

明 細 書

発明の名称：

無線端末にＩＰネットワークへのアクセスを提供するための通信システム
及び通信方法

技術分野

[0001] 本発明は、無線端末にIPネットワークへのアクセスを提供するための通信システム及び通信方法に関する。

背景技術

[0002] 近年、MVNO（仮想移動体通信事業者）の登場により無線通信回線の小売が進んでいる。従来は、無線通信インフラを有するMNO（移動体通信事業者）がエンドユーザーに直接無線通信サービスを提供してきたが、MVNOは、MNOの無線通信インフラを利用してエンドユーザーに独自の無線通信サービスを提供する。

[0003] MVNOには、自社で一切通信インフラを有しない形態と、自社でも通信インフラを有し、その通信インフラをMNOの通信インフラに接続して無線通信サービスを提供する形態（図１参照）に大別することができる。後者は前者と比較して、自社でも通信インフラを有することから、通信速度、通信容量等の通信品質に応じた価格設定が可能であり、さまざまなニーズに応えることが試みられている。たとえば、高速データ通信が可能な最大通信容量を設定することで価格を抑えた無線通信サービス用のSIMカードが量販店で販売されている。

[0004] 具体的には、このようなMVNOは、MNOからSIMカードの提供を受けて、それをさらにエンドユーザーに提供している。MVNOは、各SIMカードに設定された速度制限、容量制限等の通信品質を各SIMカードの識別番号とともにデータベースで管理しており、インターネットへのアクセスを要求してきたスマートフォン、タブレットなどの携帯端末の契約内容に応じて無線通信サービスを提供している。

- [0005] MN0とMVN0の間に、MVN0が円滑な事業を行うための支援サービスを提供するMVNE（仮想移動体通信サービス提供者）が介在し、MVNEがMN0からSIMカードの提供を受けて、それをさらにMVN0に提供する場合もある。
- [0006] MVN0又はMVNEが、コアネットワークのエンドポイントであるゲートウェイ（3GではGGSN、LTEではP-GW）を自社の通信インフラとして保有する場合には、MN0の通信インフラ（3GのSGSN、LTEのS-GW）とL2接続（full MVN0 connection（LTEではS5/S8 interface、3GではGn interface））が可能となり、エンドユーザーのSIMカードが装着された無線端末とゲートウェイとの間にL2トンネルが形成される。L2接続は、L3接続（light MVN0 connection）と比較して無線通信サービスの設計自由度が高まり、注目を集めている。

発明の概要

発明が解決しようとする課題

- [0007] L2接続のためにMVN0又はMVNEが保有する必要のあるゲートウェイは、1台数億円の高価なハードウェアであり、その性能として、同時接続可能機器数が重視されるところ、さまざまなニーズに応えるべくゲートウェイにおいて付加的なデータ処理を行うと、ゲートウェイが実装されるサーバの物理的限界により、同時接続可能機器数の低下を必然的に招いてしまう。
- [0008] そして、MVN0又はMVNEがMN0の通信インフラと接続可能なゲートウェイの数は、MN0により予め定められていて台数制限が設けられている場合が大半で、コストを度外視しても、制限台数より多くハードウェアを追加して同時接続可能機器数の増加を受容するといったことはできない。
- [0009] たとえば、無線通信サービスに対するニーズとして近年顕著に増加しているのが、あらゆるモノに無線通信機能を加えてインターネットにつなげるIoTの動きである。以下、インターネットを含めコンピュータネットワークに接続可能な無線機器を「IoT機器」と呼ぶ。IoT機器は、ヒトが保持するスマートフォン、タブレットなどの携帯端末も含むが、ヒトよりも高速に移動する自動車等の移動手段、逆にあまり特定の位置から移動しないセンサーなど、ヒトとは異なる形での通信を必要とする機器に拡がりを見せており、それに

に伴い、求められる通信速度や通信の頻度は多岐に渡る。産業機械にSIMカードを組み込んでIoT化し、遠隔で他の機器と通信させることを考えた場合には、各IoT機器は低容量の通信を低頻度で行うのみであっても、無数の機器が通信を行うことを考慮すれば、多くの同時接続数が求められる。このようなニーズに応える上では、同時接続可能機器数の増加を抑制することなく、無線通信サービスの高付加価値化を実現する必要がある。

[0010] 類似の問題は、エンドユーザーの無線端末とL3接続されるゲートウェイにおいても発生し得る。また、MNOが保有するゲートウェイに関しても、同様のことが問題となり得る。

[0011] さらに、ゲートウェイにおいて付加的なデータ処理を行わなくとも、同時接続可能機器数は、GTPヘッダのトンネリング処理によっても制限を受けており、こうしたデータ処理の負担も同時接続可能機器数の増加を可能とする上で障害となっている。

[0012] 本発明は、このような問題点に鑑みてなされたものであり、その目的は、無線端末にIPネットワークへのアクセスを提供するための通信システム及び通信方法において、従来よりも同時接続可能機器数の増加を可能とすることにある。

課題を解決するための手段

[0013] このような目的を達成するために、本発明の第1の態様は、無線端末にIPネットワークへのアクセスを提供するための通信システムであって、前記無線端末からのデータを前記IPネットワークに通過させるためのゲートウェイを備え、前記ゲートウェイは、前記無線端末からのデータを受信する受信部と、受信した前記データのヘッダを元に複数の送信先アドレスのいずれかから1つを選択する選択部と、前記データを前記選択部が選択した送信先アドレスに転送する転送部とを有する第1のサーバ群を有し、前記第1のサーバ群のそれぞれは、前記複数の送信先アドレスのいずれかに対応する第2のサーバ群を構成するサーバに前記データを転送することを特徴とする。

[0014] また、本発明の第2の態様は、第1の態様において、前記ゲートウェイがG

GSN又はP-GWであることを特徴とする。

- [0015] また、本発明の第3の態様は、第1又は第2の態様において、前記通信システムは、前記無線端末からの接続開始要求を受信するCプレーン・サーバ群をさらに備え、前記第1のサーバ群は、前記Cプレーン・サーバ群が受けた接続開始要求を契機に第1又は第2の態様に記載の処理を行うことを特徴とする。
- [0016] また、本発明の第4の態様は、第1から第3のいずれかの態様において、前記第1のサーバ群を構成するサーバ数が予め定められた数以下であることを特徴とする。
- [0017] また、本発明の第5の態様は、第4の態様において、前記予め定められた数がMN0により定められていることを特徴とする。
- [0018] また、本発明の第6の態様は、第1から第5のいずれかの態様において、前記第2のサーバ群を構成するサーバ数が前記第1のサーバ群を構成するサーバ数を超えることを特徴とする。
- [0019] また、本発明の第7の態様は、第1から第6のいずれかの態様において、前記第1及び第2のサーバ群は、それぞれ、前記ヘッダが指示していた前記送信先アドレスと、前記複数の送信先アドレスとの間の対応テーブルを前記無線端末が確立したトンネル毎に有することを特徴とする。
- [0020] また、本発明の第8の態様は、第7の態様において、前記対応テーブルが前記ヘッダ内のGTPヘッダが有する識別子と関連づけられていることを特徴とする。
- [0021] また、本発明の第9の態様は、第1から第8のいずれかの態様において、前記選択部は、前記ヘッダが指示する送信先アドレスを書き換えることで転送先の複数の送信先アドレスのいずれかを特定することを特徴とする。
- [0022] また、本発明の第10の態様は、第1から第9のいずれかの態様において、前記複数の送信先アドレスは、同一の送信先アドレスを割り当てないように決定されることを特徴とする。
- [0023] また、本発明の第11の態様は、第1から第9のいずれかの態様において

、前記複数の送信先アドレスのうちの少なくとも一部が同一であることを特徴とする。

[0024] また、本発明の第12の態様は、第1から第11のいずれかの態様において、前記無線端末と前記第1のサーバ群との間にL2トンネルが通信経路として形成されることを特徴とする。

[0025] また、本発明の第13の態様は、第1から第12のいずれかの態様において、前記第2のサーバ群を構成する各サーバがクラウド上のインスタンスであることを特徴とする。

[0026] また、本発明の第14の態様は、第13の態様において、前記第1のサーバ群を構成する各サーバがクラウド上のインスタンスであることを特徴とする。

[0027] また、本発明の第15の態様は、第14の態様において、前記第2のサーバ群が前記第1のサーバ群の各サーバで共用されることを特徴とする。

[0028] また、本発明の第16の態様は、第13から第15のいずれかの態様において、前記第2のサーバ群と接続された第3のサーバ群を備え、前記第3のサーバ群は、前記第2のサーバ群を構成するサーバ数を制御することを特徴とする。

[0029] また、本発明の第17の態様は、第16の態様において、前記第3のサーバ群が前記第2のサーバ群を構成するインスタンスにより構成されることを特徴とする。

[0030] また、本発明の第18の態様は、第13から第17のいずれかの態様において、前記第2のサーバ群を構成するインスタンス数が前記第2のサーバ群の処理負荷に応じてゼロまで減少可能であることを特徴とする。

[0031] また、本発明の第19の態様は、第13から第18のいずれかの態様において、前記ゲートウェイがマルチテナントであることを特徴とする。

[0032] また、本発明の第20の態様は、第1から第19のいずれかの態様において、前記第2のサーバ群が前記データに対してデータ処理を行うことを特徴とする。

- [0033] また、本発明の第21の態様は、第20の態様において、前記データ処理は、前記データのペイロードの暗号化、前記データのプロトコル変換、前記無線端末からのデータ送信のスループット制御、及び、前記データのペイロードに対するクレデンシャル付与のうちの少なくとも1つであることを特徴とする。
- [0034] また、本発明の第22の態様は、第1から第21のいずれかの態様において、前記第2のサーバ群は、前記ゲートウェイから前記無線端末に送信されるデータに対して下り時のデータ処理を行うことを特徴とする。
- [0035] また、本発明の第23の態様は、第22の態様において、前記下り時のデータ処理は、前記データに含まれる画像又は動画の解像度変換、前記無線端末のデータ受信のスループット制御、前記無線端末のデータ受信の優先度制御、及び、HTML5に準拠した機能の実行のうちの少なくとも1つであることを特徴とする。
- [0036] また、本発明の第24の態様は、無線端末にIPネットワークへのアクセスを提供するための通信方法であって、前記無線端末からのデータを前記IPネットワークに通過させるためのゲートウェイにおいて、前記ゲートウェイが有する第1のサーバ群が、前記無線端末からのデータを受信する受信ステップと、前記第1のサーバ群が、受信した前記データのヘッダを元に複数の送信先アドレスのいずれかから1つを選択する選択ステップと、前記第1のサーバ群が、前記データを前記選択された送信先アドレスに転送する転送ステップとを含み、前記転送ステップは、前記第1のサーバ群のそれぞれが、前記複数の送信先アドレスのいずれかに対応する第2のサーバ群を構成するサーバに前記データを転送することを特徴とする。
- [0037] また、本発明の第25の態様は、無線端末からのデータをIPネットワークに通過させるためのゲートウェイに、前記無線端末に前記IPネットワークへのアクセスを提供するための通信方法を実行させるためのプログラムであって、前記通信方法は、前記ゲートウェイにおいて、前記ゲートウェイが有する第1のサーバ群が、前記無線端末からのデータを受信する受信ステップと

、前記第1のサーバ群が、受信した前記データのヘッダを元に複数の送信先アドレスのいずれかから1つを選択する選択ステップと、前記第1のサーバ群が、前記データを前記選択された送信先アドレスに転送する転送ステップとを含み、前記転送ステップは、前記第1のサーバ群のそれぞれが、前記複数の送信先アドレスのいずれかに対応する第2のサーバ群を構成するサーバに前記データを転送することを特徴とする。

発明の効果

[0038] 本発明の一態様によれば、無線端末からのデータをIPネットワークに通過させるためのゲートウェイにおいて、無線端末からのデータを受信する第1のサーバ群と、付加的なデータ処理又は第1のサーバ群で行われなかったデータ処理を行う部分である第2のサーバ群とを切り分けることによって、第1のサーバ群の同時接続可能数の上限を引き上げつつ、第2のサーバ群においてより多くのコンピューティングリソースを用いたデータ処理を行うことが可能となる。

図面の簡単な説明

[0039] [図1]自社の通信インフラをMNOの通信インフラに接続して無線通信サービスを提供するMVNOを模式的に示す図である。

[図2]本発明の一実施形態におけるゲートウェイを備える通信システムの概要を示す図である。

[図3]無線端末からゲートウェイに送信されるパケットを例示的に示す図である。

[図4]本発明の一実施形態におけるGTPヘッダの書き換えを示す図である。

[図5]本発明の別の実施形態におけるゲートウェイを備える通信システムの概要を示す図である。

[図6]第1のサーバ群と第2のサーバ群との対応関係の決定プロセスの一例を示す図である。

[図7]第1のサーバ群と第2のサーバ群との対応関係の一例を示す図である。

[図8]本発明の一実施形態におけるペイロードに対する付加的なデータ処理を

示す図である。

[図9]本発明の一実施形態におけるプロトコル変換のデータ処理の流れを示す図である。

発明を実施するための形態

[0040] 以下、本発明の実施形態を図面を参照して説明する。

[0041] 本明細書では、MVNOがゲートウェイを保有する場合を例に説明するが、本発明は、MVNE又はMNOが保有する場合にも適用できることを付言しておく。

[0042] また、MNO、MVNE、MVNOという用語は、国によりその定義が異なることがある。以下では、日本におけるように、MNOは3GのSGSN、LTEのS-GWを通信インフラとして保有し、MVNOが3GのGGSN、LTEのP-GWを通信インフラとして保有する場合を例に説明する。いずれにしても、設置可能台数に制約のあるゲートウェイの下でも同時接続可能機器数の増加を可能とする本発明を適用可能であることは変わらない。

[0043] (本発明の概要)

図2に、本発明にかかるゲートウェイを備える通信システムの概要を示す。

[0044] 本発明にかかる通信システム200は、無線端末にIPネットワークなどへのアクセスを提供するための通信システムであって、当該無線端末からのデータをIPネットワークに通過させるためのゲートウェイ210を備える。

[0045] ゲートウェイ210は、第1のサーバ群211を有する。たとえば、MNOの通信インフラ（コアネットワーク）と接続できるサーバの台数には制限があり、これは一般に、MNOにより定められており、超えることができない。たとえば、第1のサーバ群211を構成するサーバのIPアドレスを、予めMNOのデータベースに登録し、MNOが、登録されたIPアドレスのリストを参照して、無線端末からのデータを第1のサーバ群211のうちのいずれのサーバに伝送するかを決定している。たとえば、ラウンドロビン方式で割り振ることができる。登録する情報は、IPアドレス以外にMACアドレスなどの送信先を識別するための送信先アドレスとしてもよい。

- [0046] 第1のサーバ群211を構成する各サーバは、無線端末からのデータを受信する受信部と、受信したデータのヘッダを元に複数の送信先アドレスのいずれかから1つを選択する選択部と、受信したデータを選択部が選択した送信先アドレスに転送する転送部とを有する。
- [0047] 転送部では、本発明の一実施形態においては、無線端末からの受信した当該データのヘッダが指示するIPアドレスを複数のIPアドレスのいずれかに書き換える。データは、図3に示すように、ペイロード及びヘッダを有する。図3では、HTTP層よりも上位の部分ペイロードとして示しているが、たとえばGTPヘッダに着目して、GTP層よりも上位（図3では左側）の部分ペイロード、この場合にはGTPペイロードとみることできる。
- [0048] 書き換えは、図4に示すように、ヘッダ内のIPヘッダにより指示されるIPアドレスを書き換えることにより行うことができる。具体的には、第1のサーバ群211が、到着したIPパケットに含まれるGTPヘッダの中の識別子（TEID）からIMSI、ICCIDなどのSIMカードの識別番号を判別することで、どの無線端末からのデータであるかを識別し、当該識別子に関連づけられた第2のサーバ群212のうちのいずれかのサーバのIPアドレスに書き換えることができる。
- [0049] また、書き換えは、第1のサーバ群211のそれぞれが受信したデータを第2のサーバ群212を構成するサーバに転送する際に必要な処理の一例であり、他のトンネリング（IPIP、GRE、IPsec、GTPなど）を行って転送することもできる。
- [0050] 選択部においては、転送先の第2のサーバ群212に対応する複数のIPアドレスは、プライベートIPアドレス又はグローバルIPアドレス空間の中から、第2のサーバ群212の複数のサーバに同時に同一のIPアドレスを割り当てないように決定することができる。また、複数のIPアドレスのうちの少なくとも一部を同一のものとして、これを複数のサーバに同時に割り当てることもでき、この場合には、同一のIPアドレスが与えられた割り振られたサーバのうちの1つだけをアクティブにしてその他をスタンバイにすればよい。

また、契約内容に応じて、一部のIPアドレスを特定の識別子（TEID）の場合にのみアクセス可能とすることもできる。そして、第1のサーバ群211及び第2のサーバ群212は、それぞれ、ヘッダが指示していた書き換え前のIPアドレスと、書き換え後の複数のIPアドレスとの間の対応テーブルをL2又はL3トンネル毎に保持することができる。各対応テーブルは、GTPヘッダ内の識別子等によって、各無線端末との関連づけが可能である。また、対応テーブルを第2のサーバ群212にのみ持たせ、第1のサーバ群211は第2のサーバ群212に問い合わせることもできる。

[0051] ここでは、ゲートウェイ210の動作として、携帯端末との間にL2トンネルが通信経路として形成された後のIPネットワークへのアクセスのためにGTPのUプレーン上で通信を行う場面を主に想定して説明しているが、L2トンネルを形成するためにGTPのCプレーン上で通信を行う場面においても、同様にIPアドレスの書き換えを行うことも考えられる。また、L2トンネルのほか、L3トンネルなどのトンネルが形成されている状況においても適用は可能である。

[0052] 本発明の上述の例では、ゲートウェイ210を、MN0の通信インフラと接続する部分である第1のサーバ群211と、付加的なデータ処理等を行う部分である第2のサーバ群212とを切り分けることによって、ゲートウェイ210における付加的なデータ処理を追加しても、ゲートウェイ210の同時接続可能機器数の増加を制限するコンピューティングリソースの占有を回避し、ゲートウェイ210にリッチな機能、複雑な機能を実装することが可能となっている。

[0053] また、付加的なデータ処理を追加しなくとも、たとえば第1のサーバ群211では受信したデータのヘッダが指示するIPアドレスの複数のIPアドレスのいずれかへの書き換えのみを実質的に行うようにするなど、第1のサーバ群211におけるデータ処理を限定することで、第1のサーバ群211の限られたコンピューティングリソースを主にそのような制限のない第2のサーバ群212への転送に用い、MN0の通信インフラと接続可能なゲートウェイの

台数制限の中で同時接続可能台数の向上も可能である。一例として、第2のサーバ群212において、GTPヘッダのトンネリング処理を行うようにしてもよい。

[0054] また、たとえば、お正月又は新年の通信制限は、通信を処理するサーバのCPU、メモリ、ストレージ又はネットワークトラフィックの設計容量をどれか一つでも超過してしまうことを主な理由とするものであるが、本実施形態によれば、第2のサーバ群212の処理能力のメトリックスに基づき、負荷の低いサーバのIPアドレスを割り当てるか、第2のサーバ群212全体の容量が足りない場合には、新しくサーバを起動させることにより処理を分散させることで、同時接続可能数及び通信システム全体の容量を増加させることができる。

[0055] 特に、第2のサーバ群212を構成する各サーバを、クラウド上のインスタンスとすることができ、このようにすることで、必要に応じてインスタンスの数を増やしてスケールアウトしていくことができるため、実質的に無尽蔵にコンピューティングリソースを用いることができる。

[0056] ここで、本明細書において「クラウド」とは、ネットワーク上で需要に応じてCPU、メモリ、ストレージ、ネットワーク帯域などのコンピューティングリソースを動的にプロビジョニングし、提供できるシステムを言う。たとえば、AWS等によりクラウドを利用することができる。

[0057] さらに、第1のサーバ群211を構成する各サーバを、クラウド上のインスタンスとすることもでき、このようにすることで、各サーバのIPアドレスを仮想化して、耐障害性を向上させることができる。仮にあるサーバがダウンしても、第1のサーバ群211と第2のサーバ群212との対応関係を、たとえば後述する第3のサーバ群500のデータベース上にリアルタイムに最新の状態で保持しておくことにより、別のサーバにIPアドレスを付け替えることで瞬時に障害復旧が可能である。また、クラウド化することでスケールアップ及びスケールダウンが自在にできるため、処理負荷に応じた最適なCPU、メモリ、ストレージ、ネットワーク帯域を有するインスタンスを選ぶこ

とができる。これにより、ピーク時の最大需要に合わせたハードウェアを常に用意する必要がなくなり、コストの最適化が可能となる。このことは、第2のサーバ群212のクラウド化においても同様である。

[0058] 本実施形態により可能となる無線端末によるアクセス先のIPネットワークなどの例としては、インターネットのほか、プライベートシステムが挙げられる。オンプレミスのプライベートシステム、ゲートウェイ210の少なくとも一部をクラウド上で実現する場合には同一クラウド上の異なるIPネットワーク又は異なるクラウド上のIPネットワークなどがプライベートシステムに含まれ、通信システム200又はゲートウェイ210とこれらのプライベートシステムとは専用線又は仮想的な専用線によって接続可能である。

[0059] (第2のサーバ群の制御)

本発明の一実施形態においては、図5に示すように、通信システム200が第3のサーバ群500をさらに備え、第2のサーバ群212を構成するサーバ数を処理能力のメトリックスに基づいて制御し、スケールアウト・スケールインさせることができる。第3のサーバ群500は、第1のサーバ群211、第2のサーバ群212で利用されるインスタンスの処理能力のメトリックスを監視し、無線端末から接続要求が来た都度、処理負荷の低いインスタンスに接続を割り振ることができる。

[0060] 監視するメトリックスとしては、例として、CPU負荷、メモリ利用率、ディスク読み出し、ディスク書き込み、ネットワークトラフィック受信量、ネットワークトラフィック送信量等が挙げられる。

[0061] また、第1のサーバ群211と第2のサーバ群212との対応関係を、第3のサーバ群500を用いて選択することができる。図6に示すように、第1のサーバ群211が無線端末のIMSI、TEIDなどを含むGTP接続開始要求(GTP-C)を受信した際に、CreateSessionメッセージを第3のサーバ群500に送信し、それを受けた第3のサーバ群500が、第2のサーバ群212のコンピューティングリソースのCPU、メモリ、ストレージ、ネットワークトラフィックなどの利用状況を継続的又は断続的にモニタリングすることで作成さ

れた処理能力のメトリックスに基づいて転送先の第2のサーバ群212を選択しておくことができる。GTP接続開始要求時に、L2トンネルの識別子、無線端末の識別番号、第1のサーバ群211のIPアドレス、第2のサーバ群212のIPアドレスの対応を第3のサーバ群500で記録することにより、第1のサーバ群211や第2のサーバ群212の障害時にL2トンネルのセッションを復旧することが可能になる。

[0062] 図6では、第1のサーバ群211において接続開始要求を受信しており、第1のサーバ群211がCプレーン上の機能も果たしているが、ゲートウェイ210の一部として、又は通信システム200の一部として、Cプレーン上の機能を果たすCプレーン・サーバ群を別途備えることができる。このようにすることで、第1のサーバ群211のコンピューティングリソースをより多くの機器との同時接続に用いることが可能となる。この場合には、第1のサーバ群211は、接続開始要求を受信せず、Cプレーン上の機能を有しない。

[0063] また、第3のサーバ群500から、第2のサーバ群212の中でも利用状況が少なくGTP接続開始時点でのパフォーマンスメトリックスの小さいサーバへプロビジョニングメッセージを送信することで、第2のサーバ群212の中から処理負荷が小さいサーバを選択してGTPの接続先サーバを割り当てることができる。

[0064] そして、図7に示すように、第1のサーバ群211を構成するサーバは、書き換え後の複数のIPアドレスに対応する第2のサーバ群212のいずれかに無線端末からのデータを転送する。図7では、第1のサーバ群211及び第2のサーバ群212がそれぞれ3つのインスタンスを有し、第2のサーバ群212が第1のサーバ群211の各サーバで共用されるように示しているが、このような例に限るものではない。

[0065] 第2のサーバ群212を構成するサーバ数は、第1のサーバ群211を構成するサーバ数を超えることが望ましい。第2のサーバ群212を構成するサーバ数が第1のサーバ群211を構成するサーバ数以下であると、サーバ又はインスタンスの数で決まってしまうセッション数、IPフロー数等のリソ

ースは増えないため、十分なスケーラビリティを発揮できないことがある。

[0066] 第2のサーバ群212のインスタンスグループ全体の処理負荷が設定した水準を超えた場合、新たにインスタンスを起動してサーバを追加することができる。一方、第2のサーバ群212のインスタンスグループ全体の処理負荷が設定した水準を下回った場合、起動中のインスタンスのいずれかを選択し、通信中のL2トンネルなどがあればそれを別のインスタンスに移した後、選択したインスタンスを停止させることができる。たとえば、第2のサーバ群212を構成するインスタンスの数をゼロにすることもできる。

[0067] 第3のサーバ群500は、前記第2のサーバ群212をスケールアウト・スケールインさせるインスタンス数を、予め決められたスケジュールで設定することもできる。たとえば、通信需要がピークを迎えるお昼休み時間の少し前の11:30からインスタンス数を自動的に増加させておき、ピークを過ぎた13:30にインスタンス数を自動的に減少させるということができる。

[0068] なお、上述の例では、第3のサーバ群500をゲートウェイ210の外に設けているが、ゲートウェイ210が第3のサーバ群500を有していてもよく、第2のサーバ群212を構成するインスタンスにより構成することもできる。

[0069] また、第3のサーバ群500は共通のデータベースを持ち、複数のインスタンスで並列に処理をすることができる。

[0070] (付加的なデータ処理)

本発明によりゲートウェイ210で可能となる付加的なデータ処理について具体例を説明する。

[0071] 暗号化

図8を参照して、ペイロードの暗号化のデータ処理を説明する。第1のサーバ群211から、ヘッダが付加されたペイロードを受け取った第2のサーバ群212の対応するサーバは、まず、すべてのヘッダを一旦取り除く。そして、HTTPのペイロード部分(HTTPペイロード)を暗号化した上でHTTPSのヘッ

ッダを新しく付与し、TCP/IPのヘッダもさらに付け直す。

[0072] このようなデータ処理によって、第1のサーバ群211に負荷を加えることなく、無線端末の代わりに、いわば肩代わりしてペイロードの暗号化を行い、通信の安全性を高めることができる。特に、比較的性能の低いIoT機器等の無線端末であっても、通信システム（通信プラットフォーム）側で所要の付加価値を加えることができる。

[0073] なお、図示の例ではTCPのペイロード部分（TCPペイロード）を暗号化しているが、IPのペイロード部分（IPペイロード）から暗号化することもできる。この場合には、GTPペイロードに含まれるIPヘッダを一旦取り除き、IPペイロードを暗号化すればよい。

[0074] クレデンシャル付与

第1のサーバ群211からヘッダが付加されたペイロードを受け取った第2のサーバ群212のうちの対応するサーバは、ヘッダを取り除く過程で、GTPヘッダに含まれる識別子を識別することにより、無線端末の識別番号などの加入者情報を特定することができる。そのため、たとえば第3のサーバ群500のデータベースに記録されている加入者情報に付属するパスワードなどのクレデンシャルを取得することができる。第2のサーバ群212は、無線端末からの特定のHTTPサーバを宛先とした通信を捕捉して、そのHTTPサーバに必要なクレデンシャルで署名を生成した後、HTTPSで送信することができる。ここでいう「クレデンシャル」とは、IDやパスワードをはじめとする、ユーザーの認証に用いられる情報の総称である。

[0075] このようなデータ処理により、IoT機器等、非力で記憶容量がない無線端末でもパスワードなどのクレデンシャルを付与してインターネット上のサーバと通信をおこなうことができる。

[0076] 代理ストレージ

無線端末から送信されたHTTP、MQTT、TELNET、FTP、TCPなどのプロトコルによるデータを第1のサーバ群211が受信する。第1のサーバ群211からヘッダが付加されたペイロードを受け取った第2のサーバ群212のうち

の対応するサーバは、まず、すべてのヘッダを一旦取り除く。そして、クラウド上のコンピューティングストレージに保存するために必要なクレデンシャルやJSON等のデータ記述言語による受け渡しデータを付与した上でHTTPのヘッダを付け直すか、HTTPのペイロード部分を暗号化した上でHTTPSのヘッダを新しく付与し、クラウド上のコンピューティングストレージやデータベースにペイロード内のデータを格納することができる。

[0077] このようなデータ処理により、IoT機器等の非力で記憶容量のない無線端末でも、通信機能を具備するのみで、データの保存が容易に可能になる。

[0078] プロトコル変換

図9を参照して、プロトコル変換のデータ処理の流れを説明する。無線端末からエンドポイントに向けて送信されたHTTP、HTTPS、MQTT、TELNET、FTP、TCPなどのプロトコルによって運ばれるデータを第1のサーバ群211経由で受け取った第2のサーバ群212のインスタンスは、すべてのヘッダを一旦取り除く際に、GTPヘッダに含まれる識別子を識別することにより、送信元の無線端末の加入者情報を判別することができる。ペイロードに対して、インスタンス自身のIPアドレスが外部との通信の受信先になるようにIPヘッダ中の送信元IPアドレスを置き換えて、たとえば第3のサーバ群500のデータベースから加入者情報に対応するパスワードなどの情報を取得して、無線端末の加入者情報に含まれるID、パスワードなどのクレデンシャルを付与した上で、無線端末から送信されてきた時とは別のプロトコル（HTTP、HTTPS、MQTT、TELNET、FTP、TCP、IP等）のヘッダを付け直し、ペイロード部分を暗号化あるいは非暗号化してインターネット上の任意のサーバ（ターゲット）に送信することができる。

[0079] このような付加的なデータ処理によって、IoT機器等、無線端末側の処理能力が非力で具備が困難なプロトコルがあっても、第2のサーバ群212が中継役となってプロトコル変換を実施することにより、インターネット側のサーバと通信を行うことが可能になる。

[0080] スループット制御

無線端末とIPネットワークで送受信されるデータ量を、第2のサーバ群212で計測することにより、単位時間あたりでのデータ通信量を制御することができる。無線端末からの送信データ量、あるいは受信データ量が制限値を超える場合には、第2のサーバ群212でIPパケットを廃棄または一定時間バッファすることで、データの流量を制御することができ、結果としてスループット制御が可能になる。

[0081] このようなデータ処理によって、第1のサーバ群211に負荷を加えることなく、例えば受信能力が限られた多数のIoT機器に対して、適切なスループットにコントロールしてデータを送信することができる。

[0082] 総データ通信量制御

無線端末とIPネットワークで送受信されるデータ量を、第2のサーバ群212で計測し、第3のサーバ群500へ送信し、第3のサーバ群500のデータベースで個別の無線端末の送受信通信量を管理することにより、総データ通信量を制御することができる。無線端末と第2のサーバ群212との対応は、GTPが接続される際に上述したように決定できるため、ある特定のサーバだけでは第2のサーバ群212を通過する総データ通信量を計測することはできない。しかし、第3のサーバ群500にてリアルタイムに集計することにより、特定の無線端末が送受信するデータ量が一定期間内に設定値を超えた場合には、その無線端末への送受信データ通信を止めるか、一定値以下の流量にスループット制御することができる。

[0083] このようなデータ処理によって、第1のサーバ群211に負荷を加えることなく、第3のサーバ群500のデータベースによる個別の無線端末の総データ通信量をリアルタイムで合算することができるため、IoT機器のように数が膨大な無線端末を1つのグループとしてまとめて、総データ通信量を把握することができる。

[0084] レスポンス及び優先度制御

インターネットから無線端末へ送信されるデータに優先度をつけた処理ができる。例えば、第2のサーバ群212の送信待ち行列の内部で、優先度の

高いSIMカードに向けたデータを優先度の高いインスタンスの送信待ち行列に配置し、その他の優先度の低いSIMカードのデータを優先度の低いインスタンスの送信待ち行列に配置し、優先度の高い送信待ち行列のデータから先に処理を行なうことで、優先度に応じた優先送信処理を行なうことでレスポンス（呼応）性能を制御することができる。あるいは、優先度の低いSIMカードへのデータを第2のサーバ群212で廃棄することによって優先度の高いSIMカードへのデータを優先的に処理し、レスポンスを向上させることができる。

[0085] このようなデータ処理によって、緊急通報が必要なIoT機器は優先度を高くして、レスポンスを短くデータ通信をすることが可能になる。

[0086] 時間帯制限

第1のサーバ群211からヘッダが付加されたペイロードを受け取った第2のサーバ群212のうちの対応するサーバは、特定の無線端末からのデータ通信を判別し、時間帯によってデータ通信を止めるか、一定値以下の流量にスループット制御することができる。例えば、深夜帯の時間のみ通信を許可する無線端末の場合、許可された時間以外は第1のサーバ群211で新しいリクエストを受け付けないか、すでに通信路が確立している場合は第2のサーバ群212で通信処理を止めることができる。

[0087] このようなデータ処理により、無線データ通信がピークを迎えたときにはIoT機器からのデータを止めることができ、逆に無線データ通信がオフピークである時間帯を有効に活用することができる。

[0088] 通知とプログラム実行

無線端末から送信されたHTTP、MQTT、TELNET、FTP、TCPなどのプロトコルによるデータを第1のサーバ群211が受信する。第1のサーバ群211から、ヘッダが付加されたペイロードを受け取った第2のサーバ群212の対応するサーバは、まず、すべてのヘッダを一旦取り除く。そして、第3のサーバ群500に保存されている情報に基づき、宛先に対してSMS、Eメールを送信して通知するか、あらかじめ用意してあるプログラムヘッペイロードを入力し、特定のプログラムを実行する。

[0089] このようなデータ処理により、障害等の特定の事象が起こった場合にEメールで通知を行ったり、決められたプログラムを自動的に実行したりする処理が容易に実現できることになる。

[0090] 画像・動画の解像度変換

無線端末のCPU処理負荷を低減するために、ゲートウェイ210から無線端末に伝送される下り時のデータに含まれる画像又は動画の解像度を変更することができる。第2のサーバ群212は、個別の無線端末に対する単位時間当たりのデータ送信量を計測することによってスループットを割り出すことが可能である。第2のサーバ群212はスループットが一定値以下に下がった場合には、送信データの中に含まれている画像・動画データを抽出し、現在利用可能なスループットに適合するコーデックのビットレートに変換したのちに、無線端末へデータを送信することができる。

[0091] このようなデータ処理により、監視カメラなどのリアルタイム性が必要なIoT機器に対して、画像が動画の伝送遅延無く送信することが可能になる。

[0092] シンクライアント

スマートフォン、タブレットなどのモバイル端末用に、HTML5に準拠したアプリが増加しているところ、このようなアプリの実行を、モバイル端末のウェブブラウザ上で行うのではなく、第2のサーバ群212において行って、画面表示のみをモバイル端末側で行うことができる。

[0093] 具体的には、ゲートウェイ210側で、ネットワーク機能（HTTP又はHTTPS）、画像又は動画のデータ圧縮機能（モバイル端末の画面サイズに最適化して縮小等）、画像又は動画のコーデック変換機能、DRMの追加・変換機能、Javascript（登録商標）のインタプリタ（又はコンパイラ）機能、データキャッシュ機能（ユーザーが次に読み込むページを予測する機能付き）、レンダリング機能（ユーザーが次に読み込むページを予測する機能付き）、データストレージ機能等を代理することができる。

[0094] モバイル端末側では、極限としては表示装置としての機能を果たせばよく、必要なアプリの機能は、通信システム（通信プラットフォーム）側で代理

することが可能である。

[0095] (パブリッククラウド)

通信システム200は、クラウド上で実装することができ、特に、パブリッククラウド上で実装することができる。ここで、「パブリッククラウド」とは、複数のテナントが利用可能なクラウドを言う。

[0096] パブリッククラウドでは、統計多重により、単一の事業者が利用するプライベートクラウドと比較すると、複数の事業者が異なったパターンでコンピューティングリソースを利用するため、統計多重効果によりピーク時と平均時の利用率の差が小さくなる。そのため、パブリッククラウド事業者は大規模なコンピューティングリソースを効率的に運用することができる。本件発明においては、パブリッククラウド上で通信事業以外の事業者とコンピューティングリソースを共有して利用することにより、ピーク時に必要なコンピューティングリソースを無制限にパブリッククラウドから確保することができる。

符号の説明

[0097] 200 通信システム（通信プラットフォーム）
210 ゲートウェイ
211 第1のサーバ群
212 第2のサーバ群
500 第3のサーバ群

請求の範囲

- [請求項1] 無線端末にIPネットワークへのアクセスを提供するための通信システムであって、
前記無線端末からのデータを前記IPネットワークに通過させるためのゲートウェイを備え、
前記ゲートウェイは、
前記無線端末からのデータを受信する受信部と、
受信した前記データのヘッダを元に複数の送信先アドレスのいずれかから1つを選択する選択部と、
前記データを前記選択部が選択した送信先アドレスに転送する転送部と
を有する第1のサーバ群を有し、
前記第1のサーバ群のそれぞれは、前記複数の送信先アドレスのいずれかに対応する第2のサーバ群を構成するサーバに前記データを転送することを特徴とする通信システム。
- [請求項2] 前記ゲートウェイは、GGSN又はP-GWであることを特徴とする請求項1に記載の通信システム。
- [請求項3] 前記通信システムは、前記無線端末からの接続開始要求を受信するCプレーン・サーバ群をさらに備え、
前記第1のサーバ群は、前記Cプレーン・サーバ群が受けた接続開始要求を契機に請求項1又は2に記載の処理を行う通信システム。
- [請求項4] 前記第1のサーバ群を構成するサーバ数は、予め定められた数以下であることを特徴とする請求項1から3のいずれかに記載の通信システム。
- [請求項5] 前記予め定められた数は、MNOにより定められていることを特徴とする請求項4に記載の通信システム。
- [請求項6] 前記第2のサーバ群を構成するサーバ数は、前記第1のサーバ群を構成するサーバ数を超えることを特徴とする請求項1から5のいずれ

かに記載の通信システム。

- [請求項7] 前記第1及び第2のサーバ群は、それぞれ、前記ヘッダが指示していた前記送信先アドレスと、前記複数の送信先アドレスとの間の対応テーブルを前記無線端末が確立したトンネル毎に有することを特徴とする請求項1から6のいずれかに記載の通信システム。
- [請求項8] 前記対応テーブルは、前記ヘッダ内のGTPヘッダが有する識別子と関連づけられていることを特徴とする請求項7に記載の通信システム。
- [請求項9] 前記選択部は、前記ヘッダが指示する送信先アドレスを書き換えることで転送先の複数の送信先アドレスのいずれかを特定することを特徴とする請求項1から8のいずれかに記載の通信システム。
- [請求項10] 前記複数の送信先アドレスは、同一の送信先アドレスを割り当てないように決定されることを特徴とする請求項1から9のいずれかに記載の通信システム。
- [請求項11] 前記複数の送信先アドレスのうちの少なくとも一部が同一であることを特徴とする請求項1から9のいずれかに記載の通信システム。
- [請求項12] 前記無線端末と前記第1のサーバ群との間にL2トンネルが通信経路として形成されることを特徴とする請求項1から11のいずれかに記載の通信システム。
- [請求項13] 前記第2のサーバ群を構成する各サーバは、クラウド上のインスタンスであることを特徴とする請求項1から12のいずれかに記載の通信システム。
- [請求項14] 前記第1のサーバ群を構成する各サーバは、クラウド上のインスタンスであることを特徴とする請求項13に記載の通信システム。
- [請求項15] 前記第2のサーバ群が前記第1のサーバ群の各サーバで共用されることを特徴とする請求項14に記載の通信システム。
- [請求項16] 前記第2のサーバ群と接続された第3のサーバ群を備え、
前記第3のサーバ群は、前記第2のサーバ群を構成するサーバ数を

制御することを特徴とする請求項 13 から 15 のいずれかに記載の通信システム。

[請求項17] 前記第3のサーバ群は、前記第2のサーバ群を構成するインスタンスにより構成されることを特徴とする請求項 16 に記載の通信システム。

[請求項18] 前記第2のサーバ群を構成するインスタンス数は、前記第2のサーバ群の処理負荷に応じてゼロまで減少可能であることを特徴とする請求項 13 から 17 のいずれかに記載の通信システム。

[請求項19] 前記ゲートウェイは、マルチテナントであることを特徴とする請求項 13 から 18 のいずれかに記載の通信システム。

[請求項20] 前記第2のサーバ群は、前記データに対してデータ処理を行うことを特徴とする請求項 1 から 19 のいずれかに記載の通信システム。

[請求項21] 前記データ処理は、前記データのペイロードの暗号化、前記データのプロトコル変換、前記無線端末からのデータ送信のスループット制御、及び、前記データのペイロードに対するクレデンシャル付与のうちの少なくとも1つであることを特徴とする請求項 20 に記載の通信システム。

[請求項22] 前記第2のサーバ群は、前記ゲートウェイから前記無線端末に送信されるデータに対して下り時のデータ処理を行うことを特徴とする請求項 1 から 21 のいずれかに記載の通信システム。

[請求項23] 前記下り時のデータ処理は、前記データに含まれる画像又は動画の解像度変換、前記無線端末のデータ受信のスループット制御、前記無線端末のデータ受信の優先度制御、及び、HTML5に準拠した機能の実行のうちの少なくとも1つであることを特徴とする請求項 22 に記載の通信システム。

[請求項24] 無線端末にIPネットワークへのアクセスを提供するための通信方法であって、

前記無線端末からのデータを前記IPネットワークに通過させるため

のゲートウェイにおいて、前記ゲートウェイが有する第1のサーバ群が、前記無線端末からのデータを受信する受信ステップと、

前記第1のサーバ群が、受信した前記データのヘッダを元に複数の送信先アドレスのいずれかから1つを選択する選択ステップと、

前記第1のサーバ群が、前記データを前記選択された送信先アドレスに転送する転送ステップと

を含み、

前記転送ステップは、前記第1のサーバ群のそれぞれが、前記複数の送信先アドレスのいずれかに対応する第2のサーバ群を構成するサーバに前記データを転送することを特徴とする通信方法。

[請求項25]

無線端末からのデータをIPネットワークに通過させるためのゲートウェイに、前記無線端末に前記IPネットワークへのアクセスを提供するための通信方法を実行させるためのプログラムであって、

前記通信方法は、

前記ゲートウェイにおいて、前記ゲートウェイが有する第1のサーバ群が、前記無線端末からのデータを受信する受信ステップと、

前記第1のサーバ群が、受信した前記データのヘッダを元に複数の送信先アドレスのいずれかから1つを選択する選択ステップと、

前記第1のサーバ群が、前記データを前記選択された送信先アドレスに転送する転送ステップと

を含み、

前記転送ステップは、前記第1のサーバ群のそれぞれが、前記複数の送信先アドレスのいずれかに対応する第2のサーバ群を構成するサーバに前記データを転送することを特徴とするプログラム。