために頻繁ヒッター検出器160によって保守されることができる。この情報は、次に、経路ランカー(path ranker)150に送信され、経路ランカー150は、クエリ・マネージャ140および頻繁ヒッター検出器160によって受け取られたすべての情報を利用し、関連スコアを推定して、ランク付けされたリストを編集し、このリストが推奨事項となる。次いで、この推奨事項が応答としてクライアント・ネットワーク要素180に送信される。クライアント・ネットワーク要素180は、送信元ネットワーク要素の名前で要求を送信する、送信元ネットワーク要素またはDNSサーバなどのプロキシとすることができる。
10
20
30
30

【0036】

情報取り出しサブシステム112は、図2により詳細に示されているが、2つのサブシステムすなわち左手側の測定取り出しサブシステムと右手側の計算処理サブシステムとを備える。2つのサブシステムのそれぞれは、アクティブ・モードまたはパッシブ・モードのいずれかとすることができます、すなわち、プッシュ・モードまたはプル・モードのいずれかで動作することができます。以下では、各モードの各サブシステムの機能について説明する。

40**【0037】**

図2の左手側には、測定取り出しサブシステム210のためのプッシュ・プロセスに関与できるアクティブ測定情報送信元またはシード(seed)211、212、および213の第1の例が示されており、測定取り出しサブシステム220のためのプル・プロセスに関与できる受動測定情報送信元またはシード221、222、および223の第2の例も示されている。能動測定は、ネットワーク構成要素のハードウェアの状態、ネットワークのトポロジのためのリスナ、およびルーティング・プロトコルが使用される間にネットワークで交換されるメッセージのアクティブ・リスナに関するシードを含みうるが、こ
50

れらに限定されない。上述の情報の更新が、更新をリスンする準備ができている一連のフィーダに送信される。これらのリスナは、物理的フィード・プロセッサ 232、モニタリングフィード・プロセッサ 234、ネットワーク・ポリシーフィード・プロセッサ 236、およびメタ情報フィード・プロセッサ 238である。これらのフィーダは、更新を保持する。パッシブ・モードでは、上述のフィーダは、図示の実施形態では例として手動トポロジ情報 221、構成エクスポート 222、およびテスト環境セットアップ 223が挙げられているがこれらに限定されない受動測定送信元にクエリを行うことによって、使用可能な更新を要求する。

【0038】

図 2 の右手側には、計算処理サブシステムのプッシュ・プロセスおよびプル・プロセスそれぞれにより実現されるアクティブ・モードおよびパッシブ・モードの 2 つの例が示されている。プッシュ・モードでは、フィーダのそれぞれ、すなわち物理的フィード・プロセッサ 232、モニタリングフィード・プロセッサ 234、ネットワーク・ポリシーフィード・プロセッサ 236、およびメタ情報フィード・プロセッサ 238 は、測定取り出しサブシステムのプッシュ手順を使用して更新を受け取ると、ネットワーク・マップ・ジェネレータ 120 にトリガを送信して更新の認識を報告したフィードから更新をプルする更新ハンドラとして機能する情報ウォッチドッグ 240 に更新通知を送信する。プル・モードでは、ネットワーク・マップ・ジェネレータは、上述のフィーダから情報を定期的にプルしようとする。

【0039】

図 3 は、ネットワーク・プロバイダのネットワーク・マップを構築および保守する目的で図 2 に示すフィーダから更新された情報をプルするために使用されるネットワーク・マップ・ジェネレータ 120 によって実行されるステップを含む流れ図を概略的に示す。図 2 に関して説明したプッシュ・モードでは、ウォッチドッグ 240 は、ネットワーク・マップ・ジェネレータ 120 に更新通知を提供する役割を果たす。プル・モードでは、ネットワーク・マップ・ジェネレータ 120 は、フィーダのための情報をプルする。いずれの場合にも、ステップ 301において、ネットワークの状態の微細な変更のリストが、プル手順またはプッシュ手順を使用してすべてのフィーダから得られる。次に、ステップ 302において、微細な変更の 1つ1つが処理される。ステップ 303 では、それぞれの微細な変更が取り出される。ステップ 304 で解析される保留中の更新がある場合、ステップ 305 では、その更新が有効な更新であるかどうか確認される。その更新が有効な更新である場合、保守されるトポロジの変更がステップ 306 でマークされ、その変更は、ステップ 307 でトランザクションとしてデータベースに適用される。その変更が有効でない場合、その更新はステップ 308 で捨てられる。データベースを変更した更新がある場合、これは、ステップ 309 で確認される。このような場合、トポロジの変更があるかどうかがステップ 310 でチェックされる。そうである場合、ステップ 311 でルーティング・テーブルが再計算され、パス・キャッシュもステップ 313 で再計算される。パス伝搬の変更がある場合は、再計算がパス・キャッシュにプッシュされ、変更がない場合は、ステップ 314 でトリガ復元ポイントが作成される。トリガ復元ポイントは、パス・キャッシュが再計算された場合にも作成される。トリガ復元ポイントから、フェイルオーバ手順 320 が呼び出される。

【0040】

図 4 は、システムに送出された要求を変換するためのステップを含む流れ図を概略的に示す。要求トランスレータ・システム 170 は、2 つの目的に使用される。第 1 に、プロトコル間の通信のためであり、第 2 に、適切なランク付け機能の判定を補助するためである。ステップ 400 でネットワークが要求を受信すると、次に、その要求が要求トランスレータ・プロセスに送信される。要求はステップ 401 で取り出され、ステップ 402 では、既知のフォーマットであるかどうかチェックされる。既知のフォーマットでない場合、要求は、ステップ 403 で破棄される。既知のフォーマットである場合、要求は、ステップ 404 で復号され、ステップ 405 で適切なランク付け機能を選択するために使用で

10

20

30

40

50

きると復号情報から判断され、次いでステップ 406 にて統一要求が作成され、ステップ 407 において、図 4 では P a D I S サーバという名前であるクエリ管理サーバに要求が送られ、ステップ 408 で、この要求が、ネットワークまたは I P C または共有メモリにあるクエリ・プロセッサに転送される。要求がクエリ・プロセッサによって処理されると、ステップ 409 で、リプライが受信される。そうでない場合、ステップ 413 において、サードパーティのプロトコル・エラーがあるかどうかチェックされる。そのようなエラーがない場合、ステップ 415 で、修正されていない要求がクライアント・ネットワーク要素に返送される。サードパーティのエラーが見つかった場合、ステップ 414 でサードパーティのエラー・メッセージが生成され、ステップ 416 においてサードパーティのエラー応答が送信される。リプライが有効なリプライである場合、ステップ 410 で統一要求が復号され、次にステップ 411 でリプライが有効かどうかチェックされる。そうでない場合は、ステップ 413 と同じ手順が使用される。一方、リプライが有効なリプライである場合、ステップ 412 で、サードパーティが順序を変更したリストが生成され、サードパーティのプロトコル応答が送信される。
10

【 0041 】

図 5 は、システムがクエリ・マネージャ・サブシステムで返信する前にシステムに到着した送出された要求を管理するためのステップを含む流れ図を概略的に示す。要求が要求トランスレータ 170 によって変換されると、ステップ 501 で統一要求が復号されて検証される。ステップ 501 では、承認制御も実行される。あるいは、承認制御は、クエリ・マネージャ・サブシステムによって実行されることもできる。ステップ 502 で空の注釈付きリストが作成され、ステップ 503 では送信元 - 宛先の 1 ペアが抽出され、ステップ 504 で注釈付き経路がトポロジ情報と共に編集される。これには、ステップ 505 でネットワーク・マップ・データベースにクエリを行うことと、ステップ 506 で送信元および / または宛先について頻繁ヒッター検出器にクエリを行うことが必要である。トポロジ情報のある注釈付き経路が取り出されると、この情報は、ステップ 507 において、ランク付け機能の選択を決定するために使用できるメタ情報と共に、選択された機能により経路ランカーを起動するために使用される。次に、ステップ 508 で、送信元 - 宛先ペアが注釈付きリストに追加される。ステップ 509 で、検討中のペアが最後のペアかどうかチェックされる。そうでない場合、ステップ 503 で次の送信元 - 宛先ペアが抽出される。送信元 - 宛先ペアが最後の場合、ステップ 510 で送信元 - 宛先ペアに関連する注釈付き重みが並べ替えられ、ステップ 511 で情報が頻繁ヒッター検出器サブシステムに渡され、ステップ 512 で統一応答が符号化され、この符号化が要求トランスレータ 170 に送信される。クライアント・ネットワーク要素 180 から直接受信された直接統一要求 181 は、類似のように処理される。
20
30

【 0042 】

図 6 は、障害復旧サブシステムのためのステップを含む流れ図を概略的に示す。このサブシステムの主な目的は、障害前の最後の状態を取り戻し、その間に発生した変更を推論することである。2つの構成要素がある。右手側にはバックアップ取り出しがあり、左手側には障害発生間隔更新の推論および取り出しがある。バックアップ取り出しは、適切なバックアップ・プロセスをブートストラップすることによって、ステップ 601 でバックアップ構成をロードする役割を果たす。次に、ステップ 602 では、バックアップが存在するかどうかチェックされる。バックアップが存在する場合は、ステップ 603 でバックアップが取り出される。バックアップがない場合は、ステップ 604 でプロセスが終了し、生のネットワーク情報の収集がリセットされる。バックアップが取り出された場合、復旧以降に受信され、ネットワークの前回のセーブ状態に対応するすべての微細な変更がステップ 701 で送られ、ステップ 702 でクラッシュセーフ (c r a s h - s a f e) なコンテナに再保存される。次に、ステップ 703 でバックアップ・キャッシュがあるかどうかがチェックされる。バックアップ・キャッシュがある場合、そのキャッシュがステップ 704 でコンテナにダンプされ、次いで、ステップ 705 でタイムスタンプのバックアップが行われ、ステップ 706 では、システムは次の通知を待機する。キャッシュ・バッ
40
50

クアップがない場合、ステップ 705 でタイムスタンプのバックアップが行われる。

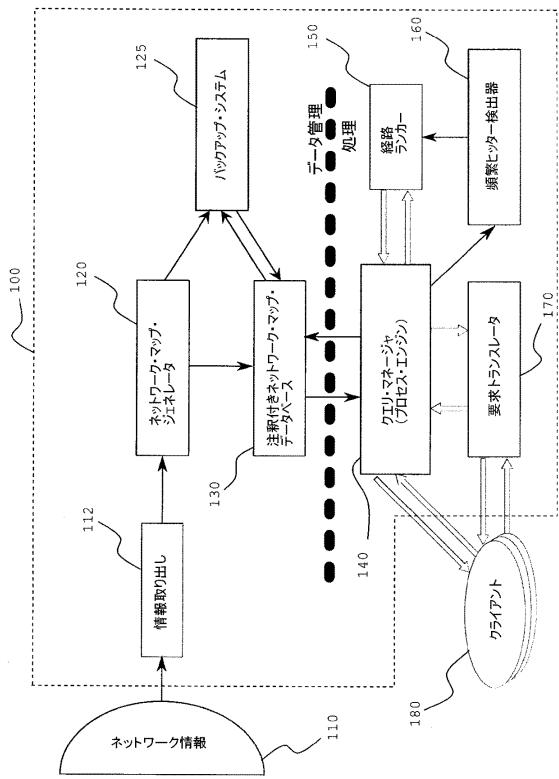
【0043】

図 7 a は、5 つのノード 801 ~ 805 からなるネットワークのトラヒックを例示的に示す。図 7 b は、システムによって達成されるトラヒック・エンジニアリングの同じ例を示す。目的は、ネットワークの全輻輳およびリンクごとの最大負荷および方向を減少させるように送信元 - 宛先のフローを再割り当てすることである。700 のノードからなるネットワークに関して、ノードのうちの 5 つずなわちノード 801 ~ 805 を接続する最も輻輳したアップロード・リンクおよびダウンロード・リンクに焦点を当てて考える。本発明を用いることによって、ネットワークを通過する全トラヒックが 560 . 35 ギガバイトから 455 . 36 ギガバイトに減少する。図 8 a にも示されるように、要求が本発明のシステムに送出されると、平均ダウンロード速度が 406 キロビット / 秒から 520 キロビット / 秒に増加する。図 8 b にも示されるように、1 接続ずなわち送信元 - 宛先ペアあたりの平均ホップカウントは、4 . 65 から 3 . 97 に減少する。いくつかのホップが図 7 a および 7 b に示されていないことに留意されたい。図 7 a および図 7 b は、負荷の高いリンクのトラヒックが減少することも示す。たとえば、28_17 から 65_21 へのリンクの負荷は 100 % から 47 % に減少し、13_17 から 65_21 へのリンクの負荷は 91 % から 70 % に減少し、65_21 から 54_16 へのリンクの負荷は 77 % から 54 % に減少する。
10

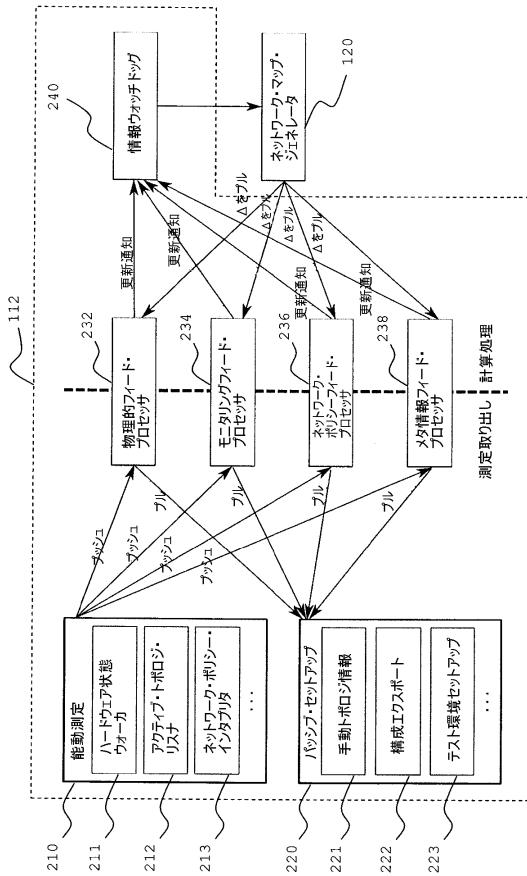
【0044】

現在、情報の収集、処理、および保守は一元化されておらず、ネットワーク・プロバイダにおけるアクティブなトラヒック・エンジニアリングおよびサーバまたはピア選択に使用されていない。これらの構成要素のそれぞれは、独立して動作し、主にネットワークの計画、モニタリング、および課金に使用される。送信元 - 宛先のマッチングを向上させようとする試みとしては、サードパーティを含む生のネットワーク情報の配信、またはネットワーク・プロバイダ側からの生の情報の感度に起因して広く展開されていない、ユーザによる直接的なプッシュ / ブル手順がある。送信元 - 宛先のマッチングを向上させようとする他の試みとしては、コンテンツ・プロバイダおよびコンテンツ配信業者への類似の要求の一元化された保存ならびにユーザの近接度を推論する送信元の推奨事項の相関がある。これは、重要なネットワーク・プロバイダ情報もない。対照的に、本発明は、有利には、機密性の高いネットワーク情報を明らかにすることなくネットワークの詳細図に基づいて推奨事項を提供する。
20
30

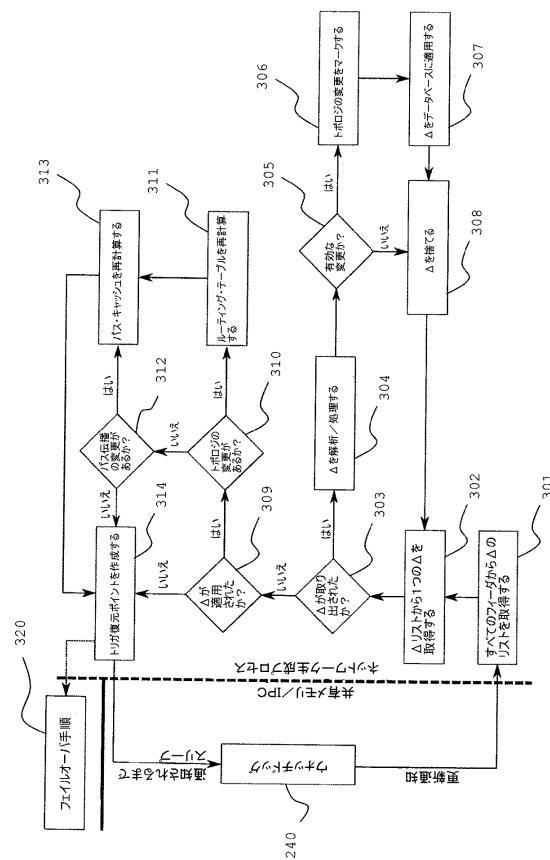
【 図 1 】



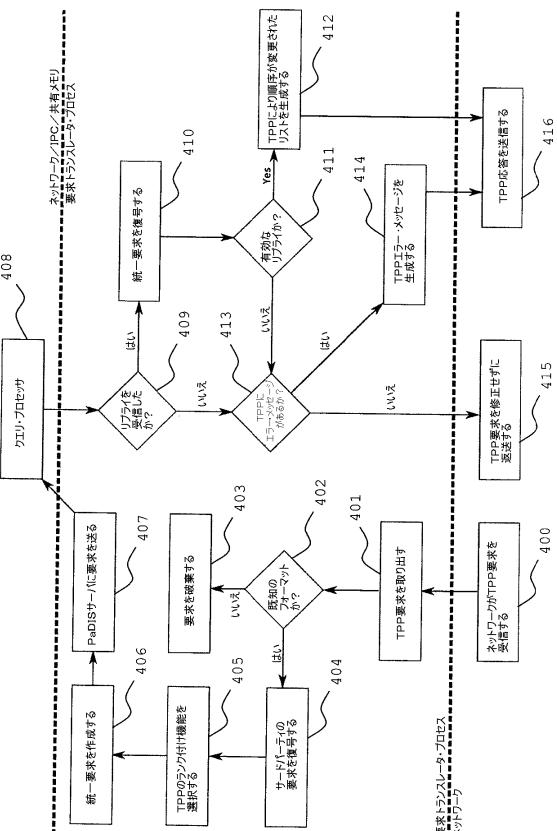
【 四 2 】



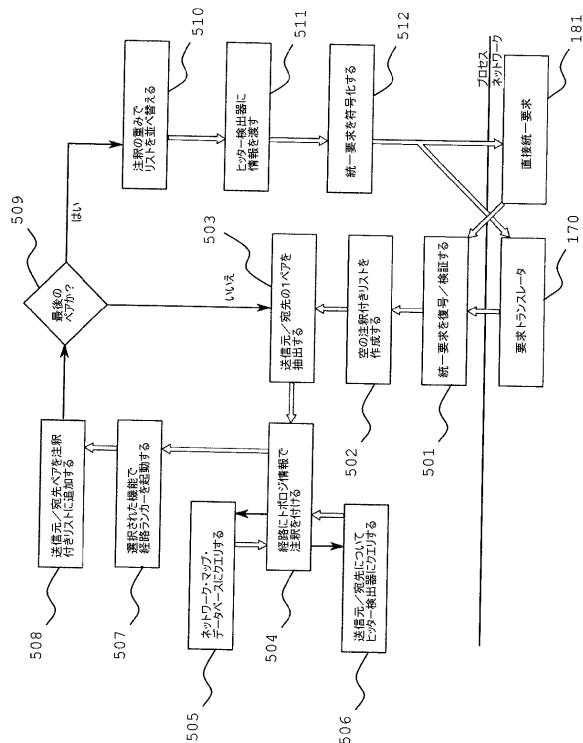
【図3】



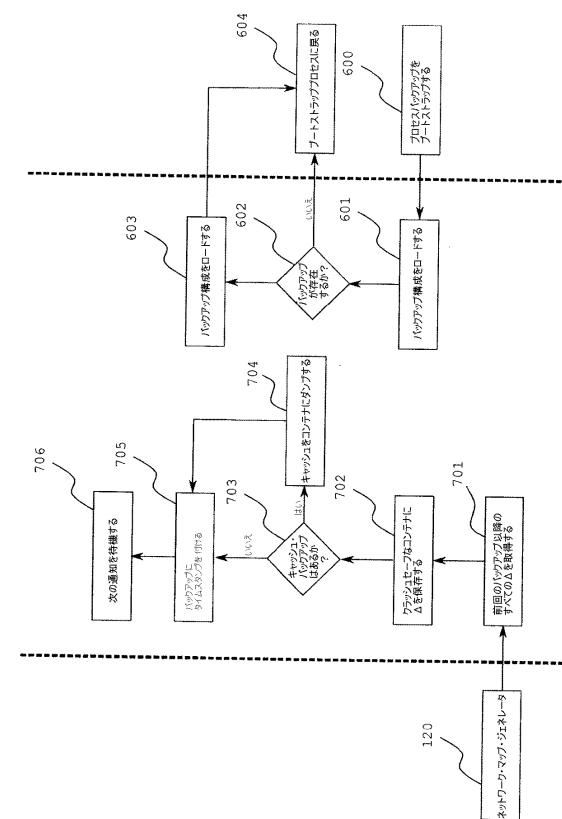
〔 四 4 〕



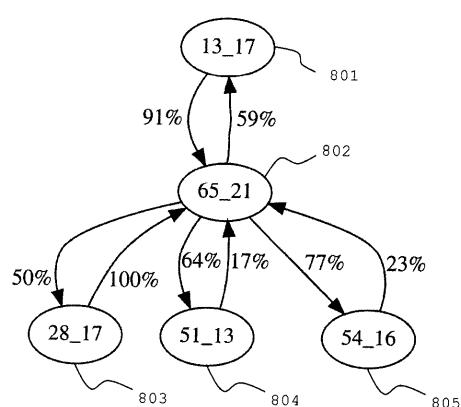
【図5】



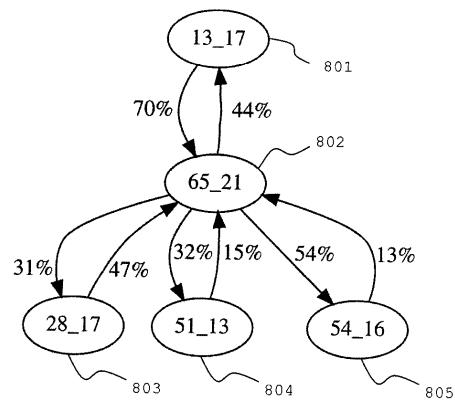
【図6】



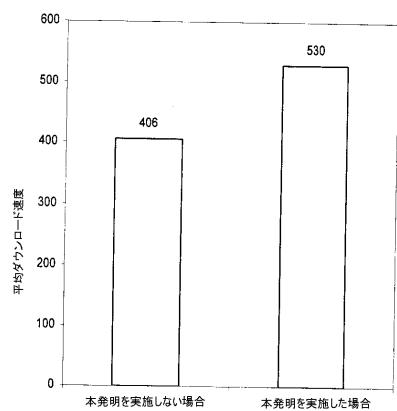
【図7 a】



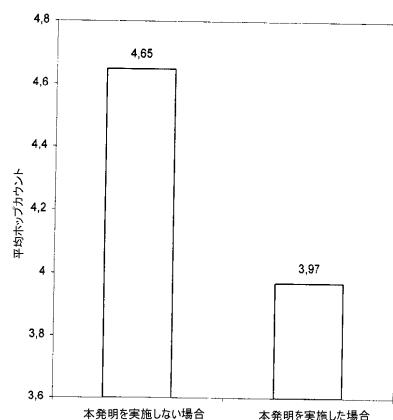
【図7 b】



【図 8 a】



【図 8 b】



フロントページの続き

(72)発明者 ポエセ, イングマー
　　ドイツ 10553 ベルリン, ロイドリンシュトラーセ 2a
(72)発明者 フランク, ペンヤミン
　　ドイツ 12555 ベルリン, バーンホーフシュトラーセ 2D
(72)発明者 スマラグダキス, ジョルジオス
　　ドイツ 14052 ベルリン, ライヒスシュトラーセ 7
(72)発明者 フェルドマン, アンヤ
　　ドイツ 10717 ベルリン, ナッサウイッシェ シュトラーセ 56

審査官 安藤 一道

(56)参考文献 國際公開第2009/092440 (WO, A1)
米国特許出願公開第2009/0168768 (US, A1)
米国特許第07103664 (US, B1)
米国特許第07254626 (US, B1)

(58)調査した分野(Int.Cl., DB名)

H04L 12/70
H04L 12/721