

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5608794号
(P5608794)

(45) 発行日 平成26年10月15日(2014.10.15)

(24) 登録日 平成26年9月5日(2014.9.5)

(51) Int.Cl. F 1
G 0 6 F 9/46 (2006.01) G 0 6 F 9/46 3 5 0

請求項の数 16 外国語出願 (全 30 頁)

(21) 出願番号	特願2013-144451 (P2013-144451)	(73) 特許権者	392026693
(22) 出願日	平成25年7月10日 (2013.7.10)		株式会社NTTドコモ
(65) 公開番号	特開2014-21979 (P2014-21979A)		東京都千代田区永田町二丁目11番1号
(43) 公開日	平成26年2月3日 (2014.2.3)	(74) 代理人	100099623
審査請求日	平成25年7月10日 (2013.7.10)		弁理士 奥山 尚一
(31) 優先権主張番号	12176591.1	(74) 代理人	100096769
(32) 優先日	平成24年7月16日 (2012.7.16)		弁理士 有原 幸一
(33) 優先権主張国	欧州特許庁 (EP)	(74) 代理人	100107319
			弁理士 松島 鉄男
		(74) 代理人	100114591
			弁理士 河村 英文
		(74) 代理人	100125380
			弁理士 中村 綾子
		(74) 代理人	100142996
			弁理士 森本 聡二

最終頁に続く

(54) 【発明の名称】 複数の仮想マシンを管理するための階層システム、方法、及びコンピュータプログラム

(57) 【特許請求の範囲】

【請求項1】

少なくとも2つの物理マシン(100a、100b)を有する第1のグループに接続可能であり、少なくとも2つの物理マシンを有する前記第1のグループ内の1つの物理マシンにある第1の仮想マシンの仮想マシン識別情報と、該1つの物理マシンの物理マシン識別情報とを含むデータセット(110a、110b)を記憶する第1のローカルマイグレーションアンカポイント(110)と、

少なくとも2つの物理マシン(120a、120b)を有する第2のグループに接続可能であり、少なくとも2つの物理マシンを有する前記第2のグループ内の1つの物理マシンにある第2の仮想マシンの仮想マシン識別情報と、該1つの物理マシンの物理マシン識別情報とを含むデータセット(130a、130b)を記憶する第2のローカルマイグレーションアンカポイント(130)と、

前記第1のローカルマイグレーションアンカポイントと前記第2のローカルマイグレーションアンカポイントとに接続され、前記第1の仮想マシンによって実行されるアプリケーションに関する第1のサービス識別情報と、前記第1の仮想マシンに関連付けられた識別情報と、前記第1のローカルマイグレーションアンカポイントの識別情報とを第1のデータレコード(140a)として記憶するとともに、前記第2の仮想マシンによって実行されるアプリケーションの第2のサービス識別情報と、前記第2の仮想マシンに関連付けられた識別情報と、前記第2のローカルマイグレーションアンカポイントの識別情報とを第2のデータレコード(140b)として記憶するグローバルマイグレーションアンカポ

イント(140)と、

前記第1のサービス識別情報と、前記第1の仮想マシンの識別情報と、前記第1のローカルマイグレーションアンカポイントの識別情報とを含む、前記第1の仮想マシンに関する第1のデータエントリ(150a)を記憶するとともに、前記第2のサービス識別情報と、前記第2の仮想マシンの識別情報と、前記第2の仮想マシンが位置づけられている物理マシンが接続することのできる前記第2のローカルマイグレーションアンカポイントの識別情報とを含む第2のデータエントリ(150b)を有する仮想マシンロケーションレジスタ(150)と、

中央ネットワーク管理システム(160)と、

物理マシンの各グループのグループマネージャ(101、121)と

10

を備え、

前記中央ネットワーク管理システム(160)は、物理マシンの前記第1のグループから物理マシンの前記第2のグループ内の第1の物理マシンへと前記第1の仮想マシンをマイグレーションさせる旨の決定を受信する(315)か又は該決定を行うものであり、

前記第2のローカルマイグレーションアンカポイントは、前記第1の仮想マシンが物理マシンの前記第2のグループ内の前記第1の物理マシンにあるという情報を、物理マシンの前記第2のグループ内の前記第1の物理マシンから受信する(320)ものであり、

前記第2のローカルマイグレーションアンカポイントが、前記第1の仮想マシンが物理マシンの前記第2のグループにあるというメッセージを前記グローバルマイグレーションアンカポイントに送信する(325)か、前記グローバルマイグレーションアンカポイントが、以前のローカルマイグレーションアンカポイントに関する情報を受信する前記仮想マシンロケーションレジスタにアクセスする(330)か、又は前記第2のローカルマイグレーションアンカポイントが、前記以前のローカルマイグレーションアンカポイントに関する情報を得るために、前記仮想マシンロケーションレジスタに対して、あるメッセージを送信(335)し、

20

前記第1のローカルマイグレーションアンカポイントは、前記第1の仮想マシンを対象としたデータメッセージを、該データメッセージの宛先エントリを前記第2のローカルマイグレーションアンカポイントとすることにより該第2のローカルマイグレーションアンカポイントへと送信する(340)ものである、複数の仮想マシンを管理するための階層システム。

30

【請求項2】

第3のローカルマイグレーションアンカポイントと第4のローカルマイグレーションアンカポイントとに接続された第2のグローバルマイグレーションアンカポイントを更に備え、

前記仮想マシンロケーションレジスタ(150)が更に、各データエントリにおいて、該データエントリにより特定されるローカルマイグレーションアンカポイントが接続しているグローバルマイグレーションアンカポイントのグローバルマイグレーションアンカポイント識別情報を記憶する、請求項1に記載の階層システム。

【請求項3】

前記グローバルマイグレーションアンカポイントは、前記サービス識別情報によって特定されるあるサービスについてのデータメッセージをあるクライアントから受信する(200)ものであり、前記データメッセージは、前記クライアントを特定する送信元エントリと、前記グローバルマイグレーションアンカポイントを特定する宛先エントリとを有するものであり、

40

前記グローバルマイグレーションアンカポイントは、前記サービス識別情報を有する、記憶しているデータレコードに基づいて、前記送信元エントリが前記グローバルマイグレーションアンカポイントを特定するものとなり、かつ前記宛先エントリが前記ローカルマイグレーションアンカポイントと前記仮想マシンとを特定するものとなるように前記データメッセージを処理する(205)ものである、請求項1又は2に記載の階層システム。

【請求項4】

50

前記ローカルマイグレーションアンカポイントは、前記グローバルマイグレーションアンカポイントからデータメッセージを受信し(210)、該データメッセージが示す仮想マシン識別情報を含む、記憶しているデータセットに基づいて、前記データメッセージ内のローカルマイグレーションアンカポイント識別情報を物理マシン識別情報へと置き換える(215)ものである、請求項1～3のいずれか1項に記載の階層システム。

【請求項5】

前記ローカルマイグレーションアンカポイントは、仮想マシン識別情報を送信元として含むとともにグローバルマイグレーションアンカポイントを宛先として含むデータメッセージ(806)を受信し、当該仮想マシンの識別情報を含むデータエントリにおいて特定されるグローバルマイグレーションアンカポイントに向けて前記データメッセージ(807)を転送するものである、請求項1～4のいずれか1項に記載の階層システム。

10

【請求項6】

前記グローバルマイグレーションアンカポイントは、あるローカルマイグレーションアンカポイントからデータメッセージ(807)を受信するものであり、該データメッセージは、送信元として仮想マシン識別情報を有するとともに宛先として当該グローバルマイグレーションアンカポイントを有するものであり、

前記グローバルマイグレーションアンカポイントは、前記仮想マシンの識別情報を含むデータレコードに基づいて、前記データメッセージを処理し、前記宛先としての前記グローバルマイグレーションアンカポイントをクライアント識別情報へ置き換え、前記送信元としての前記仮想マシンを当該グローバルマイグレーションアンカポイントのグローバルマイグレーションアンカポイント識別情報へ置き換えるものである、請求項1～5のいずれか1項に記載の階層システム。

20

【請求項7】

前記中央ネットワーク管理システム(1100)は、前記第1のグループ内の第1の物理マシンから前記第1のグループ内の第2の物理マシンへと前記第1の仮想マシンをマイグレーションさせる要求を受信する(300)ものであり、

前記第1のローカルマイグレーションアンカポイントは、前記決定に応じて前記グループマネージャ(1020、1110)から前記第2の物理マシンの識別情報を受信する(305)ものであり、

前記ローカルマイグレーションアンカポイントは、前記データセット内の前記第1の物理マシンの識別情報を前記第2の物理マシンの識別情報へと置き換える(310)ものである、請求項1に記載の階層システム。

30

【請求項8】

前記第1のローカルマイグレーションアンカポイント及び前記第2のローカルマイグレーションアンカポイントと、前記グローバルマイグレーションアンカポイントと、前記仮想マシンロケーションレジスタとのうちの少なくとも1つが、前記仮想マシンロケーションレジスタか、前記第1のローカルマイグレーションアンカポイント及び前記第2のローカルマイグレーションアンカポイントか、前記グローバルマイグレーションアンカポイントかにおけるデータが有効ではない場合にページング処理を行うものである、請求項1～7のいずれか1項に記載の階層システム。

40

【請求項9】

前記仮想マシンロケーションレジスタと、前記グローバルマイグレーションアンカポイントと、前記第1のローカルマイグレーションアンカポイント及び前記第2のローカルマイグレーションアンカポイントとのうちの少なくとも1つが、仮想マシンが過去においてのみ登録されていた全てのローカルマイグレーションアンカポイントにページング処理を行うよう依頼するか、又は

前記仮想マシンロケーションレジスタが、前記グローバルマイグレーションアンカポイントにページング処理を行うよう依頼し、該グローバルマイグレーションアンカポイントが、当該グローバルマイグレーションアンカポイントに接続されているローカルマイグレーションアンカポイントにページング処理を行うよう依頼する、請求項8に記載の階層シ

50

ステム。

【請求項 1 0】

前記ローカルマイグレーションアンカポイントが、ある仮想マシンを特定したロケーション登録又は更新要求を、接続可能な物理マシンのグループ内の全ての物理マシンへ送信し(350)、

前記ローカルマイグレーションアンカポイントは、前記ある仮想マシンを有している物理マシンから返答を受信し(355)、

前記ローカルマイグレーションアンカポイントは、前記仮想マシンロケーションレジスタ、あるいはこれに加えて前記グローバルマイグレーションアンカポイントに対し、前記ある仮想マシンが存在している物理マシンを通知する(360)ものである、請求項1~9のいずれか1項に記載の階層システム。

10

【請求項 1 1】

各物理マシンが、前記第1のローカルマイグレーションアンカポイントと通信するように構成されているマイグレーション管理制御機能と仮想マシン機能とを有するものである、物理マシンの前記第1のグループを更に備えている請求項1~10のいずれか1項に記載の階層システム。

【請求項 1 2】

前記第1のローカルマイグレーションアンカポイントと前記第2のローカルマイグレーションアンカポイントとがそれぞれ満了期間を示すタイマ(110c、130c)を有し、

20

前記物理マシン又は前記仮想マシンは、前記満了期間内にロケーション登録メッセージを前記ローカルマイグレーションアンカポイントに送信し、対応するデータセットが更なる期間にわたり延長されるか、又は、

前記ロケーション登録メッセージを前記満了期間内に受信しなかった場合には、前記第1のローカルマイグレーションアンカポイントと前記第2のローカルマイグレーションアンカポイントとが、ある仮想マシンを特定した前記対応するデータセットを削除する、請求項1~11のいずれか1項に記載の階層システム。

【請求項 1 3】

前記グローバルマイグレーションアンカポイントが、あるクライアントから受信したサービス識別情報を仮想マシンのIPアドレスへと変換する、請求項1~12のいずれか1項に記載の階層システム。

30

【請求項 1 4】

少なくとも2つのグローバルマイグレーションアンカポイントを備え、

第1のグローバルマイグレーションアンカポイントが、サービス識別情報を有するクライアント要求を受信し(370)、

前記第1のグローバルマイグレーションアンカポイントは、前記サービス識別情報に関してデータレコードを調べ(375)、

前記第1のグローバルマイグレーションアンカポイントが記憶しているデータレコードにおいて前記サービス識別情報が見つからない場合には、前記第1のグローバルマイグレーションアンカポイントは、前記サービス識別情報と関連付けられているグローバルマイグレーションアンカポイントの識別情報を前記仮想マシンロケーションレジスタに要求し(380)、

40

前記グローバルマイグレーションアンカポイントは、前記サービス識別情報をデータレコード内に有する第2のグローバルマイグレーションアンカポイントの識別情報を受信し(385)、

前記第1のグローバルマイグレーションアンカポイントは、前記仮想マシンロケーションレジスタによって特定された前記第2のグローバルマイグレーションアンカポイントに対し、前記サービス識別情報によって特定されたサービスを要求しているクライアントを通知する(390)、請求項1~13のいずれか1項に記載の階層システム。

【請求項 1 5】

50

少なくとも2つの物理マシンを有する第1のグループ内の1つの物理マシンにある第1の仮想マシンの仮想マシン識別情報と、該1つの物理マシンの物理マシン識別情報とを含むデータセットを記憶する第1のローカルマイグレーションアンカポイント(110)を、少なくとも2つの物理マシンを有する前記第1のグループに接続するステップと、

少なくとも2つの物理マシンを有する第2のグループ内の1つの物理マシンにある第2の仮想マシンの仮想マシン識別情報と、該1つの物理マシンの物理マシン識別情報とを含むデータセットを記憶する第2のローカルマイグレーションアンカポイント(130)を、少なくとも2つの物理マシンを有する前記第2のグループに接続するステップと、

前記第1の仮想マシンによって実行されるアプリケーションに関する第1のサービス識別情報と、前記第1の仮想マシンに関連付けられた識別情報と、前記第1のローカルマイグレーションアンカポイントの識別情報とを第1のデータレコードとして記憶するとともに、前記第2の仮想マシンによって実行されるアプリケーションの第2のサービス識別情報と、前記第2の仮想マシンに関連付けられた識別情報と、前記第2のローカルマイグレーションアンカポイントの識別情報とを第2のデータレコードとして記憶するグローバルマイグレーションアンカポイント(140)を、前記第1のローカルマイグレーションアンカポイントと前記第2のローカルマイグレーションアンカポイントとに接続するステップと、

仮想マシンロケーションレジスタ(150)が、前記第1のサービス識別情報と、前記第1の仮想マシンの識別情報と、前記第1のローカルマイグレーションアンカポイントの識別情報とを含む、前記第1の仮想マシンに関する第1のデータエントリを記憶するとともに、前記第2のサービス識別情報と、前記第2の仮想マシンの識別情報と、前記第2の仮想マシンが置かれている物理マシンが接続することのできる前記第2のローカルマイグレーションアンカポイントの識別情報とを含む第2のデータエントリを有するステップと、

中央ネットワーク管理システム(160)が、物理マシンの前記第1のグループから物理マシンの前記第2のグループ内の第1の物理マシンへと前記第1の仮想マシンをマイグレーションさせる旨の決定を受信する(315)か又は該決定を行うステップと、

前記第2のローカルマイグレーションアンカポイントが、前記第1の仮想マシンが物理マシンの前記第2のグループ内の前記第1の物理マシンに置かれているという情報を、物理マシンの前記第2のグループ内の前記第1の物理マシンから受信する(320)ステップと、

前記第2のローカルマイグレーションアンカポイントが、前記第1の仮想マシンが物理マシンの前記第2のグループにあるというメッセージを前記グローバルマイグレーションアンカポイントに送信する(325)か、前記グローバルマイグレーションアンカポイントが、以前のローカルマイグレーションアンカポイントに関する情報を受信する前記仮想マシンロケーションレジスタにアクセスする(330)か、又は前記第2のローカルマイグレーションアンカポイントが、前記以前のローカルマイグレーションアンカポイントに関する情報を得るために、前記仮想マシンロケーションレジスタに対して、あるメッセージを送信する(335)ステップと、

前記第1のローカルマイグレーションアンカポイントが、前記第1の仮想マシンを対象としたデータメッセージを、該データメッセージの宛先エントリを前記第2のローカルマイグレーションアンカポイントとすることにより該第2のローカルマイグレーションアンカポイントへと送信する(340)ステップと

を含む、複数の仮想マシンを管理する方法。

【請求項16】

請求項15に記載の複数の仮想マシンを管理する方法をコンピュータに実行させるプログラムコードを含むコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

10

20

30

40

50

本発明はコンピュータシステムに関し、特に種々の物理マシン (physical machine) にある仮想マシン (virtual machine) の管理に関する。

【背景技術】

【0002】

仮想化、仮想マシン、マイグレーション管理、クラウドコンピューティングは、ますます重要となってきた手段である。仮想マシンの管理は、特に、クラウドサービス、ネットワークベースのマイグレーション管理、災害管理、又はエネルギー節約の目的で有用であるとともに適用可能である。

【0003】

基本的に、仮想マシンコンピューティングは、種々のマシンすなわち物理マシンにおいて特定のサービスを実行することを可能にする。物理マシンは、あるロケーションに置かれているコンピュータである。仮想マシンは、あるサービスを実行するためにインプリメントされるものであるが、ある物理マシンから別の物理マシンへと移行又は移動 (マイグレーション) することができるように設計される。特にこのことは、仮想マシンをインプリメントするためある物理マシンが提供する計算リソースをその仮想マシンが第1の期間において用いることができ、ある物理マシンから別の物理マシンへのマイグレーションの後には、それまでの物理マシンが提供する計算リソースを他のサービス用に自由に利用でき、仮想マシンは、新たな物理マシンの計算リソースを、新たなサービスの実行又は現在動作中のプロセスの継続のために利用するということを意味する。

10

【0004】

ある物理マシンから別の物理マシンへの仮想マシンのマイグレーションは、セッション継続の観点で問題であり、仮想マシンのロケーションに関するネットワーク全体の更新に関しても問題である。特に、「クラウド」とも呼ばれる物理マシンのグループが幾つかあって別々に制御されている場合には、あるクラウドから別のクラウドへの仮想マシンのマイグレーションも難しい処理である。

20

【0005】

レイヤ2仮想プライベートネットワーク (layer 2 virtual private network, L2VPN) のワーキンググループが存在する。このワーキンググループは、プロバイダから提供されるレイヤ2仮想プライベートネットワークをサポートするための限られた数の解決策を規定し、特定することを担当する。クラウド内のマイグレーションの管理の場合、L2VPNが主に用いられる解決策である。L2VPNにおいては、レイヤ2スイッチは、仮想マシンがどのポートを通じて到達可能であるのかを記憶している。仮想マシンがある物理マシンから別の物理マシンに移動すると、その仮想マシン用のポートが変化する。しかし、現在のL2スイッチは学習機能を有しており、あるポートを通じて到着したパケットのMACアドレスを調べることができる。仮想マシンのMACアドレスはマイグレーションするまで変わらないため、L2スイッチは、仮想マシンから別のポートを通じて到着したパケットをスヌーピングする (snoop) ことにより、その仮想マシンを特定することができる。特に、L2スイッチは、仮想マシンをそのMACアドレスによって特定するとともに、その仮想マシンがどのポートを通じて到達可能であるのかを特定する。しかし、現在のクラウドの非常に大規模な展開を考慮すると、L2VPNは、拡張性 (scalability) という点でみると全く拡張させることができない。L2VPNは手動で設定され、VLANタグは12バイトの長さしかないゆえVLANを4096個しか作成できないからである。加えて、この解決策は、クラウド間マイグレーションのシナリオに適用することもできない。

30

40

【0006】

この研究分野における別の主な解決策は、オープンフロー (Open Flow) に基づいた解決策である。クラウド内のシナリオの場合、この解決策はL2VPNと同じである。特に、マイグレーションするまでフローをある仮想マシンへと再ルーティングするのは、オープンフローコントローラ (Open Flow controller) である。仮想マシンのマイグレーションは、オープンフローコントローラによって監視することができる。マイグレーションの

50

後、オープンフローコントローラは、オープンフロースイッチの転送用テーブルを書き換えて、そのスイッチが適切なポートを通じてパケットを転送することができるようにする。しかし、この解決策も、クラウド間マイグレーションのシナリオには適用できない。

【0007】

特許文献1は、サーバ間の仮想マシンマイグレーションを開示している。仮想マシンのマイグレーションは、2つのサーバ間で行われる。第1のサーバでは、仮想マシンに関係した全てのファイルが記憶されているボリュームがディスマウントされる。第2のサーバでは、仮想マシンに関係した全てのファイルが記憶されている該ボリュームがマウントされ、第2のサーバは、仮想マシンをホストすることができるようになる。このようにして、第1のサーバから第2のサーバに全てのファイルをコピーする必要はなく、仮想マシンのマイグレーションを行うことができる。仮想マシンに関係したファイルは、ストレージエリアネットワーク(storage-area network, SAN)に記憶されている。しかし、この解決策によりクラウド間マイグレーションをサポートするのは、ある1つのクラウドのSANに対し別のクラウドがアクセスできることが想起されるため現実的ではない。もしこれがインプリメントされたとしても、仮想マシンの新たなロケーションへのルートを変更することに対処しなければならない。

10

【0008】

特許文献2は、各データセンタとWANとの協調により、仮想マシンのマイグレーションが、その処理が中断することなく、効果的な電力節約の実施、負荷分散、又は障害対応処理が可能となるように行われることを開示している。WANと別のネットワークとの境界ポイントに置かれている各ノードには、仮想マシンのマイグレーションに起因したアドレス重複を回避するために動的に設定可能なネットワークアドレス変換(network address translation, NAT)機能が設けられる。あるいは、WAN内の各ノードにネットワーク仮想化機能を設け、マイグレーションの前にある仮想マシンを含めるためにデータセンタに接続された仮想ネットワークと、マイグレーションの後にその仮想マシンを含めるためにデータセンタに接続された仮想ネットワークとがインプリメントされ、それによって、同一アドレスの共存した提供(coexistent provision)が可能となる。したがって、仮想マシンのマイグレーション時にネットワークルーティング情報を変更する必要性をなくすことができ、マイグレーションのための設定変更を迅速に行うことができる。

20

【先行技術文献】

30

【特許文献】

【0009】

【特許文献1】米国特許第8,042,108号公報

【特許文献2】米国特許出願公開第2011/0161491号公報

【発明の概要】

【発明が解決しようとする課題】

【0010】

本発明の目的は、複数の仮想マシンを管理するための改善された階層のコンセプトを提供することにある。

【課題を解決するための手段】

40

【0011】

上記目的は、請求項1に記載の複数の仮想マシンを管理するための階層システムと、請求項15に記載の複数の仮想マシンを管理する方法と、請求項16に記載のコンピュータプログラムとのいずれかによって達成される。

【0012】

本発明は、複数の仮想マシンを管理するための階層システムに関する。この階層システムは、少なくとも2つの物理マシンを有する第1のグループに接続可能であり、少なくとも2つの物理マシンを有する前記第1のグループ内の1つの物理マシンにある第1の仮想マシンの仮想マシン識別情報と、該1つの物理マシンの物理マシン識別情報とを含むデータセットを記憶する第1のローカルマイグレーションアンカポイント又はローカルモビリ

50

ティアンカポイントと、少なくとも2つの物理マシンを有する第2のグループに接続可能であり、少なくとも2つの物理マシンを有する前記第2のグループ内の1つの物理マシンにある第2の仮想マシンの仮想マシン識別情報と、該1つの物理マシンの物理マシン識別情報とを含むデータセットを記憶する第2のローカルマイグレーションアンカポイント又はローカルモビリティアンカポイントと、前記第1のローカルマイグレーションアンカポイントと前記第2のローカルマイグレーションアンカポイントとに接続され、前記第1の仮想マシンによって実行されるアプリケーションに関する第1のサービス識別情報と、前記第1の仮想マシンに関連付けられた識別情報と、前記第1のローカルマイグレーションアンカポイントの識別情報とを第1のデータレコードとして記憶するとともに、前記第2の仮想マシンによって実行されるアプリケーションのサービス識別情報と、前記第2の仮想マシンに関連付けられた識別情報と、前記第2のローカルマイグレーションアンカポイントの識別情報とを第2のデータレコードとして記憶するグローバルマイグレーションアンカポイント又はグローバルモビリティアンカポイントと、前記第1のサービス識別情報と、前記第1の仮想マシンの識別情報と、前記第1のローカルマイグレーションアンカポイントの識別情報とを含む、前記第1の仮想マシンに関する第1のデータエントリを記憶するとともに、前記第2のサービス識別情報と、前記第2の仮想マシンの識別情報と、前記第2の仮想マシンが置かれている物理マシンが接続することのできる前記第2のローカルマイグレーションアンカポイントの識別情報とを含む第2のデータエントリを有する仮想マシンロケーションレジスタとを備えている。

10

【0013】

20

本発明の別の態様は、これに対応する方法及び対応するコンピュータプログラムに係している。

【0014】

本発明は、セッション継続の観点から及び仮想マシンのロケーションに関してネットワーク全体を更新するという問題からも、ある物理マシンから別の物理マシンへの仮想マシンのマイグレーションを実行する問題に対処するものである。特に、本発明は、仮想マシンが物理マシンのあるグループすなわちクラウドから物理マシンの別のグループすなわちクラウドへとマイグレーションする状況にも有用である。

【0015】

本発明の実施の形態は、マイグレーション管理のための3層(3-tier)アーキテクチャに関する。1つのクラウドは、1つのローカルマイグレーションアンカポイント(local migration anchor point, LP)によって管理され、複数のLPは、グローバルマイグレーションアンカポイント(global migration anchor point, GP)によって管理される。さらに、仮想マシンロケーションレジストラ(virtual machine location registrar, VM LR)があり、これは、仮想マシンのロケーションを示すデータベースを維持するものである。すなわち、仮想マシンは、LP及びGPを通じて到達可能である。特に、仮想マシンロケーションレジスタすなわち仮想マシンロケーションレジストラは、データベース内にデータエントリを有している。マイグレーションの際又はその後、仮想マシンのロケーション情報は、関連するLP、GP、及びVM LRへのシグナリングを通じて更新されるため、仮想マシンのロケーション情報は常に利用可能である。好ましい実施の形態は、正確なデータパスのセットアップ及び正確な変更手順に関する。

30

40

【0016】

本発明の実施の形態は、システムが技術に依存しないという利点を有する。本発明は、例えば、オープンフロー(Open Flow)において用いられるような特定のルーティング、転送方法を前提としていない。さらに、本発明は、ある特定の実施の形態に関して管理が容易である。なぜならば、少数(20個未満等)のグローバルマイグレーションアンカポイント(GP)しか必要としないか、又は必要でありかつ更新する必要があるGPは1つでもよい。このシステムは、クラウド内及びクラウド間のマイグレーションの管理を同時にサポートすることができるため、2つの別々のマイグレーション管理方式は必ずしも必要ではない。

50

【 0 0 1 7 】

さらに、実施の形態は、セルラネットワークに対して親和性を有する。なぜならば、アーキテクチャ及びマイグレーション（マイグレーション管理）の手順が、高いレベルにおいてではあるが、セルラネットワークの手法に類似しているからである。したがって、セルラネットワークの手法を実施する際に得られた経験は、複数の仮想マシンを管理するための階層システムを実施する際にも使用及び適用することができる。本発明によって、自然災害の前、その間、又はその後のネットワークの再構成が可能となる。仮想マシンをより安全なロケーションにマイグレーションすることができ、これによって、サービス継続性が確保されるため、顧客の満足が確保される。仮想マシンをある特定のロケーションにマイグレーションし、残りのもの、すなわち不要なリソースをシャットダウンする等のネットワーク再構成が、例えば夜間において容易に行うことができる。これによって、エネルギー消費も削減され、グリーンネットワーク化が実現されることになる。以下の説明において、物理マシンの1つのグループをクラウドとも呼び、このクラウドは、ウェブサーバ、ビデオサーバ等の、仮想マシンベースのアプリケーションサービスを提供する単一の管理エンティティとして表現されるように編成された複数の物理マシンとして見るともできる。

10

【 0 0 1 8 】

本発明とは対照的に、特許文献2の構想は集中型方式である。本発明は分散型方式である。実施の形態では、仮想マシンは自らを、関連したエンティティ、例えば、ローカルモビリティアンカポイント、グローバルモビリティアンカポイントに登録する。ルートをV

20

【 0 0 1 9 】

本発明の方式による中央ネットワーク管理システムは、マイグレーションそのものを管理せず、ルートをVMの新たなロケーションに変更することもしない。中央ネットワーク管理システムは、クラウド又はVMに、リソースが利用可能な別のクラウドにマイグレーションするように通知するだけである。それ以外のものは、本発明の実施の形態では自律的に行われる。

【 0 0 2 0 】

上記従来技術の文献とは異なり、実施の形態は、WAN内の各ノードを仮想化しない。その仮想化には非常に多くのコストを要する。実施の形態によれば、限られた数のノード

30

【 0 0 2 1 】

さらに、LAN又はサブネットのルーティング情報をWAN内に広めることはおよそあり得ず、拡張性のあるシナリオでないと言うべきである。この情報をどこまで広めなければならないかという問題が依然として残る。WANには数百ものルータあるいはスイッチがある。したがって、本発明の実施の形態では、数個のアンカポイントだけが定められる。

【 0 0 2 2 】

実施の形態によれば、パフファリングを行わない。音声電話のようなリアルタイムのアプリケーションの場合は、パフファリングは何ら利点をもたらさない。

40

【 0 0 2 3 】

さらに、従来技術の文献によれば、VMのマイグレーションは、マネージャにより中央で制御される。これは、拡張性を欠いている。これは、VMのマイグレーションの数が多くなり、例えば数千になると、拡張できなくなる。これとは対照的に、実施の形態は、自己管理されるとともに分散されたVMのマイグレーションを有する。

【 0 0 2 4 】

従来技術では、切換え（changeover）の命令が、VMのロケーションの変更をノードに知らせる。これも、同様に集中型の方法である。マイグレーションの数に応じて、同じ数のノードに知らせなければならない。これも同様に拡張性の問題をもたらす。

50

【 0 0 2 5 】

さらに、影響を受けるノードは、送信元クラウド及び宛先クラウドの数に等しい。これは、拡張性の欠如を引き起こす。クラウドの数が増加するにつれて、影響を受けるノードの数も増加する。その一方で、本発明の実施の形態では、ローカルモビリティアンカポイントの数が、クラウド数に1つのグローバルモビリティアンカポイントを加えた数に等しいことが好ましい。その数は、上記従来技術の文献において必要とされる数の半分である。

【 0 0 2 6 】

実施の形態では、VMの以前のロケーションに対しVMの新たなロケーションが通知される結果、パケットを新たなロケーションに転送することができる。さらに、以前のロケーションに向かうパケットを新たなロケーションに転送することができるようなカプセル化方式が好ましい。カプセル化とは、ネットワークアドレス変換(network address translation, NAT)を実行することではない。

10

【 0 0 2 7 】

全体的には、各セッションについて、上記従来技術の文献におけるネットワークアドレス変換の回数は2(クライアント側で1つとVM側で1つ)である。他方、本発明の実施の形態では、ネットワークアドレス変換は、グローバルモビリティアンカポイントでしか実行されない。宛先アドレス(すなわちVMのアドレス)は置き換えられない。その代わりに、アドレスは、VMに到達するまで、ローカルモビリティアンカポイント等によりカプセル化される。

20

【 0 0 2 8 】

以下、添付図面を参照して本発明の好ましい実施形態を説明する。

【 図面の簡単な説明 】

【 0 0 2 9 】

【 図 1 】 複数の仮想マシンを管理するための階層システムの一実施形態のブロック図である。

【 図 2 a 】 グローバルマイグレーションアンカポイントによって実行される処理のフローチャートである。

【 図 2 b 】 ローカルマイグレーションアンカポイントによって実行される処理のフローチャートである。

30

【 図 3 a 】 内部マイグレーション用に実行されるプロセスを示すフローチャートである。

【 図 3 b 】 クラウド間マイグレーションにおいて行われる処理のフローチャートである。

【 図 3 c 】 ページングプロセスの際に行われる処理を示す説明図である。

【 図 3 d 】 複数のグローバルマイグレーションアンカポイントが存在する場合に行われるプロセスを示す説明図である。

【 図 4 】 本発明の利用のシナリオのための対象となる構成を示す説明図である。

【 図 5 】 セルラネットワークのマイグレーション管理アーキテクチャと比較した本発明のシステム及び方法の概要を示す説明図である。

【 図 6 】 詳細な初期化手順を示す説明図である。

【 図 7 】 サービス発見及びセッション確立の詳細な手順を示す説明図である。

40

【 図 8 】 セッション確立後のデータベースを示す説明図である。

【 図 9 a 】 開始モードにおけるマイグレーションサポート及びハンドオーバー手順を示す説明図である。

【 図 9 b 】 クラウド内マイグレーション用のマイグレーションサポート及びハンドオーバー手順を示す説明図である。

【 図 9 c 】 クラウド間マイグレーション用のマイグレーションサポート及びハンドオーバー手順を示す説明図である。

【 図 9 d 】 クラウド間マイグレーションの最終状態を示す説明図である。

【 図 1 0 】 ロケーション更新処理のフローチャートである。

【 図 1 1 】 ネットワーク構成プラットフォームを有する高レベルの説明図である。

50

【図12】ロケーション登録及び更新処理を示す説明図である。

【発明を実施するための形態】

【0030】

実施形態を詳細に説明する前に、仮想マシン技術に関係する幾つかの基本的事項を説明する。1つの手順は、仮想マシンのインスタンス作成 (instantiation) である。まず、ハイパーバイザ (hypervisor) へのログインが行われ、次に発行コマンドが与えられる。この発行コマンドは、ある仮想マシンのインスタンスが作成されるべきということを意味し、その仮想マシンには、ある識別情報 (ID) が与えられる。さらに、128Mbps等のあるメモリが定められる。さらに、例えば1つ以上のコアを有するCPUが定められ、w.x.y.zといったIPアドレスが与えられる。このデータは、本例では、あるハードウェア又は物理マシン上に仮想マシンのインスタンスを作成するため、すなわちインプリメントするために必要となる。仮想マシンの特定の実施態様は、本発明の範囲外である。幾つかの実施態様の例は、XEN、VMWare、KVM等である。

10

【0031】

仮想マシンのマイグレーションにおいては、このインプリメントされた仮想マシンを第1の物理サーバすなわち物理マシンAから第2の物理サーバすなわち物理マシンBへとマイグレーションする必要がある。物理サーバA上で事前にインスタンスが作成されている仮想マシンは、その仮想マシン用に定められたリソースを用いてあるセッションを実行する。通常、仮想マシンのマイグレーションは、第2の物理サーバB上で同じ仮想マシンのインスタンスを作成し、物理サーバAから物理サーバBへとメモリのコピーを開始することによって行われる。

20

【0032】

次に、仮想マシンが、実際に物理サーバAから移動して物理サーバBに置か (place) れ、続いて、セッションが物理サーバB上で実行され、その仮想マシンにより用いられていた物理サーバA上のリソースが、このとき開放される。しかし、これは、1つのクラウド内などの1つの管理ドメイン内でしか実行できない。

【0033】

続いて、図4について説明する。図4は、例えば日本のコア伝送ネットワークといったコア伝送ネットワーク400を示している。さらに、インターネットが、1つのクラウド402として示され、日本の都市である大阪及び仙台用の個々のノードクラウド404、406も同様に示されている。

30

【0034】

さらに、日本の首都である東京用の2つのサービスクラウドが、符号408及び410として示され、日本の首都用の3つのノードクラウドが、符号412、414、416として示されている。さらに、エリアA及びエリアBといった2つのエリアが、符号418及び420として示されている。基本的に、本発明の構想は、固定電話が可動性を持つ (mobile) ようになれば、固定サーバとすることができるという点に基づいている。このような手順の使用事例は災害管理である。この目的のために、例えば、サービスクラウド東京408上に置かれているアプリケーションを、サービスクラウド大阪410へとマイグレーションすることができる。別の使用事例はメンテナンスである。この目的のために、例えば、1つのアプリケーションを符号412に示したノードクラウド東京1からノードクラウド東京3へとマイグレーションすることができる。他の手順は、例えば、ノードクラウド東京2 (符号414) から符号416へとアプリケーションを移動させることとすることができる。さらに別の使用事例は、エネルギー節約である。特に災害管理のためには、マイグレーション時間が1分未満であることが求められる。

40

【0035】

地理的に分散したクラウドシステムでは、クラウド内マイグレーション (マイクロマイグレーション) の管理及びクラウド間マイグレーション (マクロマイグレーション) の管理が有用となる。課題は、仮想化技術の普及により、仮想マシンがもはやいかなる物理的ロケーションにも結び付けられていないということである。仮想マシンを完全に可動性の

50

あるものとするためには、これらの課題は特に、シームレスなセッションマイグレーションと、マイグレーション後の仮想マシンの発見と、ルート最適化、すなわちコア伝送ネットワークを通じてあるクラウドに向かい、次いで、ある仮想マシン又は(その仮想マシンが動作している)物理マシンへと向かう通信ルートの最適化とに関係する。

【0036】

本発明の基本的構想を図5に具体的に示している。本発明の構想に基づく構造を図5の右側に示している。同図では、物理マシンの第1のグループ100が、ローカルマイグレーションアンカポイント110へと接続され、物理マシンの第2のグループ120が、第2のマイグレーションアンカポイント130へと接続されている。さらに、ローカルマイグレーションアンカポイント110、130の双方は、グローバルマイグレーションアンカポイント140に接続されている一方で、さらに仮想マシンロケーションレジスタ150とも通信できるように接続されている。グローバルマイグレーションアンカポイント140はさらに、仮想マシンロケーションレジスタ(virtual machine location register, VMLR)150への通信接続を有している。続いて、図1を詳細に説明する。図1は、複数の仮想マシンを管理するための階層システムを示している。

【0037】

本システムは、少なくとも2つの別々の物理マシン100a、100b、100cを有する第1のグループに接続可能な第1のローカルマイグレーションアンカポイント110を有している。ローカルマイグレーションアンカポイント110は、別々のデータセット110a、110bを記憶するものである。各データセットは、物理マシン100bすなわちPM2に置かれているような、少なくとも2つの物理マシンを有する第1のグループ内のある物理マシンに置かれた(located)VM1等の第1の仮想マシンの仮想マシン識別情報と、その物理マシンすなわちPM2の物理マシン識別情報とを含んでいる。これと並行して、符号120a、120b、120cといった少なくとも2つの物理マシンを有する第2のグループに接続可能な第2のローカルマイグレーションアンカポイント130は、対応するデータセット130a、130bを記憶するものである。各データセット130a、130bは同様に、少なくとも2つの物理マシンを有する第2のグループ内のある物理マシンに置かれた仮想マシンの仮想マシン識別情報と、これに対応する上記物理マシンの物理マシン識別情報とを含んでいる。具体的には、仮想マシンnが、物理マシン識別情報PM4を有する物理マシン120cに置かれている場合には、データセットは、仮想マシンnが置かれている物理マシンのIDであるPM4と関連付けられたVM IDであるVMnを含むことになる。一例として、別の仮想マシンVM(n+1)が、物理マシンIDとしてPM5を有する物理マシン120bに置かれると、第2のデータセット130bは、仮想マシンVM(n+1)というIDと、それに関連する物理マシンPM5というIDとを互いに関連付けて有することになる。当然のことながら、物理マシンは、さらに多くの仮想マシンを追加的にホストすることができる。この場合、各仮想マシンは、特定の物理マシンに置かれた各仮想マシンに対して同一の物理マシンIDを含んだデータセットを有する。

【0038】

さらに、GP1として示されているグローバルマイグレーションアンカポイント140は、第1の接続ライン141aを通じて第1のローカルマイグレーションアンカポイントLP1に接続され、別の接続ライン141bを通じて第2のローカルマイグレーションアンカポイントLP2に接続されている。

【0039】

グローバルマイグレーションアンカポイントGP1は、第1の仮想マシンによって実行されるアプリケーションに関する第1のサービス識別情報をデータレコードに記憶する。この第1のサービス識別情報は、データレコード140aにおいてはID1として示しており、第2のデータレコード140bにおいてはID2として示している。さらに、データレコード140aは、第1の仮想マシンVM1に関連付けられた識別情報と、第1のローカルマイグレーションアンカポイントLP1の識別情報とを含んでいる。さらに、第2

のデータレコード140bは、物理IDとしてPM4を有する物理マシン120cにおいてVMn等の第2の仮想マシンによって実行されるアプリケーションのサービス識別情報ID2を有している。その一方で、本発明は階層的な2層構造を有するため、グローバルマイグレーションアンカポイントのデータレコードにおいて物理マシンIDは必要とされない。

【0040】

仮想マシンロケーションレジスタは、ハッチング線151a及び151bによって示しているようにローカルマイグレーションアンカポイントに接続することができるが、必ずしもそうとも限らない。その一方で、VMLR150は、接続ライン151cを通じてグローバルマイグレーションアンカポイントに接続されるとともに、接続ライン151dを

10

【0041】

VMLRは、当該VMLRに接続されたグローバルマイグレーションアンカポイントと関連付けられたいずれかの物理マシンにおいて動作する各仮想マシン用のデータエントリを有している。したがって、1つのVMLRが、複数の異なるクラウドを有するネットワーク全体に用いられ、このVMLRは、これらのクラウドのうちのいずれかにおいて動作するありとあらゆる仮想マシン用のデータエントリを有する。さらに、このVMLRは、ID1、ID2といったサービスの識別情報と、仮想マシンの識別情報と、その仮想マシンを有する物理マシンが接続されているローカルマイグレーションアンカポイントの識別情報とを有し、加えて、このVMLRは、各IDにつき、対応するグローバルマイグレーションアンカポイントを有する。仮想マシンVM1、VMnの双方がGP1に接続されているため、双方のデータエントリは、同じGP1エントリを有している。単一のグローバルマイグレーションアンカポイントしか用いない場合には、VMLR内のGPのエントリは不要となる。

20

【0042】

この階層システムは更に、中央ネットワーク管理システム160と、第1のグループ100用のグループマネージャ101と、物理マシンの第2のグループ用の別のグループマネージャ121とを有している。

【0043】

さらに、後で説明するように、各ローカルマイグレーションアンカポイントは、LP1用の符号110c及びLP2用の符号130cとして示されている、満了期間を示すタイマを備えていることが好ましい。特に、図1に示した各デバイスは、必要に応じて、あるメッセージを別の通信相手に送信し、及び/又は別の通信相手からメッセージを受信して解釈し処理をする。

30

【0044】

そして、図2aに示すように、グローバルマイグレーションアンカポイント140は、ブロック200に示すように、サービス識別情報(ID)によって特定されるサービスを求めているクライアントからデータメッセージを受信する。ブロック200の右に示したデータメッセージは、送信元エントリ201と、宛先エントリ202と、ペイロードエントリ203とを有している。送信元エントリは、あるサービスによるサービス提供を受けようとするクライアントを示し、宛先エントリは、このデータメッセージを受信するグローバルマイグレーションアンカポイントを特定するものである。次に、ステップ205に略述するように、グローバルマイグレーションアンカポイントは、受信したデータメッセージを処理して、送信元エントリ201がグローバルマイグレーションアンカポイントを特定するものとなるように、そして宛先エントリがローカルマイグレーションアンカポイントLPを特定する一方で、仮想マシンを特定するものとなるようにする。グローバルマイグレーションアンカポイントは、特定のサービス識別情報を含む、記憶されたデータレコードにより上記を行う立場にある。

40

【0045】

図2bに示すように、ローカルマイグレーションアンカポイントは、符号210に示す

50

ように、グローバルマイグレーションアンカポイントからデータメッセージを受信する。そして特に、ローカルマイグレーションアンカポイントは、宛先フィールド202に示すような、データメッセージが示す仮想マシン識別情報を含む記憶されたデータセットに基づいて、データメッセージ内におけるローカルマイグレーションアンカポイント識別情報を物理マシン識別情報へと置き換える。具体的には、宛先エントリの、特定の物理マシンへのこのような置換えをブロック215にも示している。

【0046】

続いて、図3aについて説明する。図3aは、図1に示した中央ネットワークマネージャ（central network manager, CNMS）160の1つの機能を示している。具体的には、CNMSは符号300に示すように、グループ内マイグレーションの要求又は決定を受信する。ローカルマイグレーションアンカポイントは符号305に示すように、図1の符号101等の、対応するグループマネージャから新たな物理マシンのIDを受信する。続いて、ローカルマイグレーションアンカポイントは、データセット内の第1の物理マシンの識別情報を第2の（新たな）物理マシンの識別情報に置き換える。これを符号310に示している。したがって、グループ内又はクラウド内のマイグレーションは、ローカルマイグレーションアンカポイントに記憶されているデータセットにのみ影響を与え、グローバルマイグレーションアンカポイントに記憶されているデータレコードには何ら影響を与えない。VMLRは、物理マシン識別情報を何ら記憶しておらず、LP及びGPデータ並びにサービスIDデータのみを記憶しているため、VMLRにおける変更も不要である。

【0047】

図3bは、中央ネットワークマネージャ160が、グループ間マイグレーションを決定する状況、すなわち、第1のローカルマイグレーションアンカポイントに関連付けられている第1の物理マシンから別のローカルマイグレーションアンカポイントに関連付けられている第2の物理マシンへの仮想マシンのマイグレーションによるグループ間マイグレーションを決定する状況を示している。そのために、図1のCNMS160は符号315に示すように、グループ間マイグレーションの要求又は決定を受信する。次に、マイグレーションの宛先である第2のローカルマイグレーションアンカポイントは符号320に示すように、物理マシンの第2のグループ内の第1の物理マシンから、物理マシンの第2のグループ内のその第1の物理マシンすなわち新たな物理マシンに第1の仮想マシンが置かれているという情報を受信する。さらに、第2のローカルマイグレーションアンカポイントは符号325に示すように、グローバルマイグレーションアンカポイントにメッセージを送信する。このメッセージは、第1の仮想マシンが現時点で物理マシンの第2のグループに置かれていることを示すものである。符号330に示すように、グローバルマイグレーションアンカポイントは、以前のローカルマイグレーションアンカポイント330に関する情報を受信する仮想マシンロケーションレジスタVMLR150にアクセスする。代替的に又はこれに加えて、第2のローカルマイグレーションアンカポイントは符号335に示すように、以前のローカルマイグレーションアンカポイントに関する情報を取得するためにあるメッセージをVMLRに送信する。基本的には、ステップ330及び335のうち的一方のみで十分であるが、実施態様に応じて、双方のステップを累積的に実行することもできる。

【0048】

ステップ340では、第1のローカルマイグレーションアンカポイントは、第1の仮想マシンへと指定されたデータメッセージの宛先エントリに第2のローカルマイグレーションアンカポイントを示すことによって、そのデータメッセージを第2のローカルマイグレーションアンカポイントに送信し、そのデータメッセージが、求められている仮想マシンが置かれている正しい物理マシンへとルーティングされるようにする。加えて、第1の仮想マシンは、マイグレーション後に第1のローカルモビリティアンカポイントを第2のローカルモビリティアンカポイントに知らせることができる。

【0049】

10

20

30

40

50

続いて、ある特定のページング (paging) 機能を示す図 3 c について説明する、ステップ 350 において、ローカルマイグレーションアンカポイントは、ある仮想マシンを特定したロケーション登録更新要求を、そのローカルマイグレーションアンカポイントに接続されている物理マシンのグループ内の全ての物理マシンに送信する。ローカルマイグレーションアンカポイントは、ステップ 355 において、上記仮想マシンが置かれている物理マシンから応答を受信する。ステップ 360 では、ローカルマイグレーションアンカポイントは、仮想マシンロケーションレジスタ又はこの仮想マシンロケーションレジスタに加えてグローバルマイグレーションアンカポイントに対し、上記仮想マシンが存在する物理マシンを通知する。さらに、VM は LP に直接返答することができ、そのため、全体のトラフィックは PM にとって透過的なもの (transparent) として保たれる。

10

【0050】

図 3 d は、GP1 及び GP2 といった 2 つのグローバルマイグレーションアンカポイントを有するシステムにおいて実行されることが好ましい処理を示している。ステップ 370 において、GP1 は、あるサービス ID を有するサービスのクライアント要求を受信する。次にステップ 375 において、GP1 は、そのサービス ID に関し自己のデータレコードを調べる。ステップ 380 に示すように、メッセージに含まれているサービス ID が見つからない場合には、GP1 は VMLR にアクセスする。次のステップ 385 において、GP1 は、GP2 からの ID を VMLR から受信する。次のステップ 390 において、GP1 は GP2 に対し、クライアントとそのサービス ID を有するサービスと少なくとも一方を通知する。ステップ 395 において、GP2 は、クライアントを直接、アドレス指定するか、又は通信が GP1 を通じて GP2 及びクライアントへとルーティングされる。しかし、データメッセージの受信者すなわちあるローカルマイグレーションアンカポイントが、サービス識別情報によってアドレス指定されたある仮想マシンがこの階層ネットワーク内で接続している実際のグローバルマイグレーションアンカポイントを特定するとすぐに、別の代替形態も同様に実行することができる。

20

【0051】

続いて、初期化処理のための詳細な実施形態を示すために、図 6 を詳細に説明する。

【0052】

符号 600 に示す物理マシンは、マイグレーション管理モジュール 601 を備えている。ある仮想マシンについて、その仮想マシンの ID と、IP アドレスと、メモリと、例えばコア 1 等のある特定のハードウェアリソースとを定めることによりインスタンスが作成された後、その仮想マシン 602 が物理マシンに存在することになる。次に、物理マシンコントローラ 603 は自己の物理マシン ID を送信する。この物理マシン ID を PM ID として示している。次に、仮想マシンのマイグレーション管理モジュール 604 は、この PM ID を記憶し、自己の VM ID 又は「サービス ID」を物理マシンのマイグレーション管理 601 に返信する。このサービス ID は、この分野において知られているように、アプリケーション ID 又は URL と同義であることに留意すべきである。次に、物理マシンのマイグレーション管理機能は、符号 605 に示すように、仮想マシンのサービス ID 及び物理マシンの物理マシン ID を、指定されたマイグレーションアンカポイントに送信する。次に、ローカルマイグレーションアンカポイントは、仮想マシン ID、物理マシン ID を記憶し、次いで、ステップ 606 に示すように、サービス ID と仮想マシン ID と物理マシン ID とローカルマイグレーションアンカポイント ID とをグローバルマイグレーションアンカポイントに知らせる。次に、グローバルマイグレーションアンカポイントは、サービス ID と仮想マシン ID とローカルマイグレーションアンカポイント ID とを記憶し、符号 607 に示すように、サービス ID と仮想マシン ID とローカルマイグレーションアンカポイント ID とグローバルマイグレーションアンカポイント ID とを VMLR に知らせる。次に、VMLR は、エントリを作成して、サービス ID と仮想マシン ID とローカルマイグレーションアンカポイント ID とグローバルマイグレーションアンカポイント ID とを互いに関連付けて記憶する。さらに、全体の登録プロセスは、登録情報を受信したあらゆるモジュールからの ACK (確認応答) メッセージ及び応答を用い

30

40

50

て行われることが好ましい。すなわち、LPは応答を物理マシンに返信し、GPは応答をLPに返信し、VMLRは応答をGPに返信することが好ましい。

【0053】

次に、図7を参照してサービス発見及びセッションの確立について説明する。

【0054】

まず、図7の符号700に示すクライアントが、メッセージをいわゆるDNSサーバに送信する。具体的には、このクライアントは、ある特定のサービスにアクセスしようとしており、この特定のサービスはそのクライアントが当然知らない仮想マシン上で動作している。しかし、クライアントはウェブアドレスを知っているため、クライアントは、このURLのサーバIDに関する情報を見つけるために、第1のメッセージを用いてDNSサーバ701にアクセスする(符号711)。次に、DNSサーバ702は、仮想マシンが関連付けられているグローバルマイグレーションアンカポイントのIDを示す第2のメッセージにより応答する。この情報は、ステップ712に示すように、DNSサーバが提供することができる。なぜならば、DNSサーバは、一方でグローバルマイグレーションアンカポイントとの関連付け、他方ではサービスID又はURLとの関連付けに関して更新されるからである。次に、第3のステップにおいて、クライアント700は、符号713に示すように、メッセージ712にて示されたGPにアクセスし、当該クライアントがこのURL用のセッションを確立しようとしていることを要求する。

【0055】

次に、GP1は、符号714に示すように、(クライアント700そのものではなく)GP1がこのURLのセッションを確立しようとすることをLP1に知らせることにより、関連付けられているLP1をアドレス指定する。GP1は、そのセッションがクライアントのためのものではなくGP1のためのものであることを示す。この情報は、LP1等の対応するLPを通じて第1のクラウド720へとルーティングされ、LP1は、メッセージ714に示されている仮想マシンIDが属する物理マシンID721を知る。仮想マシンIDを符号722に示されている。次に、物理マシン、特に物理マシンのマイグレーション管理部及び仮想マシンのマイグレーション管理部、又は図6において説明したマイグレーション管理要素のみが、セッション確立がOKであること及びセッションの所有者がGP1であることを伝えるメッセージ715を通じて応答する。次に、符号716によって示すように、GP1は、セッションがOKであること及びセッションがクライアント用であることをクライアントに返信して報告する。しかし、クライアントは、特定のセッションの所有者がクライアント自身ではなくGP1であることは知らない。

【0056】

続いて、図8を参照してデータパスを説明する。セッションが図7の手順によって確立された後、クライアント700はこの時点でペイロードの通信を開始する。これは、メッセージ800によって行われる。このメッセージは、クライアントから送信されるメッセージの宛先としてGP1を表す宛先セクションを有しており、送信元としてクライアントを表す送信元セクションを有しており、そして、ペイロードセクションを有している。次にGP1はメッセージ802を、特定のサービスIDと関連付けられたローカルマイグレーションアンカポイントまで送信する。この目的のために、メッセージ802に示しているように、送信元フィールドは、クライアント1からGP1へ変更され、宛先フィールドは、一方では仮想マシンに変更されるとともに他方ではローカルマイグレーションアンカポイントIDへ変更される。次に、ローカルマイグレーションアンカポイントは、メッセージ803を特定の物理マシンに送信する。この場合も、送信元フィールドは、変更されずGP1のままであるが、宛先フィールドは、メッセージ802のようなローカルマイグレーションアンカポイントID及び仮想マシンではなく、物理マシンID及び仮想マシンを表すものとなるように変更される。次に、示された物理マシンIDを有する物理マシンは、宛先フィールドによって示されている仮想マシンにメッセージ804を送信し、次いで仮想マシンはそのメッセージを処理し、その結果をメッセージ805として返信する。この場合、このメッセージは、宛先GP1と、クラウド1(符号720)内でこのメッセ

10

20

30

40

50

ージを実際に生成した仮想マシンである送信元とを有している。次に、仮想マシンをホストする物理マシンに関連付けられたマイグレーション管理マネージャは、仮想マシンに関連付けられたマイグレーションマネージャからメッセージ 805を受信する。次に、物理マシンは、メッセージ 806をローカルマイグレーションアンカポイントまで送信する。このメッセージは、送信元フィールドがVMのままであり、宛先フィールドはGP1のままであるとともに、宛先にはLP1が追加して示されている。しかし、これは、LPがクラウドの外部ゲートウェイ(outgoing gateway)として構成されていない場合にのみ必要である。その一方で、LP1がクラウド720内の全ての物理マシンのための外部ゲートウェイとなるように自動的に構成されている場合には、メッセージ806内でLP1は必要なく、メッセージ805及び806は同一となる。

10

【0057】

次に、LP1は、送信元フィールド及び宛先フィールドが、LP1の識別情報が削除されていることを除いて変更のないメッセージ807をGP1に送信する。次に、URLとVMのエントリを実際に有するGP1が、ファイルメッセージ808をクライアント700へ送信し、クライアントは、そのクライアントのサービスがGP1から提供されていることを実際に認識する。したがって、図8は、本発明の階層システムの重要な利点、すなわちクライアントが階層システム内の下層については何も気にする必要がなく、メッセージが送信されるグローバルマイグレーションアンカポイントにのみ対応するだけでよいことを示している。

【0058】

20

続いて、マイグレーションのサポート及びハンドオーバを説明するために図9aを参照する。3つのクラウド901、902、903が示されており、物理マシン904及び別の物理マシン905が示されている。さらに、VMLR150と、LP1(符号110)と、LP2(符号130)と、番号nを有する別のローカルマイグレーションアンカポイントとが示されており、さらに、2つのグローバルマイグレーションアンカポイント140及び別のグローバルマイグレーションアンカポイント910が示されている。メッセージ911は、ペイロードセクションと、VM ID及び物理マシン904の物理マシンIDを有する宛先セクションとを有している。ここで、図9bに示しているように、仮想マシンが、物理マシン904から物理マシン905へと、すなわちクラウド内でマイグレーションすることになる。このことは、物理マシンからメッセージ912を通じてローカルマイグレーションアンカポイントに通信され、次いで、ローカルマイグレーションアンカポイントLP1は、メッセージ911内の物理IDエントリをマシン904の物理IDからマシン905の物理IDへ変更する。しかし、仮想マシンがクラウド901内の物理マシン905から第2のクラウド内の物理マシン915に移動する場合は、追加の手順が必要である。続いて、これらの手順を図9cを参照して説明する。第1のステップにおいて、仮想マシンIDを有する仮想マシンが、符号920に示すように、物理マシン905から物理マシン915に移動する。次のステップでは、物理マシン915が、その関連付けられたローカルマイグレーションアンカポイント130にこの新たな状況を通知する。これは、メッセージ921により行われる。次に、ローカルマイグレーションアンカポイント130は、それに関連付けられているグローバルマイグレーションアンカポイントに対しメッセージ922によりこの新たな状況を通知する。加えて、この新たな状況は、メッセージ923によりLP2(符号130)からLP1(符号110)に通知することもできるし、メッセージ924によりGP1(符号140)からVMLRに通知することもできる。次に、図9bのメッセージ911をまだ所有しているローカルマイグレーションアンカポイント110は、このメッセージを処理しなければならない。この目的のために、それまで物理マシン905を示していた宛先フィールドは、いまでは、図9cの符号925に示すように、ローカルマイグレーションアンカポイント130を示す。次に、このメッセージは、ローカルマイグレーションアンカポイント130に実際に到達することができる。次に、図9dに示すように、LP2(符号130)は、それ自身すなわちそのローカルマイグレーションアンカポイントの物理IDを、物理マシン915の物理マシンID

30

40

50

に置き換える。次に、図 9 d に示しているように、グローバルマイグレーションアンカポイントが受信したメッセージは、ブロック 9 1 5 に現在存在する仮想マシンに最終的に到達するように、ローカルマイグレーションアンカポイント 1 1 0 にルーティングされ、そこからローカルマイグレーションアンカポイント 1 3 0 にルーティングされる。しかし、パケットの順序付け (ordering) が問題とならない場合、すなわち、仮想マシンがパケット再順序付け (packet re-ordering) 機能を有している場合には、グローバルマイグレーションアンカポイント 1 4 0 は、符号 9 3 9 に示す L P 1 を通じた間接ルートではなく、符号 9 4 0 に示す直接ルートを取ることができる。

【 0 0 5 9 】

したがって、この手順によれば、マイグレーションによるセッション切断を回避することができる。なぜならば、再ルーティングは、クライアントが何らの処理を行うことなくスムーズに行われるからである。さらに、全ての再ルーティングの手順は、利用可能な情報を用いて行われるため、L P 1 (符号 1 1 0、1 1 3) 又は G P は、上記で説明したように、送信元フィールド又は宛先フィールドに対応した処理によってメッセージを容易に転送することができる。中央コントローラのみが存在する集中型の方法と比較すると、ルート 9 3 9、9 4 0 は大幅に短い。

【 0 0 6 0 】

続いて、個々のエンティティの協調処理に関するフローチャートを説明するために図 1 0 を参照する。ブロック 1 0 0 0 において、V M のインスタンス作成又は V M のマイグレーションが行われたか否かが判断される。行われていないと判断された場合にはこの手順は終了する。他方、インスタンス作成又はマイグレーションに該当すると判断された場合にはステップ 1 0 1 0 が行われる。本ステップでは仮想マシンは自己を物理マシンに登録する。物理マシンが有効な仮想マシン情報を既に有している場合には、本ステップを省略することができる。ステップ 1 0 2 0 において、仮想マシン及び対応する物理マシンは、それらに対応するローカルマイグレーションアンカポイントに登録を行う。クラウド内のプロセスの場合には、ブロック 1 0 3 0 に示している手順が実行される。他方、ブロック 1 0 2 5 において、クラウド間処理に該当すると判断された場合には、ブロック 1 0 3 5 に示すように、仮想マシンを現在有しているローカルマイグレーションアンカポイントは、それまでのローカルマイグレーションアンカポイントとグローバルマイグレーションアンカポイントと V M L R とに対しクラウド間マイグレーションであることを知らせる。その一方で、ブロック 1 0 2 5 にてクラウド内マイグレーションと判断された場合には、ブロック 1 0 3 0 が行われる。ブロック 1 0 3 0 は、登録タイマの満了、又は L P、G P 若しくは V M L R によって発行される意図的な更新要求を示している。このタイマは、ブロック 1 1 0 c、1 3 0 c に示したように、ローカルマイグレーションアンカポイントに配置されている。このタイマが満了すると、L P、G P 若しくは V M L R 又は個々の全てのエンティティによってロケーションの更新が行われる。他方、タイマが満了していない場合には、ブロック 1 0 4 0 が、登録タイマが満了するまで何も起きないことを示している。したがって、ブロック 1 0 1 0 及び後続のブロックに示す手順は、個々のトリガに応じて行われる。トリガの 1 つは登録タイマの満了であり、別のトリガは、L P のいずれか、G P のいずれか、又は V M L R からのロケーション更新要求である。

【 0 0 6 1 】

続いて、特定の好ましいページング機能を説明する。V M L R、G P、L P において有効なエントリが何らかの理由で利用可能でない場合には (1 つの理由は、データ破損若しくはデータ送信エラー又はそれに類似のものである)、V M L R と G P と L P との少なくともいずれかが、全ての L P (又は仮想マシンが近時において最後に存在していた幾つかの L P) に対しページングを行うように依頼することができる。これは、G P を通じて行うこともでき、加えて、V M L R が単独でページングを行うこともできれば、G P にページングを行うように依頼し、その G P がその配下にある L P にページングを行うように依頼することもできる。

【 0 0 6 2 】

10

20

30

40

50

次に、各LPは、自己の各クラウド内にある全ての物理マシン(PM)に対しロケーション登録、更新要求をブロードキャストする。次に、対象となっているVMをホストしている物理マシン(又はVM自体)は、LP、具体的にはロケーション登録、更新要求に対して応答し、このとき、LPは、どの物理マシンが仮想マシンをホストしているのかわかる。次にLPはVMLRに通知し、また、GP又はより多くのグローバルマイグレーションアンカポイントに知らせることもできる。この目的のために、LPは次に、それ自身のLP IDをVMLRに転送し、次いでVMLRはクライアントからの新たなセッション要求を正しいGPを通じて正しいLPに、そしてそこから正しい物理マシンに実際に転送することができるように、サービスIDに対応するデータエントリを更新することができる。

10

【0063】

図11は、マイグレーションに関する決定を見いだすための更なる手順を示している。符号1100に示すネットワーク構成プラットフォーム(network configuration platform)NCPは、種々のクラウド、具体的には種々のクラウドコントローラ1110及び1120とのインタフェースを維持するものである。ネットワーク構成プラットフォームNCPは、これらの種々のクラウドコントローラ(クラウドO&M)とのこのようなインタフェースを維持し、それ自体が行う監視に基づいて又はクラウドコントローラ1120、1110からの信号に基づいて決定を行うことが好ましい。この決定は、仮想マシンがどのクラウドからどのクラウドにマイグレーションすべきであるかを示すものである。クラウドコントローラ1120、1110は、これらのクラウドコントローラの制御下にある物理マシン上の仮想マシンにリソースを割り当てることを担当する。これらのクラウドコントローラは、図1にてグループマネージャ101、121としても示したものである。特に、VMLRは、サービスID又は「URL」、関連付けられている仮想マシンIDと、LP IDと、GP IDとを記憶する。

20

【0064】

ルートの変更と、サービス提供と、仮想マシンの発見とは、GPと共同して行われ、これは上記で説明したとおりである。

【0065】

本発明は、次の理由から有利である。本発明の階層システムは、LP及びGPといった少数のアンカポイントしか必要としないため拡張性を有している。これによって、例えばシグナリングの観点で複雑度が減少する。さらに、本発明の手順は、セルラネットワークに対して親和的であり、世界中で広範囲にわたって稼動しているセルラネットワークの動作に基づく経験(experience)をクラウドコンピューティングにも用いることができる。本発明の好ましい実施形態は、クラウドが、又は少なくとも2つの物理マシンを有する1つのグループかを有するシステムに関する。このシステムでは、複数の物理コンピューティングマシン(PM)が、複数の仮想マシンをホストする。さらに、このシステムは、1つ以上のローカルマイグレーションアンカポイント(LP)と、1つ以上のグローバルマイグレーションアンカポイントGPと、仮想マシンロケーションレジストラ(virtual machine location registrar)を備えており、これらのエンティティのそれぞれが、特定されるための一意のIDを保持するとともに仮想マシンのロケーションへのポインタを保持する。

30

40

【0066】

本発明の1つの特徴は、VM、PM、LP及び/又はGPにおいて行われるロケーション登録ステップである。本ステップを通じて、VM、PM、LP、GP及び/又はVMLRは、前述したVMがネットワークのどこに置かれているのか、すなわちこのVMがどのPMに存在するのか、及びこのVMがサービスID又はURLによって特定されるどのような種類のサービスを提供するものであるのかという情報を受信する。

【0067】

本発明は更に、サービスIDあるいはURL等のアプリケーションプログラム用アクセスIDと、それをホストする仮想マシンと、その仮想マシンが置かれている物理マシンと

50

、物理マシンとLPの関連付けと、LPとGPの関連付けとのマッピングを保持するデータベースシステムに関する。これによれば、これらのエンティティ、すなわち物理マシンと、ローカルマイグレーションアンカポイントと、グローバルマイグレーションアンカポイントとは、それらのIDによって特定される。

【0068】

本発明の別の態様では、ローカルマイグレーションアンカポイントは、仮想マシンが同じクラウド内に存在する場合にその仮想マシンのマイグレーションをサポートし、どの仮想マシンがどの物理マシンに置かれているのかに関する情報を保持する。特に、ローカルマイグレーションアンカポイントは、仮想マシンが新たな物理マシンに移動した場合には物理マシンIDを変更する。したがって、ローカルマイグレーションアンカポイントは、その仮想マシンを宛先とするデータを、その仮想マシンが置かれている適切な物理マシンへとルーティングするものであり、これは、データヘッダの前に適切な物理マシンIDを追加することによって行われることが好ましい。

10

【0069】

ローカルマイグレーションアンカポイントは、当該ローカルマイグレーションアンカポイントが担当するクラウド内に置かれていた仮想マシンを宛先とするデータを、マイグレーション後には、例えば新たなLP-IDをデータヘッダに追加することによって、その仮想マシンの新たなローカルマイグレーションアンカポイントに転送することを担当する。

【0070】

ローカルマイグレーションアンカポイントは、VMがあるクラウドから別のクラウドにマイグレーションするか否かをVMLR及びGPに更に知らせ、加えて、以前のLPにも同様に知らせる。

20

【0071】

ローカルマイグレーションアンカポイントは、VMLR若しくはGPからの要求に応じて又は単独でブロードキャストページングメッセージをそのクラウド内の全ての物理マシンに対して発行し、このページングメッセージ内に特定の仮想マシンIDを明示的に挙げておくことによって、全ての仮想マシン又は1つ若しくは幾つかの仮想マシンについての仮想マシンのロケーションの更新を開始することができる。

【0072】

グローバルマイグレーションアンカポイント(GP)は、クラウド間での仮想マシンのマイグレーションをサポートし、どのローカルマイグレーションアンカポイントを通じてどのように仮想マシンに到達することができるのかに関する情報を保持する。GPは、アプリケーションIDと、VM等のアプリケーションマシンのホストとの間の関係を解決するためのソルバとして更に機能し、GPは、それ自身のIDを、クライアントが検索しているアプリケーションのIDとして、問い合わせたクライアントに返す。GPは、App-IDとVMとLPとGPとの情報又は少なくともその一部を保持する。

30

【0073】

GPは、クライアントに代わり、アプリケーションが置かれている仮想マシンとのセッションをセットアップ又はオープンすることが好ましく、送信元自体であるかのように振る舞う。これは、セッションの分割(session splitting)を参照して説明したものである。

40

【0074】

GPは、送信元であるクライアントIDをそれ自体のIDに置き換えることによって、クライアントからのデータを転送することが好ましい。次に、GPは、データヘッダの前に適切なLP IDを追加する。

【0075】

GPは、ローカルマイグレーションアンカポイントからある仮想マシンのロケーション更新を受信すると、以前のLPに代えて新たなLPのIDを追加することにより、継続中のセッションのルートをその仮想マシンの新たなロケーションへと変更することができる

50

【 0 0 7 6 】

G Pは、クライアントを宛先とするデータを仮想マシンから受信すると、仮想マシンの送信元IDを置き換えるように更に構成されており、G Pはこれを単独で行い、当該G Pがデータの送信元であるかのように見せかける。また、G Pは、データの宛先を自身からクライアントIDに置き換える。

【 0 0 7 7 】

仮想マシンロケーションレジスタ又は仮想マシンロケーションレジスタは、どのアプリケーションIDが、どのグローバルマイグレーションアンカポイントによってカバーされたどのローカルマイグレーションアンカポイントによってカバーされているどの仮想マシンに置かれているのかに関する情報（URL、VM、LP、GP）又は少なくともこの情報の一部を保持する。アプリケーションIDは、ウェブアプリケーション、ビデオ等のアプリケーションサービスに対する識別子を示すものである。URLが、アプリケーションIDの1つの例である。

【 0 0 7 8 】

LP、GP、VMLRのロケーションは、互いに対して任意であることに留意すべきである。これらのエンティティは、物理的又は機能的に同じ場所にデプロイ（deploy）することもできるし、機能に合わせてデプロイすることもできるし、それらの物理ロケーション又は機能的ロケーションに関して別々のままとすることもできる。

【 0 0 7 9 】

一実施形態では、GPをLPと統合することができる。この場合、GPの機能は、LPによって行われる。それにもかかわらず、この統合デバイスは、GPの機能すなわちデータレコードと、LPの機能すなわちデータセットとを有する。

【 0 0 8 0 】

セッションが分割されていない場合、すなわち、カプセル化が行われていない場合には、クライアントは、仮想マシンIDを宛先として有するデータを幅広く送信し、以前のLPは、マイグレーション後にデータを新たなLPに転送する。そのような場合、LPは、クラウドの入口又は出口ゲートウェイとして機能する。

【 0 0 8 1 】

したがって、本発明は更に、各サーバファームが複数の物理サーバマシンを有する複数のサーバファームに関する。各物理サーバマシンは、複数の仮想サーバマシンを保持し、各サーバファームは、ローカルマイグレーションアンカポイントに接続され、複数のローカルマイグレーションアンカポイントは、グローバルマイグレーションアンカポイントに接続されている。ローカルマイグレーションアンカポイント及びグローバルマイグレーションアンカポイントは、どのアプリケーションがどの仮想マシンに置かれているのか、並びにどの仮想マシンがどのLPによってカバーされているのか及びどのLPがどのGPによってカバーされているのかに関する情報を保持する仮想サーバマシンロケーションレジスタに接続されている。特に、VM、PM、LP、GPは、マイグレーション管理機能を装備しており、仮想マシンのロケーションは、GPとLPとPMのつながりを通じて突き止めることができる。

【 0 0 8 2 】

更なる実施形態によれば、種々のクラウドコントローラ又はグループマネージャ（図1の符号101及び121等）とのインタフェースを維持するネットワーク構成プラットフォームが提供される。図11の符号1100等のネットワーク構成プラットフォーム又はクラウド間マイグレーション管理モジュールは、それ自身が行う監視に基づいて又はグループマネージャからの信号に基づいて、マイグレーションをどのように行うべきであるか及び/又はどの仮想マシンがどのクラウドから他のどのクラウドにマイグレーションすべきであるかの決定を行う。各クラウドのグループマネージャ又は「クラウドO&M」は、対応するグループマネージャによって管理されている物理マシン上の仮想マシンにリソースを割り当てることを担当する。

【 0 0 8 3 】

続いて、ロケーション登録、更新プロセスをより詳細に説明する。オリジナルのIDを有する仮想マシンは、自身を、現在存在する物理マシンPMに登録する。

【 0 0 8 4 】

いずれかの仮想マシンが、ロケーション登録メッセージをローカルマイグレーションアンカポイントに送信する。この場合、この仮想マシンは、物理マシンのID及びLPのIDを、当該仮想マシンが存在している物理マシンから受信する。あるいは、PMは仮想マシンに代わってロケーション登録を行う。その場合、物理マシンは、それ自身のID及びVM IDをLPに送信し、LPが、このVMがこの特定のPMに存在していることを知るようになる。LPは、物理マシンへの仮想マシンのマッピングをそのデータベース又はその複数のデータセット内に保持する。このエントリの有効性は、LP内の対応するタイマによって定められることが好ましい所定の期間の満了を条件とする。ロケーション登録プロセスは、この期間内に仮想マシン又は物理マシンによって再び行われなければならない。

10

【 0 0 8 5 】

PMが、仮想マシン、物理マシンのエントリ用のロケーション更新メッセージを受信しない場合には、そのPMは、ロケーション更新要求を仮想マシン、物理マシンに発行する。

【 0 0 8 6 】

肯定的な返答を受信した場合には、VMとPMのエントリの有効性は、所定の期間に延長される。

20

【 0 0 8 7 】

PMは、否定的な返答、すなわち、VMが当該PMにもはや存在しないことを送信することもできるし、そのようなメッセージを無視することもできる。LPが、そのロケーション更新要求に対して否定的な返答を受け取るか又は返答を受け取らなかった場合は、LPは、この特定のエントリを複数のデータセットから削除する。LPは、この特定のVMに関するVMとLPのエントリ用のエントリがもはや有効でないことをVMLRに知らせることもできる。

【 0 0 8 8 】

ロケーション登録は、仮想マシンのインスタンスが作成されるか又は別のPMに移動するときに行われる。

30

【 0 0 8 9 】

LPはまた、その範囲内の全てのPMに対しロケーション登録、更新を行うよういつでも、例えばLPが自身をリポートしなければならず、全てのVMとPMとのマッピングを失う場合に依頼することができる。そのような場合、PMは、PM IDと全てのVM IDとを1つのメッセージに含めた単一のメッセージによってロケーション登録、更新を行うことができる。

【 0 0 9 0 】

続いて、インスタンス作成、マイグレーション、リポートの後に仮想マシンによって行われる手順を示した図12について説明する。

40

【 0 0 9 1 】

第一に、図12の符号1200に示しているように、VMはそのVM-IDを当該VMが位置づけられているPMに送信する。次に、符号1202に示すように、PMは、メッセージ内においてVM-IDとそれ自身のPM-IDを、接続されているLPに送信する。特に、LPは、受信したメッセージの送信元からPM-IDを推測することができ、これはPM-IDの暗黙的な通知であり、この場合、PM-IDはメッセージ1202において必要ない。

【 0 0 9 2 】

次に、LPは、VM-ID及びLP-IDを、接続されたGPにメッセージ1204で送信し、符号1206に示すように、この情報を以前のLPに送信し、この情報をVML

50

R に対しメッセージ 1 2 0 8 により送信する。代替的に又はこれに加えて、符号 1 2 1 0 に示すように、G P はこの情報を V M L R に、すなわちメッセージ 1 2 0 8 の代わりとして又はメッセージ 1 2 0 8 に加えて送信する。

【 0 0 9 3 】

続いて、本発明の好ましい実施形態を説明するために、セッションのセットアップに関する更なる説明を提供する。

【 0 0 9 4 】

クライアントはまず、U R L と V M - I D との変換に関して当該クライアントの D N S サーバを調べる。シナリオは、例えば G P が D N S の処理と類似した、U R L と V M - I D との変換装置として機能するというものである。したがって、全てのクライアントが、G P に対し、U R L からルーティング可能な I D への変換を依頼する。この場合、全てのクライアントは、U R L とルーティング可能な I D との解決を G P に依頼するようにプログラムされている。

【 0 0 9 5 】

他の U R L と V M - I D の変換装置は、U R L と V M - I D との解決要求を、D N S リダイレクションに相当する G P にリダイレクトすることができる。

【 0 0 9 6 】

G P は、有効な（満了していない）V M - I D と L P - I D のマッピングについてそれ自身の内部データベースを調べる。このようなマッピングが見つからない場合には、G P は V M L R に対し、U R L と G P - I D と L P - I D と V M - I D との適切なマッピングを要求する。G P が、L P がその G P 自身の範囲下にあることを確認し、かつ G P が（負荷又はオペレータのポリシー上の理由から）サービス提供したい場合には、V M L R からの応答に従って、G P は、クライアントが要求している U R L に対してそれ自身の I D を宛先 I D として返信する（そして、U R L と G P と L P と V M とのマッピングをその G P のデータベースに記憶する）。

【 0 0 9 7 】

V M がこの G P の配下でない L P にアタッチされている場合、G P は、クライアントからの解決要求を、その仮想マシンのグローバルマイグレーションアンカポイントとして機能している G P にリダイレクトする。ここで、この情報は V M L R からの応答に含まれていたものである。

【 0 0 9 8 】

G P は、データセッションを開始する前にも、仮想マシンとのセッションを確立する必要がある。クライアントは、宛先としてルーティング可能な I D （G P - I D 等）を得た後、データ送信に先立ってセッション確立の手順を開始する。セッション確立メッセージにおいて、G P は、送信元 I D （すなわちクライアント I D ）をその G P 自身の I D に置き換え、宛先 I D （すなわち G P 自身の I D ）を V M - I D に置き換える。次に、G P は、担当する L P の I D を追加し、このように処理したメッセージを転送する。

【 0 0 9 9 】

したがって、V M を宛先とするデータパケットは、まず G P に到着する。G P は、それ自身の I D を宛先となる V M - I D に置き換え、そのデータパケットを転送する。この宛先となる V M - I D は G P のみが知っている。なぜならば、送信元クライアントは、実際のアプリケーションが置かれている V M ではなく G P を宛先として認識しているからである。したがって、G P は、N A T の特徴と類似した、クライアントと G P と G P と V M とのセッションのマッピングを行うテーブルを維持するものである。

【 0 1 0 0 】

G P は、データを V M に転送する前に、このデータを L P - I D を用いてカプセル化し、データが V M に向かう途中で L P に到達するようにする。L P は、このデータを受信すると、外側にある I D 、すなわちそれ自身の I D を取り除く。L P は、次の I D として V M - I D を見つけ出す。L P は、そのデータベースを調べて、V M - I D と P M - I D とのマッピングを見つけ出す。次に L P は、宛先としての P M - I D を用いてデータをカプ

10

20

30

40

50

セル化する。

【0101】

したがって、PMはデータを受信し、次いで、PMは外側のID（それ自身のID）を取り除き、そのためVM-IDを認識できるようになる。そして、そのデータは、このとき認識可能なVM-IDによって特定される適切なVMへと配信される。

【0102】

幾つかの態様を装置との関連で説明してきたが、これらの態様は、対応する方法の記述も表していることは明らかであり、この対応する方法では、ブロック又はデバイスが、方法ステップ又は方法ステップの特徴に相当する。同様に、方法ステップとの関連で説明した態様も、対応する装置の対応するブロック又は項目又は特徴の記述を表している。

10

【0103】

ある特定の実施態様の要件に応じて、本発明の実施形態はハードウェア又はソフトウェアにおいて実施することができる。実施態様は、電子的に読取り可能な制御信号が記憶されたデジタル記憶媒体、例えばフロッピー（登録商標）ディスク、DVD、CD、ROM、PROM、EPROM、EEPROM、又はフラッシュメモリを用いて実行することができ、それらは、それぞれの方法が実行されるようにプログラム可能なコンピュータシステムと連携する（又は連携可能である）。

【0104】

本発明による幾つかの実施形態は、本明細書に記載される方法のうちの1つが実行されるように又は第1の取得信号若しくは第2の取得信号又は第1の混合信号若しくは第2の混合信号を記憶するようにプログラム可能なコンピュータシステムと連携することができる電子的に読取り可能な制御信号を有する非一時的データキャリアを含む。

20

【0105】

概して、本発明の実施形態は、プログラムコードを有するコンピュータプログラム製品としてインプリメントすることができる。該プログラムコードは、コンピュータプログラム製品がコンピュータ上で実行されると、方法のうちの1つを実行するように動作可能である。プログラムコードは、例えば機械可読キャリア上に記憶することができる。

【0106】

他の実施形態は、機械可読キャリア上に記憶された、本明細書に記載された方法のうちの1つを実行するコンピュータプログラムを含む。

30

【0107】

したがって、換言すれば、本発明の方法の一実施形態は、コンピュータプログラムがコンピュータ上で実行されるときに本明細書に記載された方法のうちの1つを実行するためのプログラムコードを有するコンピュータプログラムである。

【0108】

したがって、本発明の方法の更なる実施形態は、データキャリア（又はデジタル記憶媒体若しくはコンピュータ可読媒体）であって、該データキャリア上に記録された、本明細書に記載された方法のうちの1つを実行するためのコンピュータプログラムを含む、データキャリアである。

【0109】

40

したがって、本発明の方法の更なる実施形態は、本明細書に記載された方法のうちの1つを実行するコンピュータプログラムを表すデータストリーム又は信号シーケンスである。データストリーム又は信号シーケンスは、例えば、データ通信接続を介して、例えばインターネットを介して転送されるように構成することができる。

【0110】

更なる実施形態は、本明細書に記載された方法のうちの1つを実行するように構成又は適合された処理手段、例えばコンピュータ又はプログラム可能な論理デバイスを含む。

【0111】

更なる実施形態は、本明細書に記載された方法のうちの1つを実行するためのコンピュータプログラムがインストールされたコンピュータを含む。

50

【0112】

幾つかの実施形態では、プログラム可能な論理デバイス（例えばフィールドプログラマブルゲートアレイ）を用いて、本明細書に記載された方法の機能のうちの幾つか又は全てを実行することができる。幾つかの実施形態では、フィールドプログラマブルゲートアレイは、本明細書に記載された方法のうちの1つを実行するためにマイクロプロセッサと連携することができる。概して、本方法は任意のハードウェア装置によって実行されることが好ましい。

【0113】

上述した実施形態は、単に本発明の原理を説明するものである。本明細書において記載した構成並びに詳細の変更及び変形は当業者には明らかであることが理解される。したがって、特許請求の範囲によってのみ限定され、本明細書における実施形態の記述及び説明のために提示された特定の詳細によって限定されるものではないことが意図される。

10

【図1】

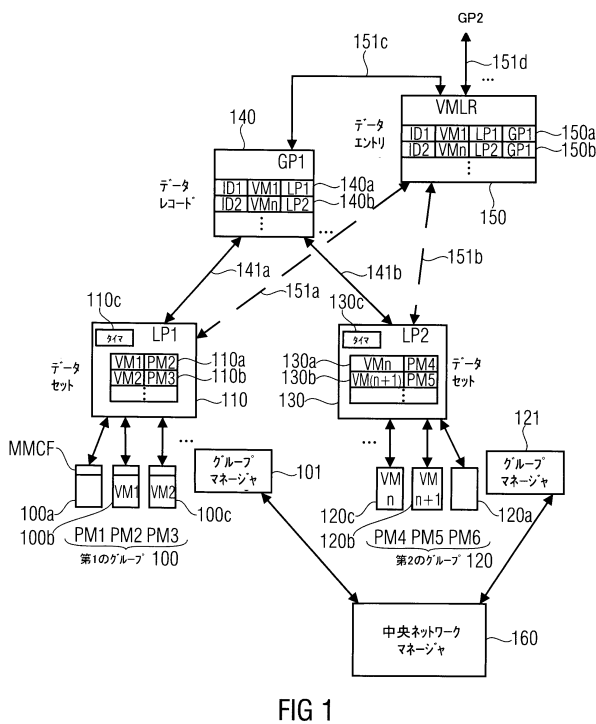


FIG 1

【図2a】

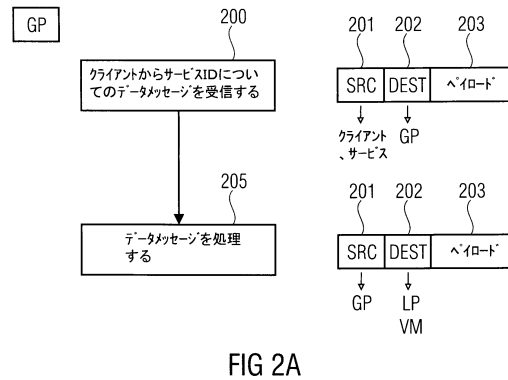


FIG 2A

【図2b】

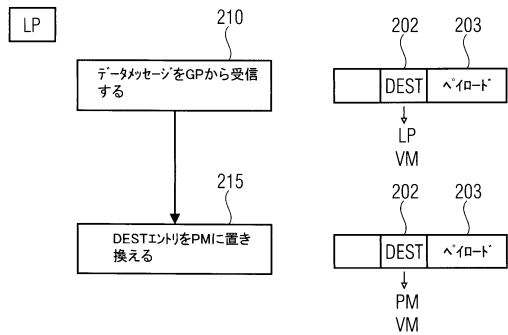


FIG 2B

【図3a】

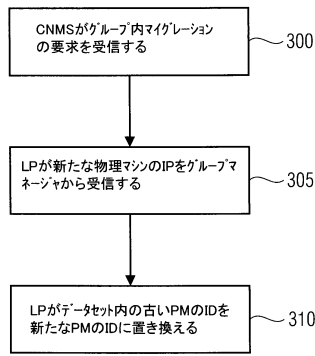


FIG 3A

【図3b】

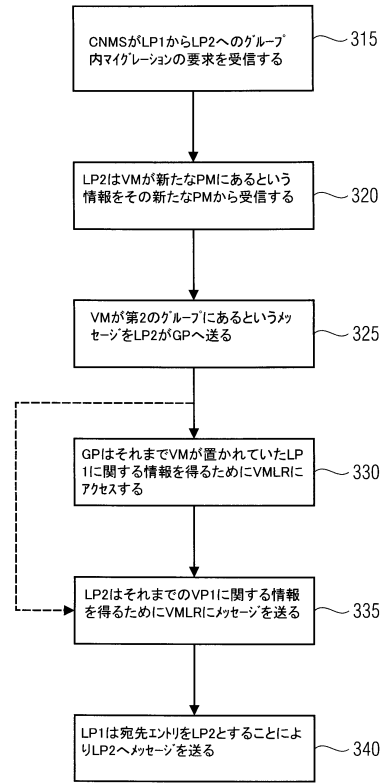


FIG 3B

【図3c】

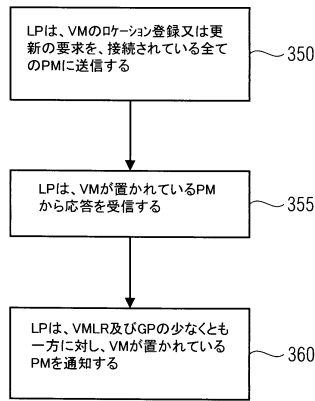


FIG 3C

【図3d】

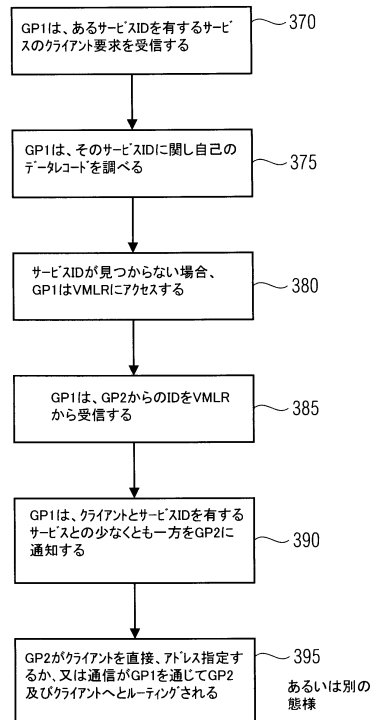
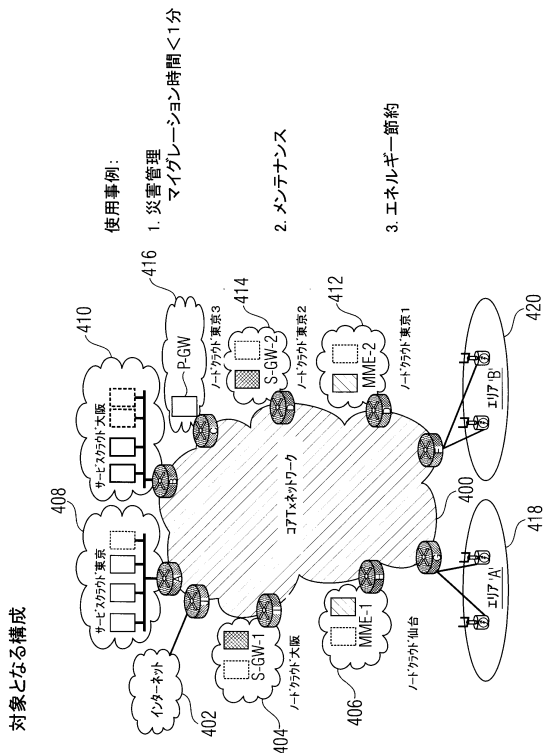


FIG 3D

【 図 4 】



【 図 5 】

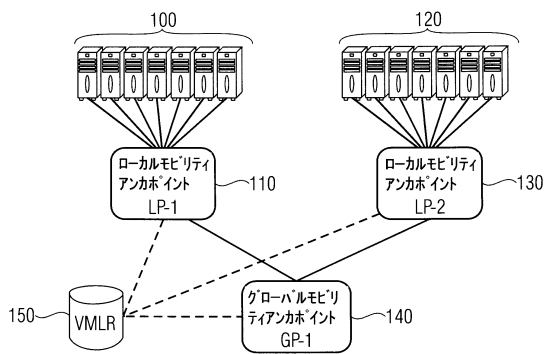


FIG 4

FIG 5

【 図 6 】

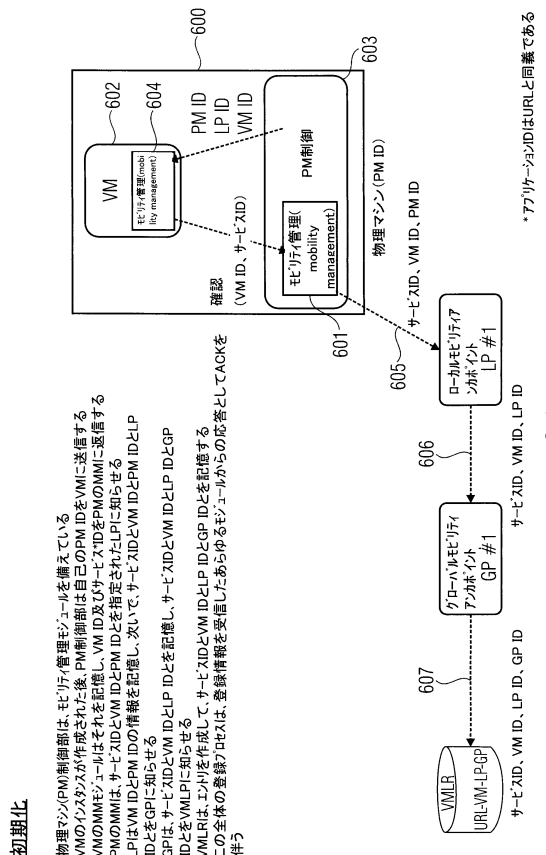


FIG 6

【 図 7 】

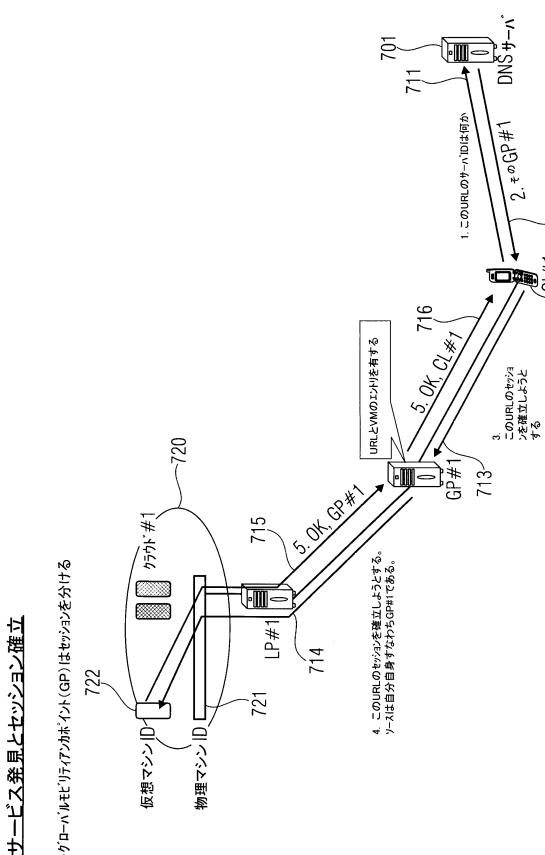


FIG 7

【図8】

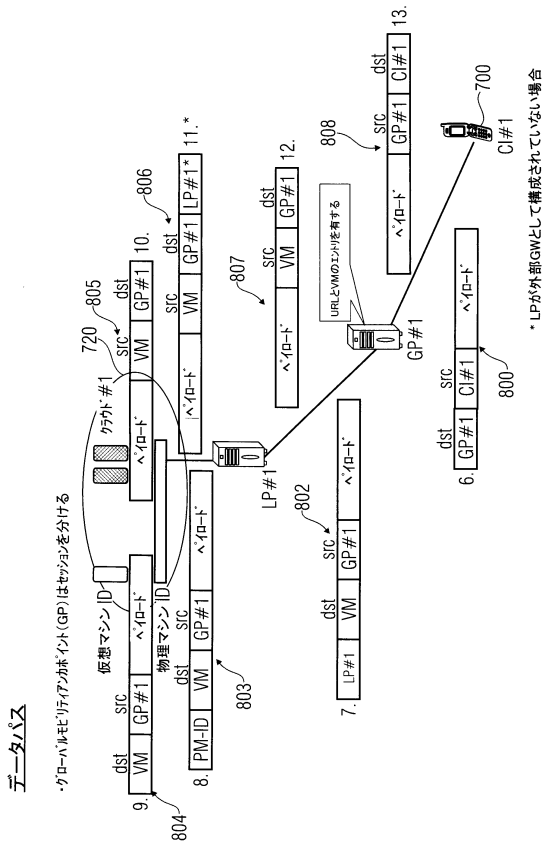


FIG 8

【図9a】

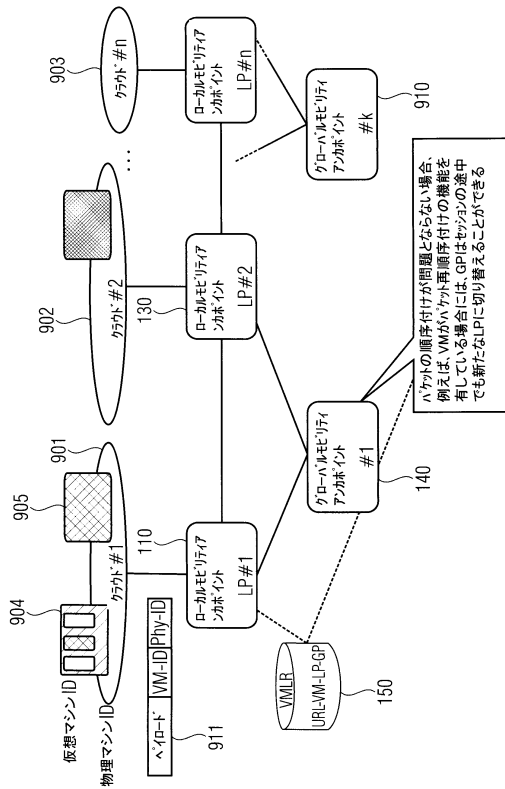


FIG 9A

【図9b】

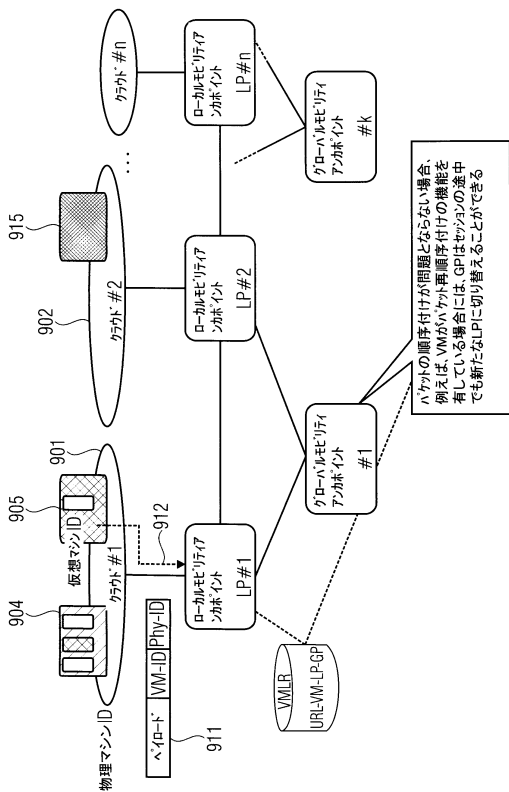


FIG 9B

【図9c】

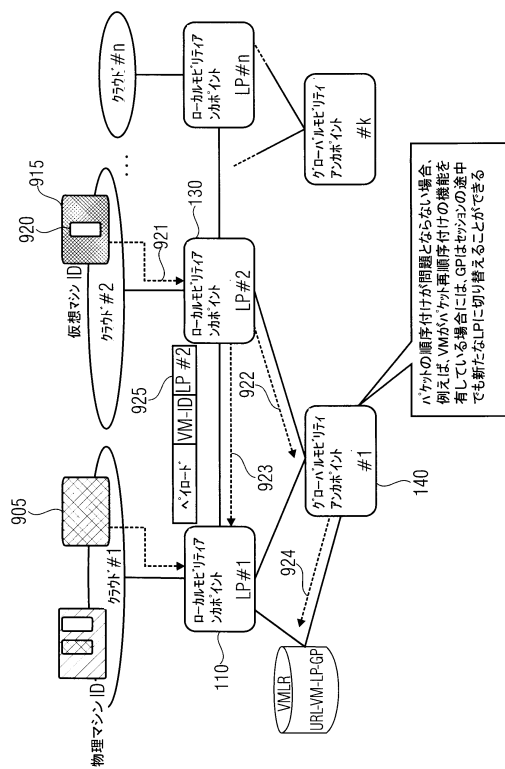
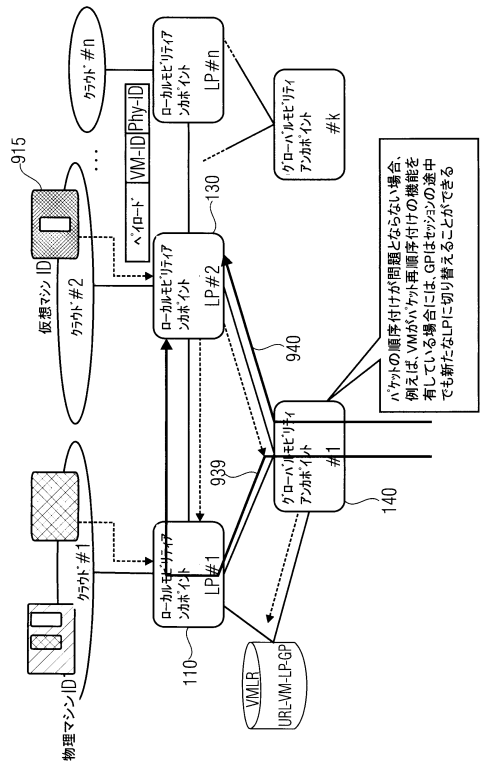


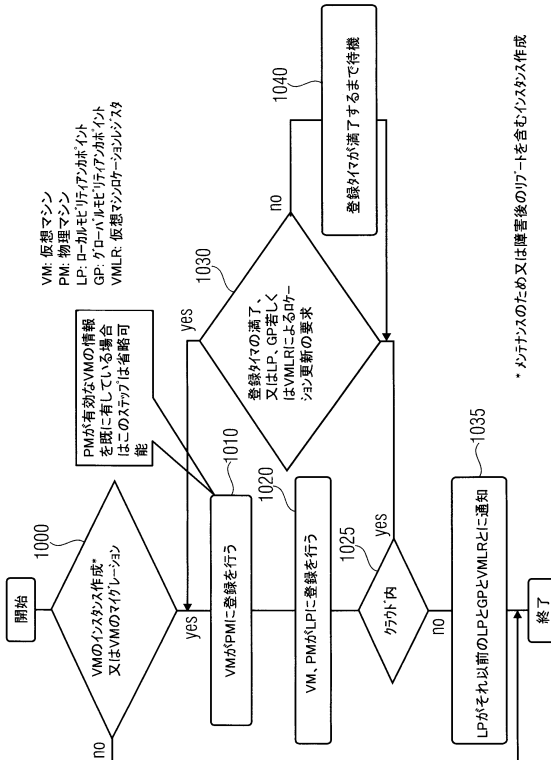
FIG 9C

【 図 9 d 】



マイグレーションによるセッション中断を回避することができる

【 図 10 】



*クラフト内のため又は購置後のワートを含むインスタンス作成

FIG 10

【 図 1 1 】

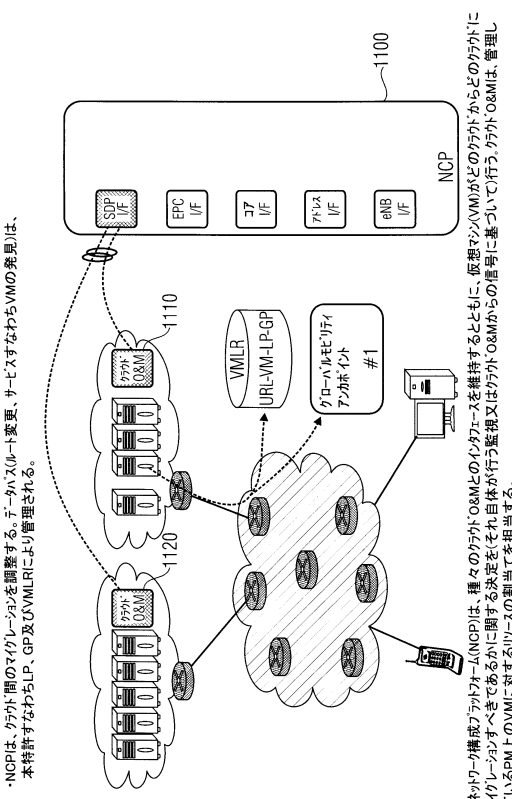


FIG 11

【 図 1 2 】

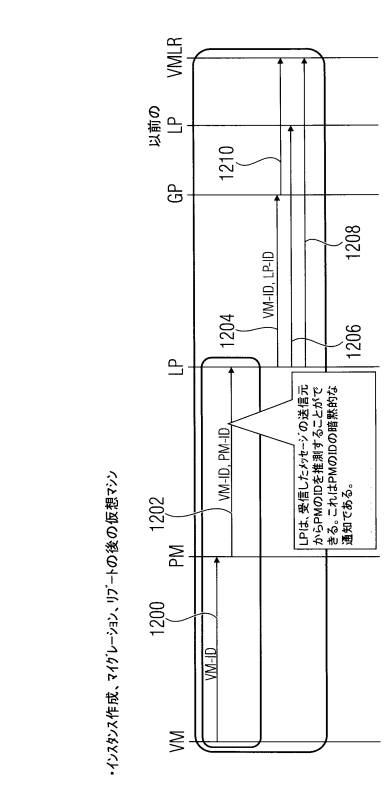


FIG 12

フロントページの続き

- (74)代理人 100154298
弁理士 角田 恭子
- (74)代理人 100166268
弁理士 田中 祐
- (74)代理人 100170379
弁理士 徳本 浩一
- (74)代理人 100161001
弁理士 渡辺 篤司
- (74)代理人 100179154
弁理士 児玉 真衣
- (74)代理人 100180231
弁理士 水島 亜希子
- (74)代理人 100184424
弁理士 増屋 徹
- (72)発明者 アシック・カーン
ドイツ連邦共和国, 8 0 6 8 7 ミュンヘン, ランズベルガー・シュトラッセ 3 0 8 - 3 1 2 ,
エヌ・ティ・ティ・ドコモ ユーロ ラボラトリーズ内
- (72)発明者 神津 和志
ドイツ連邦共和国, 8 0 6 8 7 ミュンヘン, ランズベルガー・シュトラッセ 3 0 8 - 3 1 2 ,
エヌ・ティ・ティ・ドコモ ユーロ ラボラトリーズ内
- (72)発明者 イシャーン・ヴァイシュナーヴィ
ドイツ連邦共和国, 8 0 6 8 7 ミュンヘン, ランズベルガー・シュトラッセ 3 0 8 - 3 1 2 ,
エヌ・ティ・ティ・ドコモ ユーロ ラボラトリーズ内

審査官 篠塚 隆

- (56)参考文献 国際公開第2012/033117(WO, A1)
特開2010-124129(JP, A)
特開2011-198299(JP, A)

- (58)調査した分野(Int.Cl., DB名)
G 0 6 F 9 / 4 6
H 0 4 W 4 0 / 0 0