(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4828865号
(P4828865)
(45)発行日　平成23年11月30日(2011.11.30)　(24)登録日　平成23年9月22日(2011.9.22)

(51)Int.Cl.　F I
H04L 12/56 (2006.01)　H04L 12/56　100Z

請求項の数 8　(全 36 頁)

(21)出願番号	特願2005-155178 (P2005-155178)
(22)出願日	平成17年5月27日 (2005.5.27)
(65)公開番号	特開2005-341589 (P2005-341589A)
(43)公開日	平成17年12月8日 (2005.12.8)
審査請求日	平成20年5月22日 (2008.5.22)
(31)優先権主張番号	60/575350
(32)優先日	平成16年5月28日 (2004.5.28)
(33)優先権主張国	米国 (US)
(31)優先権主張番号	11/106410
(32)優先日	平成17年4月14日 (2005.4.14)
(33)優先権主張国	米国 (US)

(73)特許権者　596092698
アルカテル－ルーセント　ユーエスエー　インコーポレーテッド
アメリカ合衆国　０７９７４　ニュージャーシィ，マレイ　ヒル，マウンテン　アヴェニュー　６００－７００
(74)代理人　100094112
弁理士　岡部　讓
(74)代理人　100064447
弁理士　岡部　正夫
(74)代理人　100085176
弁理士　加藤　伸晃
(74)代理人　100106703
弁理士　産形　和央

最終頁に続く

(54)【発明の名称】トラフィック・パターン可変性と独立な効率的で堅牢なルーティング

(57)【特許請求の範囲】
【請求項１】
リンクによって相互接続され、少なくとも１つの入口点および少なくとも１つの出口点を有するノードのネットワークを通ってデータをルーティングする方法であって、
（ａ）前記入口点と前記出口点との間でのデータのルーティングに関するサービス需要を有する、パスに関する要求を受け取る工程と、
（ｂ）前記入口点と前記出口点との間の１つまたは複数の中間ノードの集合を選択する工程と、
（ｃ）前記ネットワークの帯域幅に基づいて、前記入口点から１つまたは複数の中間ノードの前記集合の各ノードに送られる前記データのめいめいの分数を決定する工程とを含み、前記入口点および前記出口点に対応するトラフィック行列が、前記トラフィック行列を制約する行合計上限および列合計上限を有し、前記めいめいの分数の前記決定が、前記トラフィック行列の実際のトラフィックのボリュームを考慮に入れずに、前記トラフィック行列を制約する前記行合計上限および前記列合計上限の少なくとも１つを使用して行われ、さらに、
（ｄ）前記入口点から１つまたは複数の中間ノードの前記集合の各ノードへ、前記決定されためいめいの分数で前記データをルーティングする工程と、
（ｅ）１つまたは複数の中間ノードの前記集合の各ノードから前記出口点へ前記データをルーティングする工程とを含む、方法。
【請求項２】

工程（ｃ）での前記めいめいの分数の前記決定が、主解および双対解を有する線形プログラムを解くことによって行われ、前記リンクに沿ったフローが、前記主解で増やされ、前記リンクの重みが、前記双対解で乗法的な形で更新される、請求項１に記載の発明。

【請求項３】

前記主解が、

【数１】

$$\sum_{P\in P_{ij}} x(P) = \alpha_j R_i + \alpha_i C_j \ \forall i,j \in N, i \neq j$$

$$\sum_{P:e\in P} x(P) \leq u_e \ \forall e \in E$$

のもとで

【数２】

$$\sum_{i\in N} \alpha_i$$

を最大化すること、という線形計画法定式化によって表され、前記双対解が、

【数３】

$$\sum_{i:i\neq k} R_i SP(i,k) + \sum_{j:j\neq k} C_j SP(k,j) \geq 1 \ \forall k \in N$$

$$w(e) \geq 0 \ \forall e \in E$$

のもとで

【数４】

$$\sum_{e\in E} u_e w(e)$$

を最小化すること、という線形計画法定式化によって表され、

Ｎが、ソース・ノードｉ、宛先ノードｊ、および中間ノードｋを含む、前記ネットワーク内のすべてのノードの集合を表し、

Ｅが、前記ネットワーク内のすべてのリンクｅの集合を表し、

Ｐが、ノードｉからノードｊへの所与のパスを表し、

ｘ（Ｐ）が、パスＰ上のトラフィックを表し、

α_iが、ノードｉに送られるトラフィックの配分比を表し、

α_jが、ノードｊに送られるトラフィックの配分比を表し、

R_iが、任意の時にノードｉが前記ネットワークに送るトラフィックの最大総帯域幅を表し、

C_jが、任意の時にノードｊが前記ネットワークから受け取るトラフィックの最大総帯域幅を表し、

u_eが、リンクｅの使用可能な容量を表し、

ｗ（ｅ）が、リンクｅの重みの集合を表し、

ＳＰ（ｉ，ｋ）が、ノードｉからノードｋへの重みｗ（ｅ）の下での最短パスを表し、

ＳＰ（ｋ，ｊ）が、ノードｋからノードｊへの重みｗ（ｅ）の下での最短パスを表す、

請求項２に記載の発明。

【請求項４】

工程（ｃ）での前記めいめいの分数の前記決定が、不等であるめいめいの分数をもたらす、請求項１に記載の発明。

【請求項５】

工程（ｃ）での前記めいめいの分数の前記決定が、前記出口点の識別を考慮に入れずに行われる、請求項１に記載の発明。

【請求項６】

工程（ｅ）での１つまたは複数の中間ノードの前記集合の各ノードから前記出口点への前記データのルーティングが、前記入口点および前記出口点に対応するトラフィック行列に関する他の情報の知識なしで、前記出口点の識別に基づいて実行される、請求項１に記載の発明。

【請求項７】

リンクによって相互接続され、少なくとも１つの入口点および少なくとも１つの出口点を有するノードのネットワークを通ってデータをルーティングする装置であって、

（ｉ）前記入口点と前記出口点との間でのデータのルーティングに関するサービス需要を有する、パスに関する要求と、（ｉｉ）前記要求に関連する前記データと、を受け取るように適合された入力モジュールと、

前記要求の前記パスを決定するように適合された処理モジュールとを含み、前記処理モジュールは、（ａ）前記入口点と前記出口点との間の１つまたは複数の中間ノードの集合を選択し、（ｂ）前記ネットワークの帯域幅に基づいて、前記入口点から１つまたは複数の中間ノードの前記集合の各ノードに送られる前記データのめいめいの分数を決定する、ことによって前記パスを決定し、前記入口点および前記出口点に対応するトラフィック行列が、前記トラフィック行列を制約する行合計上限および列合計上限を有し、前記めいめいの分数の前記決定が、前記トラフィック行列の実際のトラフィックのボリュームを考慮に入れずに、前記トラフィック行列を制約する前記行合計上限および前記列合計上限の少なくとも１つを使用して行われ、さらに、

前記要求の前記パスに従って、前記入力モジュールからのパケットをルータの出力モジュールに転送するように適合されたルータを含み、前記ルータは、（ｃ）前記入口点から１つまたは複数の中間ノードの前記集合の各ノードへ、前記決定されためいめいの分数で前記データをルーティングし、（ｄ）１つまたは複数の中間ノードの前記集合の各ノードから前記出口点へ前記データをルーティングするように適合される、装置。

【請求項８】

リンクによって相互接続され、少なくとも１つの入口点および少なくとも１つの出口点を有するノードのネットワークであって、

（ａ）前記入口点と前記出口点との間でのデータのルーティングに関するサービス需要を有する、パスに関する要求を受け取り、

（ｂ）前記入口点と前記出口点との間の１つまたは複数の中間ノードの集合を選択し、

（ｃ）前記ネットワークの帯域幅に基づいて、前記入口点から１つまたは複数の中間ノードの前記集合の各ノードに送られる前記データのめいめいの分数を決定し、前記入口点および前記出口点に対応するトラフィック行列が、前記トラフィック行列を制約する行合計上限および列合計上限を有し、前記めいめいの分数の前記決定が、前記トラフィック行列の実際のトラフィックのボリュームを考慮に入れずに、前記トラフィック行列を制約する前記行合計上限および前記列合計上限の少なくとも１つを使用して行われ、さらに、

（ｄ）前記入口点から１つまたは複数の中間ノードの前記集合の各ノードへ、前記決定されためいめいの分数で前記データをルーティングし、

（ｅ）１つまたは複数の中間ノードの前記集合の各ノードから前記出口点へ前記データをルーティングする、ネットワーク。

【発明の詳細な説明】

【技術分野】

【０００１】

本願は、参照によって本明細書に組み込まれる、２００４年５月２８日出願の同時係属の米国仮特許出願第６０／５７５３５０号の優先権を主張するものである。

本発明は、遠隔通信システムにおけるルーティングに関し、具体的には、保証されたサービス・レベルを有するルーティングのためのネットワークのノードを通るパスの判定に

関する。

【背景技術】

【０００２】

インターネットなどのパケットベース通信ネットワークでは、パケット・フローと称するデータ・パケットの各ストリームが、ネットワークを介して、ソースから宛先へのネットワーク・パス上で転送される。各ネットワーク・パスは、リンクの集合によって相互接続されたノードの集合によって定義される。ノードに、１つまたは複数のルータが含まれる場合があり、ルータは、コンピュータの間でのデータ転送を処理する、ネットワーク内のデバイスである。

【０００３】

通信システムは、異なるサイズのネットワークが相互接続されるように構成することができ、その代わりにまたはそれに加えて、同等のサイズのネットワークが相互接続される１つまたは複数のピア構造を含めることができる。パケット・ネットワークは、入口点（ｉｎｇｒｅｓｓ　ｐｏｉｎｔ）および出口点（ｅｇｒｅｓｓ　ｐｏｉｎｔ）と称するノードを介して別のパケット・ネットワークに接続することができる。用語入口点および出口点は、別のパケット・ネットワークに接続された、パケット・ネットワークのノードを指すことができ、あるいは、その代わりに、他のパケット・ネットワークの接続するノードを指すことができる。複数の他のパケット・ネットワークの間でパケットを転送する大容量のパケット・ネットワークを、一般に「バックボーン」ネットワークと称する。

【０００４】

図１に、パケット・ネットワーク１０２から１０４の間の通信を可能にする、リンク１０１を介して相互接続されたノードｎ１からｎ９を有する従来技術のバックボーン・ネットワーク１００を示す。バックボーン・ネットワーク１００の入口点の１つが、ノードｎ１であり、このノードは、ソース（すなわちパケット・ネットワーク１０２）からパケットを受け取り、このバックボーン・ネットワークの出口点の１つが、ノードｎ４であり、このノードは、宛先（すなわちパケット・ネットワーク１０４）にパケットを送る。バックボーン・ネットワーク１００は、内部ルーティング・プロトコルをサポートして、ネットワーク・トポロジ情報を配布し、ノードｎ１からｎ９を介するベストエフォート・ルーティング（たとえば、宛先ベースの最短パス・ルーティング）に基づいて、入口点と出口点の間でパケットをルーティングする。集中ネットワーク管理システム１０５を使用して、（ｉ）バックボーン・ネットワーク１００を通る仮想回線（ｖｉｒｔｕａｌ　ｃｉｒｃｕｉｔ）またはパケット・フローを提供し、（ｉｉ）リンク１０１の容量および使用率を監視し、（ｉｉｉ）プロビジョニングされるパスの計算およびインストールを調整することができる。転送テーブルが、各ノードによって使用されて、受け取られたパケットが、その宛先に向かって次のノードに転送される。さらに、集中ネットワーク管理システム１０５を使用して、ネットワーク・トポロジ情報を収集し、配布することもできる。

【０００５】

内部ルーティング・プロトコルは、ソース／宛先対の間でバックボーン・ネットワークのノードを通るパスに沿ったパケットの転送を決定するのに使用される。ノードによって受け取られたパケットは、内部ルーティング・プロトコルに従って構成された転送テーブルまたは明示的経路プロビジョニングによってインストールされた経路に基づいて、他のノードに転送される。内部ルーティング・プロトコルは、ノードの間でのネットワーク・トポロジおよびリンク状態情報（「ネットワーク・トポロジ情報」）の交換を指定して、ノードが対応する転送テーブルを構成できるようにすることもできる。さらに、いくつかのルーティング・プロトコルは、リンク「コスト」をノード間の各リンクに関連付ける。このリンク・コストは、たとえば、平均リンク使用率またはリンクによって生成される収入ならびにネットワークでのリンクの重要さに関連付けることができる。リンク状態情報またはリンク帯域幅（たとえば、接続性または使用可能な帯域幅）がルータの間で交換される時に、ネットワーク内の各ノードは、そのネットワークのトポロジの完全な記述を有する。「ベストエフォート」ルーティングに広く使用されている内部ルーティング・プロ

トコルの例が、ＯＳＰＦ（Ｏｐｅｎ　Ｓｈｏｒｔｅｓｔ　Ｐａｔｈ　Ｆｉｒｓｔ）プロトコルである。

【０００６】

ルーティング・プロトコルは、接続性を提供するほかに、トラフィック管理も可能にすることができる。たとえば、ＭＰＬＳ（Ｍｕｌｔｉ－Ｐｒｏｔｏｃｏｌ　Ｌａｂｅｌ　Ｓｗｉｔｃｈｅｄ）標準規格は、バックボーン・ネットワークでのそのようなルーティング・プロトコルを可能にする。ＭＰＬＳ標準規格は、プロビジョニングされるサービス・レベル（保証されたサービス品質（ＱｏＳ）とも称する）を有する仮想回線（パケット・フロー）を有するネットワークに使用することができる。

【０００７】

プロビジョニングされるサービス・レベルは、たとえば、バックボーン・ネットワークを通るパケット・フローのパスの保証される最低帯域幅とすることができる。入口点と出口点の間のサービスの保証されるレベルを有するこのパスを、ネットワーク・トンネル・パス（Ｎｅｔｗｏｒｋ　Ｔｕｎｎｅｌ　Ｐａｔｈ、ＮＴＰ）と称する場合がある。当業者に明白であるように、ＮＴＰの特定の実装が、異なるネットワークのタイプについて存在する。ＮＴＰの例として、仮想回線を、ＴＣＰ／ＩＰネットワーク内のパケット・フローについて確立することができ、仮想回線を、非同期転送モード（ＡＴＭ）ネットワーク内のセルについて確立することができ、ＬＳＰ（ｌａｂｅｌ－ｓｗｉｔｃｈｅｄ　ｐａｔｈ）を、ＭＰＬＳネットワーク内のパケットについて確立することができる。ＮＴＰのルーティングが計算されたならば、ＲＳＶＰ（ＩＰネットワークおよびＭＰＬＳネットワークのＲｅｓｅｒｖａｔｉｏｎ　Ｐｒｏｔｏｃｏｌ）またはＬＤＰ（ＭＰＬＳネットワークのＬａｂｅｌ　Ｄｉｓｔｒｉｂｕｔｉｏｎ　Ｐｒｏｔｏｃｏｌ）などのシグナリング・プロトコルのパケットを使用して、リンク帯域幅を予約し、ＮＴＰを確立することができる。ＮＴＰは、バックボーン・ネットワークのノードの間の特定のパスに沿った明示的な経路としてプロビジョニングすることができ、すなわち、ＮＴＰがパケット・フローについてプロビジョニングされる時に、そのＮＴＰの入口点と出口点の間のすべての中間ノードを指定でき、これを、そのフローの各パケットが通過する。

【０００８】

ＭＰＬＳネットワークでは、パケットが、そのパケットが入口点で受け取られる時に追加情報をパケットに付加するか、パケットから形成することによってカプセル化される。ラベルと称するこの追加情報は、バックボーン・ネットワークのルータによって、パケットを転送するのに使用される。図２に、ラベル２０１をパケット２０２に付加された、そのようなカプセル化されたパケット２００を示す。ラベルは、パケット・ヘッダ内の情報を要約する。この要約は、ヘッダ・フィールドに基づくものとすることができ、入口点のアドレスを識別する起点（ソース）アドレス・フィールド（ｏ）２１０と、出口点のアドレスを識別する終点（宛先）アドレス・フィールド（ｔ）２１１を含む。いくつかの場合に、ラベルを、単に、受け取られたパケットのヘッダ内の起点アドレス・フィールドおよび終点アドレス・フィールドを識別するか他の形でこれらに関連するポインタとすることができる。ラベルに、１つまたは複数のサービス・レベル・フィールド（ｂｄ）２１２も含めることができる。サービス・レベル・フィールド２１２は、必要な最小帯域幅など、仮想回線について望まれるサービス・レベル（「デマンド」と称する）を識別することができる。いくつかのネットワークで、サービス・レベル・フィールドは、ラベル自体から暗示される。ＭＰＬＳ標準規格バージョン、内部ルーティング・プロトコル・バージョン、最大遅延、または他のタイプのサービス・レベル・パラメータなど、他のフィールド２１３をラベル２０１に含めることができる。その代わりに、ラベル２０１を、パケット２０２のパケット・ヘッダ（ＰＨ）２１４に挿入することができ、したがって、図２に示されたフィールドの順番は、例示にすぎない。バックボーン・ネットワークは、ラベルを使用して、類似するＬＳＰを有するカプセル化されたパケットをクラス（ｅｑｕｉｖａｌｅｎｃｅ　ｃｌａｓｓ）にグループ化することができ、ｅｑｕｉｖａｌｅｎｃｅ　ｃｌａｓｓを転送する方法を使用して、ＬＳＰのルーティングの計算を単純にすることができる。

【０００９】
転送テーブルを生成するために、ネットワーク・ノードを通る好ましいパスの集合が、計算され、好ましいパスの集合を計算するのに、重みを使用することができる。各好ましいパスは、ノードの間の最小の総重み（パスの総重みは、パス内のすべてのリンクの重みの合計である）を有し、この重みが、当技術分野で最短パス・ルーティングと称する技法で使用される。好ましいパスの結果の集合は、最短路木（ｓｈｏｒｔｅｓｔ－ｐａｔｈ ｔｒｅｅ、ＳＰＴ）を用いて定義することができる。ルーティング情報（たとえば、ソース／宛先対、ソース・ポート、および宛先ポート）を有する転送テーブルが、ＳＰＴから生成される。その後、このルーティング情報が、受け取ったパケットをＳＰＴの最短パスに沿ってその宛先に転送するのに使用される。ＳＰＴは、その教示が参照によって本明細書に組み込まれる、Ｅ．Ｄｉｊｋｓｔｒａ、「Ａ Ｎｏｔｅ：Ｔｗｏ Ｐｒｏｂｌｅｍｓ Ｉｎ Ｃｏｎｎｅｃｔｉｏｎ Ｗｉｔｈ Ｇｒａｐｈｓ」、Ｎｕｍｅｒｉｃａｌ Ｍａｔｈｅｍａｔｉｃｓ、ｖｏｌ．１、１９５９年、２６９～２７１頁に記載のダイクストラのアルゴリズムなどのアルゴリズムを使用して計算することができる。

【００１０】
ＬＳＰのルーティングを生成するのにルータによって使用される一般的な最短パス・ルーティング・アルゴリズムが、最小ホップ・アルゴリズムである。最小ホップ・アルゴリズムでは、各ルータが、入口点と出口点の間で、パケットのストリーム（パケット・フロー）のバックボーン・ネットワークを通るパスを計算する。各ルータは、最小の個数（「最小」）の実現可能なリンク（「ホップ」）を有する、入口点から出口点までパケット・フローをルーティングするパスを構成する（実現可能なリンクとは、そのパケット・フローをルーティングするのに十分な容量を有するリンクである）。最短パス・ルーティングなどの従来技術のルーティング方式では、宛先アドレスだけに基づいてパケットを転送し、静的でトラフィック固有の独立リンク重みを使用して、ルーティング・テーブルのパスを計算する。ある入口点／出口点対の間の最短パス上の一部のリンクが、輻輳する場合があるが、代替パス上の他のリンクは、あまり使用されない。

【００１１】
ＲＳＶＰまたはＬＤＰなどのシグナリング機構を使用して、パケット・フローに関するネットワークを介する接続の予約および確立の両方を行うことができる。シグナリング機構は、バックボーン・ネットワークをトラバースするＬＳＰのサービス品質属性を指定することができる。複数のＬＳＰの最短パス・ルーティングによって引き起こされるリンク輻輳は、わずかに長いだけである代替のあまり利用されていないパスにＬＳＰに十分なサービス・レベル（サービス品質保証）が存在する可能性がある場合であっても、シグナリング機構による予約要求の拒絶を引き起こす場合がある。最短パス・ルーティングが使用される時には、使用可能なネットワーク・リソースが、必ず効率的に使用されるわけではない。

【００１２】
ＢＧＰ（Ｂｏｒｄｅｒ Ｇａｔｅｗａｙ Ｐｒｏｔｏｃｏｌ）は、自律システム間ルーティング・プロトコルである。自律システムとは、共通の監督の下で共通のルーティング・ポリシを有するネットワークまたはネットワークのグループである。自律システム間ルーティング・プロトコルは、自律システムの間でデータをルーティングするのに使用される。ＢＧＰは、インターネットのルーティング情報を交換するのに使用され、インターネット・サービス・プロバイダ（ＩＳＰ）の間で使用されるプロトコルである。大学および会社などの顧客ネットワークは、通常、ネットワーク内でルーティング情報を交換するのに、ＲＩＰ（Ｒｏｕｔｉｎｇ Ｉｎｆｏｒｍａｔｉｏｎ Ｐｒｏｔｏｃｏｌ）またはＯＳＰＦ（Ｏｐｅｎ Ｓｈｏｒｔｅｓｔ Ｐａｔｈ Ｆｉｒｓｔ）などのＩＧＰ（Ｉｎｔｅｒｉｏｒ Ｇａｔｅｗａｙ Ｐｒｏｔｏｃｏｌ）を使用する。顧客は、ＩＳＰに接続し、ＩＳＰは、ＢＧＰを使用して、顧客経路およびＩＳＰ経路を交換する。ＢＧＰは、自律システムの間で使用することができ、あるいは、サービス・プロバイダは、ＢＧＰを使用して、自律システム内で経路を交換することができる。

【特許文献１】米国仮特許出願第６０／５７５３５０号
【非特許文献１】Ｅ．Ｄｉｊｋｓｔｒａ、「Ａ Ｎｏｔｅ：Ｔｗｏ Ｐｒｏｂｌｅｍｓ Ｉｎ Ｃｏｎｎｅｃｔｉｏｎ Ｗｉｔｈ Ｇｒａｐｈｓ」、Ｎｕｍｅｒｉｃａｌ Ｍａｔｈｅｍａｔｉｃｓ、ｖｏｌ．１、１９５９年、２６９～２７１頁
【非特許文献２】Ｎ．Ｇ．Ｄｕｆｆｉｅｌｄ、Ｐ．Ｇｏｙａｌ、Ａ．Ｇ．Ｇｒｅｅｎｂｅｒｇ、Ｐ．Ｐ．Ｍｉｓｈｒａ、Ｋ．Ｋ．Ｒａｍａｋｒｉｓｈｎａｎ、Ｊ．Ｅ．ｖａｎ ｄｅｒ Ｍｅｒｗｅ、「Ａ ｆｌｅｘｉｂｌｅ ｍｏｄｅｌ ｆｏｒ ｒｅｓｏｕｒｃｅ ｍａｎａｇｅｍｅｎｔ ｉｎ ｖｉｒｔｕａｌ ｐｒｉｖａｔｅ ｎｅｔｗｏｒｋ」、ＡＣＭ ＳＩＧＣＯＭＭ １９９９、１９９９年８月
【非特許文献３】Ｆ．Ｓｈａｈｒｏｋｈｉ ａｎｄ Ｄ．Ｍａｔｕｌａ、「Ｔｈｅ Ｍａｘｉｍｕｍ Ｃｏｎｃｕｒｒｅｎｔ Ｆｌｏｗ Ｐｒｏｂｌｅｍ」、Ｊｏｕｒｎａｌ ｏｆ ＡＣＭ、３７（２）：３１８～３３４、１９９０年
【非特許文献４】Ｒ．Ｋ．Ａｈｕｊａ、Ｔ．Ｌ．Ｍａｇｎａｎｔｉ、ａｎｄ Ｊ．Ｂ．Ｏｒｌｉｎ、「Ｎｅｔｗｏｒｋ Ｆｌｏｗｓ：Ｔｈｅｏｒｙ，Ａｌｇｏｒｉｔｈｍｓ，ａｎｄ Ａｐｐｌｉｃａｔｉｏｎｓ」、Ｐｒｅｎｔｉｃｅ Ｈａｌｌ、１９９３年２月
【発明の開示】
【発明が解決しようとする課題】
【００１３】
ネットワークでの主な問題は、ＢＧＰによって導入されるトラフィック変動である。極端なネットワーク・トラフィック変動が、さまざまな理由から発生する可能性がある。たとえば、複数の他のプロバイダとトラフィックを交換する大規模なインターネット・サービス・プロバイダの場合に、キャリアの間のトラフィック交換は、通常、長い時間期間にわたる総トラフィック・ボリュームおよび多分ピーク・レート限度（通常は物理リンク容量によって決定される）によって指定される。さまざまなネットワーク出口点への入口点に入るトラフィックの実際の分布は、先験的に既知でない可能性があり、経時的に変化し得る。これは、分布が、異なる宛先プレフィックスへのトラフィックの本質的な変動などの多数の要因と、キャリアによってローカルに行われるかキャリアが制御を有しない他の自律システム内で行われる変更に起因するルーティング変更によって決定される。トラフィック分布の本質的な変化は、特別な出来事に応答する急激な混雑の突然の出現など、多数の要因によって引き起こされる可能性がある。トラフィック分布に影響し得るローカル・ルーティング変化の例が、ＩＧＰ重み変更と「ｈｏｔ－ｐｏｔａｔｏ」ルーティングの組合せであり、これは、プレフィックスの組に宛てられたトラフィックが、そうでなければ選択するはずのネットワーク出口点を変更する可能性がある。「ｈｏｔ－ｐｏｔａｔｏ」ルーティングは、ネットワークのノードが、最終的な所定の宛先に移動する前にパケットを保管するバッファを持たず、ルーティングされる各パケットが、その最終宛先に達するまでコンスタントに転送されるようになる、ルーティングの形態である。したがって、パケットは、「ｈｏｔ ｐｏｔａｔｏ（熱いジャガイモ）」のようにはね回り、ネットワークを移動し続けなければならないので、時々、宛先から遠くに向かって移動する。もう１つの例が、ＭＥＤ（Ｍｕｌｔｉ－Ｅｘｉｔ Ｄｉｓｃｒｉｍｉｎａｔｏｒ）が使用される時のＢＧＰの変化である。ＭＥＤは、経路の「外部メトリック」とも称するが、複数の入口点を有する自律システムへの好ましいパスに関する、外部の近傍への提案である。ローカル・ルーティング変更は、キャリアの制御の下にあり、したがって計画された場合に限ってトラフィック・パターンを変更するが、他の自律システム内のルーティング変更が、下流自律システムに影響する時に、予測不能なトラフィック・シフトが発生し得る。ｈｏｔ－ｐｏｔａｔｏルーティングの広い使用に起因して、ある自律システム内のＩＧＰ重み変化（新しいリンクの追加、保守、トラフィック・エンジニアリングなどに起因する可能性がある）が、トラフィック・パターンの大きいシフトを引き起こす可能性がある。ＩＧＰコストの変化は、プレフィックスのかなりの比率に関するＢＧＰ経路に影響し、影響されるプレフィックスは、トラフィックのかなりの比率を計上する可能性がある。したがって、トラフィックの大きいシフトが、ネットワークの他の場所での変化に起因して、あ

るキャリアで発生する可能性がある。

【００１４】

高いトラフィック変動性を考慮しなければならないもう１つの理由が、ピアリング合意に達するユーザまたはキャリアが、さまざまなサイトへのトラフィックの特性をうまく表せない場合があることである。送信または受信のいずれかの総集約帯域幅だけを推定することは、はるかに簡単である。したがって、正確なトラフィック行列を知ることに頼る必要を無くし、その代わりに、トラフィック行列の部分的な詳細だけを使用することが好ましい。また、トラフィック行列が既知の時であっても、しばしば、トラフィック分布の変化を検出することは困難である。

【００１５】

ネットワーク輻輳は、通常、容量の消失（ルータまたはリンクの故障時）または容量需要の増加（トラフィックの大幅な増加によって引き起こされる）のいずれかに起因して発生する。これらの制御不能な事象に応じて、キャリアは、ネットワーク輻輳を避けるようにドメイン内ルーティングを適応させるか、ルーティング変更に頼らずに、発生し得る異なるトラフィック・パターンおよび障害パターンに対処するために、先験的に十分な容量をとっておく必要があり、繰り返してそうする。動作の複雑さおよびコストに起因し、かつ、変更が正しく実施されない場合のネットワーク不安定性の危険性に起因して、頻繁なドメイン内ルーティング変更を避けることが好ましい。さらに、上で述べたように、ある自律システム内の変更が、他の自律システム内のトラフィック変化のカスケードを引き起こし、これによって多数のインターネット・パスの総合的な安定性に影響する可能性がある。大規模なルーティング変化を避けることのトレードオフは、障害パターンまたは変化するトラフィック・パターンに対処するために行わなければならないかなりの容量オーバープロビジョニングである。理想的には、プロバイダは、（ｉ）構成パラメータのトラフィック依存の動的適応を必要とせず、（ｉｉ）障害の後の動的容量再割振りを最小にし、（ｉｉｉ）オーバープロビジョニングの必要が最小である、ほぼ固定されたルーティング方式を使用することを好む。

【００１６】

トラフィック行列が先験的に未知であるもう１つの応用例が、ネットワークベースの仮想プライベート・ネットワーク（ＶＰＮ）サービスの企業顧客への提供である。この場合に、各顧客とのサービス・レベル契約によって、ＶＰＮに属する各サイトによって送信または受信できるトラフィックの量が指定される。このシナリオでは、ユーザは、トラフィック行列を知らず、キャリアに、総トラフィック・ボリュームおよびピーク・レートだけを指定する。提供されるＶＰＮトラフィックのすべてをネットワークにトランスポートし、大きすぎる遅れを導入せずにそのトラフィックを搬送することが、キャリアの仕事である。各サイトから他のサイトへの実際のトラフィック分布は、通常、未知であり、時刻によって変化し得る。キャリア・ネットワークは、トラフィック・パターン変化の時またはノード障害もしくはリンク障害の時にネットワーク輻輳を経験せずに、提供されるＶＰＮトラフィックのすべてを搬送する仕事を割り当てられる。

【００１７】

グリッド・コンピューティング用のネットワークは、さらに、トラフィック変動が極端になる可能性があり、トラフィック行列が先験的に既知ではないさらなるシナリオを提供する。グリッド・コンピューティングでは、複雑な計算タスクが、地理的に分散する可能性があり、ネットワークによって接続された異なるコンピューティング・ノードの間で区分される。グリッド・コンピューティング・ノードの間の通信パターンは、非常に予測不能であり、高いバースト・レートも経験する可能性がある。トラフィック行列が先験的に既知でないので、１つのオプションは、基礎になるネットワークで容量を動的に予約することであるが、この手法は、多数のグリッド・コンピューティング応用例には低速すぎる。宛先の高い変化性およびトラフィックのバースト的性質のゆえに、ネットワークのオーバープロビジョニングが、ほとんどの時間の非常に低い容量使用量につながる。

【００１８】

トラフィック・パターンが制御不能に変化し得る時によいサービスを提供するために、キャリアは、ネットワーク輻輳を避けるためにドメイン内ルーティングを素早く繰り返して適応させるか、ルーティング変更に頼らずに発生し得る異なるトラフィック・パターンに対処するために先験的に十分な容量をとっておかなければならない。サービス・プロバイダは、（ｉ）動作の複雑さおよびコストと、（ｉｉ）リンク・メトリック変化が正しく実施されない場合のネットワーク不安定性の危険性に起因して、頻繁なドメイン内ルーティング変更を避けることを好む。さらに、ある自律システム内のＢＧＰアプリケーション内の変化が、他の自律システムのトラフィック変化のカスケードを引き起こし、これによって多数のインターネット・パスの総合的安定性に影響する可能性がある。ルーティング変更を避けることのトレードオフは、ルーティングを固定したままで変化するトラフィック・パターンに対処するために行うことができる大きい容量オーバ・プロビジョニングである。理想的には、プロバイダは、構成パラメータのトラフィック依存の動的適応を必要とせず、容量の必要において倹約的な、固定ルーティング方式を使用したがる。

【００１９】

さらに、ＩＰ－ｏｖｅｒ－ＯＴＮ（Ｏｐｔｉｃａｌ　Ｔｒａｎｓｐｏｒｔ　Ｎｅｔｗｏｒｋ）では、ルータが、通常はＩＰルータ・ポートより安価でない光クロスコネクト（ＯＸＣ）からなる再構成可能な交換光バックボーンまたはＯＴＮを介して接続される。ＯＸＣは、波長分割多重（ＷＤＭ）リンクを使用して、メッシュ・トポロジで相互接続される。そのようなＯＸＣからなるコア光バックボーンは、光レイヤでの交換、グルーミング、および回復の機能を引き受ける。ＩＰトラフィック・フローは、光レイヤ回路（「ライトパス」と称する）で搬送されるので、過渡的トラフィックに関するルータ・ポートのバイパスによって、ＩＰ－ｏｖｅｒ－ＯＴＮの光バックボーンを介してＩＰルータを相互接続することによって獲得される巨大な規模の経済の基礎が作られる。過渡的トラフィックをルータから光スイッチに移動することによって、増加するトラフィックに伴ってルータのＰｏＰ（Ｐｏｉｎｔ－ｏｆ－Ｐｒｅｓｅｎｃｅ）構成を更新する要件が、最小になる。というのは、光スイッチが、通常はルータより増やされたポート数を有することに起因して、よりスケーラブルであるからである。ＩＰ－ｏｖｅｒ－ＯＴＮアーキテクチャでは、ルータ・ライン・カードが、通常は、光スイッチ・カードより高価であり、したがって、ネットワーク・コストは、通常、トラフィックのほとんどを光レイヤに保つことによって減らされる。また、光スイッチは、通常、ルータより信頼性が高いので、そのアーキテクチャは、通常はより堅牢であり、信頼性がある。ルータは、交換光バックボーンを介して相互接続されるので、ルーティング処理は、トラフィックを光レイヤに保つことと、データ・トラフィックの効率的で満足な多重化を達成するためにパケット・グルーミングに中間ルータを使用することの間の妥協である。

【００２０】

高速回復機能を有する帯域幅保証されたパスの動的プロビジョニングは、ＭＰＬＳ（Ｍｕｌｔｉ－Ｐｒｏｔｏｃｏｌ　Ｌａｂｅｌ　Ｓｗｉｔｃｈｅｄ）ネットワークおよび光メッシュ・ネットワークなどの多数のネットワークの望ましいネットワーク・サービス機能である。光ネットワークでは、光トランスポート・ネットワークが、それぞれが異なる厳密な信頼性要件を有するさまざまなタイプのトラフィックを搬送するので、高速回復も望ましい。パケット化された音声、クリティカルな仮想プライベート・ネットワーク（ＶＰＮ）トラフィック、または他のサービス品質（ＱｏＳ）保証などのサービスに必要な信頼性を提供するために、類似する高速回復機能をＭＰＬＳネットワークで使用することができる。

【００２１】

ネットワーク内の接続は、パス・レベルまたはリンク・レベルで保護することができる。リンク回復（ローカル回復または高速回復とも称する）について、接続の各リンクを、保護されるリンクを排除した１つまたは複数の事前にプロビジョニングされる迂回パスの組によって保護することができる。リンクの障害時に、故障したリンクのトラフィックを、迂回パスに切り替える。したがって、リンク回復は、リンク障害を避けてルーティング

するローカル機構を提供する。パス回復では、接続の主なまたは機能するパスが、ソースから宛先への「別個の」バックアップ・パスによって保護される。動作するパスのリソースのいずれかの障害時に、トラフィックが、ソース・ノードによってバックアップ・パスに切り替えられる。リンク回復は、通常は、パス回復よりはるかに高速の回復サービスである。というのは、回復が、ローカルにアクティブ化され、パス回復と異なって、障害情報を、ネットワークを介してソースまで伝搬させる必要がないからである。

【００２２】

サービス回復は、光ネットワークの重要な要件である。ノード（光スイッチ）またはリンク（光ファイバ）などのネットワーク要素が故障した場合に、その障害は、１つまたは複数の特定の波長パスを故障させ、影響されるトラフィック・フローを、非常に短い間隔（たとえば５０ｍｓ）内で代替パスを使用して回復しなければならない。比較的高速の回復時間を達成するために、プロビジョニングは、波長パスごとに、ネットワークを通る２つのパスすなわち、主（アクティブ）パスと副（バックアップ）パスを識別する。バックアップ・パスは、主パスに関して、リンクが互いに疎（ｌｉｎｋ　ｄｉｓｊｏｉｎｔ、アクティブ・パスとバックアップ・パスがリンクを共用しない）またはノードが互いに疎（ｎｏｄｅ　ｄｉｓｊｏｉｎｔ、アクティブ・パスとバックアップ・パスがノードもリンクも共用しない）である。バックアップ・パスのリンクの容量は、対応する主パス（たとえば波長）に排他的に割り当てることができ、あるいは、ネットワーク帯域幅使用の効率のために、容量を、望まれる回復のタイプに応じて、異なる主パスのバックアップ・パスのリンクの間で共用することができる。光ネットワーク容量設計は、通常、可能な共用を有する互いに疎な副パスをルーティングする回復の必要を計上する。

【００２３】

現代の非常に動的で変化するトラフィック環境での堅牢なネットワーク・ルーティングは、（ｉ）異なる宛先に送りたがる、トラフィックにおいて予測不能ではあるが「よい」サービスを求めるユーザに対処し、（ｉｉ）洗練されたトラフィック予測機構およびトラフィック管理機構に頼らずに「ベストエフォート・ネットワーキングをよりよく」するためにネットワーク内で行われる必要がある「オーバープロビジョニング」の量を最小にし、（ｉｉｉ）ほとんど静的ルーティング構成を用い、動的ルーティング調整なしでネットワークを効率的に運営し、これによってネットワークの入口ルータと出口ルータの間のトラフィック・フローの劇的な変化によって引き起こされる輻輳を避けるために、インターネット・サービス・プロバイダが使用しなければならないネットワーク・ルーティング方法の知識がなければ行うことができない。これらの目標の達成は、困難であり、その代わりに、ネットワーク・ルーティングを変化したトラフィック需要に適応させるトラフィック管理方式を実施するという管理の複雑さを回避するために非常にオーバープロビジョニングされたネットワークがもたらされる。

【課題を解決するための手段】

【００２４】

本発明は、キャリアのドメイン内の最短パスＩＧＰ（Ｉｎｔｅｒｉｏｒ　Ｇａｔｅｗａｙ　Ｐｒｏｔｏｃｏｌ）ルーティングを、トラフィックが、やはりキャリアのドメイン内にある１つまたは複数の事前に決定された中間ノードを通過することを保証した後にトラフィックを宛先にルーティングする変更されたルーティング方式に置換するという発想に基づく方式を提供する（中間ノードの割当は、パケット再シーケンシングの問題を避けるためにフロー・レベルで行われる）。出口ノードは、それでも、ＢＧＰ（Ｂｏｒｄｅｒ　Ｇａｔｅｗａｙ　Ｐｒｏｔｏｃｏｌ）によって決定される自律システム・パスおよびｈｏｔ－ｐｏｔａｔｏルーティングなどの補助キャリア・ルーティング・ポリシに基づいて選択される。本発明の１実施形態と一貫する方式は、直接最短パスのＩＧＰパス選択を、１つまたは複数の先験的に割り当てられた中間ノードを通るパスに変更する。ＭＰＬＳネットワークでは、この１つまたは複数の事前に決定された中間ノードを通るルーティングを、指定された確率に従ってフローが割り当てられる１つまたは複数の中間ノードの選択された組と入口ノードとの間のＭＰＬＳ　ＬＳＰの事前に構成された組を使用して達成する

ことができる。純ＩＰネットワークでは、このルーティングを、まず１つまたは複数の事前に決定された中間ノードにパケットをトンネリングすることによって達成することができる。１つまたは複数の中間ノードの事前に決定された選択を用いるこのルーティングは、エッジ－リンク容量制約の対象になる、許容可能なすべてのトラフィック・パターンを処理するのに十分であり、さらに、ルータ障害および光レイヤ・リンク障害に対する保護を提供する。さらに、ルーティング適応は、トラフィック行列が変化する時に必要ではなく、この方式は、帯域幅効率が良い。

【００２５】

本発明は、さらに、ＩＰ－ｏｖｅｒ－ＯＴＮまたは他の回路交換ネットワークに適用される時に、１つの中間ルータだけでのパケット・グルーミングを用いて光レイヤでパケットをルーティングでき、大きく変化するトラフィックに関するパケット交換の望みの統計的多重化特性を提供できるルーティング方式を提供する。

【００２６】

１実施形態で、本発明は、リンクによって相互接続され、少なくとも１つの入口点および少なくとも１つの出口点を有するノードのネットワークを通ってデータをルーティングする方法を提供し、この方法には、（ａ）前記入口点と前記出口点との間でのデータのルーティングに関するサービス需要を有する、パスに関する要求を受け取る工程と、（ｂ）前記入口点と前記出口点との間の１つまたは複数の中間ノードの集合を選択する工程と、（ｃ）前記ネットワークの帯域幅に基づいて、前記入口点から１つまたは複数の中間ノードの前記集合の各ノードに送られる前記データのめいめいの分数を決定する工程と、（ｄ）前記入口点から１つまたは複数の中間ノードの前記集合の各ノードへ、前記決定されためいめいの分数で前記データをルーティングする工程と、（ｅ）１つまたは複数の中間ノードの前記集合の各ノードから前記出口点へ前記データをルーティングする工程とが含まれる。

【００２７】

もう１つの実施形態で、本発明は、リンクによって相互接続され、少なくとも１つの入口点および少なくとも１つの出口点を有するノードのネットワークを通ってデータをルーティングする装置であって、入力モジュールと、処理モジュールと、ルータとを含む装置を提供する。入力モジュールは、（ｉ）前記入口点と前記出口点との間でのデータのルーティングに関するサービス需要を有する、パスに関する要求と、（ｉｉ）前記要求に関連する前記データとを受け取るように適合される。処理モジュールは、前記要求の前記パスを決定するように適合され、前記処理モジュールは、（ａ）前記入口点と前記出口点との間の１つまたは複数の中間ノードの集合を選択することと、（ｂ）前記ネットワークの帯域幅に基づいて、前記入口点から１つまたは複数の中間ノードの前記集合の各ノードに送られる前記データのめいめいの分数を決定することとによって前記パスを決定する。ルータは、前記入力モジュールからのパケットを前記要求の前記パスに従ってルータの出力モジュールに転送するように適合され、前記ルータは、（ｃ）前記入口点から１つまたは複数の中間ノードの前記集合の各ノードへ、前記決定されためいめいの分数で前記データをルーティングし、（ｄ）１つまたは複数の中間ノードの前記集合の各ノードから前記出口点へ前記データをルーティングするように適合される。

【００２８】

もう１つの実施形態で、本発明は、複数の命令を保管されたコンピュータ可読媒体であって、前記複数の命令は、プロセッサによって実行される時に、前記プロセッサに、リンクによって相互接続され、少なくとも１つの入口点および１つの出口点を有するノードのネットワークを通ってデータをルーティングする方法であって、（ａ）前記入口点と前記出口点との間でのデータのルーティングに関するサービス需要を有する、パスに関する要求を受け取る工程と、（ｂ）前記入口点と前記出口点との間の１つまたは複数の中間ノードの集合を選択する工程と、（ｃ）前記ネットワークの帯域幅に基づいて、前記入口点から１つまたは複数の中間ノードの前記集合の各ノードに送られる前記データのめいめいの分数を決定する工程と、（ｄ）前記入口点から１つまたは複数の中間ノードの前記集合の

各ノードへ、前記決定されためいめいの分数で前記データをルーティングする工程と、（ｅ）１つまたは複数の中間ノードの前記集合の各ノードから前記出口点へ前記データをルーティングする工程とを含む方法を実行させる、コンピュータ可読媒体を提供する。

【００２９】

もう１つの実施形態で、本発明は、リンクによって相互接続され、少なくとも１つの入口点および１つの出口点を有するノードのネットワークを通ってデータをルーティングするシステムを提供し、前記システムは、（ｉ）前記入口点と前記出口点との間でのデータのルーティングに関するサービス需要を有する、パスに関する要求と、（ｉｉ）前記要求に関連する前記データとを受け取る手段を含む。前記システムは、さらに、（ａ）前記入口点と前記出口点との間の１つまたは複数の中間ノードの集合を選択することと、（ｂ）前記ネットワークの帯域幅に基づいて、前記入口点から１つまたは複数の中間ノードの前記集合の各ノードに送られる前記データのめいめいの分数を決定することとによって前記要求の前記パスを決定する手段を含む。前記システムは、さらに、（ｃ）前記入口点から１つまたは複数の中間ノードの前記集合の各ノードへ、前記決定されためいめいの分数で前記データをルーティングし、（ｄ）１つまたは複数の中間ノードの前記集合の各ノードから前記出口点へ前記データをルーティングすることによって前記要求の前記パスに従ってパケットを転送する手段を含む。

【発明を実施するための最良の形態】

【００３０】

図３に、本発明による、サービス・レベル保証を有するルーティングの方法の例示的実施形態を使用する、相互接続されたノードｎ１からｎ１０のネットワーク３００を示す。このルーティング方法では、ＬＳＰ（ｌａｂｅｌ－ｓｗｉｔｃｈｅｄ　ｐａｔｈ）などのネットワーク・トンネル・パスの要求に関する、ネットワーク３００を通るパスを決定する。ノードｎ１からｎ１０のそれぞれに、１つまたは複数のルータが含まれ、このルータが、本発明のルーティング方法に従って決定されたパスから構成された転送テーブルに基づいてパケットを転送する。この例示的ルーティング方法は、要求されたＬＳＰのパケットを２フェーズでルーティングし、これによって、着信トラフィックが、まず、所定の比率で１つまたは複数の中間ノードの組に送られ、次に、その中間ノードから最終宛先に送られて、ネットワーク・スループットが最大になる。そのようなルーティング方法は、必ずしも、ネットワークを通る最短パスまたは最小ホップ数に沿って、要求されたＬＳＰのパケットをルーティングしない場合がある。

【００３１】

本発明の例示的実施形態を、本明細書で、ＬＳＰ要求などの関連するサービス・レベルを有するパス要求と共にＭＰＬＳ標準規格を使用するネットワークに関して説明するが、本発明は、それに制限されない。本発明は、入口点と出口点の間の保証されたサービス・レベルを有するネットワーク・トンネル・パス（ＮＴＰ）に関する要求が受け取られる場合など、他のコンテキストでも使用することができる。ＮＴＰは、たとえば、ＴＣＰ／ＩＰネットワーク内のパケット・フローの仮想回線、非同期転送モード（ＡＴＭ）ネットワークのセルの接続、およびＬＳＰ（ＭＰＬＳネットワーク内のパケットの）とすることができる。本発明は、ＯＸＣからなる再構成可能交換光バックボーンを介して接続され、これによってコア光バックボーンが光レイヤでの交換、グルーミング、および回復の機能を利用するルータのコンテキストでのＩＰ－ｏｖｅｒ－ＯＴＮ（または他の回路交換ネットワーク）にも特に役立つ可能性がある。

【００３２】

ネットワーク３００などの相互接続されたノードのネットワークは、Ｇ（Ｎ，Ｅ）と定義され、ここで、Ｎは、ノードｎ１からｎ１０の集合、Ｅは、ノードを相互接続するリンク（弧）または（有向）辺の集合である。本明細書に記載の例示的実施形態では、サービス・レベルなどの使用可能なリソースの値が、リンクまたはパスの帯域幅容量であるが、その代わりにまたはそれに加えて、他の実施形態でのサービス・レベル値に、遅延、パケット消失の確率、収入、または他のサービス品質パラメータなどの１つまたは複数のリン

ク・パラメータを含めることができる。当技術分野で既知のように、これらのさまざまなサービス・レベル値の１つまたは複数を、有効帯域幅と称する量によって表すことができる。リンクの集合Ｅのリンクe_{ij}は、リンクe_{ij}によって接続されるノードn_iおよびn_jを表す２つの添字ｉおよびｊ（０＜ｉ，ｊ≦Ｎ）を有する。一般性を失わずに、各リンクe_{ij}は、有向である（パケット・フローが、ノードn_iからノードn_jに向かう）。

【００３３】

図３のソースS_1、S_2、およびS_3は、ノードｎ１、ｎ２、ｎ３、ｎ５、およびｎ９内のルータにパケット・フローを集合的に供給するパケット・ネットワークとすることができ、これらのノードは、たとえば他のキャリアなどの外部ネットワークへの接続の潜在的な入口点である。同様に、宛先D_1、D_2、およびD_3は、ノードｎ３、ｎ４、ｎ８、ｎ９、およびｎ１０内のルータからパケット・フローを集合的に受け取るパケット・ネットワークとすることができ、これらのノードは、外部ネットワークへの接続の潜在的な出口点である。あるネットワーク内のどのノードでも、入口点および／または出口点になることができることを理解されたい。ソースS_1、S_2、およびS_3は、は、入口点に接続され、宛先D_1、D_2、およびD_3は、出口点に接続される。ソース／宛先対は、（S_1，D_1）、（S_1，D_2）、（S_1，D_3）、（S_2，D_1）、（S_2，D_2）、（S_2，D_3）、（S_3，D_1）、（S_3，D_2）、および（S_3，D_3）と定義され、各ノードは、１つまたは複数のソースおよび／あるいは１つまたは複数の宛先をサポートすることができる。ノードｎ１からｎ１０は、現在のネットワーク・トポロジ情報およびリンク状況情報（以下では「ネットワーク・トポロジ」と称する）を有するか、これへのアクセスを有することができ、このネットワーク・トポロジは、配布プロトコルを使用して（たとえば、ＯＳＰＦプロトコルに従う制御パケットによって）ネットワークを介して供給し、配布することができる。

【００３４】

ソースS_1、S_2、およびS_3は、ネットワーク３００内の新しいＬＳＰまたは現在提供されているＬＳＰに関するパケットを生成し、このパケットには、入口点／出口点対を識別するフィールド（たとえば、ソースS_1、S_2、またはS_3のいずれかのアドレスと、宛先D_1、D_2、およびD_3のいずれかのアドレス）が含まれる。たとえばＲＳＶＰまたはＬＤＰのシグナリング・パケットを使用して、帯域幅などのサービス品質（ＱｏＳ）属性またはＱｏＳ保証をネットワーク要素（たとえばルータまたはノード）に通信することができるが、ＬＳＰのパケットには、ＱｏＳ属性またはＱｏＳ保証に対応する１つまたは複数のサービス・レベル・パラメータの値も含めることができる。ネットワーク３００を介して転送されるＬＳＰのこれらのパケットは、ＭＰＬＳ標準規格に従うものとすることができ、図２に示し、これに関して説明したものに類似するフォーマットを有することができる。

【００３５】

図３に示されたネットワーク３００について、９つの潜在的な入口点／出口点対（ソース／宛先対）が存在する。次の議論では、ノードn_iおよびn_jを相互接続する各リンク（ｉ，ｊ）（本明細書ではe_{ij}とも称する）が、残存帯域幅（ｒｅｓｉｄｕａｌ ｂａｎｄｗｉｄｔｈ）と称する関連する使用可能な容量u_{ij}（またはu_e）を有する。あるリンクの残余帯域幅u_eは、そのリンクの総帯域幅と、現在そのリンクに割り当てられているＬＳＰの帯域幅需要の合計との間の差である。ネットワークは、リンクの残余容量に関する情報を交換することができ（ＱｏＳＰＦ（ＱｏＳ ｓｈｏｒｔｅｓｔ－ｐａｔｈ ｆｉｒｓｔ）ネットワークなど）、この情報を、経路の分散計算に使用することができる。残余帯域幅は、一般に、たとえばｋｂｉｔ／秒またはＭｂｉｔ／秒単位で表すことができ、あるいは、リンクの総容量の比率として表すことができる。ノードn_iおよびn_jを相互接続する各リンク（ｉ，ｊ）は、関連するリンク・コストc_{ij}（またはc_e）すなわち、特定のリンクの相対的な使用量、重要性、または他のコストに対応することができる関連するスカラ重みも有することができる。リンク・コストは、特定の入口点／出口点

対のスカラ重みを指す場合もある。リンク・コストを特定のリンクに割り当てて、ルーティング・アルゴリズムが、たとえば遅延、帯域幅提供のコスト、他のトラフィック・エンジニアリング考慮事項、または他の物理リンク－レイヤ考慮事項のゆえに特定のリンクを通る経路を好むか嫌うようにすることができる。

【００３６】

一般に、要求は、入口点ｏと出口点ｔの間の、ｂｄ（「需要」ｂｄ）の要求サービス・レベルを有するパスを提供し、ルーティングするために、ネットワーク３００に達する。図３の例示的ネットワークでは、これを、要求された帯域幅ｂｄ　Ｍｂ／秒を有するソース／宛先対の間のパス、たとえば（S_1，D_1）の提供を求めるＬＳＰ要求または他の形のＮＴＰ要求とすることができる。ＬＳＰ要求は、将来のＬＳＰ要求による帯域幅に関する需要の特性の先験的知識なしに、一時に１つが到着する可能性がある。さらに（ｉ）ＱｏＳ属性またはＱｏＳ保証の特性、（ｉｉ）接続到着、ホールド時間、または出発、および（ｉｉｉ）他のトラフィック・エンジニアリング情報の先見的知識は、必ずしも入手可能でない。需要ｂｄは、パケット・フローのパケットが変化する帯域幅必要に関する統計学的プロセスを表すことができるので、「等価」帯域幅値または「有効」帯域幅値とすることができる。当技術分野で既知のように、サービス・レベル（たとえばＱｏＳ）の属性または要件を、等価帯域幅値または有効帯域幅値に変換することができる。等価帯域幅値または有効帯域幅値は、たとえばピーク・パケット・レートおよび平均パケット・レート、到着時間およびホールド時間、ならびに接続持続時間などに基づく統計学的変数を近似する決定的な値である。

【００３７】

本発明によるルーティング方法は、入口点／出口点対の間のネットワークを通る１つまたは複数のパスに沿ったＬＳＰを評価し、ルーティングする。集合Ｐは、ネットワークＧ（Ｎ，Ｅ）に含まれる特定の（区別可能な）ノード入口点／出口点対の集合であり、ネットワークＧ（Ｎ，Ｅ）は、潜在的なソース／宛先対（S_1，D_1）、（S_1，D_2）、…、（S_3，D_3）である。集合Ｐの要素を、（ｓ，ｄ）と表し（すなわち、（ｓ，ｄ）∈Ｐ）、ｓおよびｄは、それぞれ、ソース・ネットワークおよび宛先ネットワークに対応する。要素（ｓ，ｄ）の間に、複数のＬＳＰを提供することができる。

【００３８】

ネットワーク３００へのＬＳＰ要求は、集中ネットワーク管理システム（図３に図示せず）を介して、または分散プロトコルに従ってネットワーク３００のノードｎ１からｎ１０に供給される制御メッセージによってのいずれかで実施することができる。集中ネットワーク管理システムおよび／または各ネットワーク・ルータは、要求されたＬＳＰに対応するネットワークを通って提供されるパスをＬＳＰ要求が決定する例示的なルーティング方法を実施する。集中ネットワーク管理システムおよび／または各ネットワーク・ルータによるプロビジョニングによって、ＲＳＶＰ制御（たとえば、ＲＳＶＰシグナリング・プロトコルのＱｏＳ要求）が、たとえば必要な帯域幅または他のタイプのサービス・レベルを有する１つまたは複数の接続（パケット・フロー）を確立できるようになる。

【００３９】

ノード－弧接続行列Ｍは、（ｎ×ｅ）行列（ｎは、集合Ｎの要素数と等しく、ｅは、集合Ｅの要素数と等しい）と定義され、各行が、集合Ｎの異なるノードｎに対応し、各列が、集合Ｅの異なるリンクｅに対応する。各列は、ノードn_iとn_jの間の対応するリンクe_{ij}の２つの非０項目（ｉ，ｊ）を有する。リンクe_{ij}に対応する列は、行ｉの「＋１」の値と、行ｊの「－１」の値と、他のすべての行に対応する各位置の「０」の値を有する。

【００４０】

ネットワークの入口（または出口）ノードに入る（または出る）トラフィックの総量は、そのノードのすべての外部入口（または出口）リンク（たとえば、顧客ネットワークまたは他のキャリアへのライン・カード）の総容量によって上限を定められる。所与のノードｉについて、ノードｉから出るトラフィックの総量に対する上限（たとえば、帯域幅ま

たは他のサービス・レベル）が、R_iであり、ノードｉに入るトラフィックの総量に対する上限（たとえば、帯域幅または他のサービス・レベル）が、C_iである。これらのリンク容量限度は、ルータのシャーシ内に物理的に置かれたハードウェアの最大容量などの要因に基づいてモデル化されるが、ネットワーク内のトラフィックのポイントツーポイント行列を制約する。これらの制約だけが、そのネットワークによって搬送されるトラフィックの既知の態様である場合があり、これを知ることは、トラフィック行列の行および列の合計上限を知ることと同等である、すなわち、最大の可能な行の合計は、最大の可能な発信トラフィックを示し、最大の可能な列の合計は、最大の可能な着信トラフィックを示す。したがって、ネットワークのすべての許容可能なトラフィック行列$T=<t_{ij}>$は、次の式（１）および（２）に従い、ここで、Ｔは、（ｎ×ｎ）行列であり（ｎはノードの個数）、（ｉ，ｊ）番目の要素は、ノードｉからノードｊへのトラフィックを表す。

【数１】

$$\sum_{j:j\neq i}^{n} t_{ij} = R_i \quad \forall i \in N \qquad \text{かつ} \qquad (1)$$

$$\sum_{j:j\neq i} t_{ji} = C_i \quad \forall i \in N \qquad (2)$$

【００４１】

前述の式（１）および（２）では、同等性（≦ではなく）を検討することで十分である。というのは、そのすべての行または列の合計が所与の上限より小さいすべての行列Ｔ’∈Ｔ（Ｒ，Ｃ）を、非負（非対角）項目を有する行列Ｔ”の加算によって、行列Ｔ＝Ｔ’＋Ｔ”∈Ｔ（Ｒ，Ｃ）に変換できるからである。Ｔ（Ｒ，Ｃ）は、すべての可能なトラフィック行列の集合を表す。したがって、Ｔをルーティングするすべてのルーティング方式は、Ｔ’もルーティングすることができる。

所与のR_iおよびC_iの値について、行および列の合計だけによって指定されるすべてのそのような行列の集合Ｔ（Ｒ，Ｃ）を、次の式（３）によって表すことができる。

【数２】

$$\mathcal{T}(\mathcal{R},\mathcal{C}) = \{ < t_{ij} > \text{ただし} \sum_{j\neq i} t_{ij} = R_i \text{かつ} \sum_{j\neq i} t_{ji} = C_i \ \forall i \} \quad . \qquad (3)$$

トラフィック分布Ｔを、Ｔ（Ｒ，Ｃ）の任意の行列とすることができ、経時的に変更できることに留意されたい。本発明のある実施形態と一貫するルーティング・アーキテクチャで、Ｔに関して行う必要がある唯一の仮定が、行および列の合計上限だけによってＴが指定されることであることが望ましい。したがって、本発明の１実施形態と一貫するルーティング戦略は、望ましくは、（ｉ）Ｔ（Ｒ，Ｃ）のすべての行列のルーティングを許容しなければならず、（ｉｉ）既存の接続の再構成を必要としてはならない、すなわちＴ（Ｒ，Ｃ）に属する限りトラフィック行列Ｔの変化を忘れなければならず、（ｉｉｉ）帯域幅効率が良くなければならない、すなわち、ノードｉからノードｊへの需要のｍｉｎ（R_i，C_j）量の提供の普通の戦略よりはるかに多い帯域幅を使用してはならない。

【００４２】

ＶＰＮの帯域幅要件を指定する方法の１つの既知のモデルが、その教示が参照によって本明細書に組み込まれる、Ｎ．Ｇ．Ｄｕｆｆｉｅｌｄ、Ｐ．Ｇｏｙａｌ、Ａ．Ｇ．Ｇｒｅｅｎｂｅｒｇ、Ｐ．Ｐ．Ｍｉｓｈｒａ、Ｋ．Ｋ．Ｒａｍａｋｒｉｓｈｎａｎ、Ｊ．Ｅ．ｖａｎ ｄｅｒ Ｍｅｒｗｅ、「Ａ ｆｌｅｘｉｂｌｅ ｍｏｄｅｌ ｆｏｒ ｒｅｓｏｕｒｃｅ ｍａｎａｇｅｍｅｎｔ ｉｎ ｖｉｒｔｕａｌ ｐｒｉｖａｔｅ ｎｅｔｗｏｒｋ」、ＡＣＭ ＳＩＧＣＯＭＭ １９９９、１９９９年８月に記載のｈｏｓｅモデルである。このモデルでは、トラフィック行列が、部分的にのみ指定され、ＶＰＮ端点ｉごとに、R_iおよびC_iだけが指定され、R_iは、ｉが任意の時にネットワークに送り込むトラ

フィックの最大総帯域幅であり、C_iは、ｉが任意の時にネットワークから受け取るトラフィックの最大総帯域幅である。ＶＰＮ用に予約されるネットワーク容量は、R_iおよびC_iの値と一貫するすべての可能なトラフィック・パターンに十分なものでなければならない。

【００４３】

本発明のある実施形態と一貫するルーティング方式は、洗練されたトラフィック・エンジニアリング機構または追加のネットワーク・シグナリングを必要とせずに、ネットワークが任意の（多分素早く変化する）トラフィック需要を満たすことを可能にする。実際に、ネットワークは、トラフィック分布の変化を検出する必要さえない。必要になる可能性がある、トラフィックに関する唯一の知識は、ネットワークの端で外部インターフェースに接続されるすべてのライン・カードの総容量によって課せられる限度である。

【００４４】

図４を参照すると、本発明の１実施形態と一貫する例示的２フェーズ・ルーティング方式が、物理ビューおよび論理ビューの両方で示されている。フェーズ１（４０１）では、任意のノードｉでネットワークに入るトラフィックの所定の分数α_kが、トラフィックの最終的な宛先と独立に、１つまたは複数の中間ノードｋに分配される。フェーズ２（４０２）では、各ノードｋが、異なる宛先宛のトラフィックを受け取り、受け取ったトラフィックをめいめいの宛先にルーティングする。このルーティング方式を実施する方法の１つが、一部がフェーズ１トラフィックを搬送し、残りがフェーズ２トラフィックを搬送する、ノードの間の固定帯域幅トンネルを形成することである。この２フェーズ・ルーティング戦略が働くのは、これらのトンネルに必要な帯域幅が、トラフィック行列の個々の項目ではなく、ＲおよびＣだけに依存するからである。フェーズ１で、α_1、α_2、…、α_nが、次の式（４）を満足するものであることに留意されたい。

【数３】

$$\sum_{i=1}^{n} \alpha_i = 1 \quad . \qquad (4)$$

【００４５】

この２フェーズ・ルーティング方法を、これから詳細に説明する。最大発信トラフィックR_iを有する所与のノードｉについて、ノードｉは、各ｋ∈Ｎについて、フェーズ１中に中間ノードｋにこのトラフィックの$\alpha_k R_i$量を送る。したがって、フェーズ１の結果としてのノードｉからノードｋへの需要は、$\alpha_k R_i$である。フェーズ１の終りに、ノードｋが、各ノードｉから$\alpha_k R_i$を受け取っている。行上限の合計が、列上限の合計と等しくなければならないので、すべてのソースｉからノードｋで受け取られる総トラフィックは、

【数４】

$$\sum_{i=1}^{n} \alpha_k R_i = \sum_{j=1}^{n} \alpha_k C_j$$

である。同一宛先へのトラフィックが所定の比率で分割されると仮定すると、フェーズ１の後に、ノードｋで受け取られたトラフィックのうちで、ノードｊ宛のトラフィックは、$\alpha_k t_{ij}$である。したがって、フェーズ２中にノードｋからノードｊにルーティングされる必要がある総トラフィックすなわち、ノードｋからノードｊへのトラフィック需要は、次の式（５）で示されるものである。

【数５】

$$\sum_{i \in N} \alpha_k t_{ij} = \alpha_k C_j \quad . \qquad (5)$$

したがって、フェーズ１で、ｋは本質的にｊと同一であり、フェーズ２で、ｋが本質的にｉなので、フェーズ１および２でのルーティングの結果としてのノードｉからノードｊへの総需要は、（$\alpha_j R_i + \alpha_i C_j$）であり、これは、行列Ｔ∈Ｔ（Ｒ，Ｃ）の知識なしで導出することができる。次の３つの特性が、この２フェーズ・ルーティング方式の特性を表す。

【００４６】

（ｉ）ルーティングが、トラフィック変動に気付かない。フェーズ１および２中にルーティングされる必要がある需要は、特定のトラフィック行列Ｔ∈Ｔ（Ｒ，Ｃ）に依存するのではなく、Ｔ（すなわち集合Ｔ（Ｒ，Ｃ））を制約する行および列の合計上限だけに依存する。

【００４７】

（ｉｉ）ルーティングされる需要が、トラフィック行列に独立である。フェーズ１および２でのルーティングの結果としてのノードｉとｊの間の総需要は、$t_{ij}' = \alpha_j R_i + \alpha_i C_j$であり、特定の行列Ｔ∈Ｔ（Ｒ，Ｃ）に依存しない。

（ｉｉｉ）提供される容量が、完全に使用される。行列Ｔ∈Ｔ（Ｒ，Ｃ）のそれぞれについて、このルーティング方式は、フェーズ１および２で関連するポイントツーポイント需要を完全に利用する。

【００４８】

特性（ｉｉ）は、この方式が、行および列の合計上限と配分比α_1、α_2、…、α_nだけに依存し、特定の行列Ｔ∈Ｔ（Ｒ，Ｃ）に依存しない変換された行列$T' = \langle t_{ij}' \rangle$を効果的にルーティングすることによってトラフィック行列Ｔ∈Ｔ（Ｒ，Ｃ）の変動性を扱い、これによって、このルーティング方式が、トラフィック分布の変化に気付かないことを暗示する。

【００４９】

トラフィック分布が行または列の合計上限に従うことの保証は、行または列の合計上限を、ノードで外部インターフェースに接続されるライン・カード容量の合計と等しくなるようにし、これによって、物理レイヤで厳しい形で制約を実施することによって達成することができる。その代わりに、ＤｉｆｆＳｅｒｖ（ｄｉｆｆｅｒｅｎｔｉａｔｅｄ ｓｅｒｖｉｃｅ）タイプのポリシ設定方式（ネットワークに入るトラフィックが、ネットワークの境界で分類され、多分条件付けられ、異なる挙動の集団に割り当てられる）によって、各入口ノードでネットワークに入る総トラフィックをレート制限し、各ノードがオーバーサブスクライブ（ｏｖｅｒ－ｓｕｂｓｃｒｉｂｅ）されないことを保証することができる。

【００５０】

したがって、本発明の１実施形態と一貫するルーティング方式で、フェーズ１中の各ソース・ノードでのルーティング決定は、ネットワーク側の状態情報（たとえば、他方のピアリング・ポイントのトラフィックがどのように変化しているか）を全く必要とせず、フェーズ２中のルーティング決定は、パケット宛先だけに基づく。さらに、このネットワークは、入口点／出口点がオーバーサブスクライブされない限り、すべてのトラフィック分布を満たすことができ、他のキャリアに接続されるライン・カードのハード・レート保証によって、またはノードでネットワークに入るトラフィックのレート制限用のｄｉｆｆｅｒｅｎｔｉａｔｅｄ ｓｅｒｖｉｃｅタイプのポリシ設定方式の実施によって、輻輳を避けることができる。さらに、このルーティング方式は、トラフィック分布の変化に気付かず、これに対して堅牢であり、エンドツーエンド帯域幅保証の提供が、リアルタイムでのネットワークの再構成を必要としない。

【００５１】

図５の流れ図に示されているように、本発明の１実施形態と一貫するルーティング・アーキテクチャは、次の例示的な方法で実施することができる。ステップ５０１で、この方法は、自律システム間ピアリング合意および／または他のキャリアに接続された各ノードのライン・カードのレートを使用して、行（または列）の上限R_i（またはC_i）の計算

を開始する。次に、ステップ５０２で、トラフィック配分比α_1、α_2、…、α_nを計算する（下で詳細に説明する、必要なネットワーク帯域幅を最適化する例示的アルゴリズムを使用して）。次に、ステップ５０３で、ノード対ｉ，ｊのそれぞれについて、フェーズ１のノードｉから１つまたは複数の中間ノードへの帯域幅$\alpha_j R_i$の１組と、フェーズ２の１つまたは複数の中間ノードからノードｊへの帯域幅$\alpha_i C_j$の１組の、２組の接続（たとえば、ＭＰＬＳ　ＬＳＰ、ＩＰトンネル、または光レイヤ回路）を提供する。次に、ステップ５０４で、フェーズ１および２に従ってトラフィックをルーティングする（上で詳細に説明したように、ソース・ノードと中間ノードでのローカル動作だけを必要とする）。次に、ステップ５０５で、ｄｉｆｆｅｒｅｎｔｉａｔｅｄ　ｓｅｒｖｉｃｅタイプのポリシ設定機構を使用して、各ノードでネットワークに入る総トラフィックをレート制限する。次に、ステップ５０６で、たとえば新しいピアリング合意または既存のピアリング合意に対する変更の結果として、行（または列）の上限R_i（またはC_i）が変化したかどうかを判定する。上限が変化していない場合には、この方法は、ステップ５０４に戻って、ルーティング動作を継続する。上限が変化した場合には、ステップ５０７で、α_i配分比を再最適化し、ステップ５０８で、フェーズ１および２中のルーティングに関するＬＳＰ（または光レイヤ回路あるいはＩＰトンネル）の帯域幅をそれ相応に調整し、その後、ステップ５０４に戻る。

【００５２】

前述の方法では、同一の接続内でトラフィックが分割される場合に、同一のエンドツーエンド接続に属するパケットが、出口ノードに順序はずれで到着する可能性がある。この状況は、この方式のフェーズ１でフローごとの分割を使用することによって回避することができる。それに加えておよび／またはその代わりに、下でさらに説明するように、トラフィックのソース・ノードおよび／または宛先ノードに依存するように、トラフィック分割比α_iを一般化することができる。

【００５３】

入口／出口トラフィックに関するリンク容量および制約R_i，C_iを有するネットワークで、ネットワーク内のすべてのリンクの最大使用率を最小にするようにルーティングすることが望ましい。リンクの使用率は、リンクのトラフィックをその容量で割ったものと定義することができる。λ・Ｔ（Ｒ，Ｃ）が、その項目にλを乗じられたＴ（Ｒ，Ｃ）のすべてのトラフィック行列の集合を表す場合に、線形プログラムを使用して、λ・Ｔ（Ｒ，Ｃ）のすべての行列をルーティングできる最大の乗数λ（スループット）を見つけることができる。

【００５４】

等しい分割比すなわち、$\alpha_i = 1/n \quad \forall i \in N$の場合に、ノードｉとｊの間の需要は、$(R_i + C_j)/n$であり、この問題は、その教示が参照によって本明細書に組み込まれる、Ｆ．Ｓｈａｈｒｏｋｈｉ　ａｎｄ　Ｄ．Ｍａｔｕｌａ、「Ｔｈｅ　Ｍａｘｉｍｕｍ　Ｃｏｎｃｕｒｒｅｎｔ　Ｆｌｏｗ　Ｐｒｏｂｌｅｍ」、Ｊｏｕｒｎａｌ　ｏｆ　ＡＣＭ、３７（２）：３１８～３３４、１９９０年で説明されているように、最大同時フロー（ｍａｘｉｍｕｍ　ｃｏｎｃｕｒｒｅｎｔ　ｆｌｏｗ）問題になる。

【００５５】

本発明の１実施形態での例示的なリンク・フローベースの線形計画法定式化を、これから説明するが、この定式化では、フローが、主問題の解で増やされ、重みが、対応する双対問題に対する解で乗法的な形で更新される。主問題および双対問題とそれらの解は、次のように特性を表すことができる。

【００５６】

１．主問題が、ｎ個の変数およびｍ個のリソース制約を有する場合に、双対問題は、ｍ個の変数およびｎ個のリソース制約を有する。したがって、双対問題の制約行列は、主問題の制約行列の転置行列である。

２．主制約と双対変数の間に１対１対応がある、すなわち、双対問題の変数は、主問題の不等式と対になり、主変数と双対制約も同様である。

３．双対問題の目的関数は、主制約の右辺によって決定され、主問題の目的関数と双対制約の右辺も同様である。

【００５７】

下の例示的な線形計画法定式化では、コモディティ・インデックス（ｃｏｍｍｏｄｉｔｙ　ｉｎｄｅｘ）ｋが与えられ、用語「コモディティ」は、ソースと宛先の間のフローを指し、コモディティｋのソース・ノードはｓ（ｋ）、宛先ノードはｄ（ｋ）、コモディティｋに対応するフローの量はｆ（ｋ）によって表される。ベクトルx^k（ｅ）は、ネットワーク内のリンクｅのコモディティｋのフローの量を表し、δ^-（ｉ）およびδ^+（ｉ）は、それぞれノードｉでの着信エッジおよび発信エッジの集合を表す。式（６）および（７）ならびに不等式（８）の制約を有する例示的なリンク・フローベースの線形計画法定式化は、次のように表される。

【数６】

$$\sum_{e\in\delta^-(i)} x^k(e) = \sum_{e\in\delta^+(i)} x^k(e) \qquad \forall i \neq s(k), d(k), \forall k \qquad (6)$$

,

$$\sum_{e\in\delta^+(i)} x^k(e) = a_{s(k)}C_{d(k)} + a_{d(k)}R_{s(k)} \qquad i = s(k), \forall k \qquad (7)$$

,

$$\sum_k x^k(e) \le u_e \quad \forall e \in E \qquad (8)$$

のもとで

【数７】

$$\sum_{i\in N} \alpha_i$$

を最大化すること。

上の線形プログラムに、２組の決定変数すなわち、トラフィック分離比α_iおよび、x^k（ｅ）によって表される、コモディティｋのリンクｅのフローが含まれる。コモディティｋに関する需要が、$\alpha_{s(k)}C_{d(k)}+\alpha_{d(k)}R_{s(k)}$によって与えられることに留意されたい。上の線形プログラムの最適解のα_i値は、α_i^*によって表され、最適の目的関数値は、λ^*によって表され、

【数８】

$$\lambda^* = \sum_i \alpha_i^*$$

である。$\lambda^* \geqq 1$の場合に、この問題は実行可能である、すなわち、所与の需要をネットワーク上でルーティングすることができる。α_i^*値をλ^*倍だけ減らして、実際の分割比を得ることができ、それに沿って需要がルーティングされる明示的なパスを、上の問題の解から決定することができる。値$\lambda^* < 1$の場合に、この問題は実行不能である。この場合には、出口（または入口）制約R_i（C_i）を、$1/\lambda^*$で割ることによってスケール・ダウンすることができ、その場合に、この問題は、所与のリンク容量の下でのルーティングに関して実行可能になる。その代わりに、$1/\lambda^*$を乗ずることによってリンク容量をスケール・アップして、すべての所与の需要のルーティングに対処することができる

。

【００５８】

本発明の１実施形態の例示的なパス・フローベースの線形計画法定式化（高速組合せＦＰＴＡＳ（Ｆｕｌｌｙ　Ｐｏｌｙｎｏｍｉａｌ　Ｔｉｍｅ　Ａｐｐｒｏｘｉｍａｔｉｏｎ　Ｓｃｈｅｍｅ）アルゴリズムを開発するのに使用することができる）を、これから説明する。次の例の定式化では、P_{ij}が、ノードｉからノードｊへのすべてのパスの集合を表し、ｘ（Ｐ）が、パスＰ上のトラフィックを表す。式（９）および不等式（１０）の制約を有する例示的なパス・フローベースの線形計画法定式化は、次のように表される。

【数９】

$$\sum_{P\in P_{ij}} x(P) = \alpha_j R_i + \alpha_i C_j \;\forall i,j \in N, i \neq j \qquad (9)$$

$$\sum_{P:e\in P} x(P) \le u_e \;\forall e \in E \qquad (10)$$

のもとで

【数１０】

$$\sum_{i\in N} \alpha_i$$

を最大化すること。

【００５９】

ネットワークは、一般に、指数関数的な個数のパス（ネットワークのサイズに関して）を有することができるので、前述の（主）線形プログラムは、多分、指数関数的な個数の変数を有する可能性があり、その双対（下で詳細に示す）は、指数関数的な個数の制約を有する可能性がある。したがって、これらのプログラムは、中程度から大きいサイズのネットワークでの実行によく適さない可能性がある。それでも、そのような主／双対定式化は、下で説明するように、この問題の高速多項式時間組合せアルゴリズムに有用である。

【００６０】

高速組合せ近似アルゴリズムを使用して、任意のε＞０について、最適目的関数の（１＋ε）倍までの分割比を計算することができる。εの値は、解の所望の度合の最適性を提供するように選択することができる。このアルゴリズムは、ＰＦＴＡＳ方式であることが好ましく、入力サイズおよび１／εの多項式である時間で動作する。このアルゴリズムでは、各ステップで主解および双対解が維持されるので、主目的関数値と双対目的関数値の比を計算することによって、最適性ギャップを推定することができる。

【００６１】

上で式（９）および不等式（１０）に示された線形プログラムの双対定式化では、変数ｗ（ｅ）を不等式（１０）で各リンク容量制約に関連付け、変数π_{ij}を式（９）の需要制約に関連付ける。ＳＰ（ｉ，ｊ）は、次の式（１１）に示されているように、重みｗ（ｅ）の下での最短パスＰ∈P_{ij}を表す。

【数１１】

$$SP(i,j) = \min_{P\in P_{ij}} \sum_{e\in P} w(e) \qquad (11)$$

単純化および双対変数π_{ij}の消去の後に、双対線形計画法定式化を、不等式の制約（１２）から（１３）を用いて、次のように書くことができる。

【数１２】

$$\sum_{i:i\neq k} R_i SP(i,k) + \sum_{j:j\neq k} C_j SP(k,j) \geq 1 \ \forall k \in N \quad (12)$$

$$w(e) \geq 0 \ \forall e \in E \quad (13)$$

のもとで

【数１３】

$$\sum_{e\in E} u_e w(e)$$

を最小化すること。

【００６２】

所与のノードｋについて、Ｖ（ｋ）は、不等式（１２）の制約の左辺を表す。重みｗ（ｅ）を与えられれば、２つの最短パス計算すなわち、ノードｋをルートとし、すべての宛先に達する最短路木の計算およびすべての他のノードからノードｋに達する逆の最短路木の計算によって、Ｖ（ｋ）を多項式時間で計算できることに留意されたい。

重みｗ（ｅ）を与えられれば、次の不等式（１４）が満足される場合に限って、双対問題の実行可能な解が存在する。

【数１４】

$$\min_{k\in N} V(k) \geq 1. \quad (14)$$

【００６３】

このアルゴリズムは、等しい初期重みｗ（ｅ）＝δ（量δは、εに依存し、後で導出される）から始まる。次に、下のステップ（１から５）を、双対実行可能制約が満足されるまで繰り返す。

（１）図６に示されているように、そのＶ（ｋ）が最小になるノード

【数１５】

$$\bar{k}$$

を計算し、これによって、リンク

【数１６】

$$\bar{k}$$

を識別し、すべてのｉについて、ノードｉからノード

【数１７】

$$\bar{k}$$

へのパスＰ$_i$を識別し、すべてのｊについてノード

【数１８】

$$\bar{k}$$

からノードｊへのパスＱ$_i$を識別する。

【００６４】

（２）ｅ∈Ｅのそれぞれについて、Ｎｐ（ｅ）を、Ｐ$_i$がリンクｅを含むノードｉの集合と定義し、Ｎ$_Q$（ｅ）を、Ｑ$_j$がリンクｅを含むノードｊの集合と定義する。分数αを、次の式（１５）を使用して計算する。

【数１９】

$$a = \min_{e \in E} \frac{u_e}{\sum_{i \in N_P(e)} R_i + \sum_{j \in N_Q(e)} C_j} \quad . \qquad (15)$$

【００６５】

（３）すべてのｉについて、量αＲ$_i$のフローをパスＰ$_i$で送り、すべてのｊについて、量αＣ$_j$のフローをパスＱ$_j$で送り、リンクｅで送られる総フローΔ（ｅ）を、すべてのｅ∈Ｅについて計算する。リンクｅのフローを、Δ（ｅ）だけ増分する。

（４）すべてのｅ∈Ｅについて、ｗ（ｅ）←ｗ（ｅ）（１＋εΔ（ｅ）／ｕ$_e$）として重みｗ（ｅ）を更新する。

（５）ノード

【数２０】

$\alpha_{\bar{k}}$

に関連する分割比

【数２１】

$\bar{k}$

を、αだけ増分する。

【００６６】

前述の手順が終了する時に、双対実行可能制約が満足される。しかし、各ステージの元の（残余ではない）リンク容量がこの手順で使用されるので、各リンクの主容量制約を侵害している場合がある。これを直すために、分割比を均一にスケールダウンし、その結果、容量制約に従わせることができる。

【００６７】

上述の例示的方法を実施するのに使用できる例示的アルゴリズムの擬似コードを、下で示す。この擬似コードでは、配列ｆｌｏｗ（ｅ）によって、リンクｅのトラフィックが記憶される。変数Ｇは、０に初期化され、双対制約が満足されないままである限り、１未満になる。ループ全体が終了した後に、各リンクｅの容量制約を侵害した倍率が、計算され、配列ｓｃａｌｅ（ｅ）に格納される。最後に、α$_i$の値を、最大の容量侵害倍率で割り、結果の値を、最適値として出力する。

【００６８】

この例示的アルゴリズムに関する２つの定理を次に示す。

定理１：Ｌ＝（ｎ－１）（Σ$_{i \in N}$Ｒ$_i$＋Σ$_{j \in N}$Ｃ$_j$）であり、Ｌ’が、Ｒ$_i$’およびＣ$_j$’の最小の非０の値であり、εおよびδの値が、下で示すアルゴリズムの近似係数保証に関係すると考えるならば、任意の所与のε’＞０について、このアルゴリズムは、次の式（１６）および（１７）に関して最適値の（１＋ε’）倍以内の目的関数値を有する解を計算する。

【数２２】

$$\delta = \frac{1+\varepsilon}{L'\left[(1+\varepsilon)\frac{L}{L'}\right]^{1/\varepsilon}} \quad , \qquad (16)$$

$$\varepsilon = 1 - \frac{1}{\sqrt{1+\varepsilon'}} \quad . \qquad (17)$$

定理２：定理１に従って所望の近似係数保証を提供するように選択された任意の所与のε＞０について、このアルゴリズムは、入力サイズおよび１／εの多項式であり、

【数２３】

$$O(\frac{nm}{\varepsilon}(m+n\log n)\log_{1+\varepsilon}\frac{L}{L'})$$

である。

次の例示的擬似コードを使用して、上で示した例示的アルゴリズムを実施することができる。

$\alpha_k \leftarrow 0 \; \forall k \in N$；

$w(e) \leftarrow \delta \; \forall e \in E$；

$flow(e) \leftarrow 0 \; \forall e \in E$；

$G \leftarrow 0$；

while G＜1 do

リンク・コストw（ｅ） ∀ｉ，ｊ∈Ｎの下でｉからｊへのコストＳＰ（ｉ，ｊ）の最短パスを計算する；

$V(k) \leftarrow \Sigma_{i \neq k} R_i SP(i,k) + \Sigma_{j \neq k} C_j SP(k,j)$；

$G \leftarrow \min_{k \in N} V(k)$；

if G≧1 break；

【数２４】

$\bar{k}$

が、ｇ（ｋ）が最小になるノードであるものとする；

P_iが、すべてのｉについて、ｉから

【数２５】

$\bar{k}$

への最短パスであるものとする；

Q_jが、すべてのｊについて、

【数２６】

$\bar{k}$

からｊへの最短パスであるものとする；

$N_P(e) \leftarrow \{i : e \text{を含む} P_i\}$ for all e；

$N_Q(e) \leftarrow \{j : e \text{を含む} Q_j\}$ for all e；

【数２７】

$$\alpha \leftarrow \min_{e \in E} \frac{u_e}{\sum_{i \in N_P(e)} R_i + \sum_{j \in N_Q(e)} C_j};$$

すべてのｉについてパスP_iでαR_iフローを送り、すべてのｊについてパスQ_jでαC_jフローを送り、すべてのｅについてリンクｅでの結果の容量使用量Δ（ｅ）を計算する；

$flow(e) \leftarrow flow(e) + \Delta(e)$ for all e；

$w(e) \leftarrow w(e)(1 + \varepsilon \Delta(e) / u_e)$ for all e；

【数２８】

$$\alpha_k \leftarrow \alpha_k + \alpha;$$

ｅｎｄ　ｗｈｉｌｅ
ｓｃａｌｅ（ｅ）←ｆｌｏｗ（ｅ）／ｕ$_e$　ｆｏｒ　ａｌｌ　ｅ∈Ｅ；
ｓｃａｌｅ＿ｍａｘ←ｍａｘ$_{e \in E}$　ｓｃａｌｅ（ｅ）；
α$_k$←α$_k$／ｓｃａｌｅ＿ｍａｘ　ｆｏｒ　ａｌｌ　ｋ∈Ｎ；
最適トラフィック分割比としてα$_k$を出力する；

【００６９】

定理１および２の証明および基礎になる補助定理は、次の通りである。

双対重みｗ（ｅ）の集合に対して、Ｄ（ｗ）が双対目的関数値を表し、Γ（ｅ）が、すべてのノードｋ∈Ｎに関する不等式（１２）に示された双対プログラム制約の左辺の最小値を表すものとすると、この双対プログラムを解くことは、Ｄ（ｗ）／Γ（ｗ）が最小になる重みｗ（ｅ）の集合を見つけることと同等である。Ｄ（ｗ）／Γ（ｗ）の最適目的関数値を、θによって表す、すなわち、θ＝ｍｉｎ$_w$Ｄ（ｗ）／Γ（ｗ）である。上のｗｈｉｌｅループの反復ｔの初めの重み関数を、ｗ$_{t-1}$によって表し、ｆ$_{t-1}$は、反復ｔ－１の終りまでのΣ$_{j \in N}$α$_j$（主目的関数）の値である。上で定義したように、Ｌ＝（ｎ－１）（Σ$_{i \in N}$Ｒ$_i$＋Σ$_{j \in N}$Ｃ$_j$）であり、Ｌ’は、Ｒ$_i$’およびＣ$_j$’の最小の非０の値である。このアルゴリズムは、反復Ｎの後に終了する。

補助定理１：アルゴリズムのすべての反復ｔ　∀１≦ｔ≦Ｋの終りに、次の不等式（１７．１）が満足される。

【数２９】

$$\Gamma(w_t) \le \delta L \prod_{j=1}^{t} [1 + \frac{\varepsilon}{\theta}(f_j - f_{j-1})]$$

証明：Ｖ（ｋ）が最小になるノードは、

【数３０】

$$k = \bar{k}$$

であり、反復ｔ中にフローが増やされる対応するパスは、上で定義したように、Ｐ$_i$，Ｑ$_j$によって表される。重みが、ｗ$_t$（ｅ）＝ｗ$_{t-1}$（ｅ）（１＋εΔ（ｅ）／ｕ$_e$）∀ｅ∈Ｅとして更新され、Δ（ｅ）は、反復ｔ中にリンクｅで送られる総フローである。これを使用すると、次の式（１７．２）に示されているように、Ｄ（ｗ$_t$）を導出することができる。

【数３１】

$$D(w_t)=\sum_{e\in E}u_e w_t(e)$$
$$=\sum_{e\in E}u_e w_{t-1}(e)+\varepsilon\sum_{e\in E}\Delta(e)w_{t-1}(e)$$
$$=D(w_{t-1})+\varepsilon\sum_{e\in E}w_{t-1}(e)[\sum_{i\in N_P(e)}\alpha R_i+\sum_{j\in NQ(e)}\alpha C_j]$$
$$=D(w_{t-1})+\varepsilon\alpha[\sum_i R_i\sum_{e\in P_i}w_{t-1}(e)+\sum_j C_j\sum_{e\in Q_j}w_{t-1}(e)]$$
$$=D(w_{t-1})+\varepsilon\alpha\Gamma(w_{t-1}) \quad . \qquad (17.2)$$

最初の反復まで戻って各反復に上で導出された式を使用すると、Ｄ（ｗ$_t$）を、次の式（１７．３）のように定義することができる。

【数３２】

$$D(w_t)=D(w_0)+\varepsilon\sum_{j=1}^{t}(f_j-f_{j-1})\Gamma(w_{j-1}) \quad . \qquad (17.3)$$

ここで、重み関数ｗ$_t$－ｗ$_0$を考慮すると、Ｄ（ｗ$_t$－ｗ$_0$）＝Ｄ（ｗ$_t$）－Ｄ（ｗ$_0$）であることがわかり、パスＰ$_i$，Ｑ$_j$のどれもが、多くともｎ－１ホップの長さなので、Γ（ｗ$_0$）≦Σ$_i$（ｎ－１）δＲ$_i$＋Σ$_j$（ｎ－１）δＣ$_j$＝δＬであることもわかる。したがって、Γ（ｗ$_t$－ｗ$_0$）≧Γ（ｗ$_t$）－δＬである。θは、最適の双対目的関数値なので、次の不等式（１７．４）および（１７．５）が成り立つ。

【数３３】

$$D(w_t)=\sum_{e\in E}u_e w_t(e)$$
$$=\sum_{e\in E}u_e w_{t-1}(e)+\varepsilon\sum_{e\in E}\Delta(e)w_{t-1}(e)$$
$$=D(w_{t-1})+\varepsilon\sum_{e\in E}w_{t-1}(e)[\sum_{i\in N_P(e)}\alpha R_i+\sum_{j\in NQ(e)}\alpha C_j]$$
$$=D(w_{t-1})+\varepsilon\alpha[\sum_i R_i\sum_{e\in P_i}w_{t-1}(e)+\sum_j C_j\sum_{e\in Q_j}w_{t-1}(e)]$$
$$=D(w_{t-1})+\varepsilon\alpha\Gamma(w_{t-1}) \quad . \qquad (17.2)$$

不等式（１７．５）を式（１７．３）と比較することによって、次の不等式（１７．６）を導出することができる。

【数３４】

$$\Gamma(w_t)\le\delta L+\frac{\varepsilon}{\theta}\sum_{j=1}^{t}(f_j-f_{j-1})\Gamma(w_{j-1}) \quad . \qquad (17.6)$$

【００７０】

補助定理１の特性を、不等式（１７．６）および反復回数に対する数学的帰納法を使用して証明することができる。ｗ$_0$（ｅ）＝δ　∀ｅ∈Ｅかつ Γ（ｗ$_0$）≦δＬなので、この帰納法の基礎になる事例（反復ｔ＝１）が成り立つことに留意されたい。次に、リンク

容量制約が侵害されないことを保証するために、このアルゴリズムが終了した時に主解の目的関数値ｆ$_K$をスケーリングする必要がある倍率の推定を行うことができる。

補助定理２：このアルゴリズムが終了する時に、主実現可能性を保証するために、主解を、多くとも次の値の倍率だけスケーリングしなければならない。

【数３５】

$$\log_{1+\varepsilon}\frac{1+\varepsilon}{\delta L'} \quad .$$

【００７１】

証明：リンクｅおよびそれに関連する重みｗ（ｅ）を考慮すると、ｗ（ｅ）の値は、フローがエッジｅで増やされる時に更新される。リンクｅ上のフロー増大のシーケンス（反復ごとの）は、Δ$_1$、Δ$_2$、…、Δ$_r$であり、ｒ≦Ｋである。リンクｅでルーティングされる総フローは、その容量をκ倍だけ超える、すなわち、

【数３６】

$$\sum_{t=1}^{r}\Delta_t = \kappa u_e$$

である。このアルゴリズムは、Γ（ｗ）≧１の時に終了し、双対重みは、各反復の後に多くとも１＋ε倍だけ更新されるので、Γ（ｗ$_K$）≦１＋εである。上で述べた増大の直前に、少なくともＬ’の係数を有する重みｗ（ｅ）が、Γ（ｗ）の合計成分の１つである。したがって、Ｌ’ｗ$_K$（ｅ，ｆ）≦１＋εであり、ｗ$_K$（ｅ，ｆ）の値は、次式（１７．７）によって与えられる。

【数３７】

$$w_K(e,f) = \delta\prod_{t=1}^{r}(1+\frac{\Delta t}{u_e}\varepsilon)\,. \qquad (17.7)$$

∀ｘ≧０かつすべての≦β≦１について（１＋βｘ）≧（１＋ｘ）$^{\beta}$であるという事実と、設定ｘ＝εかつβ＝（Δ$_t$／ｕ$_e$）≦１を使用すると、次の不等式（１７．８）および（１７．９）が成り立つ。

【数３８】

$$\frac{1+\varepsilon}{L'} \geq w_K(e,f) \geq \delta\prod_{t=1}^{r}(1+\varepsilon)^{\Delta_t/u_e}$$

$$\geq \delta(1+\varepsilon)^{\sum_{t=1}^{r}\Delta_t/u_e}$$

$$\geq \delta(1+\varepsilon)^{\kappa} \quad , \qquad (17.8)$$

$$\kappa \leq \log_{1+\varepsilon}\frac{1+\varepsilon}{\delta L'}\,. \qquad (17.9)$$

定理１の証明：補助定理１および１＋ｘ≦ｅ^{x}　∀ｘ＞０という事実を使用すると、次の不等式（１７．１０）を導出することができる。

【数３９】

$$\Gamma(w_t) \le \delta L \prod_{j=1}^{t} e^{\frac{\varepsilon}{\theta}(f_j - f_{j-1})}$$

$$\le \delta L e^{\varepsilon f_t/\theta} \quad . \qquad (17.10)$$

前述のステップの単純化は、ｊに関する合計（ｆ$_{j}$－ｆ$_{j-1}$）のテレスコーピック・キャンセレーション（ｔｅｌｅｓｃｏｐｉｃ　ｃａｎｃｅｌｌａｔｉｏｎ）を使用する。このアルゴリズムは、反復Ｋの後に終了するので、Γ（ｗ）≧１である。したがって、次の不等式（１７．１１）および（１７．１２）が成り立つ。

【数４０】

$$1 \le \Gamma(w_K) \le \delta L e^{\varepsilon f_t/\theta} \qquad (17.11)$$

$$\frac{\theta}{f_K} \le \frac{\varepsilon}{\ln(1/\delta L)} \quad . \qquad (17.12)$$

補助定理２から、スケーリングの後の実現可能な主解の目的関数値は、少なくとも次の値である。

【数４１】

$$\frac{f_K}{\log_{1+\varepsilon} \frac{1+\varepsilon}{\delta L'}}$$

【００７２】

主解の近似係数は、多くとも、主解と双対解の間の（比）ギャップである。不等式（１７．１２）を使用すると、このギャップを、次の不等式（１７．１３）によって与えることができる。

【数４２】

$$\frac{\theta}{f_K} \le \frac{\varepsilon \log_{1+\varepsilon} \frac{1+\varepsilon}{\delta L'}}{\ln(1/\delta L)}$$

$$\le \frac{\varepsilon}{\ln(1+\varepsilon)} \frac{\ln \frac{1+\varepsilon}{\delta L'}}{\ln(1/\delta L)} \quad . \qquad (17.13)$$

量

【数４３】

$$\ln \frac{1+\varepsilon}{\delta L'} / \ln(1/\delta L)$$

は、

【数４４】

$$\delta = \frac{1+\varepsilon}{L'} / [(1+\varepsilon)\frac{L}{L'}]^{1/\varepsilon}$$

に関して１／（１－ε）と等しい。δのこの値を使用すると、近似係数は、次の不等式（１７．１４）によって上限を与えられる。

【数４５】

$$\frac{\varepsilon}{\ln(1+\varepsilon)}\frac{1}{(1-\varepsilon)} \le \frac{\varepsilon}{(\varepsilon-\varepsilon^2/2)(1-\varepsilon)} \le \frac{1}{(1-\varepsilon)^2}. \qquad (17.14)$$

１＋ε’＝１／（１－ε）^{2}とし、εについて解くことによって、定理１で述べたεの値が得られる。

定理２の証明：まず、このアルゴリズムの各反復の実行時間は、ノード

【数４６】

$$\bar{k}$$

およびそれに関連するパスＰ$_i$，Ｑ$_j$がフローの増大のために選択されている間と考えられる。このノードおよびパスの選択は、参照によって本明細書に組み込まれる、Ｒ．Ｋ．Ａｈｕｊａ、Ｔ．Ｌ．Ｍａｇｎａｎｔｉ、ａｎｄ　Ｊ．Ｂ．Ｏｒｌｉｎ、「Ｎｅｔｗｏｒｋ　Ｆｌｏｗｓ：Ｔｈｅｏｒｙ，Ａｌｇｏｒｉｔｈｍｓ，ａｎｄ　Ａｐｐｌｉｃａｔｉｏｎｓ」、Ｐｒｅｎｔｉｃｅ　Ｈａｌｌ、１９９３年２月で説明されているように、フィボナッチ・ヒープと共にダイクストラの最短パス・アルゴリズムを使用してＯ（ｎｍ＋ｎ^2　ｌｏｇ　ｎ）時間で実施できる全対最短パス計算を伴う。１つの反復中の他のすべての演算が、この全対最短パス計算が占める時間に吸収され（一定の倍率まで）、反復ごとに合計Ｏ（ｎ（ｍ＋ｎ　ｌｏｇ　ｎ））の時間につながる。

【００７３】

次に、各反復で、フローがパスＰ$_i$，Ｑ$_j$に沿って増やされ、値が、その反復中にリンクｅで送られる総フローΔ（ｅ）が多くともｕ$_e$になるものであるという事実に関して、このアルゴリズムが停止する前の反復の回数を推定する。したがって、少なくとも１つのリンクｅで、Δ（ｅ）＝ｕ$_e$であり、ｗ（ｅ）が、１＋ε倍だけ増える。

ここで、固定されたｅ∈Ｅの重みｗ（ｅ）を検討する。ｗ$_0$（ｅ）＝δかつｗ$_K$（ｅ）≦（１＋ε）／Ｌ’なので、この重みを反復に関連付けることができる回数の最大値は、次式（１８）によって定義することができる。

【数４７】

$$\log_{1+\varepsilon}\frac{1+\varepsilon}{\delta L'} = \frac{1}{\varepsilon}(1+\log_{1+\varepsilon}\frac{L}{L'}) = O(\frac{1}{\varepsilon}\log_{1+\varepsilon}\frac{L}{L'}). \qquad (18)$$

合計ｍ個のｗ（ｅ）があるので、反復の総数は、

【数４８】

$$O(\frac{m}{\varepsilon}\log_{1+\varepsilon}\frac{L}{L'})$$

を上限とする。これに反復当たりの実行時間を乗じると、全体的なアルゴリズム実行時間を、

【数４９】

$$O(\frac{nm}{\varepsilon}(m+n\log n)\log_{1+\varepsilon}\frac{L}{L'})$$

として得ることができる。ｌｏｇ （Ｌ／Ｌ’）が、ｌｏｇ ｎとＲ$_i$およびＣ$_j$値を表すのに使用されるビット数の多項式であることに留意されたい。

【００７４】

次に、上で示したルーティング方式の容量性能を検討すると、ルーティング方式の「ルーティング保証」を、まず定義しなければならず、次に、Ｔ（Ｒ，Ｃ）のすべての行列をルーティングできるすべての方式のクラスの最良の可能な方式のルーティング保証と比較しなければならない。トラフィック行列に対するリンク容量および上限Ｒ$_i$，Ｃ$_i$を有するネットワークについて、最適目的関数を、λ＊と表し、これは、本発明の１実施形態と一貫するルーティグ方式によってλ＊－Ｔ（Ｒ，Ｃ）のすべての行列をルーティングできることの保証を提供する、上で説明した線形問題定式化の出力である。図７からわかるように、行合計

【数５０】

$$\hat{R}$$

および列合計

【数５１】

$$\hat{C}$$

のルーティング保証の概略は、すべてのルーティング方式によって許されるλの最高の可能な値が

【数５２】

$$\hat{\lambda}$$

であるとするならば

【数５３】

$$\lambda^* \leq \hat{\lambda}$$

であり、ルーティング方式の効率は、量

【数５４】

$$\lambda^* / \hat{\lambda} (\leq 1)$$

によって測定することができる。

値

【数５５】

$$\hat{\lambda}$$

は、計算が困難である可能性がある。それでも、単一の行列Ｔ∈Ｔ（Ｒ，Ｃ）が存在すると仮定し、最大の乗数λ（Ｔ）を計算する、たとえば、λ（Ｔ）・Ｔを所与のリンク容量を有するネットワーク内で実行可能にルーティングできるように最大並列フロー定式化を使用すると、

【数５６】

$$\hat{\lambda} \leq \lambda(T)$$

であり、したがって、次の不等式（１９）が満足される。
【数５７】

$$\frac{\lambda^*}{\lambda(T)} \leq \frac{\lambda^*}{\hat{\lambda}} \leq 1 \quad . \qquad (19)$$

【００７５】
したがって、任意のトラフィック行列Ｔ∈Ｔ（Ｒ，Ｃ）について、量λ＊／λ（Ｔ）は、ルーティング方式の効率の下限である。ルーティング効率に関するより近い下限を得るために、λ（Ｔ）が最小である行列Ｔ∈Ｔ（Ｒ，Ｃ）を識別しなければならないが、１つの行列がルーティングすべきかなりの帯域幅をとるので、これは計算が困難である可能性がある。次の例示的なヒューリスティック手法を使用して、ルーティングすべき最大の帯域幅をとる行列を近似することができ、ここで、Ｃ（Ｔ）は、行列Ｔ∈Ｔ（Ｒ，Ｃ）をルーティングするための最小帯域幅を表す。Ｃ（Ｔ）を最小にする行列Ｔ∈Ｔ（Ｒ，Ｃ）は、線形計画法定式化を使用することによって、多項式時間で計算可能である。この問題は、容量制約を有しないので、ノードｉからノードｊへのトラフィックが、単一の最短パスに沿って分割不能にルーティングされると仮定することができる。d_{ij}が、ノードｉからノードｊへの最短パスのホップ・カウントを表すならば、ルーティングに最大の帯域幅をとるトラフィック行列Ｔ∈Ｔ（Ｒ，Ｃ）の判定という問題は、式（２０）および（２１）と不等式（２２）の制約を有する、次の例示的な線形プログラムとして定式化することができる。
【数５８】

$$\sum_{j \in N, j \neq i} t_{ij} = R_i \, \forall i \in N \quad , \qquad (20)$$

$$\sum_{j \in N, j \neq i} t_{ji} = C_i \, \forall i \in N \quad , \qquad (21)$$

$$t_{ij} \geq 0 \ \forall i, j \in N \quad . \qquad (22)$$

のもとで、
$\Sigma_{i,j \in N} d_{ij} t_{ij}$を最大化すること。
必要な帯域幅は、この線形プログラムの目的関数であり、行および列の合計上限（Ｔ（Ｒ，Ｃ）を定義する）が、制約を形成する。この定式化を使用して、行列Ｔおよび値λ（Ｔ）を計算して、本発明の１実施形態と一貫するルーティング方式の効率の下限を提供することができる。
【００７６】
本発明の１実施形態と一貫する２フェーズ・ルーティング方式の分割比の２つの変形が、次のようなルーティング方式の一般化をもたらす。
（Ｉ）ソースと宛先の両方に依存する分割比：この方式では、ノードｉから発し、宛先がノードｊであるトラフィックの分数
【数５９】

$$\alpha_k^{ij}$$

が、中間ステージでノードｋにルーティングされると仮定する。第２フェーズでノードｉとｊの間に必要な容量を計算する。第１フェーズで、ノードｉとｊの間に必要な容量は、次の不等式（２３）によって定義される。

【数６０】

$$\sum_k \alpha_j^{ik} t_{ik} \le \max_k \alpha_j^{ik} \sum_k t_{ik} \le \max_k \alpha_j^{ik} R_i \quad . \qquad (23)$$

第２フェーズについて、ノードｉとｊの間に必要な容量は、次の不等式（２４）によって定義される。

【数６１】

$$\sum_k a_i^{kj} t_{kj} \le \max_k \alpha_i^{kj} \sum_k t_{kj} \le \max_k \alpha_i^{kj} C_j \quad . \qquad (24)$$

したがって、この２つのフェーズでノードｉとｊの間に必要な容量は、次の不等式（２５）によって定義することができる。

【数６２】

$$C_{ij} \ge \alpha_j^{ik} R_i + a_i^{mj} C_j \; \forall k \forall m \quad . \qquad (25)$$

（ＩＩ）ソースだけに依存する分割比：この方式では、

【数６３】

$$\alpha_k^i$$

が、ノードｉからノードｋに入るトラフィックの分数を表す。フェーズ１でのノードｉからノードｊへのトラフィックの量は、

【数６４】

$$\alpha_j^i R_i$$

によって与えられ、これは、フェーズ１でのノードｉとｊの間の必要な容量である。ノードｉとｊの間の必要な容量は、次の不等式（２６）によって定義することができる。

【数６５】

$$\sum_k \alpha_k^i t_{kj} \le \max_k \alpha_k^i \sum_k t_{kj} \le \max_k \alpha_k^i C_j \quad . \qquad (26)$$

したがって、ノードｉとｊの間の必要な総容量Ｃ$_{ij}$は、次の不等式（２７）によって与えられる。

【数６６】

$$C_{ij} \ge \alpha_j^i R_i + \alpha_i^k C_j \; \forall k \qquad . \qquad (27)$$

どちらの場合でも、制約が、線形であり、トラフィック行列の個々の要素と独立であり、行合計および列合計だけに依存することに留意されたい。

【００７７】

上で示した実施形態を使用するシミュレーションを、キャリア・バックボーン・ネットワークを表す２つのネットワーク・トポロジに対して、そのサイズ範囲で実行した。図８からわかるように、第１のネットワークは、ノードｎ１からｎ１５を含み、２８個の両方向リンクを有する１５ノード・ネットワークである。第２のネットワークは、３３個の両方向リンクを有する２０ノード・ネットワークである（図示せず）。異なる実行について、各ネットワーク・リンクの容量を、集合｛ＯＣ－３，ＯＣ－１２，ＯＣ－４８，ＯＣ－１９２｝から選択した。結果について、Ｒ$_i$’およびＣ$_i$’が等しいと仮定し、１に正規

化した、すなわち、$R_i = C_i = 1 \forall i$である。行列Ｔは、上の２つのトポロジすなわち、（Ｉ）ソースと宛先の両方に依存する分割比および（ＩＩ）ソースだけに依存する分割比のそれぞれについて計算した。量λ（Ｔ）は、均等（λ_{equal}と表す）および不等（$\lambda_{unequal}$と表す）の両方のトラフィック分割比のλ値の上限であり、λ（Ｔ）≧$\lambda_{unequal} \geqq \lambda_{equal}$である。

【００７８】

図９および１０に、５つの異なる実行に関する上の３つのλ値のプロットを示すが、λ値の相対的な順序付けは、期待通りである。図からわかるように、本発明の１実施形態と一貫する方法のルーティング効率は、プロットされた事例のすべてで、１．０に非常に近い。両方のネットワークのプロットについて、ルーティング効率は、０．９から０．９９まで変化し、したがって、本発明の１実施形態と一貫する方法が、最適に近い動作が可能であることが示される。この結果から、分割比α_iが不等になることを許容される時に、ネットワーク・スループットが増加することも示される。１５ノード・トポロジ実行について、スループットの増加比率（$\lambda_{unequal} - \lambda_{equal}$）／$\lambda_{equal}$は、１０％から８５％まで変化する。２０ノード・トポロジでは、増加率が、２％から５３％まで変化する。この結果から、次の２つの結論が引き出される：（１）本発明の１実施形態と一貫するルーティング方式は、トラフィック分布から選択された単一の行列よりかなり低くはないネットワーク・スループットで、トラフィックの不確かさ（定義されたトラフィック変動モデルの下の）がある状態で効率的にルーティングすることができる場合があり、（２）トラフィック分割比を不等にすることを許容することによって、ネットワーク・スループットを、均等分割比の場合より大きく高めることができる。

【００７９】

したがって、本発明の１実施形態と一貫するルーティング戦略が、動的ルーティング変更を必要とせず、高い容量オーバープロビジョニングを必要とせずに、ネットワークの極端なトラフィック変動性の処理に関する複数の既知の問題に対処できることがわかった。ルーティング適合なしでトラフィック変動を処理する能力は、より安定した堅牢なインターネット挙動につながることができる。本発明の１実施形態と一貫するルーティング方式を使用することによって、サービス・プロバイダが、（ｉ）単一行列のルーティングのネットワーク・スループットにかなり近いネットワーク・スループットで、（ｉｉ）リアル・タイムでトラフィック変動を検出する必要もネットワークを再構成する必要もなしに、すべてのトラフィック分布（定義されたモデルの下の）をルーティングできるようになる。したがって、本発明は、最短パス・ルーティングより実施がそれほど複雑ではない単純なネットワーク・ルーティング方式を提供し、この方式は、次の追加の有利な特性を有する：（ｉ）この方式は、入口－出口リンクの容量制約の中で許容可能なすべてのトラフィック・パターンを効率的に処理でき、（ｉｉ）この方式は、ルーティング・パラメータ（リンク重みまたはルーティング・ポリシなど）の動的再構成を必要とせずに、高いトラフィック変動性の下でネットワーク輻輳を避けることができ、（ｉｉｉ）この方式は、帯域幅効率がよい可能性があり、この方式は入口－出口容量制約を受けるすべての可能なトラフィック・パターンを処理できるが、この方式の容量要件は、単一の悪いトラフィック・パターンに対処するのに必要な方式の容量要件に近いものとすることができる。

【００８０】

本発明の１実施形態と一貫するルーティングの方法は、ネットワーク・サービス・レベル容量のより効率的な使用率、ネットワーク・ノードでのルータの輻輳の低減、およびネットワークのより高いパケット・スループットという長所の１つまたは複数を提供することができる。この方法は、集中ネットワーク管理システムによって、ネットワークの各ノードによって、またはこの両方によって、要求されたＬＳＰについて実施することができる。結果をネットワーク・ノードに分配する集中ネットワーク管理システムを使用する実施形態は、新しいパスのプロビジョニングの調整に好ましい可能性がある。ネットワークの各ノードでの分散実施形態は、集中ネットワーク管理システムが存在しない時、および／または要求されたＬＳＰが、ネットワークを介してルーティングされる制御パケットを

用いて実施される分散要求である場合に、好ましい可能性がある。

【００８１】

本発明の１実施形態と一貫するルーティングの方法のさまざまな機能は、回路要素を用いて実施することができ、あるいは、ディジタル領域で、ソフトウェア・プログラムの処理ステップとして実施することもできる。そのようなソフトウェアは、たとえば、ディジタル信号プロセッサ、マイクロコントローラ、または汎用コンピュータで使用することができる。

【００８２】

本明細書で使用される用語「ルータ」が、単一のハードウェア・デバイスもしくはスイッチ・ファブリックなどの複数の相互接続されたハードウェア・デバイス、ソフトウェア要素とハードウェア要素の組合せ、またはソフトウェア・プログラムを指すことができることを理解されたい。

【００８３】

本発明は、方法およびこれらの方法を実践する装置の形で実施することができる。本発明は、フロッピ・ディスケット、ＣＤ－ＲＯＭ、ハード・ドライブ、または他の機械可読記憶媒体などの有形の媒体内で実施されるプログラム・コードの形で実施することもでき、この場合に、そのプログラム・コードがコンピュータなどの機械にロードされ、これによって実行される時に、その機械が、本発明を実践する装置になる。本発明は、たとえば記憶媒体に保管されるか、機械にロードされかつ／またはこれによって実行されるか、電気配線を介する、光ファイバを介する、または電磁放射を介するなどの媒体を介して送信される、プログラム・コードの形で実施することもでき、この場合に、プログラム・コードがコンピュータなどの機械にロードされ、これによって実行される時に、その機械が、本発明を実践する装置になる。汎用プロセッサで実施される時に、プログラム・コード・セグメントは、プロセッサと組み合わされて、特定の論理回路に類似して動作する独自の装置を提供する。

【００８４】

本明細書に記載のルーティングの例示的方法の工程が、必ずしも記載の順序で実行される必要がないことを理解されたく、そのような方法の工程の順序が、単に例示的であると理解されたい。同様に、本発明のさまざまな実施形態と一貫するルーティング方法で、追加の工程をそのような方法に含めることができ、ある工程を省略しまたは組み合わせることができる。

【００８５】

請求項で表される本発明の趣旨および範囲から逸脱せずに、当業者が、本発明の性質を説明するために説明され、図示された部分の詳細、材料、および配置のさまざまな変更を行えることを、さらに理解されたい。

【図面の簡単な説明】

【００８６】

【図１】他のパケット・ネットワークの間の通信を可能にするリンクを介して相互接続されたノードを有する従来技術の例示的バックボーン・ネットワークを示す図である。

【図２】図１のバックボーン・ネットワークによって、入口点から出口点へパケットをルーティングするのに使用されるカプセル化されたパケットを示す図である。

【図３】本発明の１実施形態と一貫するラベル交換パスをルーティングするサービス・レベル保証を有するルーティングの方法を使用する、相互接続されたノードのネットワークを示す図である。

【図４】本発明の１実施形態と一貫する例示的２フェーズ・ルーティング方式の物理ビューおよび論理ビューを示す図である。

【図５】本発明の１実施形態と一貫するルーティング・アーキテクチャの例示的な方法を示す流れ図である。

【図６】本発明の１実施形態と一貫する例示的な主－双対線形プログラムの１工程を示す図である。

【図７】本発明の１実施形態と一貫するアルゴリズムの行合計
【数６７】

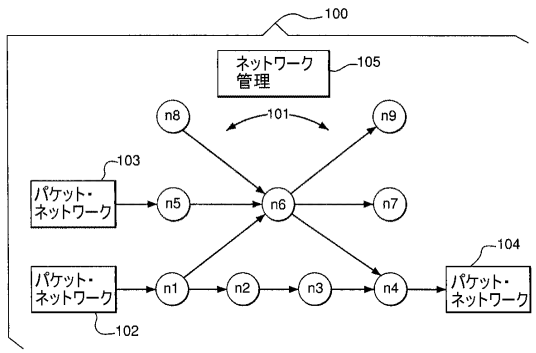
$\hat{R}$

および列合計
【数６８】

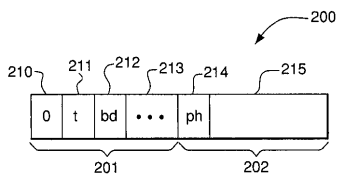
$\hat{C}$

のルーティング保証の概略を示す図である。
【図８】本発明の例示的実施形態のシミュレーションに使用された、キャリア・バックボーン・ネットワークを表すネットワーク・トポロジ内の２８個の双方向リンクを有する例示的な１５ノード・ネットワークを示す図である。
【図９】図８の例示的ネットワークのスケーリング係数を比較するシミュレーション結果のグラフである。
【図１０】例示的な２０ノード・ネットワークのスケーリング係数を比較するシミュレーション結果のグラフである。

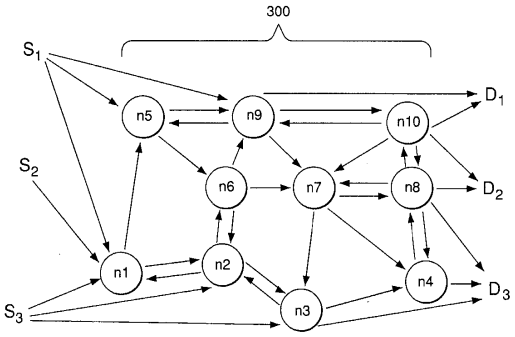
【図１】



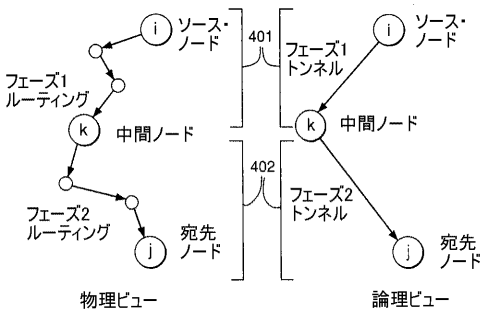
【図２】



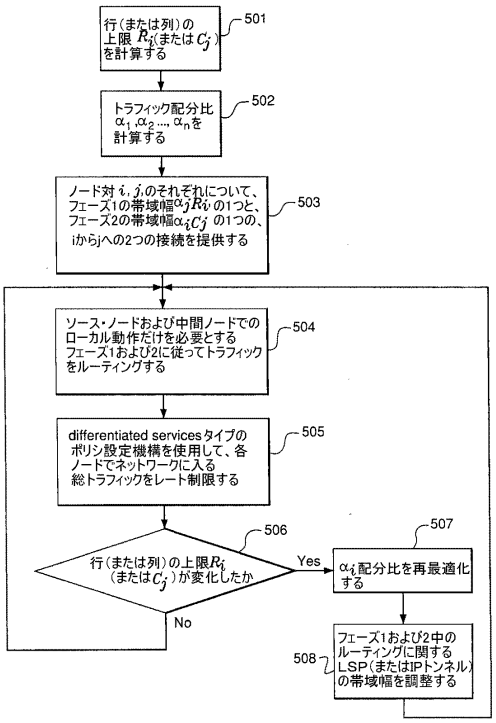
【図３】



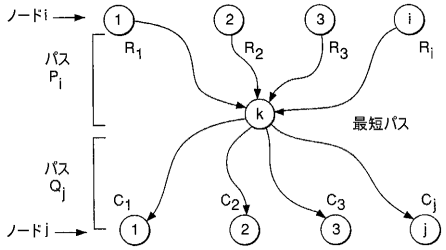
【図４】



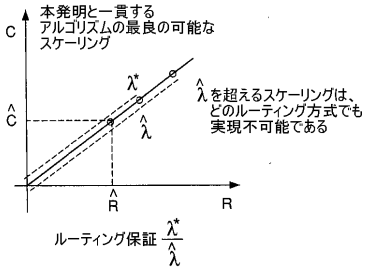
【図５】



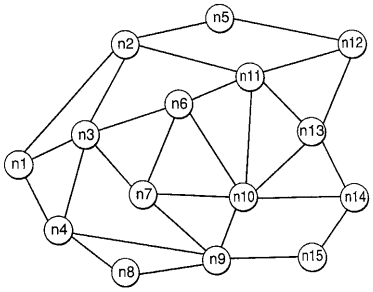
【図６】



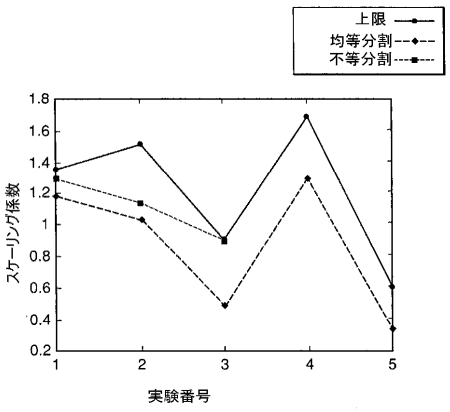
【図７】



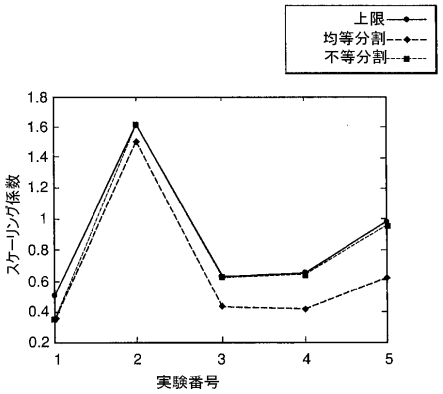
【図８】



【図９】



【図１０】



フロントページの続き

(74)代理人 100096943
弁理士 臼井 伸一
(74)代理人 100101498
弁理士 越智 隆夫
(74)代理人 100096688
弁理士 本宮 照久
(74)代理人 100104352
弁理士 朝日 伸光
(74)代理人 100128657
弁理士 三山 勝巳
(72)発明者 ムラリダラン サンパス コディアラム
アメリカ合衆国 ０７７４６ ニュージャーシィ，マールボロ，ベラ ヴィスタ コート ５
(72)発明者 ティルネル ヴィー．ラクシュマン
アメリカ合衆国 ０７７５１ ニュージャーシィ，モーガンヴィル，レアド ドライヴ １１５
(72)発明者 スディプタ センングプタ
アメリカ合衆国 ０７７４７ ニュージャーシィ，アバディーン，ダンバートン ヒル コート １２０

審査官 安藤 一道

(56)参考文献 特開２００４－０４８３３０（ＪＰ，Ａ）
特開２００４－１４７０６０（ＪＰ，Ａ）
羽賀 太，ＭＰＬＳ網管理システムにおける経路設計支援機能の開発，電子情報通信学会技術研究報告，社団法人電子情報通信学会，２００１年１１月１５日，第101巻 第442号，pp.25～28
武田 知典，コア網における公平性を考慮した動的負荷分散方式，電子情報通信学会論文誌，社団法人電子情報通信学会，２００３年 2月 1日，第J86-B巻 第2号，pp.174～186

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｈ０４Ｌ １２／５６