(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5551749号
(P5551749)
(45)発行日　平成26年7月16日(2014.7.16)　(24)登録日　平成26年5月30日(2014.5.30)

(51)Int.Cl.　F I
G06F 11/20 (2006.01)　G06F 11/20 310C

請求項の数 14 (全 32 頁)

(21)出願番号　特願2012-219878(P2012-219878)
(22)出願日　平成24年10月1日(2012.10.1)
(62)分割の表示　特願2009-522059(P2009-522059)の分割
原出願日　平成19年2月19日(2007.2.19)
(65)公開番号　特開2013-33493(P2013-33493A)
(43)公開日　平成25年2月14日(2013.2.14)
審査請求日　平成24年10月31日(2012.10.31)
(31)優先権主張番号　11/498,802
(32)優先日　平成18年8月4日(2006.8.4)
(33)優先権主張国　米国(US)

(73)特許権者　509034030
ティーエスエックス　インコーポレイテッド
ＴＳＸ　ＩＮＣ.
カナダ，オンタリオ州　エム５エックス　１ジェイ２，トロント，ザ　エクスチェンジ　タワー，キング　ストリート　ウェスト　１３０
１３０ King Street West, The Exchange Tower, Toronto, Ontario M5X 1J2, Canada
(74)代理人　100116872
弁理士　藤田　和子

最終頁に続く

(54)【発明の名称】フェイルオーバシステムおよび方法

(57)【特許請求の範囲】
【請求項１】
一次サーバ（６２－１）および少なくとも１つのバックアップサーバ（６２－２）を備えており、
上記一次サーバ（６２－１）が、複数の処理対象となる入力を確定するように、かつ、上記複数の処理対象となる入力を処理する上記一次サーバ（６２－１）の処理に先立って、上記バックアップサーバ（６２－２）に上記複数の処理対象となる入力を送信するように構成され、
上記複数の処理対象となる入力は、上記サーバ（６２－１、６２－２）の両方による上記入力の処理の決定性を保証する入力を含む、フェイルオーバのためのシステム（５０）。
【請求項２】
上記複数の処理対象となる入力は、外部リソースに対する呼び出し結果を含む、請求項１に記載のシステム（５０）。
【請求項３】
上記複数の処理対象となる入力は、タイムスタンプの取得を要求して得た結果であるタイムスタンプ値を含む、請求項１または請求項２に記載のシステム（５０）。
【請求項４】
上記複数の処理対象となる入力は、共有リソース（１２２）をさらに含む、請求項１から３の何れか１項に記載のシステム（５０）。

【請求項５】
上記共有リソース（１２２）が、個々のサーバ（６２）におけるランダムアクセスメモリ内に維持されている、請求項４に記載のシステム（５０）。
【請求項６】
上記複数の処理対象となる入力は、上記一次サーバ（６２－１）により生成されたシーケンスナンバをさらに含む、請求項１から５の何れか１項に記載のシステム（５０）。
【請求項７】
上記バックアップサーバ（６２－２）は、複製エージェント（１２６－２）を含み、上記複製エージェント（１２６－２）は、上記一次サーバ（６２－１）から上記バックアップサーバ（６２－２）への情報のミラーリングを助けるために、上記一次サーバ（６２－１）のライブラリ（１０２－１）と交信するように構成されている、請求項１から６の何れか１項に記載のシステム（５０）。
【請求項８】
上記バックアップサーバ（６２－２）は、クライアント（５４）からの複製メッセージを認識するように構成されている、請求項１から７の何れか１項に記載のシステム（５０）。
【請求項９】
上記バックアップサーバ（６２－２）は、上記クライアント（５４）からの上記複製メッセージが受信されたとき、リカバリプロトコルを実行するように構成されている、請求項８に記載のシステム（５０）。
【請求項１０】
上記リカバリプロトコルは、複製された情報に基づいて、上記バックアップサーバ（６２－２）からの返答の送信を行うことを含む、請求項９に記載のシステム（５０）。
【請求項１１】
上記リカバリプロトコルは、ギャップリカバリであって、当該ギャップリカバリは、バックアップサーバ（６２－２）を用いて一次単独状態にフェイルオーバーする間、上記複数の処理対象となる入力をギャップリカバーする、請求項９または１０に記載のシステム（５０）。
【請求項１２】
上記一次サーバ（６２－１）および上記バックアップサーバ（６２－２）は、同時に動作する、請求項１から１１の何れか１項に記載のシステム（５０）。
【請求項１３】
システム（５０）におけるフェイルオーバのための方法であって、
一次サーバ（６２－１）において、複数の処理対象となる入力を確定するステップと、
上記一次サーバ（６２－１）からバックアップサーバ（６２－２）に上記複数の処理対象となる入力を送信するステップと、
上記一次サーバ（６２－１）において、上記複数の処理対象となる入力を送信した後に、上記複数の処理対象となる入力を処理するステップと、を含み、
上記複数の処理対象となる入力は、上記サーバ（６２－１、６２－２）の両方による上記入力の処理の決定性を保証する入力を含む、フェイルオーバのための方法。
【請求項１４】
ネットワーク（５８）を介して相互接続された少なくとも２つのサーバ（６２）のうちのいずれか１つのサーバ上で実行可能な、一組のプログラミング命令を保存するコンピュータ読み取り可能な記録媒体であって、
上記一組のプログラミング命令は、
一次サーバ（６２－１）において、複数の処理対象となる入力を確定するための命令と、
上記一次サーバ（６２－１）からバックアップサーバ（６２－２）に上記複数の処理対象となる入力を送信するための命令と、
上記一次サーバ（６２－１）において、上記複数の処理対象となる入力を送信した後に

、上記複数の処理対象となる入力を処理するための命令と、を含み、
上記複数の処理対象となる入力は、上記サーバ（６２－１、６２－２）の両方による上記入力の処理の決定性を保証する入力を含む、コンピュータ読み取り可能な記録媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、コンピュータおよびネットワークアーキテクチャに関する。より具体的には、フェイルオーバシステムおよび方法に関する。
【背景技術】
【０００２】
相互交流やビジネスを行う上で、社会はコンピュータおよびネットワークにますます依存している。重要なシステムにおいて求められる高レベルの可用性を確保するために、ソフトウエアおよびハードウエアの欠陥によって起こる不測のダウンタイムを最小限に抑える必要がある。
【０００３】
金融サービス産業は、高可用性を有するシステムを必要とする産業の一例である。実際、今日の金融産業における多くのデータ処理活動がコンピュータシステムに支えられている。特に興味深いのは、いわゆるリアルタイムおよびニアリアルタイムオンライントランザクション処理（OLTP）のアプリケーションであり、それらは通常、多数のビジネス取引を長期間にわたって高速かつ低遅延で行う。これらのアプリケーションは、概して下記の特徴を示す：（１）高度かつ高速のデータ処理（２）信頼性の高い不揮発性データ記憶（３）高レベルの可用性、すなわち、実質的に継続可能な基本原理に基づいてサービスを支える能力。しかし、実行する場合、下記に概要を詳述するように、システム動作に与える相反する影響のため、３つの特徴全てを同時に完全に満たすことができる設計は存在しない。
【０００４】
第一に、高度なデータ処理とは、タイムリーな方法で、多数のコンピュータ処理、データベース検索／アップデートなどを実行する能力のことを指す。並列処理を介して上記処理が実施され、同様の物理的機械または分散型ネットワーク上で同時に作業の統括ユニット（multiple unit of work）が実行される。あるシステムにおいて、各トランザクションの結果は既に完了したトランザクションの結果によって決まる。そのようなシステム並行の形態は、本質的に非決定性を有する。すなわち、乱調状態、オペレーティングシステムのスケジュールタスク、または可変ネットワーク遅延（variable network delay）が原因で、メッセージのシーケンスおよびスレッド実行（thread execution）が予測できず、単に複製システムに入力メッセージのコピーを送信するだけでは、それらを並列処理することができない。非決定性のシステムは非同一性のアウトプットを有するので、障害が起きた場合に代替物と代えようとしても、２つの異なるコンピュータ機器上で並列にそれらを動作させることができない。
【０００５】
第二に、信頼性の高い不揮発性データ記憶とは、多数のシステムのソフトウエアまたはハードウエア機器が不測の障害に晒された場合であっても、処理済みのデータを持続して保存できる能力のことを指す。共有データにアクセスするか、または変更するときに、アトミック性（Atomic）、一貫性（Consistent）、独立性（Isolated）、および永続性（Durable）（”ACID”）トランザクションを使用することによって、通常、上記処理を実施することができる。作業単位が完了すると、ACIDトランザクションがデータ保全とデータ持続性とを直ちに確保することができる。コミットメントされた全ACIDを非揮発性コンピュータメモリ（ハードディスク）に書き込み、それによってデータ持続性が確保される。しかし、パフォーマンスの点と一般に全システムを減速させる点とを鑑みると、上記は非常にコストがかかる。
【０００６】

第三に、高可用性のシステムとは、既定のコンピュータシステムにおける可用性率を時間あたり１００％に可能な限り近づけることを確実にすることを目的とする。冗長ソフトウエアおよび／またはハードウエアを介してそのような可用性が実施可能であり、それらのソフトウエアまたはハードウエアが、機器障害が検出された場合に機能性を引き継ぐ。引き継ぐために、ファイルオーバがデータだけでなく処理状態をも複製する。当業者であればわかるように、非決定性システム（すなわち、同じ一連のイベントのコンピュータ処理が、それらのイベントの処理順序に従って１つ以上の結果を有し得るシステム）において、状態の複製は特に困難な問題をはらんでいる。

【０００７】

高可用性のソフトウエアアプリケーションは冗長環境上に配備されることが多く、基盤となるハードウエアと共通する障害の一点を低減、および／または、除去する。２つの一般的なアプローチとしては、ホットフェイルオーバおよびウォームフェイルオーバが知られている。ホットフェイルオーバとは、複合システムにおいて同じ入力を同時に処理し、基本的に、これらシステムのうちの一つにおける障害のイベントにおいて完全な冗長化を提供することを指す。ウォームフェイルオーバとは、バックアップシステムにおけるアプリケーション（すなわち、データ）の状態を複製することを指す。このとき、バックアップシステムにおいて当該データの複製処理を行うことはないが、一次システムにおける障害のイベントにおいて、読み込み済みでスタンバイ状態のデータを処理することができるアプリケーションを有する。コールドフェイルオーバとは、単にバックアップシステムを立ち上げ、一次システムからの処理負担を想定してバックアップシステムを準備することを指すが、一般に高可用性を有するとは考えられていない。

【０００８】

ホットフェイルオーバ設定におけるアプリケーションの二つの例としては、同時に、２種類のハードウエア機器において実行するものと、同じ入力のコピーを処理するものとがある。そのうちの一つが重大な障害を起こした場合、補助同期システムは、もう一方が継続して作業負荷をサポートするように確保することができる。ウォームフェイルオーバ設定において、システムのうちの一つ、すなわち、一次に指定されたシステムが、アプリケーションを起動させる。障害の場合は、スタンバイ状態で待機している二次システム、すなわちバックアップに指定されたシステムが起動し、引き継いで上記機能性を再開する。

【０００９】

ホットフェイルオーバアプローチに関する従来技術は、少なくとも２つの欠点を有する。第一に、２つのシステムの同期を維持するために、補助ソフトウエアを起動する必要がある。非決定性システムの場合では、この同期化のための注力によって、イベントの発生順序の同一性を保証するというパフォーマンスおよび高度性において、許容範囲外の（または、不要な）低下を招き得る。また、そのようなアプリケーションにおいて使用される従来技術に関する並行システムによって、通常は、多重スレッドが同時に実行できるようになるので、当該スレッドは本来的に非決定性を有する。また、非決定性は、サーバおよび地理的に離れたクライアントを含むシステムであり、当該システムにおいて、可変ネットワーク遅延によって、不測のシーケンスでサーバにメッセージが分配される。

【００１０】

ホットフェイルオーバに関する問題を克服するのにウォームフェイルオーバを使用することができる。ウォームフェイルオーバは、冗長性のバックアップシステムにシステムデータを複製し、二次システムにアプリケーション機能性を保存することによって、非決定性システムのフェイルオーバを実行する他の方法であり得る。このアプローチは、まず安定状態にデータを修復し、それからアプリケーションを機能状態にし、最終的に中断した処理地点にアプリケーションを戻すのに要する時間という点で欠点を有する。この処理には通常数時間かかり、手動介入が必要であり、一般的にはインフライトトランザクション（in-flight transaction）を復旧させることができない。

【００１１】

多数の特許が、上記問題のうちの少なくとも一部に対処しようとしている。US特許第５

，３０５，２００では、要約すると、買い手／売り手と販売人（値付け業者）との間の協議による貿易シナリオにおけるコミュニケーションのための非冗長性機構について提案している。障害発生のイベントにおいて非冗長性機構の作業を確保するために、冗長性が提供される。上記特許は、非決定性環境におけるオンライン・トランザクション・アプリケーションのフェイルオーバに対処していない。簡単に言うと、US特許第５，３０５，２００は、ネットワーク障害が起こった後で「命令が送信されたか、されていないか」という問いに対する明確な答えを提供することを目的としている。

【００１２】

US特許第５，３８１，５４５では、データのアップデートをしている間に、（データベース内に）保存されたデータをバックアップするための技術を提案している。US特許第５，９８７，４３２は、地域的分布に関する世界規模の金融市場データを収集するための耐障害性市場データ相場表示装置システム（fault-tolerant market data ticker plant system）に対処している。これは決定性環境であって、解決策は、データを消費者に送信する連続した一方向のフローを提供することに焦点を当てている。US特許第６，１５４，８４７は、従来の不揮発性記憶装置上のトランザクションログと揮発性記憶装置内のトランザクションリストとを組み合わせることによって、トランザクションのロールバックを行うことについての改善方法を提供する。US特許第６，１９９，０５５は、無認証（unsecured）のコミュニケーションリンクを介して、システムとポータブルプロセッサとの間のトランザクション分配を扱うための方法を提案している。US特許第６，１９９，０５５は、遠隔装置との完全なトランザクションを確保する認証について扱っており、障害のイベントにおける遠隔装置のリセットについて扱っている。概して、上記は非決定性環境におけるオンライン・トランザクション・アプリケーションのフェイルオーバに対処していない。

【００１３】

US特許第６，２０２，１４９は、タスクを自動的に再分配してコンピュータ停止の影響を低減させるための方法および装置を提案している。装置は、１つ以上のコンピュータシステムからなる少なくとも１つの冗長性グループを含んでおり、当該コンピュータシステムはコンピュータパーティションからなる。パーティションは、各コンピュータシステムパーティションにおいて複製されたデータベーススキーマのコピーを含む。冗長性グループはコンピュータシステムおよびコンピュータシステムパーティションの状態をモニタし、モニタされたコンピュータシステムの状態に基づいて当該コンピュータシステムにタスクを割り当てる。US特許第６，２０２，１４９の問題は、バックアップシステムが処理トランザクションの負荷を想定する場合のワークフロー復旧方法について教示しておらず、代わりに、非効率的および／または低速になり得る、全データベースの複製を目的としている点である。さらに、そのような複製によって、インフライトにおける重要なトランザクション情報が失われる可能性がある。特に、一次システムにおいて障害が発生するか、または、一次とバックアップシステムとを相互接続させるネットワークにおいて障害が発生することによって、一次とバックアップとの間で非一貫性（inconsistent）状態が生じ得る。概して、US特許第６，２０２，１４９は、オンライントランザクションなどの処理に必要な特徴、特に、非決定性システムのフェイルオーバに必要とされる特徴を欠いている。

【００１４】

US特許第６，３０８，２８７は、機器トランザクションの障害を検知し、障害を除去し（back out）、システム障害後に回復可能になるように障害表示器を確実に保存し、この障害表示器がさらなるトランザクションに対して可用性を有するようにするための方法を提案している。当該特許は非決定性環境におけるトランザクションアプリケーションのフェイルオーバに対処していない。US特許第６，５７４，７５０は、分配および複製されたオブジェクトに関するシステムを提供しており、この場合、オブジェクトは非決定性を有する。当該特許は、複製されたオブジェクトにおいて障害が発生した場合に、一貫性を保証するとともにロールバックを制限するための方法を提案している。この方法には、オブ

ジェクトが着信中のクライアント要求（incoming client request）を受信すること、および、要求IDと、オブジェクトのレプリカによって前処理された全要求のログとを比較することについて記載されている。適合がわかると、関連のレスポンスがクライアントに返ってくる。しかし、この方法だけでは、従来技術における様々な問題を解決するには不十分である。

【００１５】

他の問題として、US特許第６，５７４，７５０は同期起動チェーン（synchronous invocation chain）を想定しており、当該チェーンは高性能オンライントランザクション処理（OLTP）アプリケーションには不適であることが挙げられる。同期起動の場合、クライアントは応答またはタイムアウトのいずれかを待ってから続行する。続いて、起動したオブジェクトが他のオブジェクトのクライアントとなって同期コールチェーンを伝搬させてもよい。結果は、広域同期オペレーションとなり得るものであり、このとき、発信元クライアントにおいて構成されたロングタイムアウトを処理および必要とするクライアントをブロックする。

【発明の概要】

【００１６】

本発明の形態は、ネットワークを介して相互接続された少なくとも２つのサーバのうちのいずれか１つを選択可能であって、選択した方と接続する少なくとも１つのクライアントを備える、フェイルオーバのためのシステムを提供する。通常状態において、サーバのうちの１つがクライアントに接続される場合に当該サーバが一次サーバに指定され、もう一方のサーバがクライアントに接続されない場合に当該もう一方のサーバがバックアップサーバに指定される。少なくとも１つのクライアントが一次サーバにメッセージを送信するように構成されている。サーバは、各サーバにおいて同一の少なくとも１つのサービスを用いてメッセージ処理を行うように構成されている。サービスは、サービスに関するサーバが一次サーバとして動作しているのか、バックアップサーバとして動作しているのか認識しない。サーバは、ライブラリまたは他の固有の有効コードを維持するよう構成されており、当該ライブラリまたは有効コードは、サーバが一次サーバであるかバックアップサーバであるかの表示を含む様々なタスクを実行するように構成されている。各サーバ内部にあるサービスは、その個々のライブラリに対して外部呼び出しを行うものである。一次サーバにおけるライブラリは、外部呼び出しを完了し、一次サーバにおけるサービスに外部呼び出しの結果を戻し、バックアップサーバにおけるサービスに外部呼び出しの結果を転送するように構成されている。二次サーバにおけるライブラリは、二次サーバにおけるサービスに要求されたとき、外部呼び出しを行わず、単に、一次サーバから受信したとおりに外部呼び出しの結果を二次サーバにおけるサービスに転送する。

【００１７】

ライブラリは、１つ以上の固有の有効コードとして実施され得る。

【００１８】

サーバは、サービスがメッセージ処理の結果を保存することが可能な共有リソースを維持するようにそれぞれ構成されている。性能上の理由から、個々のサーバにおけるランダムアクセスメモリ内に共有リソースを維持することが好ましい。しかし、必ずしもランダムアクセスメモリ内に共有リソースを維持する必要はない。

【００１９】

外部呼び出しは、（例示の非制限リストとして）タイムスタンプを要求するか、または、同サーバ上に配置された他のサービスを呼び出すか、または、物理的に離れた機械上に配置された別のサービスを呼び出すかであり得る。

【００２０】

システムは電子取引システムの一部であり得るので、メッセージはセキュリティの売買注文であり得る。この場合、外部呼び出しは、セキュリティの値についての市場供給相場を要求するものであり得る。システムが電子取引システムである場合、少なくとも１つのサービスが、注文受付サービス；注文取消サービス；注文変更サービス；注文適合サービ

ス；予め実行された取引を行うためのサービス；またはクロス取引を行うためのサービスのうちのいずれか１つを含み得る。

【００２１】

一次サーバにおけるサービスは、外部呼び出しが完全にバックアップサーバに転送されたことをバックアップサーバが認証した場合に限って、メッセージの処理が行われたことをクライアントが認証するように構成され得る。

【００２２】

一次サーバにおけるサービスは、外部呼び出しがバックアップサーバに完全に転送されたという結果をバックアップサーバが認証するかどうかに関わらず、メッセージの処理が行われたことをクライアントが認証するように構成され得る。バックアップサーバが、所定の期限内に外部呼び出しが当該バックアップサーバに完全に転送されたという結果を認証しない場合、一次サーバは、バックアップサーバにおいて障害が発生したとみなし得る。

【００２３】

本発明の他の形態は、ネットワークを介して相互接続された少なくとも２つのサーバのうちのいずれか１つを選択可能であって、選択した方と接続する少なくとも１つのクライアント；クライアントに接続される場合に一次サーバに指定されるサーバのうちの１つ、および、クライアントに接続されない場合にバックアップサーバに指定されるもう一方のサーバ；および、一次サーバにメッセージを送信するように構成された少なくとも１つのクライアントを有する、システムにおけるフェイルオーバのための方法である。

【００２４】

当該方法は、サーバの各々と同一で、かつ、サービスに対応するサーバが、一次サーバかバックアップサーバのいずれで動作しているかを認識しない、少なくとも１つのサービスを用いてメッセージを処理するようにサーバを構成する工程、
サーバが一次サーバであるか、バックアップサーバであるかを表示するライブラリを維持するようにサーバを構成する工程、
その各ライブラリに対して外部呼び出しを行うようにサービスを構成する工程、および、外部呼び出しを完了させて一次サーバにおけるサービスに外部呼び出しの結果を戻し、バックアップサーバにおけるサービスに外部呼び出しの結果を転送するように、一次サーバにおけるライブラリを構成する工程を含む。

【００２５】

本発明の他の形態は、選択された少なくとも１つのクライアントと接続可能なネットワークを介して相互接続された少なくとも２つのサーバのうちのいずれか１つのサーバ上で実行可能な一連のプログラミングインストラクションを保存するコンピュータ読み取り可能な記録媒体であって、当該相互接続されたサーバが、少なくとも１つのクライアントと選択的に接続可能である記録媒体を提供する。サーバのうちの１つがクライアントと接続されたときに、当該サーバが一次サーバに指定され得、このとき、クライアントに接続されていないもう一方のサーバが、バックアップサーバに指定され得る。少なくとも１つのクライアントが一次サーバにメッセージを送信するように構成されている。プログラミングインストラクションは、
サーバの各々と同一で、かつ、サービスに対応するサーバが一次サーバまたはバックアップサーバとしてのいずれで動作するかを認識しない、少なくとも１つのサービスを用いてメッセージを処理するようにサーバを構成するためのインストラクション、
サーバが一次サーバであるか、バックアップサーバであるかを表示するライブラリを維持するようにサーバを構成するためのインストラクション、
その各ライブラリに対して外部呼び出しを行うようにサービスを構成するためのインストラクション、および、
外部呼び出しを完了させて外部呼び出しの結果を一次サーバにおけるサービスに戻し、バックアップサーバにおけるサービスに外部呼び出しの結果を転送するようにライブラリを構成するためのインストラクション、を含む。

【図面の簡単な説明】
【００２６】
本発明は、添付の図面を参照し、例示することによってのみ説明される。
【図１】本発明の実施形態に係るフェイルオーバのためのシステムの概略図である。
【図２】通常状態で動作する場合の図１のシステムの概略図であって、システムにおけるサービスを実行する様々なソフトウエア部材の例示の詳細を含む。
【図３】本発明の他の実施形態に係る通常状態でのフェイルオーバのためのシステムの起動方法を示すフローチャートである。
【図４】図３の方法に関するパフォーマンス中の図２のシステムを示す。
【図５】図３の方法に関するパフォーマンス中の図２のシステムを示す。
【図６】図３の方法に関するパフォーマンス中の図２のシステムを示す。
【図７】図３の方法に関するパフォーマンス中の図２のシステムを示す。
【図８】図３の方法に関するパフォーマンス中の図２のシステムを示す。
【図９】図３の方法に関するパフォーマンス中の図２のシステムを示す。
【図１０】図３の方法に関するパフォーマンス中の図２のシステムを示す。
【図１１】図３の方法に関するパフォーマンス中の図２のシステムを示す。
【図１２】図３の方法に関するパフォーマンス中の図２のシステムを示す。
【図１３】図３の方法に関するパフォーマンス中の図２のシステムを示す。
【図１４】図３の方法に関するパフォーマンス中の図２のシステムを示す。
【図１５】本発明の他の実施形態に係るフェイルオーバのための方法を示すフローチャートである。
【図１６】本発明の他の実施形態に係る一次単独（primary-only）状態で動作するサーバのうちの一つについての図２のシステムを示す。
【図１７】本発明の他の実施形態に係る一次単独（primary-only）状態で動作するサーバのうちのもう一方についての図１６のシステムを示す。
【図１８】本発明の他の実施形態に係る一次単独状態でサーバのうちの一つを動作させるための方法を示すフローチャートである。
【図１９】通常状態から、本発明の他の実施形態に係る一次単独状態において動作するバックアップサーバまでフェイルオーバするための方法を示すフローチャートである。
【発明を実施するための形態】
【００２７】
ここで図１を参照すると、フェイルオーバのためのシステムは主に符号５０で示されている。システム５０は複数の遠隔クライアント５４－１および５４－２を備える（ここで言う「クライアント５４」とは、総じて集合的な「クライアント５４」を指す。図面における他の部材を指すときも同様に用語を定義する）。クライアント５４をネットワーク５８に接続する。ネットワーク５８は、インターネットやローカルエリアネットワーク、広域ネットワークまたはその組合せなど、あらゆるタイプのコンピュータネットワークであり得る。続いて、第一サーバ６２－１および第二サーバ６２－２にネットワーク５８を接続する。その結果、下記に詳述するように、クライアント５４がネットワーク５８を介してサーバ６２－１および６２－２にそれぞれ交信することができる。
【００２８】
クライアント５４は、個々のクライアント５４を用いてサーバ６２－２にリクエストを送信する個人および／または団体にそれぞれ属する。便宜上、そのような個人または団体をここでは取引者Tとし、クライアント５４－１を使用するものを取引者T－１、クライアント５４－２を使用するものを取引者T－２とする。各クライアント５４は、典型的には、キーボードおよびマウス（または他の入力装置）と、モニタ（または他の出力装置）ならびに上記キーボード、マウス、およびモニタに接続されたデスクトップモジュールとを有するパーソナルコンピュータなどのコンピュータデバイスであって、１つ以上の中央演算処理装置と、揮発性メモリ（すなわち、ランダムアクセスメモリ）と、非揮発性メモリ（すなわち、ハードディスクデバイス）と、ネットワーク５８を介してクライアント５４

の交信を可能にするネットワークインターフェイスとを内蔵するパーソナルコンピュータなどのコンピュータデバイスである。しかし、クライアント５４は、電子手帳、携帯電話、ノート型パソコン、e-mailページング装置などのあらゆるタイプのコンピュータデバイスであり得ることを理解されたい。

【００２９】

サーバ６２は、UNIX（登録商標）オペレーティングシステムを実行する、カリフォルニア州パロアルト（Palo Alto Calif）に住所を有するサンマイクロシステムズ株式会社（Sun Microsystems Inc.）製のSun Fire V480など、クライアント５４からメッセージを受信および処理するように動作可能なあらゆるタイプのコンピュータデバイスであり得る。さらに、サーバ６２は、約９００メガヘルツで各々動作する４つの中央演算処理装置、および、４ギガバイトのランダムアクセスメモリおよびハードディスクドライブなどの非揮発性記憶装置を有する。サーバ６２に適する他のタイプのコンピュータデバイスとしては、アメリカ合衆国　コロラド州　８０５３７，ラブランド，サウス　タフト　８００（800 South Taft，Loveland，CO 80537）に住所を有するヒューレットパッカード株式会社（Hewlett-Packard Company）製のHP ProLiant BL25pサーバが挙げられる。しかし、これらの具体的なサーバは単なる例示であり、サーバ６２－１および６２－２について他の様々なタイプのコンピュータ環境が本発明の範囲内であることを強調したい。サーバ６２－１によって受信および処理されるメッセージのタイプは特に限定されていないが、本実施形態において、サーバ６２－１はオンライン取引システムを実行するので、オンラインで取引可能なセキュリティを購入する、売る、取り消すなどのリクエストを含むメッセージを処理することができるものである。より具体的には、サーバ６２－１は中央適合エンジン（不図示）を維持するように動作可能であり、そこで互いに対してリクエストを実行し、さらに、注文を行う中央保存装置に対してリクエストを実行し、セキュリティ取引を処理する。

【００３０】

サーバ６２－２は、通常、サーバ６２－１と同一の（または、少なくとも実質的に同一の）コンピュータ環境を有する。下記にさらに説明するように、ハードウエア、オペレーティングシステム、アプリケーションなどを含むコンピュータ環境が、サーバ６２－２として選択される。このとき、当該サーバ６２－２は、サーバ６２－１の障害発生時にサーバ６２－１の機能性を代替するように動作可能である。

【００３１】

システム５０はまた、サーバ６２－１とサーバ６２－２とを相互接続する複製リンク７８を備える。本実施形態において、複製リンク７８それ自体がメインリンク８２とフェイルセーフリンク８６とを備えており、サーバ６２－１とサーバ６２－２との間の交信において、より高いロバスト性を有する。

【００３２】

一次サーバ６２－１、バックアップサーバ６２－２、および、複製リンク７８の機能についてのさらなる詳細、ひいては、サーバ６２－１および６２－２を実行するのに使用可能な多種のハードウエアについて、下記の説明で明らかとなるだろう。

【００３３】

図２において、サーバ６２－１およびサーバ６２－２をより詳細に示す。また、図１においてシステム５０における様々な物理的な接続を実線で表すのに対し、図２においてはシステム５０における様々なネットワーク上の接続を点線で示す。したがって、図２に示されたような接続は、サーバ６２－１が一次サーバに指定されるとともに、サーバ６２－２がバックアップサーバに指定されている通常状態で動作するシステム５０を表すことを意図している。当該通常状態において、一次サーバ６２－１がクライアント５４からのリクエストに応答する。通常状態、および、システム５０が動作可能な他の状態についての更なる詳細を以下に記載する。

【００３４】

さらに図２を参照すると、サーバ６２－１およびサーバ６２－２がそれぞれ複数のソフ

トウエア構成要素を含んでおり、当該ソフトウエア構成要素は個々のハードウエア環境で実行され、クライアントからのリクエストに応答し、フェイルオーバ機能性を有する。

【００３５】

サーバ６２－１およびサーバ６２－２は、フェイルオーバエージェント９０－１および９０－２をそれぞれ備える。フェイルオーバエージェント９０は相互交信し、リンク７８および相互のインテグリティを定期的にテストするように動作可能である。本実施形態において、通常状態では、フェイルオーバエージェント９０－１が、フェイルオーバエージェント９０－２にキープアライブ信号（例えば、「アライブ状態になっているか？」）を定期的に送り、フェイルオーバエージェント９０－２が定期的に応答する（例えば、「はい、アライブ状態です」）とされる。フェイルオーバエージェント９０－２がそのようなリクエストに応答するとし、さらに、一次サーバ６２－１が通常通り動作し続けるとすると、システム５０は図２に示した通常状態を維持する。したがって、フェイルオーバエージェント９０－１はまた、サーバ６２－１における他のソフトウエア構成部材と交信し、通常状態が有効であることを示すように動作可能である。

【００３６】

フェイルオーバ９０が、適切に、または、所望のように、リンク７８を備えるメインリンク８２とフェイルセーフリンク８６の両方を利用するよう動作可能であることは明白である。この場合、メインリンク８２とフェイルセーフリンク８６のうちの少なくとも１つが動作している限り、システム５０は通常状態を維持する。

【００３７】

サーバ６２は、１つ以上のクライアント５４からの多種の要求を受信および処理することができる１つ以上のサービスをそれぞれ含む。サービスのタイプは特に限定されておらず、フェイルオーバ保護が必要とされるあらゆるタイプのサービス、アプリケーション、または処理などを含み得る。システム５０がオンライン取引システムであるという、この単なる例示の実施形態において、サーバ６２は注文受付サービス９４および注文取消サービス９８をそれぞれ含む。注文受付サービス９４とは、その名前からもわかるように、特定のセキュリティの売り注文または買い注文を受け付けるためにクライアント５４からの要求を受信するように構成されている。注文取消サービス９８とは、その名前からもわかるように、特定のセキュリティの売り注文または買い注文を取り消すためにクライアント５４からの要求を受信するように構成されている。このとき、当該セキュリティは、サービス９４によってすでに受け付けられているが、当該特定の注文が実際に実行される前のものである。電子取引の分野の当業者であれば想定し得る実行可能な他のタイプのサービスは、これらに制限されるものではないが、注文適合、注文変更、取引開始、または、クロス開始を含む。本実施形態において、要件ではないものの、サービス９４およびサービス９８はマルチスレッドである。（ここで使用するとき、マルチスレッドとは限定的な意味で使用するものではなく、複合メッセージが同時に処理されるという並行処理の様々な形態を指す。これは、さらに、システムの非決定性特性の一因となる。例えば、複合処理、または、単一処理を用いた実行のマルチスレッドを用いて、マルチスレッドが実行され得る。）

サーバ６２はまた、各々に付属する対応のサービス９４およびサービス９８にアクセス可能なライブラリ１０２をそれぞれ有する。各ライブラリ１０２はシーケンサ１０６およびキャッシュメモリ１１０を有する。下記に詳述するように、シーケンサ１０６が、シーケンスナンバを発生させ、ライブラリ１０２に関してサービス９４または９８からの要求に応答する。シーケンサ１０６－２は通常状態で非アクティブであり、そのような非アクティブ性は、オーバル型シーケンサ１０６－２を介してハッシュすることによって図２に示されている。（既定の特定状態において構成要素がアクティブであるか、非アクティブであるかを表示する他の構成要素において、ハッシュが使用される。）キャッシュメモリ１１０は、ライブラリ１０２によってなされる外部機能呼び出し結果の保存領域である。

【００３８】

各ライブラリ１０２はまた、システム５０が並列処理を行う状態を維持する状態記録装

置（state register）１１４を有する。当該状態記録装置１１４は、システム５０が並行処理している状態を適合するために、その各フェイルオーバエージェント９０と絶えず交信する。図２において、システム５０は通常状態で作動する。それに従って、サーバ６２－１が現時点で一次サーバに指定されていることを状態記録装置１１４－１が表示し、一方で、サーバ６２－２が現時点でバックアップサーバに指定されていることを情報記録装置１１４－２が表示する。しかし、下記に詳述するように、システム５０の状態は、システム５０における様々な構成材の動作状態によって変化し得る。

【００３９】

各サーバ６２はまた、サービス９４および９８に代わって外部リソースに外部呼び出しを行う機能を果たす外部リソースエージェント１１８を備えているが、当該呼び出しはライブラリ１０２を経由して行われる。外部リソースは、サービス９４および９８に外付けのリソースであって、各サーバ６２上に設置されているリソースを備える。当該外部リソースは、各サーバに外付けの、オペレーティングシステムクロック（不図示）からのタイムスタンプ、および／または、リソースなどであり、電子取引システムの場合では、マーケットフィード（不図示）などである。当該マーケットフィードは、注文受付サービス９４を経由して受け付けられた売り注文または買い注文のサブジェクトであり得る様々なセキュリティの市場価格についての最新情報を維持するものである。ここで、当業者であれば、サービス９４および９８が行うそのような外部リソースに対する呼び出しは、システム５０の有する非決定性特性の一因となることがわかるだろう。通常状態において、外部リソースエージェント１１８－１のみがアクティブであり、外部リソースエージェント１１８－２は非アクティブである。オーバル型の外部リソースエージェント１１８－２を介してハッシュすることによって、外部リソースエージェント１１８－２の非アクティブ性が図２に示される。

【００４０】

各サーバ６２はまた、共有リソース１２２を維持しており、当該共有リソース１２２は、サービス９４および９８によって実行された処理ステップの結果、および／または、サービス９４および９８がアクセス可能であることを要求するデータを維持する。例えば、本発明に係る電子取引システムにおいて、共有リソース１２２が注文票を維持する。当該票は、サービス９４によって受け付けられた注文の記録に過ぎない。それゆえ、注文サービス９４は、例えば、共有リソース１２２において買い注文についての記録を作成してもよい。後に、注文を取り消してそれを共有リソース１２２に表示するために、注文取消サービス９８が上記買い注文にアクセスする必要性が生じる可能性もある。同様に、サーバ６２上で実行される適合サービス（不図示）が上記買い注文にアクセスする必要性が生じる可能性もある。当該適合サービスは、上記買い注文と、適する対応の売り注文とを市場法則に従って適合させ、上記買い注文および売り注文をアップデートし、適合が有効で取引が完了可能であることを示す。

【００４１】

各サーバ６２はまた、複製エージェント１２６を維持する。通常状態において、複製エージェント１２６－２のみがアクティブであって、複製エージェント１２６－１は非アクティブである。オーバル型エージェント１２６－１を介してハッシュすることによって、複製エージェント１２６－１の非アクティブ性が図２に示されている。下記に詳述するように、アクティブ複製エージェント１２６が、カウンターパートサーバ６２におけるライブラリ１０２と交信し、一次サーバからバックアップサーバへの情報のミラーリングを助ける。

【００４２】

ここで図３を参照すると、本発明の他の形態に係る、通常状態の間に要求を処理するための方法は、概して符号３００に示されている。上記方法の説明を容易にするために、図２に示された、通常状態でのシステム５０を用いて方法３００を実行すると想定する。さらに、方法３００についての詳細な説明によって、システム５０およびその様々な構成材をよりよく理解できるだろう。しかし、便宜上の理由のみであるが、方法３００の様々な

処理ステップが、システム５０の特定の構成部材内で起こるように図３に示されている。そのような表示は、限定的な意味で解釈すべきものではない。しかし、当然のことながら、システム５０および／または方法３００は異なったものである可能性があり、本発明において互いに組み合わせて説明されたように機能する必要はなく、さらに、方法３００におけるステップは、図示されたような注文において実施される必要はない。そのような変形は本発明の範囲内である。そのような変形は、ここに説明された他の方法およびシステム図にも適用される。

【００４３】

まず、クライアントからメッセージを受信するステップ３１０から始める。メッセージタイプは特に限定されておらず、概して、サーバ上で実行するサービスのうちの１つについてのインプットの想定タイプを補完するものである。システム５０上で実行されるとき、メッセージは、注文受付サービス９４のためのインプットを意図する売り注文または買い注文か、または、注文取消サービス９８のためのインプットを意図する取消注文であり得る。例えば、取引者T－１による買い注文が、クライアント５４－１からのメッセージの内に受信され、当該メッセージが、ネットワーク５８を介して注文サービス９４－１に送信されることを想定する。この場合、ステップ３１０によると、注文受付サービス９４－１が当該メッセージを受信する。図４に、ステップ３１０の例示の性能を示す。メッセージM（O_1）はクライアント５４－１から発信され、注文受付サービス９４－１においてサーバ６２－１内に受信されるように示されている。表１は、注文受付メッセージM（O_1）の例示のフォーマットを示す。

【００４４】

【表１】

メッセージM(O_1)

フィールド番号	フィールド名	例示の内容
1	取引者	取引者 T-1
2	セキュリティ名	ABC Co.
3	トランザクションタイプ	買い
4	量	1,000 ユニット

【００４５】

より具体的には、取引者と名付けられた表１のフィールド１は、メッセージM（O_1）の発信元となる取引者が取引者T－１であると認識する。セキュリティネームと名付けられた表２のフィールド２は、取引のサブジェクトである特定のセキュリティの名前を、この例では、ABC Co.と認識する。トランザクションタイプと名付けられた表１のフィールド３は、フィールド２で認識されたセキュリティが買い注文なのか、売り注文なのか、などを認識する。この例では、トランザクションタイプは買い、であり、これは、買い注文であることを示している。量と名付けられた表１のフィールド４は、必要なセキュリティ量を認識する。この例では、量は１，０００ユニットであり、ABC Co.の１，０００ユニット買いを意図する。ここで、当業者であれば、注文の価格が、フィールド２のセキュリティに関する現行市場価格を基礎としていることから、表１の注文が市場注文であることがわかるだろう。

【００４６】

ステップ３１０でメッセージを受信した後、方法３００はステップ３１５に進み、この時点で、関連するサービスが、メッセージのさらなる処理に使用される外部データに対して、いずれかの呼び出しを行う。引き続き上記例でみていくと、ステップ３１５において、注文受付サービス９４－１が、そのような外部呼び出しを一次ライブラリ１０２－１に対して行う。この例では、そのような呼び出しは下記ｉ）、ｉｉ）を目的とすると想定さ

れている。
ｉ）注文が受信された時間を認識するメッセージＭ（0_1）において注文に割り当てられるタイムスタンプ、
ｉｉ）メッセージＭ（0_1）における注文で認識されるセキュリティに対する現行市場価格。
【００４７】
図５に、ステップ３１５のパフォーマンスを点線で示す。点線１３０は、注文受付サービス９４－１から一次ライブラリ１０２－１への呼び出しを示す。
【００４８】
続いて、ステップ３２０において、一次ライブラリ１０２－１が呼び出しを行う。一次ライブラリ１０２－１はフェイルオーバエージェント１１４－１から情報を得て（consult）、サーバ６２－１が一次サーバに指定されており、かつ、システム５０が通常状態であることを認証する。上記認証後、下記ｉ）、ｉｉ）によって、サービス９４－１が行う呼び出しに対して、一次ライブラリ１０２－１が応答する。
ｉ）タイムスタンプを得るために、外部リソースエージェント１１８－１に対して外部呼び出しを行う。
ｉｉ）現行市場価格を得るために外部リソースエージェント１１８－１にさらなる外部呼び出しを行う。
【００４９】
その結果、ステップ３２５において、外部リソースエージェント１１８－１がタイムスタンプを得るためにシステムクロック（不図示）に外部呼び出しを、現行市場価格を得るために市場供給（不図示）に外部呼び出しを、それぞれ行う。
【００５０】
図６に、ステップ３２０およびステップ３２５のパフォーマンスを点線で示す。点線は、外部リソースエージェント１１８－１を介するタイムスタンプのための呼び出しを１３２に、外部リソースエージェント１１８－１を介する市場価格のための呼び出しを１３４に、それぞれ示す。
【００５１】
ここで、当業者であれば、外部呼び出し１３２および１３４によって、特にシステム５０に非決定性特性を有するようになることがわかり、それゆえ、フェイルオーバの発生において、復旧中に、当該復旧が取引者Ｔに自明であるように、システムにおける非決定性特性に対処しているフェイルオーバシステムを提供するという点で、本発明独自の取り組みが理解できるだろう。（さらなる説明によって、両サーバ６２が各メッセージに対する呼び出しを行うようにシステム５０が変更されたと想定される。しかし、既定のメッセージＭのいずれに関しても、市場の公正さを確保するためには、タイムスタンプに対する呼び出しが行われるその瞬間が非常に重要であって、両サーバ６２が、同時に同じメッセージのためにタイムスタンプに対して呼び出しを行う確率は非常に低いと考えられる。それゆえ、各サーバ６２が同じメッセージＭについて異なる時間優先を割り当てることが可能であり、その結果、同じ機械処理の結果が異なることになる。フェイルオーバの間、各サーバ６２が一貫したビジネスデータを有することはなく、フェイルオーバは無意味であり得る。）さらに読み進めると、当業者であれば、そのようなチャレンジに対する対処法および本発明の他の形態が理解できるだろう。
【００５２】
ステップ３３０において、外部呼び出し１３２および外部呼び出し１３４の結果が一次ライブラリ１０２－１に戻ってくる。ステップ３３５において、キャッシュメモリ１１０－１内に呼び出し１３２および１３４の全ての結果が保存され、サーバ９４－１に戻される。
【００５３】
引き続き上記例でみていくと、呼び出し１３２の結果はタイムスタンプ２０００年１月５日午後１２時であると考えられる。さらに、呼び出し１３４の結果は＄２．００の市場

価格であると考えられる。表２および図７に、キャッシュメモリ１１０－１におけるこれらの結果の保存を示す。

【００５４】

【表２】

ステップ335の後の、キャッシュメモリ110-1における例示の内容

記録番号	フィールド番号	フィールド名	例示の内容
1	1	メッセージ	$M(O_1)$
1	2	タイムスタンプ	2000年1月5日午後12時
1	3	市場価格	$2.00

【００５５】

ステップ３４０で、サービスが呼び出し結果を受信する。上記例で続けると、図７にも示されているように、表２で保存された呼び出し結果はサービス９４－１に戻される。

【００５６】

続いて、ステップ３４５において、サービスが共有リソースのための要求を行う。上記例において、サービス９４－１がライブラリ１０２－１に対して要求を行う。次に、ステップ３５０において、ライブラリ１０２－１が共有リソース１２２－１に対して、当該リソース１２２－１をロックするように指示を発行し、他のあらゆるサービス（例えば、サービス９８－１またはサービス９４－１内の他のスレッド）が共有リソース１２２－１にアクセスしないようにする。（下記に詳述するように、共有リソース１２２－１がすでにロックされている場合、共有リソース１２２－１が解除されるまで、方法３００はステップ３４５で停止する。）ステップ３４５およびステップ３５０のパフォーマンスを図８に点線で示す。このとき、共有リソース要求を符号１４０で示す。共有リソース１２２－１のロックをパッドロック１３８で表す。

【００５７】

続いて、ステップ３５５において、共有リソースシーケンスナンバーが戻される。シーケンサ１０６－１を利用するライブラリ１０２－１によって、このステップが実行され得る。当該シーケンサ１０６－１は、メッセージＭ（0_1）に関するシーケンスナンバーを生成する。上記例で続けると、シーケンスナンバー１が生成されると考えられる。表３および図８に、キャッシュメモリ１１０－１内の結果の保存を示す。表３は表２をアップデートしたものであることに留意されたい。

【００５８】

【表３】

ステップ355の後の、キャッシュメモリ110-1における例示の内容

記録番号	フィールド番号	フィールド名	例示の内容
1	1	メッセージ	$M(O_1)$
1	2	タイムスタンプ	2000年1月5日午後12時
1	3	市場価格	$2.00
1	4	シーケンスナンバ	1

【００５９】

続いて、ステップ３６０において、複製が要求される。サービス９４－１が上記例におけるステップ３６０を実行し、当該サービス９４－１がライブラリ１０２－１に複製実行の指示を送信する。ステップ３６５において、メッセージ、呼び出し結果およびシーケン

スナンバーの複製が開始される。上記例において、ライブラリ１０２－１が表３の内容を複製する。ステップ３６５については後に説明する。

【００６０】

ステップ３７０において、呼び出し結果およびロックされた共有リソースを用いてメッセージを処理する。上記例において、サービス９４－１がステップ３７０を実行する。サービス９４－１は、表３の結果を生み出すために、表３の内容を用いて、サービス９４－１に関する処理ステップを実行する。サービス９４－１は注文受付サービスであり、メッセージＭ（0_1）は買い注文を示すので、ステップ３７０において、サービス９４－１は、共有リソース１２２－１に記録される買い注文を作り出し、例えば取引者Ｔ－２からの買い注文に対して買い注文を続けて適合させるか、または、サービス９８－１を用いた注文の取消など他の取引処理を行う。

【００６１】

上記例の目的に関して、メッセージＭ（0_1）が適合し得る共有リソース１２２－１に注文がない場合、ステップ３７０の結果は、単に、メッセージＭ（0_1）に関する買い注文の詳細についての完全な記録を作り出すことである。表４に、ステップ３７０のパフォーマンスの例示結果を示す。

【００６２】

【表４】

ステップ370におけるパフォーマンスの例示の結果

記録番号	フィールド番号	フィールド名	例示の内容
1	1	タイムスタンプ	2000年1月5日午後12時
1	2	市場価格	$2.00
1	3	シーケンスナンバ	1
1	4	取引者	取引者T-1
1	5	セキュリティ名	ABC Co.
1	6	トランザクションタイプ	買い

【００６３】

続いて、ステップ３７５において、ステップ３７０のパフォーマンス結果が共有リソースに書き込まれ、その後、共有リソースが解除される。ステップ３７０におけるサービス９４－１による表４の生成と、ステップ３７５における共有リソース１２２－１内の結果の保存を図９に示す。

【００６４】

続いて、ステップ３８０では、ステップ３７５において結果が書き込まれたこと、および、ステップ４００において複製が実行されたことについての認証を、サービスが認証する。上記例では、ステップ３８０において、サービス９４－１が、表４が共有リソース１２２－１に書き込まれたという共有リソース１２２－１からの認証を待つ。同様に、ステップ３８０では、サービス９４－１は、ステップ３６５によって開始された複製が完了したというステップ４００からの認証を待つ。ステップ３６５および４００については下記に詳述する。

【００６５】

（代替的実施例において、ステップ３８０は、ステップ３９０に進む前に、ステップ４００からの認証を実際に待つ必要はない。しかし、ステップ３８０は、ステップ４００から最終的にはそのような認証を受信することを想定しており、そのような認証が得られない場合には、ステップ３８０はサーバ６２－２が障害を起こしたとみなす。このとき、以下に説明するように、イベントサーバ６２－１が方法６００の実行を開始する。ここで、

当業者であれば、これが非同期モードの動作であり、サーバ６２－２の状態を認証する速度が早いことが好まれる特定の状況において好適である可能性があることがわかるだろう。）

続いて、ステップ３９０において、クライアントに認証が戻される。上記例において、ステップ３９０で、サービス９４－１は、取引者Ｔ－１が要求した通りにメッセージＭ（0_1）が処理されたという認証メッセージをクライアント５４－１に送信する。

【００６６】

繰り返し述べるが、方法３００におけるステップ３９０（すなわち、通常状態中の動作）は、ステップ３８０で初めて完了する。続いて、ステップ３８０は、ステップ３６５において開始される複製が完了して初めて完了する。ここで、ステップ３６５に戻ると、メッセージ、呼び出し結果、および、共有リソースシーケンスナンバーが複製される。本例では、ステップ３６０においてサービス９４－１からの要求に応答するライブラリ１０２－１が、ステップ３６５を実行する。それゆえ、ライブラリ１０２－１が表３の内容を一括し、それを複製エージェント１２６－２に伝達する。

【００６７】

図１０に、ステップ３６５、３７０、３７５、３９５、４００および３９０のパフォーマンスを示す（図１０は、図９に示されたステップ３７０および３７５のパフォーマンスに関する表示に基づく）。ステップ３６５、すなわち、ライブラリ１０２－１のキャッシュメモリ１１０－１から複製エージェント１２６－２への表３における伝達のステップを、線１４２で表す。図９に関連して上述したように、図１０にステップ３７０および３７５を示す。ステップ３９５、すなわち、メッセージ、呼び出し結果、および、共有リソースシーケンスナンバーの待機するステップを、複製エージェント１２６－２内部に表された表３のように楕円形で示す。ステップ４００、すなわち、複製エージェント１２６－２から（ライブラリ１０２－１を介して運ばれる）サービス９４－１への複製の認証に戻ることを１４４の線で示す。ステップ３９０、すなわち、サービス９４－１からクライアント５４－１までの認証に戻るステップを、点線１４６で示す。

【００６８】

上記記述で、通常状態での動作中に一次サーバが行う１つのメッセージ処理の説明は実質的に終わりである。ここで、ステップ４００を介するステップ３１０の上述に従って、一次サーバ６２－１が直列、および／または略並行かのいずれかで、複合メッセージを処理し得ることを理解すべきである。例えば、サービス９４－１が、あるメッセージＭを扱っているとき、同様に、サービス９８－１もまた、実質的に上記のように他のメッセージＭを処理することができる。このとき、ライブラリ１０２－１が両サービス９４－１、９８－１と交信する。加えて、サービス９４－１におけるあるスレッドが、あるメッセージＭを扱うとき、サービス９４－１における他のスレッドがまた、実質的に上記のように他のメッセージＭを処理することができる。このとき、ライブラリ１０２－１がサービスにおける両スレッドと交信する。ステップ３５０は、サービス９４－１と９８－１（または、そのスレッド）との間のコンテンションを避けるように共有リソース１２２－１を確実にロックし、一度に共有リソース１２２－１と交信し得るサービスが、確実にそれらサービスのうちの一つだけであるようにする。（交信とは、リーディング、ライティング、削除に留まらず、それらを含むあらゆるタイプの機能を含み得ることに留意されたい。）回避すべきコンテンションの例として、共有リソース１２２－１が、与えられた注文を取り消すためにロックされているときに、注文取消サービス９８－１が共有リソース１２２－１からの読み取り、および、リソース１２２－１への書き込みを行い、それによって、取り消された注文と適合サービス（不図示）とが適合できないようになることが挙げられる。

【００６９】

同じ特徴から、ステップ３３５がシーケンサ１０６－１を利用し、サービス９４－１または９８－１（またはそのスレッド）のいずれかがメッセージＭを扱っているのに関わらず、各メッセージに固有のシーケンスナンバーを生成する。したがって、共有リソース１

２２－１がロックされているときに、特定のサービス９４－１または９８－１（またはそのスレッド）がステップ３４５において共有リソース１２２－１への要求を行うときがあり得る。それゆえ、ステップ３４５より後に進む前に、共有リソース１２２－１が解除されるまでは特定のサービス（またはそのスレッド）は停止する。

【００７０】

通常状態で動作中である、プライマリサーバ６２－１によるメッセージ処理について記述したが、方法３００に関して、ステップ４０５およびその先のステップにおけるパフォーマンス、ならびに二次サーバ６２－２によるメッセージ処理を説明する。

【００７１】

再び図３を参照すると、ステップ４０５において、共有リソースシーケンスナンバーに従って、メッセージ、呼び出し結果およびシーケンスナンバーが発信（dispatch）する。上記例で続けると、この時点で、メッセージＭ（0_1）（すなわち、表３からの記録１におけるフィールド１の内容）をサービス９４－２に発信し、一方で、呼び出し結果（すなわち、表３からの記録１におけるフィールド２および３の内容）およびシーケンスナンバー（すなわち、表３からの記録１におけるフィールド４の内容）を二次ライブラリ１０２－２に発信する。

【００７２】

したがって、ステップ３１０Ｓでは、ステップ３１０でサービス９４－１がクライアント５４－１からメッセージＭ（0_1）を受信したのとほぼ同じ方法で、サービス９４－２が複製エージェント１２６－２からメッセージＭ（0_1）を受信する。サービス９４－２において、クライアントからメッセージＭ（0_1）を受信していた。この段階で、サービス９４－２は、全ての面においてサービス９４－１と実質的に同一であることが明白になるだろう（同様に、サービス９８－２はサービス９８－１と実質的に同一である）。サーバ６２－２におけるサービス９４－２は、サーバ６２－１におけるサービス９４－１の動作とほぼ同じ方法で動作する。換言すれば、サーバ６２－１におけるサービス９４－１がステップ３１０、３１５、３４０、３４５、３６０、３７０、３８０、および３９０を実行するのと同じ方法で、サービス９４－２がステップ３１０Ｓ、３１５Ｓ、３４０Ｓ、３４５Ｓ、３６０Ｓ、３７０Ｓ、３８０Ｓ、および３９０Ｓを実行する。サービス９４－１およびサービス９４－２の両方ともが、それらサービスが内部で動作させている特定のサーバが、一次サーバかバックアップサーバかのいずれに指定されているかを認識しない。これは、本発明の数ある利点のうちの一つを提示する。すなわち、サービスは、一括で２つ（またはそれ以上）のサーバのためのサービスを進めることができ、一次サーバに指定されたサーバのための一連のサービス、および、バックアップサーバに指定されたサーバのための一連のサービスを進める必要がないのである。

【００７３】

しかし、各ライブラリ１０２は、個々のフェイルオーバエージェント９０および状態記録装置１１４と通信する中で、その個々のサーバ６２が一次サーバかバックアップサーバのいずれに指定されているかを認識している。したがって、サービス９４－２がステップ３１５Ｓを実行して呼び出しを行うとき、ライブラリ１０２－２は外部リソースエージェント１１８－２を利用せず、ステップ４１５において、ステップ４１０でライブラリ１０２－２が受信した呼び出し結果（すなわち、表３からの記録１におけるフィールド２および３の内容）を単に戻すだけである。

【００７４】

図１１に、ステップ４０５、３１０Ｓ、４１０のパフォーマンスを示す。図１２に、ステップ３１５Ｓ、４１５、および３４０Ｓを示す。

【００７５】

同じ特徴から、サービス９４－２がステップ３４５Ｓを実行して共有リソースに要求するとき、共有リソース１２２－２をロックすること、および、ステップ４２５で共有リソースシーケンスナンバー（すなわち、表３からの記録１におけるフィールド４の内容）を戻すことによって、ライブラリ１０２－２が応答する。当該シーケンスナンバーはステッ

プ４１０においてライブラリ１０２－２が受信するものであって、シーケンサ１０６－２を利用するものではない。

【００７６】

図１３に、ステップ３４５Ｓ、４２０、４２５のパフォーマンスを示す。

【００７７】

同じ特徴から、サービス９４－２がステップ３６０Ｓを実行して複製を要求するとき、実際の複製実行によってではなく、基本的にステップ４００を模倣して、ステップ３８０Ｓにおいてサービス９４－２に対する複製認証を戻すことによって、ライブラリ１０２－２がステップ４３０において応答する。ステップ３７０および３７５と実質的に、サービス９４－２が独立して表４の内容を生成し、共有リソース１２２－２の内部に保存するようにステップ３７０Ｓおよび４３５を実行する。

【００７８】

図１４に、ステップ３７０Ｓおよび４３５のパフォーマンスを示す。

【００７９】

同様に、ステップ３８０およびステップ３９０と同じ方法で、ステップ３８０Ｓおよび３９０Ｓを実行する。但し、上記ステップは、ステップ３９０Ｓにおいて戻された認証は、クライアント５４－１ではなく、複製エージェント１２６－２に戻される。

【００８０】

当該段階、方法３００における当該パフォーマンスの結論として、表４のように、処理メッセージＭ（0_1）の結果が共有リソース１２２－１と共有リソース１２２－２との両方に保存されることがわかるだろう。ステップ３１０Ｓ、３１５Ｓ、３４０Ｓ、３４５Ｓ、３６０Ｓ、３７０Ｓ、３８０Ｓ、３９０Ｓのパフォーマンスとステップ３１０、３１５、３４０、３４５、３６０、３７０、３８０、３９０のパフォーマンスとの間の実際の待機時間は実際ごく短いこともわかるだろう。そのような待機時間は、ステップ３６５におけるネットワークの待機時間と、ステップ３９５および４０５の処理とによって決定可能であって、それらは非常に高速であり得る。あらゆるイベントにおいて、システム５０は、待機時間がハードディスクにバックアップ情報を書き込むよりも数段速くなるように構成されており、これは本発明の他の利点である。

【００８１】

したがって、サービス９４－１（および後続のサービス９４－２）を用いてセキュリティの売買注文を受け付けるためのメッセージを処理するのに、方法３００を使用することができる。同様に、サービス９８－１（および後続のサービス９８－２）を用いてこれらの注文を取り消すのに、方法３００を使用することができる。サーバ６２－１内に付加的サービスを生成および包含することができ、当該サービスのためのロバストフェイルオーバを備えるサーバ６２－２上に当該サービスを容易に配置することができる。しかし、当該サービスは、サーバ６２－１上にあるサービス用の１組のコードを必要とせず、サーバ６２－２上にあるサービスのための他の組のコードを必要とする。このとき、特定のサービスのための１組のコードとは、両サーバにとって必要なものすべてである。おそらく、ある観点からより重要なことは、システム５０が、通常はハードディスクへの書き込みに付随して起こる減速を起こすことなく、フェイルオーバ発生時の結果を実質的に保証することができる。

【００８２】

通常状態において、サーバ６２－２がサーバ６２－１において実行される処理の最新ミラーを維持するので、サーバ６２－１がレフトオフ（left-off）であるサーバ６２－１の処理タスクをサーバ６２－２が想定することによって、サーバ６２－１の障害が速やかに復旧し得る。図１５は、１対のサーバを運用するための方法５００を表すフローチャートである。このとき、該サーバのうちの一つが一次サーバに指定されており、もう一つがバックアップサーバに指定されている。システム５０を用いて実行するとき、ステップ５０５において、両サーバが有効かどうかが決定される。フェイルオーバエージェント９０および状態記録装置１１４を用いてステップ５０５を実行する。ハイの場合、ステップ５０

５はステップ５１０に進み、そこで、システム５０が方法３００に関して上述のような通常状態で動作する。両サーバが無効であると判断されない限りは、ステップ５０５およびステップ５１０が循環し続ける。両サーバが無効であると判断された場合、方法はステップ５２０に進む。ステップ５２０において、第一サーバのみが有効かどうかを決定する。たとえば、何らかの理由で、フェイルオーバエージェント９０－１がフェイルオーバエージェント９０－２との接続を確立できない場合、ステップ５２０において第一サーバのみが有効であることを決定し、方法５００はステップ５３０に進む。当該ステップ５３０において、システム５０が一次単独状態（primary-only state）で動作する。フェイルオーバエージェント９０－１がフェイルオーバエージェント９０－２との接続が確立できない理由として考えられるのは、これらに留まるわけではないが、サーバ６２－２が致命的な損傷を被ったか、または、リンク７８が切断されたかである。

【００８３】

第一サーバが無効である場合、方法５００はステップ５２０からステップ５４０に進む。ステップ５４０では、第二サーバのみが有効であるかどうかが決定される。異なる場合、方法５００は例外として終了する。しかし、第二サーバが有効であると判断された場合、方法５００はステップ５４０からステップ５５０に進む。ステップ５５０において、第二サーバがさらなる処理を実行できるように、システム５０がフェイルオーバする。続いて、ステップ５６０において、第二単独状態においてさらなる処理を行うようにオペレーションが続行される。両サーバが再び有効になるまで、方法５００はステップ５６０と５７０との間を循環する。両サーバが再び有効になった時点で、方法５００はステップ５１０に進み、システム５０は通常状態に戻る。

【００８４】

図１６は、一次単独状態でのシステム５０の例を示すものであり、そこではサーバ６２－１が一次サーバに指定されているが、サーバ６２－２はオフライン（または、リンク７８の障害によって無効）である。図１６において、サーバ６２－１が一次単独状態で動作するので、状態記録装置１１４－１は、サーバ６２－１が現在一次サーバに指定されており、かつ、一次単独状態で動作中であることを表示する。

【００８５】

図１７は、二次単独状態におけるシステム５０の例を示すものであり、そこではサーバ６２－２が一次サーバに指定されているが、サーバ６２－１はオフラインである。図１７において、サーバ６２－２が一次単独状態で動作しているので、状態記録装置１１４－２は、サーバ６２－２が現在一次サーバに指定されており、かつ、一次単独状態で動作中であることを表示する。

【００８６】

図示されていないが、システム５０はまた、サーバ６２－２が一次サーバに指定されており、サーバ６２－１がバックアップサーバに指定されている通常状態で構成され得ることに留意されたい。

【００８７】

図１８は、サーバ６２のうちの一つのみが有効なときにメッセージを処理するための方法６００を表すフローチャートである。方法５００におけるステップ５３０にあるサーバ６２－１が方法６００を実行するか、または、方法５００におけるステップ５６０にあるサーバ６２－２が方法６００を実行する。ここで、当業者であれば、方法６００が方法３００における一次サーバの動作を実質的に反映していることがわかるだろう。より具体的には、方法３００におけるステップ３１０～３６０およびステップ３７０～３９０が、方法６００におけるカウンターパート（同じ番号を有し、後ろにＦが付記する）に対応する。しかし、方法６００におけるステップ３６５Ｆは、方法３００におけるステップ３６５とは異なる。ステップ３６５Ｆは、方法３００におけるステップ４３０に対応している。ステップ３６５Ｆでは、ライブラリ１０２は複製が成されたという認証を単に複製することによって、サービス９４（または９８）からの複製要求に応答するので、サービス９４（または９８）がステップ３８０Ｆにおいてそのような認証を受信することができ、方法

６００がステップ３９０Ｆに進むことができる。
【００８８】
図１９は、一次サーバから、方法５００におけるステップ５５０を実行するのに使用可能であるバックアップサーバにフェイルオーバするための方法７００を表すフローチャートである。フェイルオーバエージェント１１４－２がサーバ６２－１の障害（故障、または、何らかの理由でもはや有効ではないこと）を発見した場合、例えば、サーバ６２－２が方法７００を実行する。サーバ６２－２は、サーバ６２－１が障害を起こしており、当該サーバが一次サーバであるとみなすのだが、クライアント５４は既にサーバ６２－１と交信しているので、当該クライアント５４はそのままサーバ６２－１と交信し続ける。この場合、方法７００はステップ７１０から始まる。当該ステップにおいて、複製エージェントキューが解消される。上記例において、サーバ６２－１が障害を起こす前に当該サーバ６２－１において処理された全メッセージ（および関連の外部呼び出し）処理を削除および複製するために、ステップ４０５（および後続のステップ３１０Ｓ、３１５Ｓ、３４０Ｓ、３４５Ｓ、３６０Ｓ、３７０Ｓ、３８０Ｓ、３９０Ｓ、４１０、４１５、４２０、４２５、４３０、および４３５）に従って、サーバ６２－２は複製エージェント１２６－２に保存された全データの処理を続ける。サーバ６２－１がステップ３７０において障害を起こした場合は、サーバ６２－２はクライアントから複製メッセージを受信してもよい。該クライアントは、ギャップリカバリや他の例、例えば、出願人の同時係属出願であるＵＳ公開公報ＵＳ20050138461において説明されているタイプのリカバリなどのリカバリプロトコルを実行する。クライアントが、メッセージを処理するサーバ６２－１からの認証を受信することはない。この場合、サーバ６２－２は、複製メッセージを認識し、複製メッセージの再処理を試みることなく、単に同じ返答を返すように構成されている。
【００８９】
続いて、ステップ７２０において、複製エージェントが停止する。上記例において、複製エージェント１２６－２は、サーバ６２－１から受信されたデータのキューをもはや維持しないように停止するか、または、サービス９４－２および９８－２にメッセージを送信するように構成されている。ステップ７３０では、外部リソースエージェントおよびシーケンサがアクティブになる。上記例において、外部リソースエージェント１１８－２がアクティブになるので、該エージェントは、方法６００におけるステップ３２５Ｆおよびステップ３３０Ｆで示された外部機能呼び出しを行うように構成されている。シーケンサ１０６－２も同様にアクティブになるので、該シーケンサ１０６－２は、方法６００のステップ３３５において示されたシーケンスナンバーを割り当てるように構成され得る。同様に、シーケンサ１０６－２は、方法６００におけるステップ３５５Ｆにおいて示された外部機能呼び出しを行うように構成され得る。続いて、ステップ７４０において、フェイルオーバエージェントが一次単独状態を表示するように設定されている。上記例において、フェイルオーバエージェント１１４－２が一次単独状態を表示するよう設定されているので、方法６００におけるステップ３２０Ｆ、３３５Ｆ、３５０Ｆ、３５５Ｆおよび３６５Ｆに従ってライブラリ１０２－２が動作を認識する。続いて、ステップ７２０では、サーバの存在（presence）がクライアントに伝えられる。上記例において、サーバ６２－２は、サーバ６２－２がクライアント５４からのメッセージを受容および処理する準備ができたことを、ネットワーク５８を介してクライアント５４に伝える。これを行う方法は特に限定されていないが、方法３００の開始に先立ってサーバ６２－１がクライアント５４に伝えた方法と実質的に同じ方法である。セッションプロトコルがギャップリカバリを実行することができるので、各サイドは、カウンターパーティが受信しない可能性のある交信を再送信し得る。この時点で、システム５０は図１７に示された状態である。該状態では、サーバ６２－２が一次サーバに指定されており、システム５０は、一次サーバに指定されたサーバ６２－２を用いて一次単独状態で動作する準備ができている。この時点で、方法は方法５００におけるステップ５６０に戻り、そこで、クライアントからのメッセージを受信し、方法６００に従って処理が行われる。
【００９０】

本発明の様々な特徴および構成要素の特定の組合せのみをここで述べたが、当業者であれば、開示された特徴、および構成要素、および／または、これら特徴および構成要素の代替的組合せについての所望の部分を、その希望に応じて利用し得ることが明らかであろう。例えば、システム５０は２つのサーバ６２－１および６２－２を含んでいるが、あらゆる数のサーバが利用可能であると考えられる。ここに記載された方法の応用形を用いて、あるサーバを一次サーバに指定し、あらゆる数の付加的サーバをバックアップサーバに指定することができ、かつ、当該付加的サーバを直列または並列に接続することができる。そのような付加的サーバは、ここに開示されたサーバ６２と実質的に同じコンピュータ環境および構造を有する。該付加的サーバは、あらゆる場合において、ライブラリおよび他のソフトウエア構成要素と交信する同一のサービスを有しており、これらサービスに代わって（一次サーバの場合は）外部呼び出しを行うか、または、これらサービスの複製に代わって（バックアップサーバの場合は）外部呼び出しの複製を提供する。

【００９１】

また、方法３００が変更可能であることを理解されたい。例えば、方法３００は完全に同時に動作するように構成可能であって、その場合、一次および二次共有リソースの両方が、特定のサービスによる処理の結果とともに書き込みされたことを鑑みると、一次サーバは、メッセージが処理されたことをクライアントに対して認証するのみである。ステップ３８０Ｓが実行されると直ちに、ステップ４００が実行されるように方法３００を変更することによって、上記を実行することができる。

【００９２】

また、フェイルオーバシステムを以下のように構成してもよい。

【００９３】

＜１＞

ネットワークを介して相互接続された少なくとも２つのサーバのうちのいずれか１つを選択可能であって、選択した方と接続する少なくとも１つのクライアントを備えており、

上記サーバのうちの１つは、上記クライアントに接続されるときに一次サーバに指定されており、かつ、上記サーバのうちの残りは、上記クライアントに接続されないときにバックアップサーバに指定されており、

上記少なくとも１つのクライアントが、上記一次サーバにメッセージを送信するように構成されており、

上記サーバの各々が、異なるタイプの上記メッセージを処理する複数のサービスを介して上記メッセージの全てを処理するように構成されており、

上記サービスの各々が、上記メッセージの処理の結果に基づいて上記サーバの各々によって維持された共有リソースに対してアクセスおよびアップデートの少なくともいずれかを行なうように構成されており、

上記サーバの各々が、上記サービスと連結されたライブラリを維持しており、

上記ライブラリは、上記一次サーバによって維持されるとき、

ｉ）外部リソースに対して少なくとも１つの外部呼び出しを実行すること、

ｉｉ）上記個々のメッセージに関するサービスからの要求に基づいて、各メッセージにシーケンスを付けること、

ｉｉｉ）上記個々のメッセージに関する上記サービスに対する外部呼び出し結果および上記シーケンスの結果を戻すこと、並びに、

ｉｖ）上記サービスによって共有されるキャッシュメモリに、上記外部呼び出し結果および上記シーケンスの結果を保存すること、

によって、上記個々のメッセージに関する上記サービスからの要求に応答するように構成されており、かつ、

上記一次サーバおよび上記バックアップサーバは、

ｉ）上記一次サーバにおいて受信される上記少なくとも１つのクライアントからのメッセージ、および、

ｉｉ）上記保存された上記外部呼び出し結果、

が上記バックアップサーバに複製されるように、互いに接続されており、
上記ライブラリは、上記バックアップサーバによって維持されるとき、
ｉ）上記一次サーバが複製した上記保存された外部呼び出し結果の内容を利用して上記外部呼び出し結果をライブラリに戻すこと、
によって、上記個々のメッセージに関するサービスからの要求に応答するように構成されており、
上記バックアップサーバは、上記一次サーバと同じシーケンスにおける上記メッセージの処理に上記キャッシュメモリを利用し、上記一次サーバおよび上記バックアップサーバにおいて上記共有リソースは実質的に同一であり、上記一次サーバが障害を起こした場合、上記バックアップサーバが上記一次サーバに指定され、実質的にトランスペアレントな方法において、上記クライアントに代わって他のさらなるメッセージを処理し続けるフェイルオーバのためのシステム。

【００９４】
＜２＞
上記キャッシュメモリが揮発性メモリ内で維持される、＜１＞に記載のシステム。

【００９５】
＜３＞
上記システムが電子取引システムの一部であって、上記サービスは取引エンジンに含まれており、
上記システムが、２つの他のさらなるメッセージを上記一次サーバにそれぞれ送信する２つのさらなるクライアントを備えており、
上記他のさらなるメッセージの各々が上記メッセージと実質的に同じ方法で処理される、＜１＞に記載のシステム。

【００９６】
＜４＞
上記メッセージのうちの１つが買い注文を示しており、上記受信された２つの他のさらなるメッセージが、実質的に同じ回数（different but nearly identical times）受信されており、
上記２つの他のさらなるメッセージの両方が、上記買い注文に適応する売り注文を示しており、上記一次サーバが障害を起こした場合に、上記バックアップサーバが上記メッセージの処理を続けて、上記売り注文のうちの注文のタイミングが早い方が上記買い注文と適合されるようになっている、＜３＞に記載のシステム。

【００９７】
＜５＞
上記外部リソースが、オペレーティングシステムのタイムスタンプおよびマーケットフィードである、＜４＞に記載のシステム。

【００９８】
＜６＞
上記外部リソースが、オペレーティングシステムのタイムスタンプである、＜１＞に記載のシステム。

【００９９】
＜７＞
ネットワークを介して相互接続された少なくとも２つのサーバのうちのいずれか１つを選択可能であって、選択した方と接続する少なくとも１つのクライアントを備えており、
上記サーバのうちの１つは、上記クライアントに接続されるときに一次サーバに指定され、かつ、上記サーバのうちの残りは、上記クライアントに接続されないときにバックアップサーバに指定されており、
上記少なくとも１つのクライアントが、上記一次サーバにメッセージを送信するように構成されており、
上記サーバは、上記クライアントに代わって複合スレッドを用いて上記メッセージを処

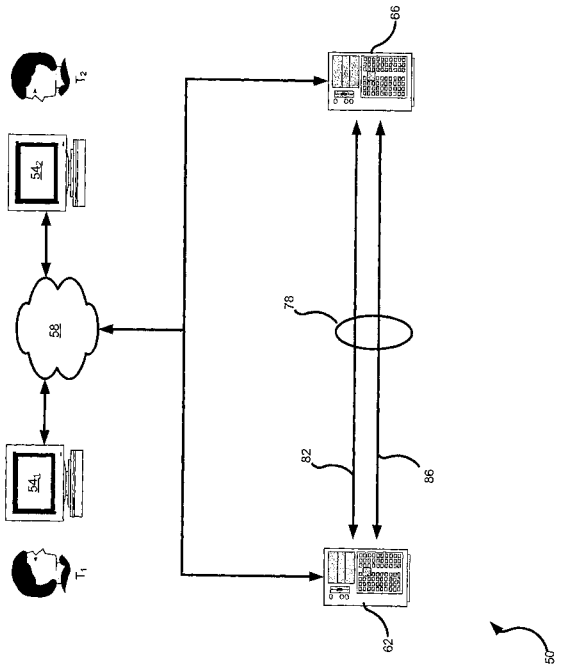
理するように構成されており、さらに、上記メッセージの処理に関して使用される上記スレッドの各々にアクセス可能な共有リソースを維持するように構成されており、
上記一次サーバは、上記１つ以上の個々のメッセージに関する外部リソースに対して少なくとも１回の外部機能呼び出しを行うことによって上記メッセージを処理するように構成されており、
上記一次サーバは、
ｉ）上記メッセージ、
ｉｉ）上記メッセージに関する上記外部機能呼び出し結果、および、
ｉｉｉ）上記メッセージを処理するためのシーケンス、
を、上記バックアップサーバに複製するように構成されており、
上記バックアップサーバは、上記シーケンスに従って上記一次サーバから受信される上記複製された外部機能呼び出し結果を用いて、上記メッセージを処理するように構成されており、上記サーバの各々による上記メッセージ処理の間、上記共有リソースが、上記一次サーバおよび上記バックアップサーバの両方において実質的に同一であるように構成されている、フェイルオーバのためのシステム。
【０１００】
＜８＞
ネットワークを介して相互接続された少なくとも２つのサーバのうちのいずれか１つを選択可能であって、選択した方と接続する少なくとも１つのクライアントを備えており、
上記サーバのうちの１つは、上記クライアントに接続されるときに一次サーバに指定され、かつ、上記サーバのうちの残りは、上記クライアントに接続されないときにバックアップサーバに指定されており、
上記少なくとも１つのクライアントが、上記一次サーバにメッセージを送信するように構成されており、
上記サーバは、上記サーバの各々において同一である少なくとも１つのサービスであって、上記個々のサービスに関するサーバが上記一次サーバか上記バックアップサーバのいずれで動作しているかを認識しない少なくとも１つのサービスを用いて、上記メッセージを処理するように構成されており、
上記サーバは、上記サーバが上記一次サーバであるか、または、上記サーバが上記バックアップサーバであるかを示すライブラリを維持するようにさらに構成されており、
各サーバ内部にて提供される上記サービスは、その個々の上記ライブラリを介して少なくとも１回の外部呼び出しを行うように構成されており、
上記一次サーバにおける上記ライブラリは、上記外部呼び出しを完了させて当該外部呼び出しの結果を上記一次サーバにおける上記ライブラリに戻すように構成されており、上記バックアップサーバにおける上記ライブラリは、上記外部呼び出しの結果を上記バックアップサーバにおける上記サービスに戻すように構成されており、
上記一次サーバにおける上記サービスと、上記バックアップサーバにおける上記サービスとが、上記外部呼び出しの結果を用いて、上記各メッセージを処理するようにさらに構成されている、フェイルオーバのためのシステム。
【０１０１】
＜９＞
上記ライブラリが、１組または異なる複数の組の固有の有効コードとして実施されている、＜８＞に記載のシステム。
【０１０２】
＜１０＞
上記サーバ各々が、上記サービスが上記メッセージの処理結果を保存することができる共有リソースを維持するように構成されている、＜８＞に記載のシステム。
【０１０３】
＜１１＞
上記共有リソースが、個々のサーバにおけるランダムアクセスメモリ内に維持されてい

る、＜１０＞に記載のシステム。
【０１０４】
＜１２＞
上記外部呼び出しがタイムスタンプ要求である、＜８＞に記載のシステム。
【０１０５】
＜１３＞
上記システムが電子取引システムの一部であって、メッセージは、セキュリティが買い注文または売り注文であるかのメッセージであって、上記外部呼び出しが、上記セキュリティの価値に関するマーケットフィード相場要求である、＜８＞に記載のシステム。
【０１０６】
＜１４＞
上記少なくとも１つのサービスが、注文受付サービス、注文取消サービス、注文変更サービス、注文適合サービス、予め実行された取引を行うためのサービス、またはクロス取引を行うためのサービス、を含んでいる、＜８＞に記載のシステム。
【０１０７】
＜１５＞
上記一次サーバにおける上記サービスは、上記外部呼び出しの結果が正しく上記バックアップサーバに送信されたことを上記バックアップサーバが確認した場合、上記メッセージが処理されたことを上記クライアントが認証できるように構成されている、＜１０＞に記載のシステム。
【０１０８】
＜１６＞
上記一次サーバにおける上記サービスは、上記外部呼び出しの結果が正しく上記バックアップサーバに送信されたことを上記バックアップサーバが確認するか否かに関わらず、上記メッセージが処理されたことを上記クライアントが認証できるように構成されている、＜１０＞に記載のシステム。
【０１０９】
＜１７＞
上記一次サーバは、所定の期限内に、上記外部呼び出しの結果が正しく上記バックアップサーバに送信されたことが、上記バックアップサーバが確認しない場合、当該バックアップサーバが障害を起こしたとみなす、＜１６＞に記載のシステム。
【０１１０】
＜１８＞
システムにおけるフェイルオーバのための方法であって、
ネットワークを介して相互接続された少なくとも２つのサーバのうちのいずれか１つを選択可能であって、選択した方と接続する少なくとも１つのクライアントを備えており、
上記サーバのうちの１つは、上記クライアントに接続されるときに一次サーバに指定され、かつ、上記サーバのうちの残りは、上記クライアントに接続されないときにバックアップサーバに指定されており、
上記少なくとも１つのクライアントが、上記一次サーバにメッセージを送信するように構成されており、
上記方法は、
上記サーバが上記一次サーバであるか、または、上記サーバが上記バックアップサーバであるかを示すライブラリを維持するように上記サーバを構成するステップと、
その個々の上記ライブラリを介して外部呼び出しを行うように上記サービスを構成するステップと、
上記外部呼び出しを完了させて当該外部呼び出しの結果を上記一次サーバにおけるサービスに戻すように、かつ、上記バックアップサーバにおける上記ライブラリに対して上記外部呼び出しの結果を送信するように上記一次サーバにおける上記ライブラリを構成するステップと、

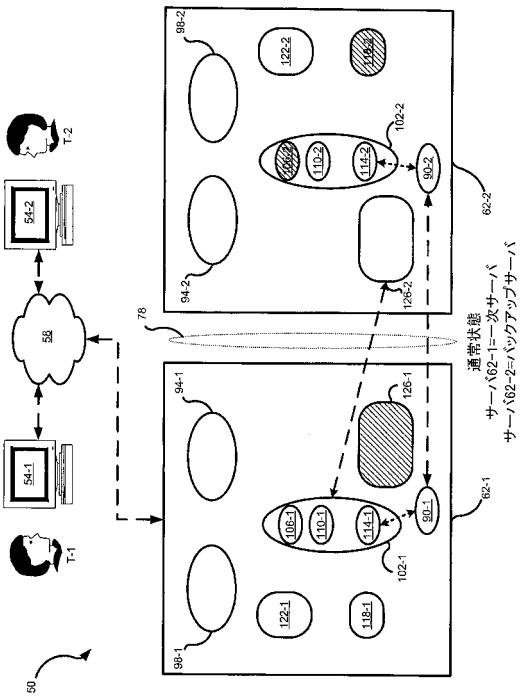
上記バックアップサーバにおけるサービスに、上記外部呼び出しの結果を送信するように上記バックアップサーバにおける上記ライブラリを構成するステップと、
上記一次サーバにおける上記ライブラリが備える上記外部呼び出しの結果を用いて、上記メッセージを処理するように上記一次サーバにおける上記サービスを構成するステップと、
上記バックアップサーバにおける上記ライブラリが備える上記外部呼び出しの結果を用いて、上記メッセージの処理をするように上記バックアップサーバにおける上記サービスを構成するステップと、を備えており、
上記サービスの各々が、上記サーバ各々において実質的に同一であって、上記サービスの各々が、上記個々のサービスに関する上記サーバが上記一次サーバであるか上記バックアップサーバかのいずれで動作しているか認識しない、フェイルオーバのための方法。
【０１１１】
＜１９＞
選択された少なくとも１つのクライアントと接続可能なネットワークを介して相互接続された少なくとも２つのサーバのうちのいずれか１つのサーバ上で実行可能な、一組のプログラミング命令を保存するコンピュータ読み取り可能な記録媒体であって、
上記サーバのうちの１つが、上記クライアントに接続されるときに一次サーバに指定されており、かつ、上記サーバのうちの残りが、上記クライアントに接続されないときにバックアップサーバに指定されており、上記少なくとも１つのクライアントが、上記一次サーバにメッセージを送信するように構成されており、
上記プログラミング命令は、
上記サーバが上記一次サーバであるか、上記サーバが上記バックアップサーバであるかを示すライブラリを維持するように、上記サーバを構成するための命令と、
上記個々のライブラリを介して外部呼び出しを行うように、上記サービスを構成するための命令と、
上記外部呼び出しを完了させて上記一次サーバにおけるサービスに対して上記外部呼び出しの結果を戻すように、かつ、上記バックアップサーバにおける上記ライブラリに対して上記外部呼び出し結果を送信するように、上記一次サーバにおける上記ライブラリを構成するための命令と、
上記バックアップサーバにおける上記サービスに対して、上記外部呼び出しの結果を送信するように上記バックアップサーバにおける上記ライブラリを構成するための命令と、
上記バックアップサーバにおける上記ライブラリが備える上記外部呼び出しの結果を用いて、上記メッセージを処理するように上記バックアップサーバにおける上記サービスを構成するための命令と、
上記サーバの各々において実質的に同一であるように上記サービスの各々を構成するための命令であって、上記個々のサービスに関する上記サーバが上記一次サーバか上記バックアップサーバのいずれで動作しているか、上記サービスの各々が認識しないように構成するための命令と、を含むコンピュータ読み取り可能な記録媒体。
【０１１２】
＜２０＞
一次サーバおよび少なくとも１つのバックアップサーバを備えており、
上記一次サーバが、複数の処理対象（processing inputs）となる入力を確定するように、かつ、上記入力を処理する上記一次サーバの処理に先立って、上記バックアップサーバに上記処理対象となる入力を送信するように構成されている、フェイルオーバのためのシステム。
【０１１３】
＜２１＞
上記処理対象となる入力は、上記サーバの両方による上記入力の処理の決定性を保証する入力を含む、＜２０＞に記載のシステム。
【０１１４】

＜２２＞
上記処理対象となる入力は、外部リソースに対する呼び出し結果を含む、＜２０＞に記載のシステム。
【０１１５】
＜２３＞
上記処理対象となる入力は、共有リソースをさらに含む、＜２０＞に記載のシステム。
【０１１６】
＜２４＞
上記処理対象となる入力は、上記一次サーバにより生成されたシーケンスナンバをさらに含む、＜２０＞に記載のシステム。

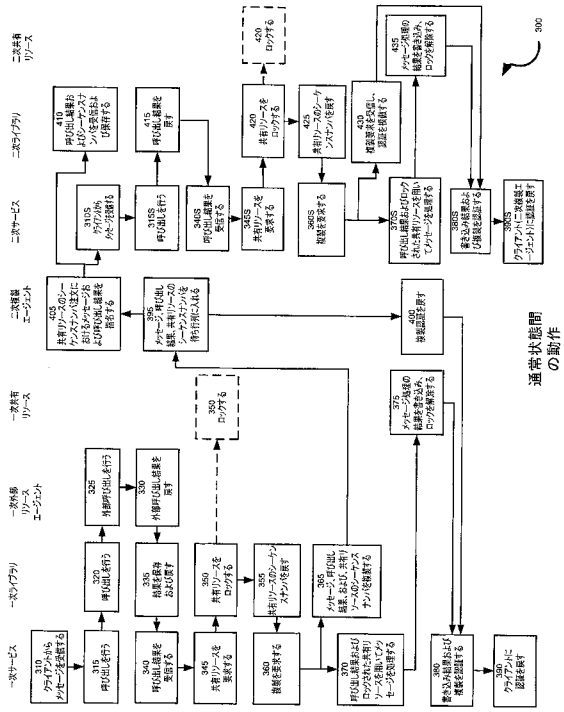
【図１】



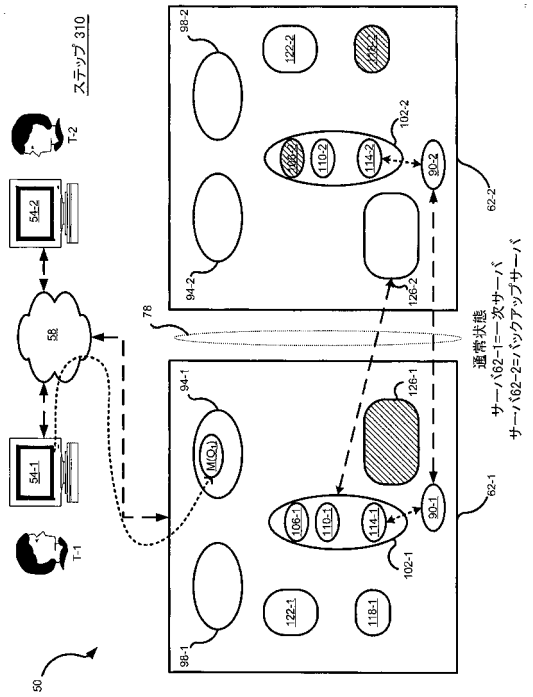
【図２】



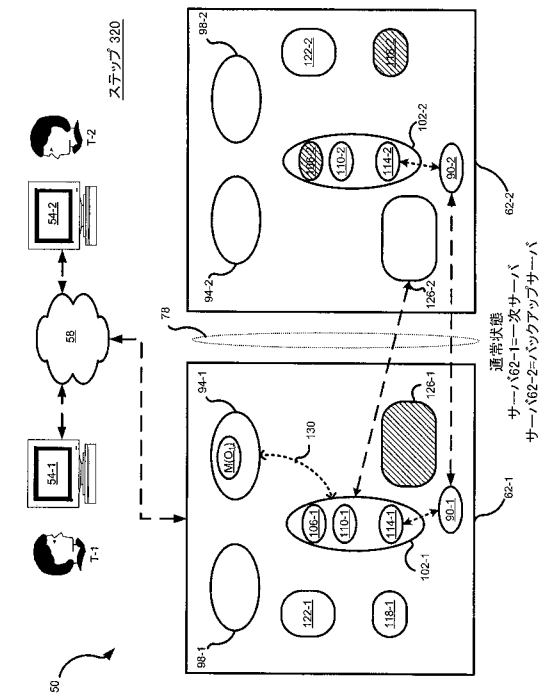
【図３】


通常状態間
の動作

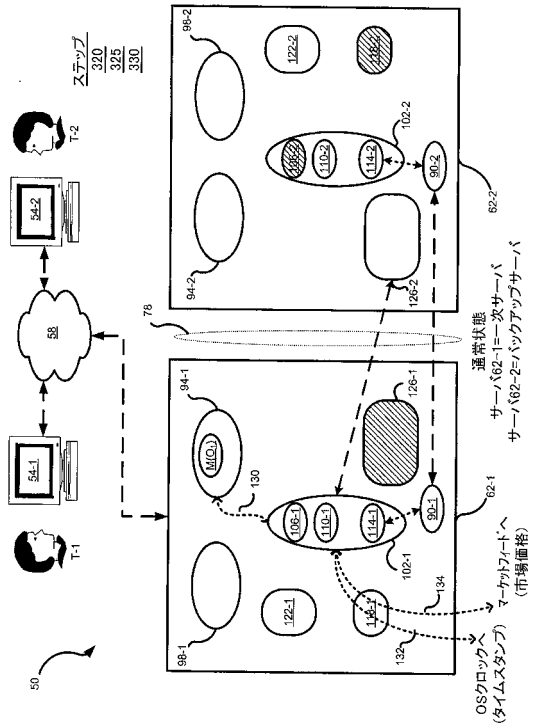
【図４】


ステップ 310
通常状態
サーバ62-1=一次サーバ
サーバ62-2=バックアップサーバ

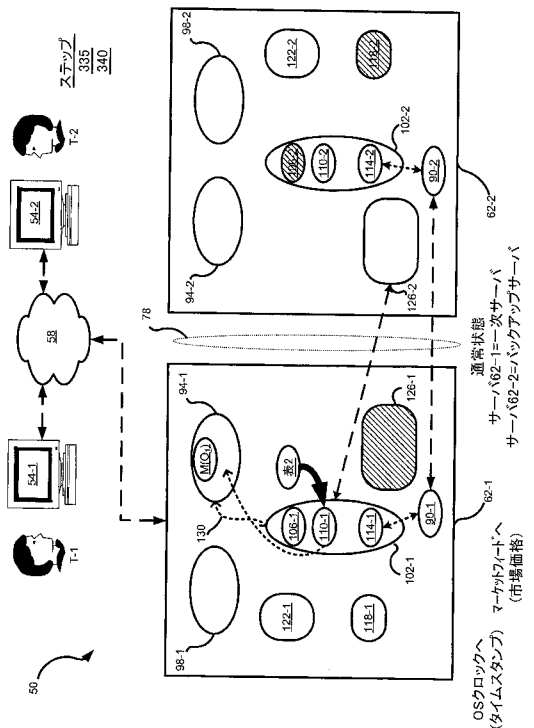
【図５】


ステップ 320
通常状態
サーバ62-1=一次サーバ
サーバ62-2=バックアップサーバ

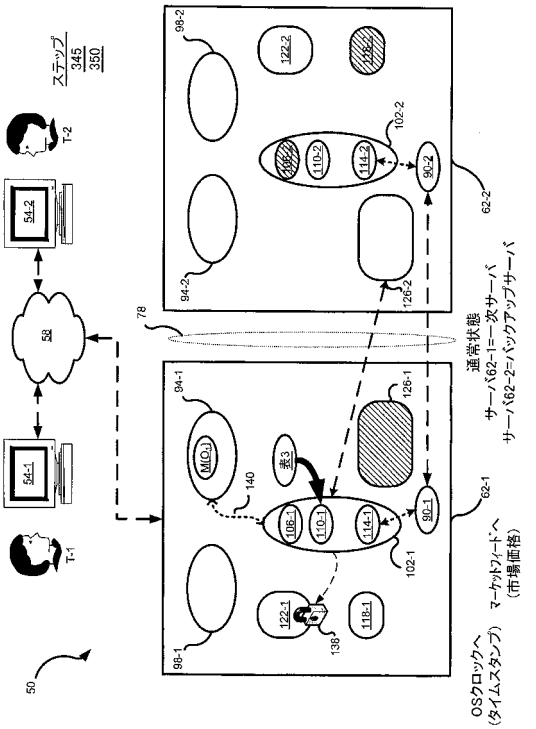
【図６】


ステップ
320
325
330
通常状態
サーバ62-1=一次サーバ
サーバ62-2=バックアップサーバ
OSクロックへ
（タイムスタンプ）
マーケットフィードへ
（市場価格）

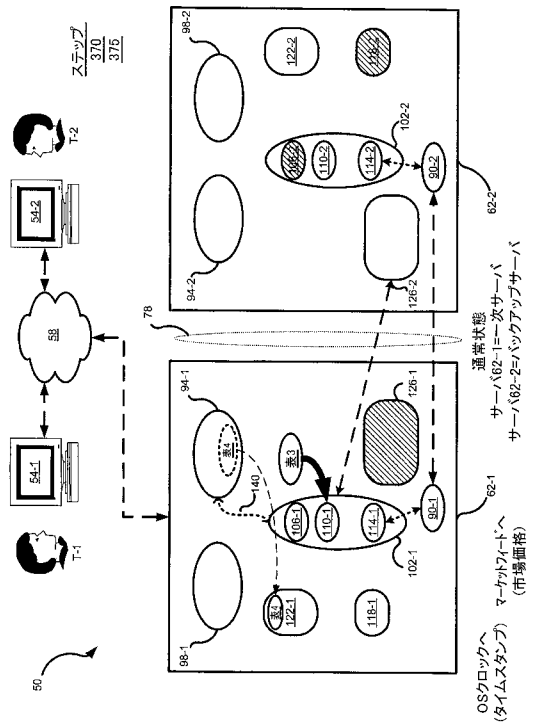
【図７】


ステップ
335
340
T-1
T-2
54-1
54-2
58
50
78
94-1
94-2
98-1
98-2
MQ-1
130
差2
106-1
110-1
114-1
102-1
126-1
90-1
62-1
122-1
118-1
106-2
110-2
114-2
102-2
126-2
90-2
62-2
122-2
118-2
通常状態
サーバ62-1=一次サーバ
サーバ62-2=バックアップサーバ
OSクロックへ
(タイムスタンプ)
マーケットフィードへ
(市場価格)

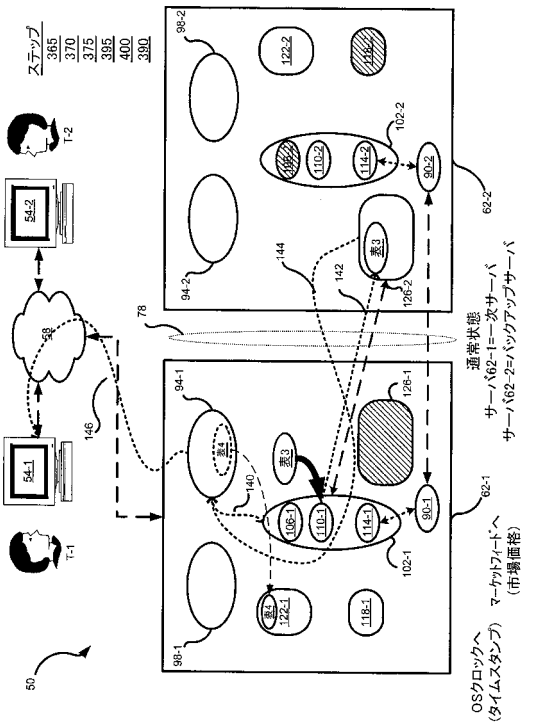
【図８】


ステップ
345
350
T-1
T-2
54-1
54-2
58
50
78
94-1
94-2
98-1
98-2
MQ-1
140
差3
106-1
110-1
114-1
102-1
126-1
90-1
62-1
122-1
138
118-1
106-2
110-2
114-2
102-2
126-2
90-2
62-2
122-2
118-2
通常状態
サーバ62-1=一次サーバ
サーバ62-2=バックアップサーバ
OSクロックへ
(タイムスタンプ)
マーケットフィードへ
(市場価格)

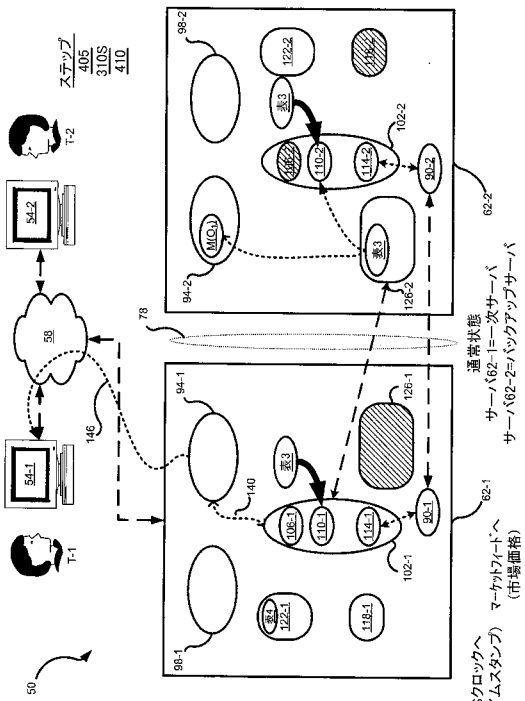
【図９】


ステップ
370
375
T-1
T-2
54-1
54-2
58
50
78
94-1
94-2
98-1
98-2
差1
140
差3
106-1
110-1
114-1
102-1
126-1
90-1
62-1
差3
122-1
118-1
106-2
110-2
114-2
102-2
126-2
90-2
62-2
122-2
118-2
通常状態
サーバ62-1=一次サーバ
サーバ62-2=バックアップサーバ
OSクロックへ
(タイムスタンプ)
マーケットフィードへ
(市場価格)

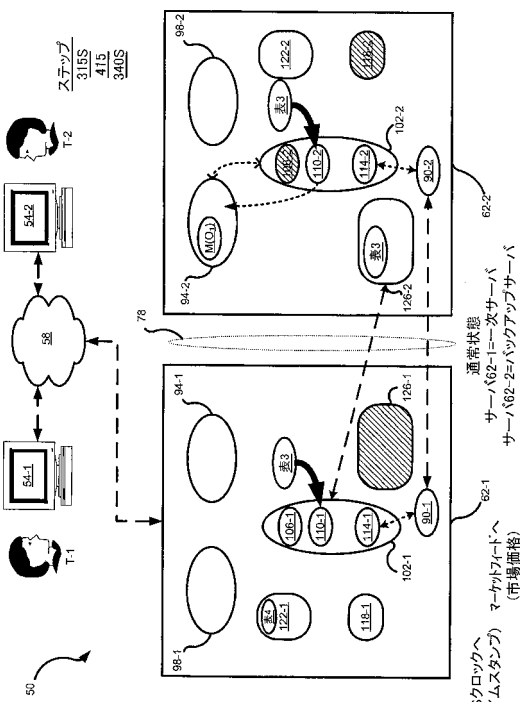
【図１０】


ステップ
365
370
375
395
400
390
T-1
T-2
54-1
54-2
58
146
50
78
94-1
94-2
98-1
98-2
差1
140
差3
106-1
110-1
114-1
102-1
126-1
90-1
62-1
差3
122-1
118-1
144
142
差3
126-2
106-2
110-2
114-2
102-2
90-2
62-2
122-2
118-2
通常状態
サーバ62-1=一次サーバ
サーバ62-2=バックアップサーバ
OSクロックへ
(タイムスタンプ)
マーケットフィードへ
(市場価格)

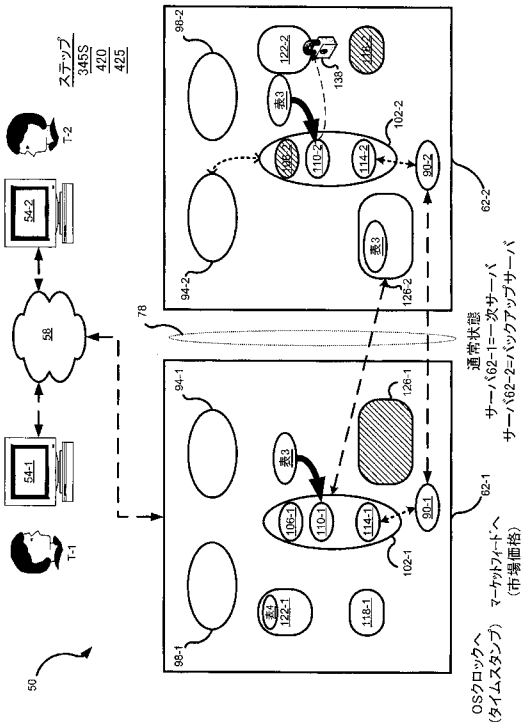
【図１１】



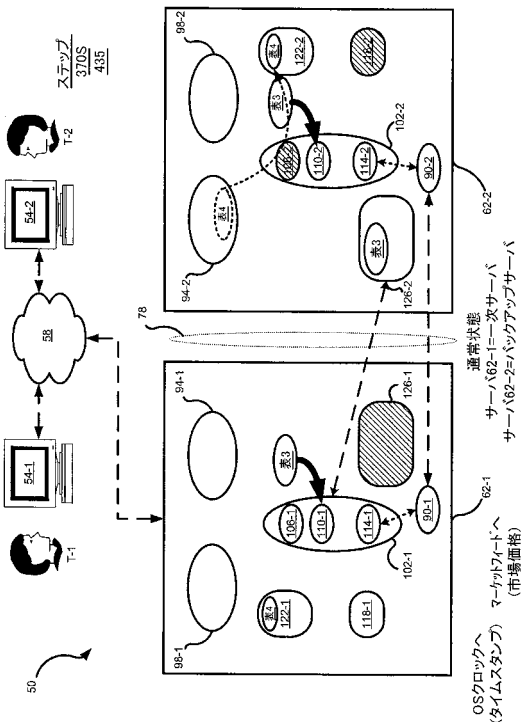
【図１２】



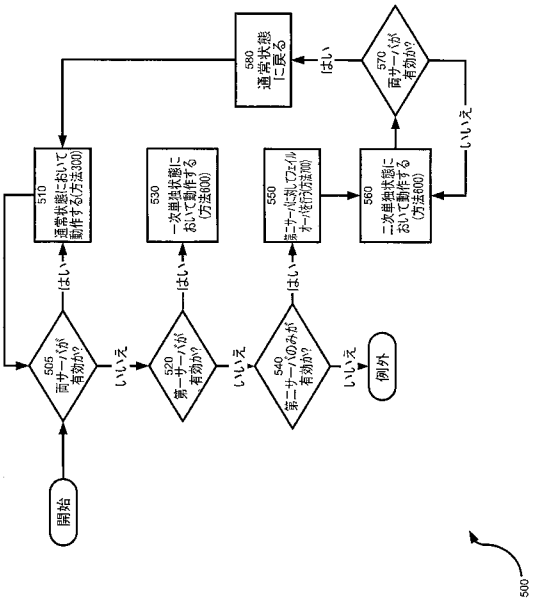
【図１３】



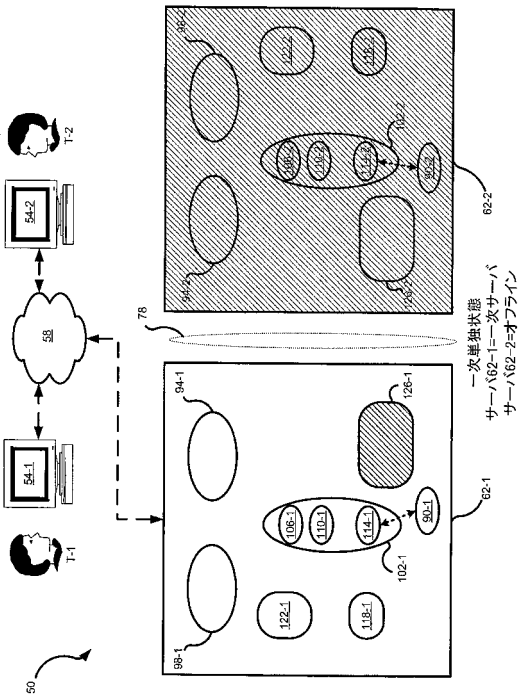
【図１４】



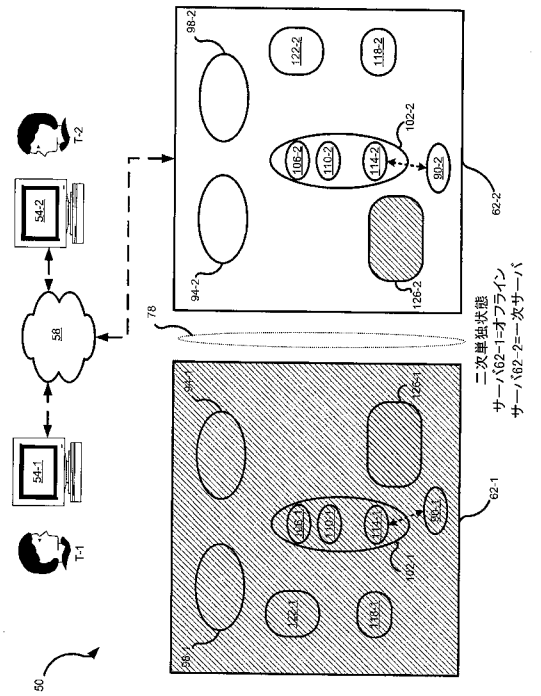
【図１５】



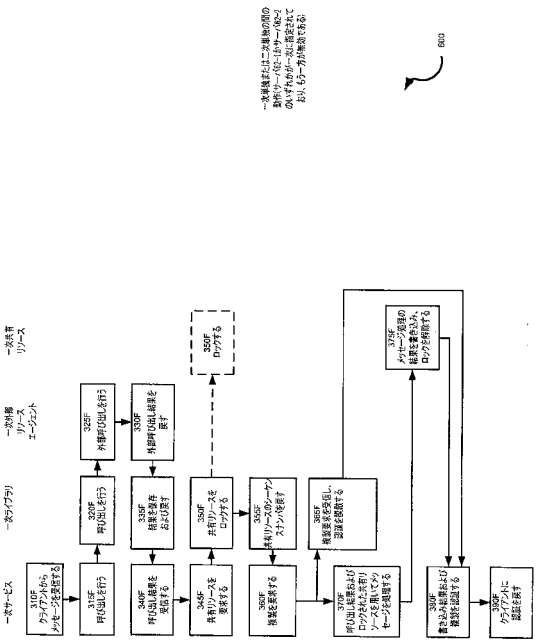
【図１６】



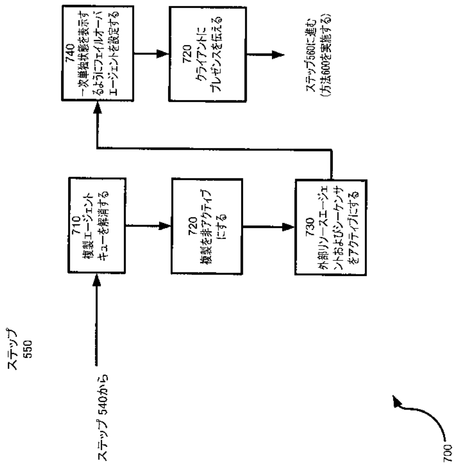
【図１７】



【図１８】



【図１９】


ステップ 550
ステップ 540から
710 複製エージェントキューを解消する
720 複製を非アクティブにする
730 外部リソースエージェントおよびシーケンサをアクティブにする
740 一次単独状態を表示するようにフェイルオーバエージェントを設定する
720 クライアントにプレゼンスを伝える
ステップ560に進む (方法600を実施する)
700

フロントページの続き

(74)代理人 100107560
弁理士 佐野 惣一郎
(72)発明者 モロサン，トゥドル
カナダ，オンタリオ州 エム２ケイ １シー１，トロント，シェパード アベニュー イースト ６４４ ＃１５２７
(72)発明者 アレン，グレゴリー，エイ.
カナダ，オンタリオ州 エル６ジェイ ２ケイ５，オークヴィル，アンソニー ドライブ ４６４
(72)発明者 パブレンコ，ヴィクター
カナダ，オンタリオ州 エル７エル ７イー８，バーリントン，オーチャード ロード ２３２１
(72)発明者 ラム，ベンソン，ゼーキット
カナダ，オンタリオ州 エル５エム ７エイ３，ミシソーガ，デル ダン ドライブ ５８６４

審査官 多賀 実

(56)参考文献 特開平０８－２７２７５３（ＪＰ，Ａ）
特表平１１－５０２６５９（ＪＰ，Ａ）
特表２００２－５２２８４５（ＪＰ，Ａ）
特開２００２－２８７９９９（ＪＰ，Ａ）
特表平０６－５０４３８９（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ１１／１６－１１／２０