(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5784644号
(P5784644)
(45)発行日　平成27年9月24日(2015.9.24)　(24)登録日　平成27年7月31日(2015.7.31)

(51)Int.Cl.　F I
H04L 12/801 (2013.01)　H04L 12/801

請求項の数 15　(全 20 頁)

(21)出願番号	特願2012-556229 (P2012-556229)
(86)(22)出願日	平成23年3月2日 (2011.3.2)
(65)公表番号	特表2013-521718 (P2013-521718A)
(43)公表日	平成25年6月10日 (2013.6.10)
(86)国際出願番号	PCT/US2011/026931
(87)国際公開番号	WO2011/109565
(87)国際公開日	平成23年9月9日 (2011.9.9)
審査請求日	平成26年2月3日 (2014.2.3)
(31)優先権主張番号	12/717,784
(32)優先日	平成22年3月4日 (2010.3.4)
(33)優先権主張国	米国(US)

(73)特許権者　314015767
マイクロソフト　テクノロジー　ライセンシング，エルエルシー
アメリカ合衆国　ワシントン州　９８０５２　レッドモンド　ワン　マイクロソフト　ウェイ
(74)代理人　100107766
弁理士　伊東　忠重
(74)代理人　100070150
弁理士　伊東　忠彦
(74)代理人　100091214
弁理士　大貫　進介

最終頁に続く

(54)【発明の名称】ネットワーク接続上の信頼性機構の選択的な無効化

(57)【特許請求の範囲】
【請求項１】
プロセッサによる実行時に、異なるネットワークに存在する複数のエンドポイント間で確立されたネットワーク接続にわたって通信する方法をコンピュータに実行させるコンピュータ実行可能命令を格納するコンピュータ読取可能記憶媒体であって、
前記方法は、
ソース・エンドポイントと宛先エンドポイントとの間に跨る前記ネットワーク接続を提供する動作を前記プロセッサに実行させるステップであって、前記ネットワーク接続が、前記ソース・エンドポイントおよび前記宛先エンドポイントが夫々存在する前記異なるネットワークを橋渡しするＴＣＰベースのトンネルとして動作する、ステップと、
前記ＴＣＰベースのトンネル上で一体的に実行されている１つまたは複数の低レベルの信頼性機構を選択的に無効化する動作を前記プロセッサに実行させるステップであって、前記１つまたは複数の低レベルの信頼性機構は、輻輳制御機構およびパケット損失機構を備える、ステップと、
前記１つまたは複数の低レベルの信頼性機構がメッセージの送信タイミングに干渉することなく、前記トンネルを通じて第１のエンドポイントと第２のエンドポイントとの間でメッセージを送信する動作を前記プロセッサに実行させるステップと、
を備えることを特徴とするコンピュータ読取可能記憶媒体。
【請求項２】
前記第１のエンドポイントと前記第２のエンドポイントとの間でメッセージを送信する

動作を前記プロセッサに実行させるステップは、
前記ＴＣＰベースのトンネルを介してＩＰパケットを前記第１のエンドポイントから前記第２のエンドポイントに送信するステップを含む、
ことを特徴とする請求項１に記載のコンピュータ読取可能記憶媒体。
【請求項３】
前記輻輳制御機構は、ＩＰパケットの送信速度を管理するように構成されることを特徴とする請求項２に記載のコンピュータ読取可能記憶媒体。
【請求項４】
前記パケット損失機構は、未配信のまたは遅延のＩＰパケットを自動的に再送信することによって前記ネットワーク接続上のパケット損失を管理するように構成されることを特徴とする請求項２に記載のコンピュータ読取可能記憶媒体。
【請求項５】
前記ネットワーク接続は、低レベルのＴＣＰベース接続の上で実行される高レベルのＴＣＰベース接続を含むことを特徴とする請求項１に記載のコンピュータ読取可能記憶媒体。
【請求項６】
１つまたは複数の高レベルの信頼性機構が、前記高レベルのＴＣＰベース接続で一体的に実行されており、前記１つまたは複数の低レベルの信頼性機構が、前記低レベルのＴＣＰベース接続で一体的に実行されていることを特徴とする請求項５に記載のコンピュータ読取可能記憶媒体。
【請求項７】
前記１つまたは複数の低レベルの信頼性機構を選択的に無効化する際に、前記１つまたは複数の高レベルの信頼性機構は有効化されたままであることを特徴とする請求項６に記載のコンピュータ読取可能記憶媒体。
【請求項８】
前記１つまたは複数の高レベルの信頼性機構は、
ＩＰパケットの送信速度を管理するように構成された輻輳制御機構と、
未配信のまたは遅延のＩＰパケットを自動的に再送信することによって前記ネットワーク接続上のパケット損失を管理するように構成されたパケット損失機構と、
を備えることを特徴とする請求項６に記載のコンピュータ読取可能記憶媒体。
【請求項９】
前記１つまたは複数の高レベルの信頼性機構は、
有効化されたときに、前記高レベルのＴＣＰベース接続上のデータ・フローを管理する１組のルールを強制し、
前記１つまたは複数の低レベルの信頼性機構は、有効化されたときに、前記低レベルのＴＣＰベース接続上のデータ・フローを管理する同一の１組のルールを強制する、
ことを特徴とする請求項６に記載のコンピュータ読取可能記憶媒体。
【請求項１０】
前記１つまたは複数の低レベルの信頼性機構は、クラウド・コンピューティング・プラットフォームのクライアントにより、部分的に、デザインされたクラウド・コンピューティング・サービス・モデルの機能として選択的に無効化され、
前記クラウド・コンピューティング・プラットフォームは、前記ソース・エンドポイントをホストするデータセンタを備える、
ことを特徴とする請求項１に記載のコンピュータ読取可能記憶媒体。
【請求項１１】
前記宛先エンドポイントは、クライアントが管理するプライベート企業ネットワーク内に配置されたリソースによりホストされることを特徴とする請求項１に記載のコンピュータ読取可能記憶媒体。
【請求項１２】
独立したネットワークに存在するエンドポイント間のデータ・フローを管理するコンピ

ュータ・システムであって、
前記コンピュータ・システムは、
ソース・エンドポイントをホストするクラウド・コンピューティング・プラットフォーム内のデータセンタであって、前記ソース・エンドポイントが、前記クラウド・コンピューティング・プラットフォームおよびプライベート企業ネットワークの両方で実行されているアプリケーションに割り当てられる、データセンタと、
前記アプリケーションに割り当てられた宛先エンドポイントをホストする前記プライベート企業ネットワーク内のリソースであって、前記ソース・エンドポイントおよび前記宛先エンドポイントは、前記データ・フローを前記ソース・エンドポイントと前記宛先エンドポイントとの間で直接転送するトンネルにより接続され、前記トンネルが、低レベルの接続の上で実行される高レベルの接続を可能とし、輻輳制御機構およびパケット損失機構が両方とも、前記高レベルの接続および前記低レベルの接続の各々に、夫々組み込まれる、リソースと、
前記トンネルを確立し前記トンネル内で前記接続を構成する、前記データセンタ内で実行されるファブリック・コントローラであって、前記接続を構成することには、前記低レベルの接続に組み込まれた前記輻輳制御機構および前記パケット損失機構を選択的に無効化することと、前記高レベルの接続に組み込まれた前記輻輳制御機構および前記パケット損失機構を選択的に有効化することを含むファブリック・コントローラと、を備える、
ことを特徴とするコンピュータ・システム。
【請求項１３】
前記高レベルの接続で前記ソース・エンドポイントに運ばれる第１のＩＰパケットを生成する、前記データセンタ内の仮想マシンをさらに備え、
前記ソース・エンドポイントが前記第１のＩＰパケットを第２のＩＰパケットにカプセル化し、前記第２のＩＰパケットを前記低レベルの接続で送信する、
ことを特徴とする請求項１２に記載のコンピュータ・システム。
【請求項１４】
前記低レベルの接続に組み込まれた前記輻輳制御機構を選択的に無効化するときに、前記ファブリック・コントローラが前記ソース・エンドポイントと通信し、
前記低レベルの接続に組み込まれた前記パケット損失機構を選択的に無効化するときに、前記ソース・エンドポイントが前記宛先エンドポイントとネゴシエーションする、
ことを特徴とする請求項１３に記載のコンピュータ・システム。
【請求項１５】
ＴＣＰベースのトンネルにわたるソース・エンドポイントと宛先エンドポイントとの間の通信を促進するコンピュータ実行方法であって、
前記方法は、
ファブリック・コントローラを使用して、前記ソース・エンドポイントおよび前記宛先エンドポイントを異なるネットワークにわたって通信可能に接続するＴＣＰベースのトンネルを確立するステップであって、前記ソース・エンドポイントの動作がデータセンタによりサポートされ、前記宛先エンドポイントの動作は、前記データセンタのクライアントが管理するプライベート企業ネットワークに存在する、遠隔に配置されたリソースによりサポートされる、ステップと、
前記データセンタ内でインスタンス化された仮想マシンから渡される第１のＩＰパケットを前記ソース・エンドポイントで受信するステップであって、前記第１のＩＰパケットは、第１の１組の信頼性機構が提供された高レベルの接続を介して運ばれる、ステップと、
前記ソース・エンドポイントで前記第１のＩＰパケットを第２のＩＰパケットにカプセル化するステップと、
前記第２のＩＰパケットを、前記高レベルの接続の下方で積層された低レベルの接続を介して前記ＴＣＰベースのトンネルを通じて送信するステップであって、前記低レベルの接続には第２の１組の信頼性機構が提供されており、前記第１の1組の信頼性機構および

前記第２の１組の信頼性機構の各々が、輻輳制御機構および損失回復機構を夫々備える、ステップと、

前記ファブリック・コントローラを使用して、前記低レベルの接続に提供された前記輻輳制御機構および前記損失回復機構を選択的に無効化するステップと、

前記ファブリック・コントローラを使用して、前記高レベルの接続に提供された前記輻輳制御機構および前記損失回復機構を受動的に有効化されたままにするステップと、

前記低レベルの接続の無効化された状態および前記高レベルの接続の有効化された状態を記憶するステップであって、前記無効化された状態は、前記低レベルの接続に提供された前記輻輳制御機構および前記損失回復機構が無効化されていることを表し、前記有効化された状態は、前記高レベルの接続に提供された前記輻輳制御機構および前記損失回復機構が有効化されていることを表す、ステップと、

を含む、

ことを特徴とする方法。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、ＴＣＰベースのトンネルを確立し、管理する方法およびシステムに関する。

【背景技術】

【０００２】

大規模ネットワークシステムは、アプリケーション実行、企業向けデータの保守および運用機能のための様々な構成で利用される一般的なプラットフォームである。例えば、データセンタ（例えば、物理クラウド・コンピューティング基盤）は、複数の消費者に対して同時に様々なサービス（例えば、ウェブアプリケーション、電子メールサービス、検索エンジンサービス等）を提供することができる。これらの大規模ネットワークシステムは一般に、データセンタ全体に分散された多数のリソースを含む。データセンタにおいては、各リソースは、物理マシン、または、物理ホスト上で実行されている仮想マシンと似ている。データセンタが複数のテナント（例えば、顧客のプログラム）をホストするとき、これらのリソースは同じデータセンタから様々なテナントに最適に割り当てられる。

【０００３】

データセンタの顧客はしばしば、プライベート企業ネットワーク（例えば、データセンタから地理的に離れた、顧客によって管理されるサーバ）内で実行されている企業アプリケーションが、データセンタ内のリソース上で実行されているソフトウェアと対話することを要求する。この事例では、企業アプリケーションのコンポーネントとデータセンタ内で実行されているソフトウェアのコンポーネントとの間のネットワーク接続が確立される。このネットワーク接続は一般に、ＴＣＰ（transmission control protocol）などのネットワーク転送プロトコルを利用して、ネットワーク接続を介した信頼性の高いパケット配信を促進する。

【０００４】

ＴＣＰベースのネットワーク接続、即ち、ＴＣＰ接続は、独立に動作するプライベート企業ネットワークおよびデータセンタにわたるエンド・ツー・エンドのメッセージ転送を管理する責任を負う。例えば、これらのツールは、エラー制御、セグメンテーション、フロー制御、輻輳制御およびアプリケーション・アドレス指定（application addressing）（例えば、ポート番号）を管理することができる。動作に際しては、これらのツールが損失パケットの再送信を要求し、パケットの送信速度を変更して輻輳を減らすことにより、ネットワーク輻輳、パケット損失などの問題あるネットワークの振舞いを検出して、改善することができる。

【０００５】

ＴＣＰベース接続が別のＴＣＰ接続の上で実行されているときは、内部ＴＣＰ接続および外部ＴＣＰ接続により使用されるこれらの機構が相互作用する結果、再送信およびフロー調整が過度に行われるおそれがあり、それが、接続全体の性能の大幅な低下の原因とな

る。このため、ＴＣＰベース通信の層に関連する１つまたは複数のツールを無効化する新技術を使用することにより、信頼性の高いパケット配信および輻輳制御が依然として保証されつつ、確立されたネットワーク接続のスループットおよび性能が高まるはずである。

【発明の概要】

【０００６】

「発明の概要」は、「発明を実施するための形態」においてさらに説明される諸概念を簡潔な形で導入するために与えられる。「発明の概要」は、特許請求する主題の主要な特徴または本質的な特徴を特定しようとするものではなく、特許請求する主題の範囲を決定するに際しての支援として使用しようとするものでもない。

【０００７】

本発明の諸実施形態は、ＴＣＰベースのトンネル内部で一体的に実行されているツール（例えば、信頼性機構）を構成し、これらのツールの重複する動作に起因する保証されていない性能劣化に対抗する役割を果たす、システム、方法およびコンピュータ読取可能媒体を提供する。一般に、ＴＣＰベースのトンネルは、別々に配置されたエンドポイント間のネットワーク接続として機能する。例えば、エンドポイントには、クラウド・コンピューティング・プラットフォームがホストするソース・エンドポイント、およびプライベート企業ネットワーク内のリソースがホストする宛先エンドポイントが含まれうる。

【０００８】

これらの構成ツールの例示的な実施形態は、ＴＣＰベースのトンネルを含む各接続で独立して信頼性機構（例えば、輻輳制御機構および損失回復機構）を選択的に有効化および無効化することに関する。１つの事例では、これらのチャネルは、低レベルのＴＣＰベース接続および高レベルのＴＣＰベース接続を含む。動作に際しては、アプリケーションのデータパケットが、低レベルのＴＣＰベース接続の上で実行されている高レベルのＴＣＰベース接続を介して送信される。

【０００９】

諸実施形態では、ファブリック・コントローラを使用して、低レベルのＴＣＰベース接続に組み込まれた輻輳制御機構および／または損失回復機構の有効化（オン）を可能とすることができる。低レベルのＴＣＰベースのチャネル内部に統合された１つまたは複数の信頼性機構の選択的な無効化を、所定の基準（例えば、クラウド・コンピューティング・サービス・モデルからの命令、ソース・エンドポイントの識別、宛先エンドポイントの識別等）により行い、ネットワーク接続ごとに行ってもよい。このように、高レベルのＴＣＰベースのチャネルのツールは、ＴＣＰベースのトンネルを通じてデータパケットが完全かつ効率的に送信されることを保証する１組の信頼性ルールを積極的に強制し、低レベルのＴＣＰベースのチャネルの１つまたは複数のツールを停止して、その冗長な動作から生ずる潜在的な性能劣化を低減する。

【図面の簡単な説明】

【００１０】

本発明の諸実施形態を、添付図面を参照して下記で詳細に説明する。

【００１１】

【図１】本発明の諸実施形態を実装する際に使用するのに適した例示的なコンピューティング環境のブロック図である。

【図２】ＴＣＰベースのトンネルに組み込まれた信頼性機構を構成するように提供された、本発明の諸実施形態を実装する際に使用するのに適した例示的なクラウド・コンピューティング・プラットフォームを示すブロック図である。

【図３】本発明の一実施形態に従って、ＴＣＰベースのトンネルをその中で確立させた例示的な分散コンピューティング環境のブロック図である。

【図４】本発明の一実施形態に従う、ＴＣＰベースのトンネルを含むＴＣＰベースのチャネルの階層の略図である。

【図５】本発明の諸実施形態に従う、ファブリック・コントローラを使用してＴＣＰベースのトンネルに組み込まれた信頼性機構を選択的に有効化／無効化する、例示的なデータ

センタのブロック図である。
【図６】本発明の一実施形態に従う、別々の位置に存在する複数のエンドポイント間で確立されたネットワーク接続にわたって通信を促進するための方法を示す流れ図である。
【図７】本発明の一実施形態に従う、別々の位置に存在する複数のエンドポイント間で確立されたネットワーク接続にわたって通信を促進するための方法を示す流れ図である。
【発明を実施するための形態】
【００１２】
法的要件を満たすために、本発明の諸実施形態の主題を本明細書で具体的に説明する。しかし、本説明自体は本特許権の範囲を限定しようとするものではない。むしろ、本発明者は、他の既存または将来の技術とともに、特許請求する主題を他の方法で具体化して、本明細書で説明したものと類似する別のステップまたはステップの組合せを含めてもよいと考えている。さらに、本明細書では「ステップ」および／または「ブロック」という用語を用いて様々な方法の要素が使用されることを示唆するが、個々のステップの順序を明示的に記載しない限り、本明細書で開示した様々なステップの間で何らかの特定の順序があることを当該用語が暗示すると解釈すべきではない。
【００１３】
本発明の諸実施形態は、接続されたネットワーク内で遠隔に配置されたネットワークまたはエンドポイントに跨るトンネルを確立し構成するための方法、コンピュータ・システムおよびコンピュータ読取可能媒体に関する。本明細書で使用する際、用語「トンネル」は限定的な意味ではなく、異なるネットワークにわたる通信を橋渡しする意図をもってソース・エンドポイントおよび宛先エンドポイントを通信可能に接続する任意のネットワーク接続を包含することができる。１つの事例では、別々のローカル・ネットワーク内でホストされるエンドポイント間でデータパケットを流すネットワーク接続としてトンネルを確立してもよく、この場合、エンドポイントは、それに割り当てられたＩＰアドレスにより発見され、識別される。さらに、ストリーミング・データパケットおよび他のトラフィックがトンネルを介して移動するとき、それらが複数のリンク、ファイアウォールおよび他のセキュリティ手段を横断してもよい。エンドポイント間のこの表面的な直接接続により、双方のエンドポイントは、あたかもそれらが共通のネットワーク内で隣接して配置されＩＰ層を介して通信しているかのように会話することができる。
【００１４】
幾つかの実施形態では、トンネルは、ＴＣＰまたはＨＴＴＰ（hypertext transfer protocol）またはＨＴＴＰＳ（HTTP Secure）を用いてそれぞれ確立され、それらにエンドポイントが参加することもあれば参加しないこともある。トンネルにより、エンドポイントが有利に接続され、１つのネットワークまたは異なるネットワークにわたる通信が可能となる。例えば、ＨＴＴＰトンネルまたはＨＴＴＰＳトンネルは、エンドポイントが２つの異なるネットワークに存在する場合であっても、直接的な仮想ＩＰレベルの接続を確立する機能をエンドポイントに提供する。換言すれば、トンネルにより、双方のエンドポイントは、あたかもそれらが共通のネットワーク内で隣接して配置されＩＰ層を介して通信しているかのように会話することができる。例えば、２つのエンドポイント上で現在実行されているアプリケーションが、そのエンドポイントが２つの異なるネットワークに存在することを認識できず、そのため、アプリケーションは自分がトンネル上で実行されていることに気付かない。このＨＴＴＰトンネルまたはＨＴＴＰＳトンネルの機能は、ファイアウォールおよびプロキシ・サーバなどの他のネットワーク・エッジ・デバイスをバイパスするためのＨＴＴＰおよびＨＴＴＰＳベースのネットワーク接続機能の産物である。
【００１５】
さらに、ＨＴＴＰトンネルまたはＨＴＴＰＳトンネルは、ＴＣＰベース接続が広範囲の様々なネットワークで利用されているので、エンド・ツー・エンドの信頼性機構が組み込まれたＴＣＰベース接続によりサポートされる。これらの信頼性機構は、損失回復および輻輳制御などの機能を実施して、これらの様々なネットワークを接続するこれらのリンク上の損失および輻輳を処理する。換言すれば、パケット損失の検出および輻輳の検出、な

らびにその各々に対応するためのビルトインの信頼性機構がＴＣＰに提供される。１つの事例では、ＴＣＰベース接続において輻輳を検出すると、信頼性機構の応答により、データパケットのトラフィックがネットワーク接続に配信される速度が減少することとなりうる。別の事例では、ＴＣＰベース接続においてパケット損失を検出すると、信頼性機構の応答により、損失したデータパケットが再送信されることとなりうる。

【００１６】

潜在的には、エンド・ツー・エンドのＴＣＰ接続の最中は、お互いの上で実行されている複数のＴＣＰベース接続が存在しうる。この状況では、パケット損失またはネットワーク輻輳が発生した場合、積層された接続の各々に統合された信頼性機構が、お互いの間で通信せずに、パケット損失およびネットワーク輻輳に対して独立に応答する。例えば、高レベルのＴＣＰベース接続の損失回復機構が、低レベルの接続の損失回復機構からの応答に加えて独自の応答の実行を試みる可能性がある。即ち、双方の損失回復機構がデータを再送し、ネットワーク接続の性能の劣化が不必要に増大することとなる。

【００１７】

また、パケット損失に輻輳問題として反応し、協調して動作するに際して送信速度の調節時に重複して動作しうる輻輳制御機構を双方のチャネルが有することがある。例えば、双方の輻輳制御機構がデータパケットのストリーミング速度を半分に減少させた場合、集約効果として速度が４分の１になり、これはパケット損失問題の解決に必要な速度よりも相当遅い。したがって、実際には、これらの重複する是正措置は原因を過度に補償し、非効率的なものとなる。この過度な補償はしばしば、エンドポイント間の通信に悪影響を及ぼし、未解決のパケット損失またはネットワーク輻輳の問題を適切に対処するのに望ましい待ち時間を超えて待ち時間を増大させる。

【００１８】

一態様では、本発明の諸実施形態は、コンピュータ実行可能命令を格納する１つまたは複数のコンピュータ読取可能媒体に関する。当該命令は、実行時に、異なるネットワークに存在する複数のエンドポイント間で確立されたネットワーク接続にわたって通信するための方法を実施する。まず、本方法は、ソース・エンドポイントと宛先エンドポイントとの間に跨るネットワーク接続を提供するステップを含む。上述したように、ネットワーク接続は、ソース・エンドポイントおよび宛先エンドポイントがそれぞれ存在する異なるネットワークを橋渡しするＴＣＰベースのトンネルとして動作する。本方法はさらに、ＴＣＰベースのトンネル上で一体的に実行されている１つまたは複数の低レベルの信頼性機構を選択的に無効化するステップ、および当該低レベルの信頼性機構がメッセージの送信タイミングに干渉することなしに、第１のエンドポイントと第２のエンドポイントとの間でメッセージを送信するステップを含む。諸実施形態では、第１のエンドポイントと第２のエンドポイントとの間でメッセージを送信するステップには、特に、ＴＣＰベースのトンネルを介してＩＰパケットを第１のエンドポイントから第２のエンドポイントへ送信するステップを含んでもよい。

【００１９】

例示的な一実施形態では、低レベルの信頼性機構が輻輳制御機構およびパケット損失機構を備える。輻輳制御機構は、ＩＰパケット内で運ばれるデータの量を管理するように構成されうる。パケット損失機構は、未配信のまたは遅延したＩＰパケットを自動的に再送信することによってネットワーク接続上のパケット損失を管理するように構成されうる。

【００２０】

別の態様では、本発明の諸実施形態は、個々のローカル・ネットワークに存在するエンドポイント間のデータの流れを管理するコンピュータ・システムに関する。まず、当該コンピュータ・システムは、クラウド・コンピューティング・プラットフォーム内のデータセンタ、プライベート企業ネットワーク内のリソース、ファブリック・コントローラおよび仮想マシンを備える。諸実施形態では、データセンタが、クラウド・コンピューティング・プラットフォームおよびプライベート企業ネットワークの双方で実行されているアプリケーションに割り当てられたソース・エンドポイントをホストしてもよい。リソースが

、やはりアプリケーションに割り当てられた宛先エンドポイントをホストしてもよい。ソース・エンドポイントと宛先エンドポイントとの間の通信を開始するに際して、ソース・エンドポイントおよび宛先エンドポイントが、それらの間で直接にデータ・フローを転送するトンネルにより接続される。ここで、トンネルは低レベルの接続の上で実行されている高レベルの接続を含んでもよい。上述したように、輻輳制御機構およびパケット損失機構は両方とも、高レベルの接続および低レベルの接続の各々に、夫々組み込まれている。

【００２１】

ファブリック・コントローラは、データセンタ内で実行され、トンネルを確立しトンネルを構成することができる。１つの事例では、トンネルを構成することには、低レベルの接続に組み込まれた輻輳制御機構およびパケット損失機構を選択的に無効化することが含まれる。別の事例では、これらの接続を構成することには、高レベルの接続に組み込まれた輻輳制御機構およびパケット損失機構を選択的に有効化することを含めてもよい。

【００２２】

データセンタ内の仮想マシンは、高レベルの接続でソース・エンドポイントに運ばれる第１のＩＰパケットを生成する。当該ＩＰパケットを受け取ると、ソース・エンドポイント（またはソース・トンネル終端エンドポイント）は第１のＩＰパケットを第２のＩＰパケットにカプセル化して、当該第２のＩＰパケットを低レベルの接続で送信する。したがって、ＴＣＰベースのトンネルのエンドポイント間でデータパケットを送信するときに、双方の接続が関係する。

【００２３】

さらに別の態様では、本発明の諸実施形態は、ＴＣＰベースのトンネルにわたるソース・エンドポイントと宛先エンドポイントとの間の通信を促進するためのコンピュータ化された方法に関する。例示的な一実施形態では、本方法は、ファブリック・コントローラを使用して、ソース・エンドポイントおよび宛先エンドポイントを１つのネットワーク上でまたは異なるネットワークにわたって通信可能に接続するＴＣＰベースのトンネルを確立することを含む。前述したように、ソース・エンドポイントの動作がデータセンタによりサポートされ、宛先エンドポイントの動作が、データセンタのクライアントが管理するプライベート企業ネットワークに存在する、遠隔に配置されたリソースによりサポートされる。本方法はさらに、第１のＩＰパケットをソース・エンドポイントで受け取ることを含む。当該第１のＩＰパケットは、データセンタ内でインスタンス化された仮想マシンから渡される。これらの第１のＩＰパケットは、第１の１組の信頼性機構が提供されている高レベルの接続を介して運ばれる。当該第１のＩＰパケットはソース・エンドポイント（または、ソース・エンドポイントと同じネットワーク内のソース・トンネル終端エンドポイント）で第２のパケットにカプセル化され、低レベルの接続を介してＴＣＰベースのトンネルを通じて宛先ネットワーク内のトンネル終端エンドポイントに送信され、最終的な宛先、即ち、リモート・ネットワーク内のエンドポイントに転送される。一般的には、低レベルの接続には第２の１組の信頼機構が提供される。諸実施形態では、これらの第１の１組の信頼性機構および第２の１組の信頼性機構の各々は、輻輳制御機構および損失回復機構をそれぞれ備える。

【００２４】

本方法は続いて、ファブリック・コントローラを使用して、低レベルの接続に提供された輻輳制御機構および損失回復機構を選択的に無効化することを実施する。また、ファブリック・コントローラを使用して、高レベルの接続に提供された輻輳制御機構および損失回復機構を受動的に有効化されたままにすることができる。低レベルの接続の無効化された状態および高レベルの接続の有効化された状態が少なくとも一時的に記憶される。明確さのため、無効化された状態は、低レベルの接続に提供された輻輳制御機構および／または損失回復機構が無効化されていることを表す。反対に、有効化された状態は、高レベルの接続に提供された輻輳制御機構および損失回復機構が有効化されていることを表す。

【００２５】

本発明の諸実施形態の概要を簡単に説明したので、本発明の諸実施形態の実装に適した

例示的な動作環境を以下で説明する。

【００２６】

図を概括的に参照する。まず、特に図１を参照すると、本発明の諸実施形態を実装するための例示的な動作環境が示され、コンピューティング装置１００として一般的に指定されている。コンピューティング装置１００は適切なコンピューティング環境の一例にすぎず、本発明の諸実施形態の使用範囲または機能範囲に関するどのような限定を示唆しようとするものでもない。また、コンピューティング環境１００が、図示したコンポーネントのうち任意の１つまたは組合せに関して何らかの依存性または要件を有するとも解すべきではない。

【００２７】

本発明の諸実施形態をコンピュータ・コードまたは機械利用可能命令の一般的なコンテキストで説明することができる。当該コードまたは命令には、コンピュータ、あるいはＰＤＡまたは他のハンドヘルド装置などの他の機械により実行されている、プログラム・コンポーネントなどのコンピュータ実行可能命令が含まれる。一般に、プログラム・コンポーネントは、ルーチン、プログラム、オブジェクト、コンポーネント、データ構造等を含み、特定のタスクを実施するかまたは特定の抽象データ型を実装するコードを指す。本発明の諸実施形態を様々なシステム構成で実施してもよく、当該システム構成には、ハンドヘルド装置、家庭用電化製品、汎用コンピュータ、特殊なコンピューティング装置等が含まれる。本発明の諸実施形態を、通信ネットワークを介して接続されたリモート処理装置によりタスクが実施される分散コンピューティング環境で実施してもよい。

【００２８】

引き続き図１を参照すると、コンピューティング装置１００は、メモリ１１２、１つまたは複数のプロセッサ１１４、１つまたは複数のプレゼンテーション・コンポーネント１１６、Ｉ／Ｏ（input/output）ポート１１８、Ｉ／Ｏコンポーネント１２０、および例示的な電源１２２を直接または間接に接続するバス１１０を備える。バス１１０は、１つまたは複数のバスでありうるもの（例えば、アドレス・バス、データ・バス、またはそれらの組合せ）を表す。図１の様々なブロックは明確さのため線で示してあるが、現実には、様々なコンポーネントはそのように明確には区別されず、比喩的に言うと、当該線はより正確にはグレーであり曖昧であるはずである。例えば、ディスプレイ装置などのプレゼンテーション・コンポーネントがＩ／Ｏコンポーネントであると考える者もいるかもしれない。また、プロセッサはメモリを有する。本発明の発明者は、それが当業界の性質であることを認識しており、図１の図は、本発明の１つまたは複数の実施形態と関連して使用できる例示的なコンピューティング装置を示すにすぎないことを繰り返しておく。「ワークステーション」、「サーバ」、「ラップトップ」、「ハンドヘルド装置」等のカテゴリ間では区別しない。なぜならば、これら全てが図１の範囲内にあり「コンピュータ」または「コンピューティング装置」を指すと考えられるからである。

【００２９】

コンピューティング装置１００は一般に、様々なコンピュータ読取可能媒体を含む。限定ではなく例として、コンピュータ読取可能媒体には、ＲＡＭ（Random Access Memory）、ＲＯＭ（Read Only Memory）、ＥＥＰＲＯＭ（Electronically Erasable Programmable Read Only Memory）、フラッシュ・メモリもしくは他のメモリ技術、ＣＤＲＯＭ、ＤＶＤ（digital versatile disk）もしくは他の光媒体もしくはホログラフィック媒体、磁気カセット、磁気テープ、磁気ディスク記憶装置もしくは他の磁気記憶装置、または所望の情報のエンコードに使用できコンピューティング装置１００がアクセスできる他の任意の媒体が含まれうる。

【００３０】

メモリ１１２は、コンピュータ記憶媒体を揮発性メモリおよび／または不揮発性メモリの形で含む。メモリは、リムーバブル、非リムーバブル、またはそれらの組合せであってもよい。例示的なハードウェア装置には、ソリッドステートメモリ、ハード・ドライブ、光ディスク・ドライブ等が含まれる。コンピューティング装置１００は、メモリ１１２ま

たはＩ／Ｏコンポーネント１２０のような様々なエンティティからデータを読み出す１つまたは複数のプロセッサを備える。プレゼンテーション・コンポーネント１１６は、データのインジケーションをユーザまたは他の装置に提示する。例示的なプレゼンテーション・コンポーネントには、ディスプレイ装置、スピーカ、印刷コンポーネント、バイブレーション・コンポーネント等が含まれる。Ｉ／Ｏポート１１８により、コンピューティング装置１００を、Ｉ／Ｏコンポーネント１２０を含む他の装置に論理的に接続することができ、それらの一部がビルトインであってもよい。例示的なコンポーネントには、マイクロフォン、ジョイスティック、ゲーム・パッド、パラボラ・アンテナ、スキャナ、プリンタ、無線装置等が含まれる。

【００３１】

図１および図２を参照すると、第１のコンピューティング装置２５５および／または第２のコンピューティング装置２６５を、図１の例示的なコンピューティング装置１００により実装してもよい。さらに、エンドポイント２０１および／またはエンドポイント２０２が、図１のメモリ１１２の一部および／または図１のプロセッサ１１４の一部を備えてもよい。

【００３２】

次に図２を参照すると、本発明の一実施形態に従うブロック図が示されている。当該ブロック図は、サービス・アプリケーションが使用するための仮想マシン２７０および２７５をデータセンタ２２５内に割り当てるように構成された、例示的なクラウド・コンピューティング・プラットフォーム２００を示す。図２に示すクラウド・コンピューティング・プラットフォーム２００は、適切なコンピューティング・システム環境の一例にすぎず、本発明の諸実施形態の使用範囲または機能範囲に関してどのような限定も示唆しようとするものではないことは理解されよう。例えば、クラウド・コンピューティング・プラットフォーム２００がパブリック・クラウド、プライベート・クラウド、専用クラウドであってもよい。また、クラウド・コンピューティング・プラットフォーム２００が、図示したコンポーネントのうち任意の１つまたは組合せに関して何らかの依存性または要件を有するとも解すべきではない。さらに、図２の様々なブロックは明確さのため線で示してあるが、現実には、様々なコンポーネントはそのように明確には区別されず、比喩的に言うと、当該線はより正確にはグレーであり曖昧であるはずである。さらに、任意数の物理マシン、仮想マシン、データセンタ、エンドポイント、またはそれらの組合せを使用して、本発明の諸実施形態の範囲内にある所望の機能を実現してもよい。

【００３３】

クラウド・コンピューティング・プラットフォーム２００は、特定のサービス・アプリケーションのエンドポイント２０１および２０２の動作をホストおよびサポートするように構成されたデータセンタ２２５を備える。用語「サービス・アプリケーション」は、本明細書で使用する際は、データセンタ２２５上で実行され、データセンタ２２５内の記憶場所にアクセスする、任意のソフトウェアまたはソフトウェアの一部を広く指す。一実施形態では、エンドポイント２０１および２０２のうち１つまたは複数が、サービス・アプリケーションに参加するソフトウェアの一部、コンポーネント・プログラム、または役割のインスタンスを表してもよい。別の実施形態では、エンドポイント２０１および２０２のうち１つまたは複数が、サービス・アプリケーションがアクセス可能な記憶データを表してもよい。図２に示すエンドポイント２０１および２０２はサービス・アプリケーションをサポートするのに適した部分の一例にすぎず、本発明の諸実施形態の使用範囲または機能範囲に関するどのような限定も示唆しようとするものではないことは理解されよう。

【００３４】

一般に、仮想マシン２７０および２７５は、サービス・アプリケーションになされた要求（例えば、処理負荷の量）に基づいてサービス・アプリケーションのエンドポイント２０１および２０２に割り当てられる。本明細書で使用する際は、用語「仮想マシン」は、限定的な意味ではなく、処理装置により実行されエンドポイント２０１および２０２の機能の基礎となる、任意のソフトウェア、アプリケーション、オペレーティング・システム

、またはプログラムを指してもよい。さらに、仮想マシン２７０および２７５が、エンドポイント２０１および２０２を適切にサポートするためのデータセンタ２２５内の処理機能、記憶位置、および他の資産を含んでもよい。

【００３５】

動作に際しては、仮想マシン２７０および２７５が、データセンタ２２５のリソース（例えば、第１のコンピューティング装置２５５および第２のコンピューティング装置２６５）内に動的に割り当てられ、エンドポイント（例えば、エンドポイント２０１および２０２）が、割り当てられた仮想マシン２７０および２７５に動的に配置されて、現在の処理負荷を満足する。１つの事例では、ファブリック・コントローラ２１０が、仮想マシン２７０および２７５を自動的に割り当て、エンドポイント２０１および２０２をデータセンタ２２５内に配置する責任を負う。例えば、ファブリック・コントローラ２１０が（例えば、サービス・アプリケーションを所有する顧客により設計された）サービス・モデルを利用して、いつどのように仮想マシン２７０および２７５を割り当ててエンドポイント２０１および２０２をそれらに配置するかについて誘導してもよい。さらに、ファブリック・コントローラ２１０が、エンドポイント２０１および２０２と遠隔に配置されたエンドポイントとの間のトンネル型のネットワーク接続に統合された信頼性機構を有効化（オン）または無効化（オフ）すべきか否かを判定するときに、クラウド・コンピューティング・サービス・モデルから命令を読み取ってもよい。これを、図３を参照して以下でより完全に論ずる。

【００３６】

上述したように、仮想マシン２７０および２７５を第１のコンピューティング装置２５５および第２のコンピューティング装置２６５の内部に動的に割り当ててもよい。本発明の諸実施形態ごとに、コンピューティング装置２５５および２６５は、例えば、サーバ、パーソナル・コンピュータ、デスクトップ・コンピュータ、ラップトップ・コンピュータ、モバイル装置、家庭用電子機器、サーバ、図１のコンピューティング装置１００等の任意の形態のコンピューティング装置を表す。１つの事例では、コンピューティング装置２５５および２６５は仮想マシン２７０および２７５の動作をホストおよびサポートし、同時に、データセンタ２２５の他のテナントをサポートするために切り出された他の仮想マシンをホストする。この場合、当該テナントは様々な顧客が所有する他のサービス・アプリケーションのエンドポイントを含む。

【００３７】

１つの態様では、エンドポイント２０１および２０２はクラウド・コンピューティング・プラットフォーム２００のコンテキスト内で動作し、したがって、仮想マシン２７０と２７５との間で動的に作成された接続を介して内部的に通信し、物理ネットワーク・トポロジを介してリモート・ネットワークのリソース（例えば、図３の企業プライベート・ネットワーク３２５のリソース３７５）と外部的に通信する。当該内部接続では、ネットワーク・クラウド（図示せず）を介して、データセンタ２２５の物理リソースにわたって分散された仮想マシン２７０および２７５を相互接続してもよい。ネットワーク・クラウドはこれらのリソースを相互接続し、エンドポイント２０１がエンドポイント２０２および他のエンドポイントの位置を認識し、それらの間の通信を確立できるようにする。さらに、ネットワーク・クラウドが、この通信を、エンドポイント２０１および２０２を論理的に接続する、第１のコンピューティング装置２５５と第２のコンピューティング装置２６５のエンドポイント間のトンネルを通じて確立してもよい。限定ではなく例として、チャネルが１つまたは複数のＬＡＮ（local area network）および／またはＷＡＮ（wide area network）を利用してもよい。かかるネットワーク環境は職場、企業規模のコンピュータ・ネットワーク、イントラネットおよびインターネットで一般的である。したがって、本明細書では当該ネットワークについてはこれ以上説明しない。

【００３８】

次に図３を参照すると、本発明の一実施形態に従う例示的な分散コンピューティング環境３００を示すブロック図が示されており、ＴＣＰベースのトンネル３３０が当該コンピ

ューティング環境３００内で確立されている。まず、分散コンピューティング環境３００は、企業プライベート・ネットワーク３２５および図２を参照して論じたクラウド・コンピューティング・プラットフォーム２００を備える。企業プライベート・ネットワーク３２５およびクラウド・コンピューティング・プラットフォーム２００を、物理ネットワークによりサポートされるネットワーク３１５を介して接続してもよい。本明細書で使用する際、用語「物理ネットワーク」は限定的な意味ではなく、有形の機構および装置（例えば、ファイバ線、配電箱、スイッチ、アンテナ、ＩＰルータ等）、並びに、地理的に離れた場所のエンドポイント間通信を促進する無形の通信および搬送波を包含することができる。例えば、物理ネットワーク（図３では図示せず）が、インターネットで利用されるか、または異なるネットワーク間の通信を推進するために利用できる、任意の有線技術または無線技術を備えてもよい。

【００３９】

一般に、企業プライベート・ネットワーク３２５は、クラウド・コンピューティング・プラットフォーム２００の顧客が管理するリソース３７５のようなリソースを含む。しばしば、これらのリソースが、顧客が所有するサービス・アプリケーションのコンポーネントの動作をホストおよびサポートする。エンドポイントＢ３８５はサービス・アプリケーションの１つまたは複数のコンポーネントを表す。諸実施形態では、図２の仮想マシン２７０のようなリソースが図２のデータセンタ２２５内に割り当てられて、遠隔に分散されたサービス・アプリケーションのコンポーネントの動作をホストおよびサポートする。エンドポイントＡ３９５は、クラウド・コンピューティング・プラットフォーム２００内部にある、これらの遠隔に分散されたサービス・アプリケーションのコンポーネントの１つまたは複数を表わす。動作に際しては、エンドポイントＡ３９５およびＢ３８５が互いに協働して、サービス・アプリケーションが正しく実行されることを保証する。１つの事例では、協働において、エンドポイントＡ３９５とＢ３８５との間でデータパケット（複数可）３１６、またはＩＰデータパケットを、物理ネットワークがサポートするネットワーク３１５にわたって送信する。

【００４０】

諸実施形態では、パケット３１６が、エンドポイントＡ３９５とＢ３８５との間で情報を交換するように動作してもよい。一般に、パケット３１６はバイト列から構成され、さらに、ヘッダを含み、続いてボディ部を含む。ヘッダはパケット３１６の宛先を記述し、場合によっては、パケット３１６がリソース３７５のようなその最終的な宛先に到達するまで転送を行うために使用される物理ネットワーク内のルータを記述する。ボディ部は、仮想マシン２７０のようなパケット３１６の起点で生成されたデータまたはペイロードを含む。

【００４１】

一般に、リソース３７５およびデータセンタ２２５は、何らかの形態のコンピューティング・ユニット（例えば、中央処理装置、マイクロプロセッサ等）を備えるか、またはそれに接続されて、当該コンピューティング・ユニット上で実行されるエンドポイントおよび／またはコンポーネントの動作をサポートする。本明細書で使用する際、用語「コンピューティング・ユニット」は一般に、処理能力および記憶メモリを有する専用コンピューティング装置を指す。当該専用コンピューティング装置は１つもしくは複数のオペレーティング・システムまたは他の基盤ソフトウェアをサポートする。１つの事例では、コンピューティング・ユニットは、リソース３７５およびデータセンタ２２５と一体かまたは動作可能に接続され各装置が様々なプロセスおよび動作を実施できるようにする、有形のハードウェア要素、またはマシンで構成されている。別の事例では、コンピューティング・ユニットが、リソース３７５およびデータセンタ２２５の各々に収容されたコンピュータ読取可能媒体に接続されたプロセッサ（図示せず）を包含してもよい。一般に、コンピュータ読取可能媒体は、少なくとも一時的に、プロセッサにより実行可能な複数のコンピュータ・ソフトウェア・コンポーネント（例えば、エンドポイントＡ３９５およびＢ３８５）を格納する。本明細書で使用する際、用語「プロセッサ」は限定的な意味ではなく、あ

る計算能力で動作するコンピューティング・ユニットの任意の要素を包含することができる。かかる能力において、プロセッサを、命令を処理する有形の製品として構成してもよい。例示的な一実施形態では、処理は、命令をフェッチする、デコード／解釈する、実行する、および書き戻すものであってもよい。

【００４２】

ＴＣＰベースのトンネル３３０（「トンネル３３０」）を確立して、エンドポイントＡ３９５およびＢ３８５を含むサービス・アプリケーションのような１つのサービス・アプリケーションに割り当てたエンドポイント間で、または、割り当てたエンドポイント間の通信を異なるネットワークにわたって橋渡しするために独立のサービス・アプリケーションの役割を果たす複数のペア間で、通信してもよい。トンネル３３０はＴＣＰを使用し、ＴＣＰはアプリケーション層とネットワーク層／ＩＰ層との間にあるトランスポート層で通信サービスを提供する。トンネル３３０は、ＨＴＴＰまたはＨＴＴＰＳを使用するときは、アプリケーション層を４３０および４４０に含めてもよい。これは図４のＴＣＰ／ＩＰスタック４３０および４４０で示されており、トンネル３３０が仮想マシン２７０におけるトランスポート層の低レベルのチャネル４２５の間で論理接続の役割を果たす。

【００４３】

動作に際しては、アプリケーション・プログラムが、大量のチャンク・データをＩＰサイズのピースに分割して一連のＩＰ要求を発行するのではなく、ＩＰを用いて大量のチャンク・データをネットワーク（例えば、インターネット）にわたって送信したいときは、当該アプリケーション・プログラムはＴＣＰを使用するトランスポート層を介して１つの要求を発行してＩＰの細部を処理することができる。したがって、当該トランスポート層を、完全な配信を保証する伝送機構、例えば、その内容またはペイロードがその宛先に安全かつ健全に到達することを保証する責任を有する運搬手段とみなすことができる。諸事例においては、配信の保証では、多数の信頼性問題を解決しデータパケット３１６の信頼できる送信を提供する、トンネル３３０に組み込まれた信頼性機構が必要である。これらの信頼性機構は高レベルで動作し、２つのエンド・システム（例えば、Ｗｅｂブラウザおよびウェブサーバ）と関係する。

【００４４】

特に、ＴＣＰは、１組のルールを強制することにより或るコンピュータ上のエンドポイント３９５から別のコンピュータ上のエンドポイント３８５へのパケット３１６のストリームの信頼できる順序付けられた配信を提供する。当該１組のルールでは、データパケット３１６が順番に到着すること、データパケット３１６にエラーがないこと（即ち、正確性）、重複するデータパケット３１６が廃棄されること、および損失／遅延したデータパケット３１６が再送されることを記述してもよい。以上の１組のルールの例を、図５の損失回復機構５０５により強制してもよい。損失回復機構５０５はデータ・ストリームを検査して任意の損失パケット３１６を特定する。当該信頼性機構はまた、ＩＰパケット内で運搬されるデータの量を管理し、一般にトンネル３３０上のトラフィックの輻輳を管理するように構成された、図５の輻輳制御機構５１５を備えてもよい。

【００４５】

一般に、トンネル３３０は、２つのエンドポイント間の経路の一部または経路全体に沿ってトランスポート層を介してエンドポイントを接続する、論理接続を表す。諸実施形態では、トンネル３３０は、ネットワーク３１５を介してＩＰネットワーク境界にわたるブリッジ接続を生成するように設計された、ＩＰ－ＨＴＴＰＳ、ＳＳＬ、ＳＳＴＰまたは別のＴＣＰベースのトンネリング技術を利用する。したがって、トンネル３３０は、基礎となる物理ネットワークから独立したエンドポイント３８５と３９５との間の論理接続を表面上確立し、それにより、エンドポイント３８５および３９５があたかもデータセンタ２２５内で隣接しているかのように対話することができる。

【００４６】

例示的な一実施形態では、トンネル３３０が、互いの上に積層された１つまたは複数のＴＣＰベース接続を含んでもよい。図４を参照すると、本発明の一実施形態に従う、トン

ネル３３０を含むＴＣＰベース接続の階層の略図が示されている。図４に示すように、互いの上で実行される２つのＴＣＰベース接続が存在する。これらのＴＣＰベース接続は双方とも、データパケット３１６の各ストリームに対して（上述の）１組のルールを強制して完全で効率的な配信を保証する。当該２つのＴＣＰベース接続は、高レベルのチャネル４１５および低レベルのチャネル４２５を含む。高レベルのチャネル４１５は、カプセル化されていないデータを仮想マシン２７０からデータセンタ２２５内のエンドポイントＡ３９５に流す。カプセル化されていないデータをエンドポイントＡ３９５でカプセル化して低レベルのチャネル４２５に置く。低レベルのチャネル４２５は、データパケット３１６にカプセル化されたデータをエンドポイントＢ３８５に運ぶ。エンドポイントＢ３８５に到達すると、データパケット３１６は低レベルのチャネル４２５で受け取られ、非カプセル化され、高レベルのチャネル４１５を介してリソース３７５に転送される。

【００４７】

トンネル３３０を含む２つの異なるチャネルを説明したが、ＩＰデータパケットを流す他種の適切な接続を使用してもよいこと、および、本発明の諸実施形態は本明細書で説明したこれらのチャネル４１５および４２５には限定されないことは理解されよう。

【００４８】

損失回復機構５０５および輻輳制御機構５１５（図５を参照）を双方とも、高レベルのチャネル４１５および低レベルのチャネル４２５の各々に組み込み、それらの上で一体的に実行してもよい。これらの機構５０５および５１５を並列かつアクティブに動作させるとしばしば、重複で冗長な動作により性能が劣化することとなる。例えば、データパケット３１６が損失した場合は、チャネル４１５および４２５の双方の中にある機構５０５と５１５の双方が夫々、損失回復および輻輳制御を実施する。したがって、図２および図５のファブリック・コントローラ２１０を使用して、チャネル４１５と４２５の何れかにある機構５０５と５１５の何れかを独立して選択的に設定（オンまたはオフ）してもよい。

【００４９】

次に図５を参照すると、本発明の諸実施形態に従って、ファブリック・コントローラ２１０を使用して、ＴＣＰベースのトンネル３３０に組み込まれた信頼性機構５０５および５１５を選択的に有効化／無効化する例示的なデータセンタ２２５を示すブロック図が示されている。示したように、データは高レベルのチャネル４１５を介して仮想マシン２７０からエンドポイントＡ３９５に流される。エンドポイントＡ３９５はストリーミング・データをカプセル化して、低レベルのチャネル４１５を介してネットワーク３１５に転送する。ファブリック・コントローラ２１０は、低レベルのチャネル４２５に組み込まれた損失回復機構５０５および／または輻輳制御機構５１５を選択的に無効化する命令５１０をエンドポイントＡ３９５に運ぶことができる。これらの命令５１０を諸基準により促してもよい。これらの命令５１０は一般に、トンネル３３０に使用されるＴＣＰベースのチャネル４１５および４２５の接続ごとに発行される。接続ごとに、または、ソケット・オプションごとに、損失回復機構５０５および／または輻輳制御機構５１５をオンまたはオフにするための命令５１０を独立に、消費者（例えば、クラウド・コンピューティング・サービス・モデル、ネゴシエーション型のサービスレベル・アグリーメント等）が提供しうるポリシのような諸基準により起動してもよい。別の実施形態では、機構５０５および５１５の選択的な無効化を、ソース・エンドポイント（エンドポイントＡ３９５）の識別および宛先エンドポイント（エンドポイントＢ３８５）の識別、運ばれているデータのタイプ、確立されているトンネルのタイプ、実行されているアプリケーション／動作のタイプ、またはポリシーベースの情報に関連する諸基準により行ってもよい。

【００５０】

例示的な一実施形態では、ファブリック・コントローラ２１０が、低レベルのチャネル４２５上で実行されている機構５０５および５１５を無効化するための命令５１０を発行し、高レベルのチャネル４１５上で実行されている機構（図示せず）を有効化させたままにしておくことにより、データ・ストリームのエンド・ツー・エンドの信頼性を提供し、遅延の原因となる二重機構の冗長性を低減する。１つの事例では、低レベルのチャネル４

２５の機構５０５および５１５を無効化する構成を、デフォルトの設定としてファブリック・コントローラ２１０内に提供する。本実施形態では、ファブリック・コントローラ２１０が機構５０５および５１５を、有効化する入力がない限り、自動的に無効化する。

【００５１】

諸実施形態では、輻輳制御機構５１５を無効化するとき、ファブリック・コントローラ２１０が、トンネル３３０の受信側（例えば、図３乃至５の宛先エンドポイント、エンドポイントＢ３８５、または図３および４のリソース３７５）に変更を加えずに、トンネル３３０の送信側（例えば、図３乃至５のソース・エンドポイント、エンドポイントＡ３９５、または図２乃至５の仮想マシン２７０）のコンポーネントに作用してもよい。したがって、受信側と何らネゴシエーションすることなく輻輳制御機構５１５をオフにすることができ、したがって、輻輳制御機構５１５は下位互換である。動作に際しては、図５を参照すると、エンドポイントＡ３９５に作用する際、このソース・エンドポイントが、高レベルのチャネル４１５の輻輳制御機構（図示せず）が指定したデータ量またはデータパケット３１６の速度を送信する。したがって、高レベルのチャネル４１５の輻輳制御機構のみがデータ量を制御し、その結果、ネットワーク３１５は、大量のデータ損失の原因となる大量のデータで溢れかえることはない。

【００５２】

諸実施形態では、損失回復機構５０５を無効化するとき、ファブリック・コントローラ２１０が、トンネル３３０の送信側（例えば、図３乃至５のソース・エンドポイント、エンドポイントＡ３９５、または図２乃至５の仮想マシン２７０）のコンポーネント、およびトンネル３３０の受信側（例えば、図３乃至５の宛先エンドポイント、エンドポイントＢ３８５、または図３および４のリソース３７５）に作用してもよい。したがって、トンネル３３０の送信側と受信側との間のネゴシエーションを介した協調により、損失回復機構５１５をオフにすることができる。即ち、受信側はネゴシエーションを通して、データパケット３１６の損失が送信側により回復されないことを認識するようになる。したがって、動作に際しては、中間データがリソース３７５まで損失するか、またはエンドポイントＢ３８５の下流であるサービス・アプリケーションの他の宛先エンドポイントまで損失した可能性があっても、受信側は配信されたデータパケット３１６を転送する。しかし、トンネル３３０の受信側が損失回復機構５０５のオフをサポートしない場合は、損失回復機構５０５は一般にアクティブで有効化されたままとなる。

【００５３】

ネゴシエーションの１つの事例では、エンドポイントＡ３９５が（ネゴシエーション情報を運ぶ）初期同期（ＳＹＮ）パケットをエンドポイントＢ３８５に送信してもよい。エンドポイントＢ３８５は、ＳＹＮパケットの受領を認識することができる。さらに、エンドポイントＢ３８５が、応答（ＳＹＮ－ＡＣＫ）パケットをエンドポイントＡ３９５に送信することによって、エンドポイントＡ３９５とのハンドシェークに返答してもよい。この時点で、返答がなされると、低レベルのチャネル４２５の損失回復機構５０５が無効化され、高レベルのチャネル４１５の損失回復機構（図示せず）がアクティブで有効化されたままとなる。それにより、ネットワーク３１５において損失または遅延した何らかのデータを再送信することで、データパケット３１６の配信中の何らかの損失または遅延からの回復が支援される。

【００５４】

ここで図６を参照すると、本発明の一実施形態に従って、異なるネットワークに存在する複数のエンドポイント間で確立されたネットワーク接続にわたって通信を促進するための方法６００を示す流れ図が示されている。ブロック６０２に示すように、ネットワーク接続が提供される。１つの事例では、ネットワーク接続は、ソース・エンドポイント（例えば、図４のエンドポイントＡ３９５）と宛先エンドポイント（例えば、図４のエンドポイントＢ３８５）との間に跨るものである。上述したように、ネットワーク接続は、ソース・エンドポイントおよび宛先エンドポイントが夫々存在する異なるネットワーク（例えば、図３のクラウド・コンピューティング・プラットフォーム２００および企業プライベ

ート・ネットワーク３２５）を橋渡しするＴＣＰベースのトンネルとして動作する。方法６００は、ブロック６０４および６０６で、ＴＣＰベースのトンネル上で一体的に実行されている１つまたは複数の低レベルの信頼性機構を選択的に無効化するステップ、および低レベルの信頼性機構がメッセージの送信タイミングに干渉することなく第１のエンドポイントと第２のエンドポイントとの間でメッセージを送信するステップを行う。諸実施形態では、第１のエンドポイントと第２のエンドポイントとの間でメッセージを送信するステップには、特に、ＴＣＰベースのトンネルを介してＩＰパケットを第１のエンドポイントから第２のエンドポイントに送信するステップを含んでもよい。

【００５５】

例示的な実施形態では、選択的に無効化する動作を、所定の基準に対応する事象が発生した際に開始してもよい。当該所定の基準がファブリック・コントローラに知らされていてもよく、当該ファブリック・コントローラが、以下の対応するイベント、即ち、所定のポートがトンネルを通じてデータパケットを送信しようとしていること、新たなネットワーク接続が所定のデータセンタ、仮想マシン、もしくはソース・エンドポイントで確立されていること、または、新たなネットワーク接続が所定のリソースもしくは宛先エンドポイントで確立されていることを検出した際に、１つまたは複数の信頼性機構（例えば、図５の損失回復機構５０５および輻輳制御機構５１５）を選択的に無効化してもよい。

【００５６】

別の実施形態では、データセンタのユーザまたはクライアントに、クラウド・コンピューティング・プラットフォームでインスタンス化された一群の信頼性機構をオン／オフするための手動制御を行う権限を付与してもよい。このように、クライアントやユーザは、ＵＩでトンネルのエンドポイントを指定することにより、１つまたは複数の信頼性機構を無効化または有効化するかどうかを動的に判定することができる。したがって、ユーザまたはクライアントは、通常は輻輳制御と損失回復を冗長かつ非効率に実施するはずである、ＴＣＰのような重複した信頼できるプロトコルの層をトンネル上で実行することに関連する性能ペナルティを回避すべきかどうかを判定することができる。

【００５７】

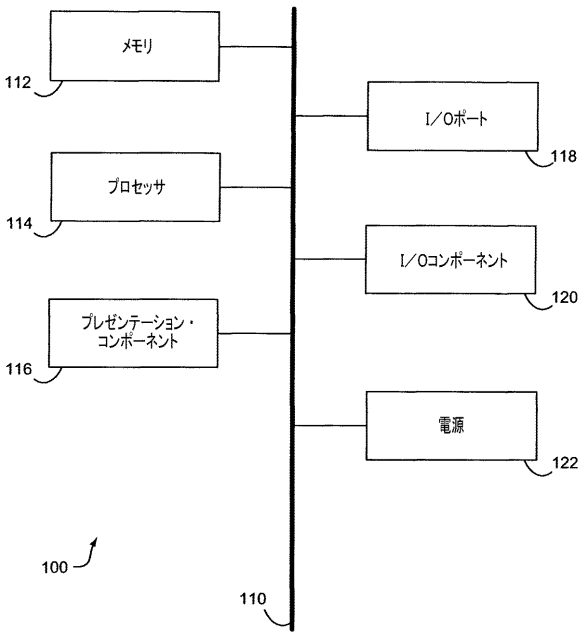
図７を参照すると、本発明の一実施形態に従って、ＴＣＰベースのトンネルにわたってソース・エンドポイントと宛先エンドポイントとの間の通信を促進するための方法７００を示す流れ図が示されている。ブロック７０２で示すように、方法７００は、ファブリック・コントローラを使用して、１つのネットワーク上でまたは異なるネットワークにわたってソース・エンドポイントおよび宛先エンドポイントを通信可能に接続するＴＣＰベースのトンネルを確立するステップを含む。諸実施形態では、（例えば、図３のデータセンタ２２５を利用して）ソース・エンドポイントの動作がデータセンタによりサポートされ、宛先エンドポイントの動作が、（例えば、図３のリソース３７５を利用して）プライベート企業ネットワークに存在する遠隔に配置されたリソースによりサポートされる。この場合、当該リソースをデータセンタのクライアントが管理／所有してもよい。方法７００はさらに、第１のＩＰパケット・ストリームをソース・エンドポイントで受け取るステップを含み、当該第１のＩＰパケット・ストリームはデータセンタ内でインスタンス化された仮想マシンから渡される。このステップをブロック７０４で示す。幾つかの実施形態では、この第１のＩＰパケット・ストリームは、第１の１組の信頼性機構が提供されている高レベルのチャネルを介して運ばれる。ブロック７０６および７０８で示すように、ソースのトンネル・エンドポイントに到着すると、第１のＩＰパケット・ストリームを第２のＩＰパケット・ストリームにカプセル化する。当該第２のＩＰパケット・ストリームは低レベルの接続を介してＴＣＰベースのトンネルを通じて送信される。ＴＣＰベースのトンネルの１つの構成では、低レベルの接続を高レベル・チャネルの下方で積層する（layer）ことで、協調してＩＰパケットを運び、その運搬の信頼性を保証する。このように、低レベルのチャネルにはしばしば第２の１組の信頼性機構が提供される。例示的な一実施形態では、第１の１組の信頼性機構および第２の１組の信頼性機構の各々が、少なくとも輻輳制御機構および損失回復機構を夫々備える。

【００５８】
方法７００は続いて、ブロック７１０で示すように、ファブリック・コントローラを使用して、低レベルのチャネルに提供された輻輳制御機構および損失回復機構を選択的に無効化するステップを有する。ブロック７１２で示すように、ファブリック・コントローラを使用して、高レベルのチャネルに提供された輻輳制御機構および損失回復機構を受動的に有効化されたままにすることもできる。ブロック７１４で示すように、低レベルのチャネルの無効化された状態および高レベルのチャネルの有効化された状態を記憶する。明確さのため、当該無効化された状態は、低レベルのチャネルに提供された輻輳制御機構および損失回復機構が無効化されていることを表す。反対に、有効化された状態は、高レベルのチャネルに提供された輻輳制御機構および損失回復機構が有効化されていることを表す。チャネルの状態を、ファブリック・コントローラがアクセス可能なデータセンタ、リソース、エンドポイント、または他の任意の場所に記憶してもよい。

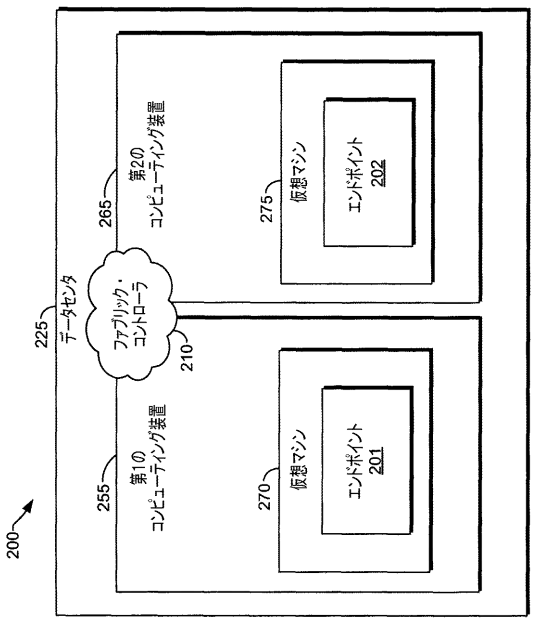
【００５９】
本発明の諸実施形態を特定の実施形態に関連して説明した。これらは全ての点で、限定的ではなく例示的であると意図している。本発明の範囲から逸脱しない本発明の諸実施形態が関連する代替実施形態は、当業者には明らかであろう。

【００６０】
以上から、本発明が、上述の目標および目的の全てを本システムおよび方法に自明で固有である他の利点とともに実現するように良く適合したものであることが分かる。特定の機能および副次的組合せが有用であり、これらを他の機能および副次的組合せを参照せずに使用できることは理解されよう。これは、特許請求の範囲により考慮されており、特許請求の範囲内にある。

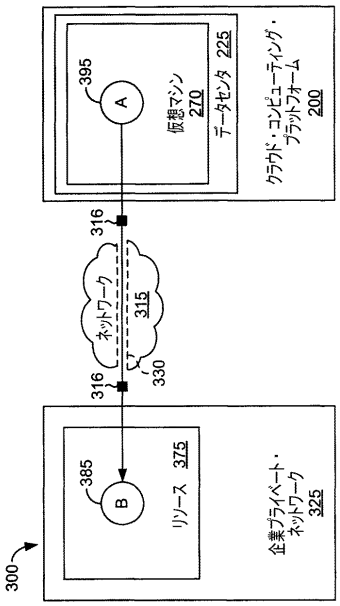
【図１】



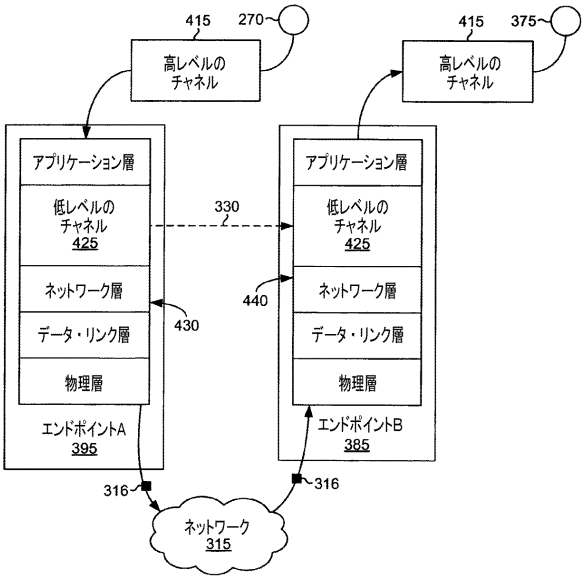
【図２】



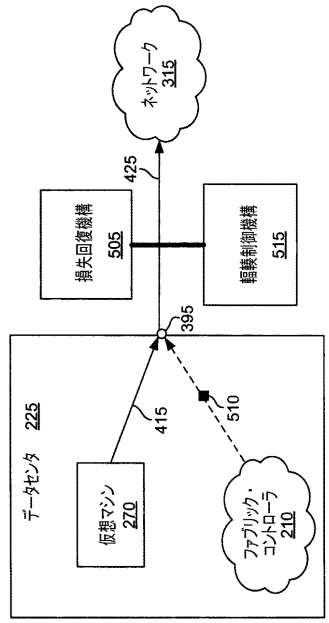
【図３】


クラウド・コンピューティング・プラットフォーム 200
データセンタ 225
仮想マシン 270
395
A
316
ネットワーク 315
330
385
B
リソース 375
企業プライベート・ネットワーク 325
300

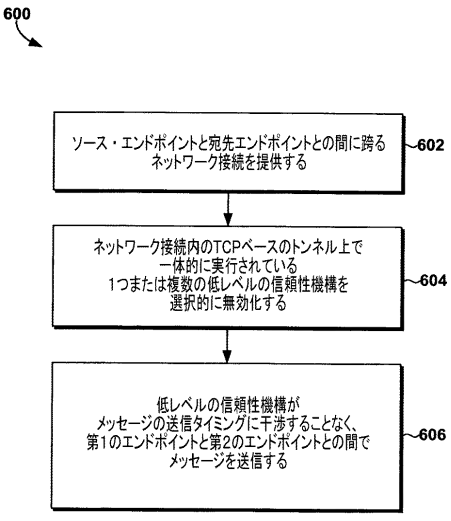
【図４】


415
270
高レベルのチャネル
415
375
高レベルのチャネル
アプリケーション層
低レベルのチャネル 425
ネットワーク層
データ・リンク層
物理層
エンドポイントA 395
330
430
440
エンドポイントB 385
316
316
ネットワーク 315

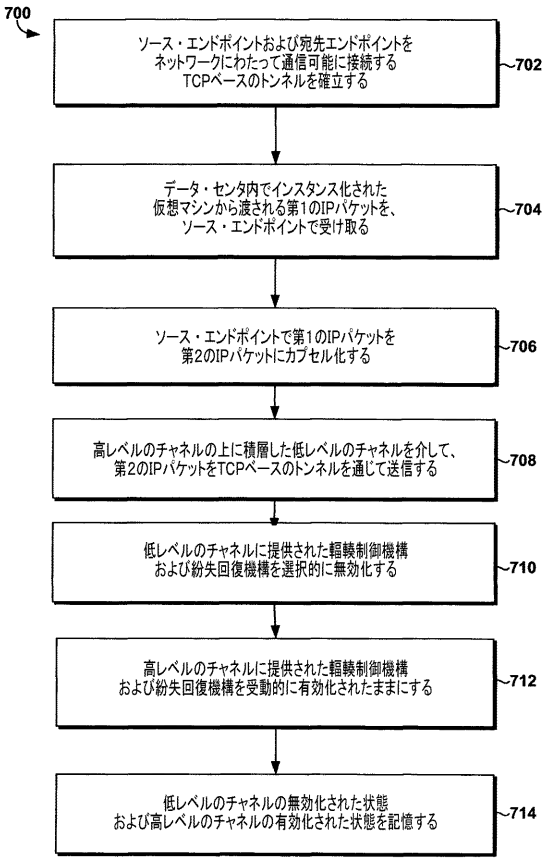
【図５】


ネットワーク 315
425
損失回復機構 505
輻輳制御機構 515
395
データセンタ 225
415
510
仮想マシン 270
ファブリック・コントローラ 210

【図６】


600
ソース・エンドポイントと宛先エンドポイントとの間に跨るネットワーク接続を提供する
602
ネットワーク接続内のTCPベースのトンネル上で一体的に実行されている1つまたは複数の低レベルの信頼性機構を選択的に無効化する
604
低レベルの信頼性機構がメッセージの送信タイミングに干渉することなく、第1のエンドポイントと第2のエンドポイントとの間でメッセージを送信する
606

【図７】


700
ソース・エンドポイントおよび宛先エンドポイントを
ネットワークにわたって通信可能に接続する
TCPベースのトンネルを確立する
702
データ・センタ内でインスタンス化された
仮想マシンから渡される第1のIPパケットを、
ソース・エンドポイントで受け取る
704
ソース・エンドポイントで第1のIPパケットを
第2のIPパケットにカプセル化する
706
高レベルのチャネルの上に積層した低レベルのチャネルを介して、
第2のIPパケットをTCPベースのトンネルを通じて送信する
708
低レベルのチャネルに提供された輻輳制御機構
および紛失回復機構を選択的に無効化する
710
高レベルのチャネルに提供された輻輳制御機構
および紛失回復機構を受動的に有効化されたままにする
712
低レベルのチャネルの無効化された状態
および高レベルのチャネルの有効化された状態を記憶する
714

フロントページの続き

(72)発明者　ディーパック　バンサル
アメリカ合衆国　９８０５２－６３９９　ワシントン州　レッドモンド　ワン　マイクロソフト　ウェイ　マイクロソフト　コーポレーション　エルシーエー－インターナショナル　パテンツ内
(72)発明者　ハサン　アルカティブ
アメリカ合衆国　９８０５２－６３９９　ワシントン州　レッドモンド　ワン　マイクロソフト　ウェイ　マイクロソフト　コーポレーション　エルシーエー－インターナショナル　パテンツ内

審査官　松崎　孝大

(56)参考文献　特開２００８－０７８９６６（ＪＰ，Ａ）
特表２００８－５０７９２８（ＪＰ，Ａ）
国際公開第２００６／０９３０２１（WO，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｈ０４Ｌ　１２／８０１