

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2017-194729

(P2017-194729A)

(43) 公開日 平成29年10月26日 (2017. 10. 26)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 11/14 (2006.01)	G06F 11/14 664	5B027
G06F 9/46 (2006.01)	G06F 9/46 350	
	G06F 11/14 669	
	G06F 11/14 671	

審査請求 未請求 請求項の数 11 O L (全 28 頁)

(21) 出願番号	特願2016-82646 (P2016-82646)	(71) 出願人	000005108
(22) 出願日	平成28年4月18日 (2016. 4. 18)		株式会社日立製作所
			東京都千代田区丸の内一丁目6番6号
(特許庁注：以下のものは登録商標)		(74) 代理人	110000279
1. UNIX			特許業務法人ウィルフォート国際特許事務所
		(72) 発明者	金 成晃
			東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内
		(72) 発明者	武田 直之
			東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内
		Fターム (参考)	5B027 AA01 CC04

(54) 【発明の名称】 計算機システムおよびシステム状態再現方法

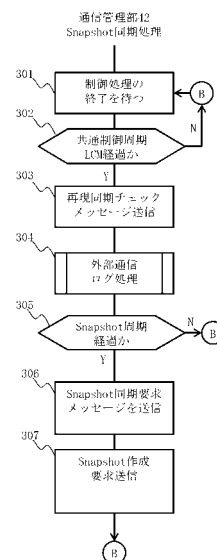
(57) 【要約】

【課題】過去の特定の時点の複数の計算機の状態を再現する。

【解決手段】再現開始処理は、複数の通信処理の因果関係に基づいて、複数の通信処理の順序を示す順序情報を作成し、順序情報を他計算機へ送信する、ことを含む。再現処理は、復元スナップショット情報に基づいて、自計算機の計算処理のスナップショットを選択し、選択されたスナップショットに基づいて、選択されたスナップショットの作成時の自計算機の計算処理の状態を復元し、順序情報に示された自計算機の計算処理の再現を示す再現同期情報を他計算機へ送信し、再現同期情報に対する応答である再現同期応答を受信し、順序情報と再現同期応答に基づいて、自計算機のログに記録された計算処理を実行する、ことを含む。

【選択図】 図 10

図10



【特許請求の範囲】

【請求項 1】

互いに接続される複数の計算機を備え、
各計算機は、他計算機との通信処理を含む計算処理を実行するように構成され、
前記複数の計算機の中の第 1 計算機は、第 1 制御対象機器に接続され、
前記第 1 計算機は、前記計算処理として、予め設定された第 1 制御周期で前記第 1 制御対象機器を制御する第 1 計算処理を実行するように構成され、
前記第 1 計算機は、選択条件を満たす場合、
現在の状態が作成条件を満たすか否かを判定し、
前記作成条件を満たすと判定された場合、スナップショットの作成を要求する作成情報
10
を他計算機へ送信する、
ことを含む、作成同期処理を繰り返し実行し、
各計算機は、前記作成情報の通信に応じて、前記計算処理の状態のスナップショットを作成する作成処理を実行し、
各計算機は、前記計算処理の実行に関するログを記録する記録処理を実行するように構成され、
前記第 1 計算機は、過去の特定時刻を示す再現要求に応じて、
前記特定時刻の直前のスナップショットを示す復元スナップショット情報を他計算機
20
へ送信し、
他計算機により記録されたログを受信し、
前記第 1 計算機により記録された第 1 ログと前記受信されたログとに示されている複数の通信処理の因果関係に基づいて、前記複数の通信処理の順序を示す順序情報を作成し、
前記順序情報を他計算機へ送信する、
ことを含む、再現開始処理を実行し、
各計算機は、前記順序情報に応じて、
前記復元スナップショット情報に基づいて、自計算機の計算処理のスナップショット
30
を選択し、
前記選択されたスナップショットに基づいて、前記選択されたスナップショットの作成時の前記自計算機の計算処理の状態を復元し、
前記順序情報に示された前記自計算機の計算処理の再現を示す再現同期情報を他計算機へ送信し、
前記再現同期情報に対する応答である再現同期応答を受信し、
前記順序情報と前記再現同期応答に基づいて、前記自計算機のログに記録された計算
処理を実行する、
ことを含む、再現処理を実行する、
計算機システム。

【請求項 2】

前記作成同期処理は、前記第 1 計算処理のアイドル状態を検出し、
前記作成条件は、現在の前記第 1 計算処理の状態が前記アイドル状態であることを含む
40
、
請求項 1 に記載の計算機システム。

【請求項 3】

前記作成同期処理は、スナップショット周期を取得し、
前記作成条件は、前回のスナップショットの作成からの経過時間がスナップショット周期以上であることを含む、
請求項 2 に記載の計算機システム。

【請求項 4】

前記複数の計算機の何れか複数のは、複数の制御計算機であり、
前記複数の制御計算機は、複数の制御対象機器に夫々接続され、
50

前記複数の制御計算機に対し、予め複数の制御周期が夫々設定され、
前記選択条件は、前記複数の制御周期の中で最も小さい制御周期を持つことを含み、
各制御計算機は、

対応する制御周期で、対応する制御対象機器を制御する計算処理を実行し、
前記選択条件を満たす制御計算機を、前記第 1 計算機として選択する、

ように構成される、

請求項 3 に記載の計算機システム。

【請求項 5】

前記第 1 計算機は、前記複数の制御周期の最小公倍数を共通周期として算出し、

前記作成条件は、前記経過時間が前記共通周期以上であることを含む、

請求項 4 に記載の計算機システム。

【請求項 6】

各計算機は、オペレーティングシステムと、前記オペレーティングシステムの機能を用いて前記作成同期処理と前記作成処理と前記再現開始処理と前記再現処理とを実行するミドルウェアと、前記ミドルウェアの機能を用いて前記記録処理を実行する管理プログラムと、前記管理プログラムの機能を用いて前記計算処理を実行するアプリケーションとを実行するように構成される、

請求項 3 に記載の計算機システム。

【請求項 7】

前記複数の計算機の中の特定計算機は、前記管理プログラムとして仮想化基盤を実行するように構成され、

前記仮想化基盤は、仮想計算機を実行するように構成され、

前記仮想計算機は、前記アプリケーションを実行するように構成される、

請求項 6 に記載の計算機システム。

【請求項 8】

前記第 1 計算機は、許容時間の入力を受け付ける第 1 画面を表示装置に表示させ、前記許容時間が入力装置へ入力される場合、前記入力された許容時間を取得し、前記許容時間に基づいて前記スナップショット周期を決定し、前記作成同期処理を実行する、

請求項 3 に記載の計算機システム。

【請求項 9】

前記第 1 計算機は、前記特定時刻と、前記特定時刻から現在時刻までの時間との何れかを示す、前記再現要求の入力を受け付ける第 2 入力画面を前記表示装置に表示させ、前記再現要求が前記入力装置へ入力される場合、前記入力された再現要求を取得し、前記再現開始処理を実行する、

請求項 8 に記載の計算機システム。

【請求項 10】

前記第 1 計算機は、現在の前記第 1 計算処理の状態が前記アイドル状態であり、且つ前記経過時間が前記共通周期以上である場合、他計算機へ前記アイドル状態を示す同期情報を送信する、

請求項 5 に記載の計算機システム。

【請求項 11】

互いに接続される複数の計算機の状態を再現するためのシステム状態再現方法であって、

各計算機が、他計算機との通信処理を含む計算処理を実行し、

前記複数の計算機の中の第 1 計算機が、前記計算処理として、予め設定された第 1 制御周期で第 1 制御対象機器を制御する第 1 計算処理を実行し、

前記第 1 計算機が、選択条件を満たす場合、

現在の状態が作成条件を満たすか否かを判定し、

前記作成条件を満たすと判定された場合、スナップショットの作成を要求する作成情報を他計算機へ送信する、

10

20

30

40

50

ことを含む、作成同期処理を繰り返し実行し、

各計算機が、前記作成情報に基づいて、前記計算処理の状態のスナップショットを作成する作成処理を実行し、

各計算機が、前記計算処理の実行に関するログを記録する記録処理を実行し、

前記第1計算機が、過去の特定時刻を示す再現要求に応じて、

前記特定時刻の直前のスナップショットを示す復元スナップショット情報を他計算機へ送信し、

他計算機により記録されたログを受信し、

前記第1計算機により記録された第1ログと前記受信されたログとに示されている複数の通信処理の因果関係に基づいて、前記複数の通信処理の順序を示す順序情報を作成し

10

、

前記順序情報を他計算機へ送信する、

ことを含む、再現開始処理を実行し、

各計算機が、前記順序情報に応じて、

前記復元スナップショット情報に基づいて、自計算機の計算処理のスナップショットを選択し、

前記選択されたスナップショットに基づいて、前記自計算機の計算処理の状態を復元し、

前記順序情報に示された前記自計算機の計算処理の再現を示す再現同期情報を他計算機へ送信し、

20

前記再現同期情報に対する応答である再現同期応答を受信し、

前記順序情報と前記再現同期応答に基づいて、前記自計算機のログに記録された計算処理を実行する、

ことを含む、再現処理を実行する、

ことを有するシステム状態再現方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、計算機システムに関する。

【背景技術】

30

【0002】

情報システムにおいては、特定時刻の計算機の状態をコピーする Snapshot 機能が知られている。計算機の処理に異常が発生した場合、取得した Snapshot から特定時刻の計算機状態に巻き戻すことで、異常状態回避の手段として捉えているシステムも存在する。ただし、Snapshot 機能は、1つの計算機内の特定メモリのデータイメージのコピーや仮想化システムにおける仮想計算機の状態のコピーに限定されるものが多い。

【0003】

制御システムにおいても、障害解析の迅速化等の高信頼性対策のために、システム障害を再現できることが望ましい。制御システムは、複数の計算機で構成されており、システム障害の再現においては従来の1台のみの計算機状態再現だけではなく、システム全体の状態再現が必要になる。また、状態再現の技術である Snapshot 機能においては、大量のメモリコピーが発生するため、システム運用中に Snapshot 処理を行うと、制御システムは、要求されるリアルタイム性を維持できない可能性が高い。そのため、システム全体の状態再現においては、複数計算機の状態の Snapshot 処理が必要であり、かつ、その処理が性能阻害にならない処理である必要がある。また、制御システムの障害解析には複数時点の Snapshot の間の期間中の処理の再現も必要となる。各計算機の処理ログを記録することにより、特定の Snapshot から再現したい時刻までの処理を、そのログで再現することができる。

40

【0004】

50

システム全体の再現処理として、特許文献 1 には、1 台の物理計算機上で、複数の仮想計算機を運用するシステム構成上で、複数の仮想計算機の Snapshot 取得とその Snapshot からのシステムの再現手段が記載されている。

【先行技術文献】

【特許文献】

【0005】

【特許文献 1】特許第 5466740 号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

10

特許文献 1 の方法は、単一の物理計算機内の複数の仮想計算機の Snapshot によるシステム再現手段であり、複数の物理計算機で構成されている制御システムの状態再現は困難である。それぞれの計算機でそれぞれの時刻での Snapshot を作成し、それらの Snapshot からの再現を行う場合、システム全体で同時刻のシステム状態再現で困難であるからである。

【課題を解決するための手段】

【0007】

上記課題を解決するために、本発明の一態様である計算機システムは、互いに接続される複数の計算機を備える。各計算機は、他計算機との通信処理を含む計算処理を実行するように構成される。前記複数の計算機の中の第 1 計算機は、第 1 制御対象機器に接続され、前記第 1 計算機は、前記計算処理として、予め設定された第 1 制御周期で前記第 1 制御対象機器を制御する第 1 計算処理を実行するように構成され、前記第 1 計算機は、選択条件を満たす場合、作成同期処理を繰り返し実行し、前記作成同期処理は、現在の状態が作成条件を満たすか否かを判定し、前記作成条件を満たすと判定された場合、スナップショットの作成を要求する作成情報を他計算機へ送信する、ことを含む。前記作成情報に基づいて、各計算機は、前記計算処理の状態のスナップショットを作成する作成処理を実行し、各計算機は、前記計算処理の実行に関するログを記録する記録処理を実行するように構成される。前記第 1 計算機は、過去の特定時刻を示す再現要求に応じて、再現開始処理を実行し、前記再現開始処理は、前記特定時刻の直前のスナップショットを示す復元スナップショット情報を他計算機へ送信し、他計算機により記録されたログを受信し、前記第 1 計算機により記録された第 1 ログと前記受信されたログとに示されている複数の通信処理の因果関係に基づいて、前記複数の通信処理の順序を示す順序情報を作成し、前記順序情報を他計算機へ送信する、ことを含む。各計算機は、前記順序情報に応じて、再現処理を実行し、前記再現処理は、前記復元スナップショット情報に基づいて、自計算機の計算処理のスナップショットを選択し、前記選択されたスナップショットに基づいて、前記選択されたスナップショットの作成時の前記自計算機の計算処理の状態を復元し、前記順序情報に示された前記自計算機の計算処理の再現を示す再現同期情報を他計算機へ送信し、前記再現同期情報に対する応答である再現同期応答を受信し、前記順序情報と前記再現同期応答に基づいて、前記自計算機のログに記録された計算処理を実行する、ことを含む。

20

30

【発明の効果】

40

【0008】

過去の特定の時点の複数の計算機の状態を再現することができる。

【図面の簡単な説明】

【0009】

【図 1】計算機システムのハードウェア構成の一例を示す。

【図 2】計算機 A 及び B のソフトウェア構成を示す。

【図 3】計算機 C のソフトウェア構成を示す。

【図 4】Snapshot 制御テーブル 46 を示す。

【図 5】ログ情報 47 を示す。

【図 6】外部通信ログフォーマットで表されたログ内容を示す。

50

【図 7】状態再現同期制御テーブル 49 を示す。

【図 8】通信管理処理を示す。

【図 9】同期メッセージ受信処理の動作を示す。

【図 10】Snapshot 同期処理を示す。

【図 11】Snapshot 同期受信処理を示す。

【図 12】Snapshot 作成処理を示す。

【図 13】外部通信ログ処理を示す。

【図 14】AP ログ制御処理を示す。

【図 15】再現同期開始処理を示す。

【図 16】再現同期開始受信処理を示す。

10

【図 17】状態再現制御処理を示す。

【図 18】再現照合処理を示す。

【図 19】状態再現同期処理を示す。

【図 20】状態再現同期受信処理を示す。

【発明を実施するための形態】

【0010】

以下、図面を用いて本発明の実施例を説明する。

【実施例 1】

【0011】

図 1 は、計算機システムのハードウェア構成の一例を示す。

20

【0012】

本実施例の計算機システムは、複数の計算機 10 を含む。計算機 10 は少なくとも、OS (Operating System) を含むプログラムを実行する演算装置 11 と、プログラムおよびその実行に必要なデータを格納する記憶装置 12 と、他の計算機 10 に対してメッセージの送受信を行う通信装置 13 と、外部ネットワーク 22 との間で制御データ等の送受信を行う通信装置 14 とを含む。ここで記憶装置 12 は、例えば RAM やハードディスクなどである。また、通信装置 13、14 は、例えば LAN アダプタなどである。演算装置 11 は 2 つ以上の CPU コアを含んでもよい。

【0013】

複数の計算機 10 は通信装置 13 を用いて通信を行う。複数の通信装置 13 は内部ネットワーク 21 で接続される。外部ネットワーク 22 は例えばネットワークスイッチなどで構成され、計算機 10 間でのデータの送受信を行う。制御システムにおいて、計算機 10 は外部ネットワーク 22 から同時に同じ制御データを受信するシステムが多いが、本実施例では限定事項ではない。また、計算機 10 間でのメッセージの送受信と制御データの送受信を同じ通信装置で行ってもよい。

30

【0014】

本実施例の計算機システムは、3 台の計算機 10 を含み、それぞれ計算機 A、計算機 B と計算機 C とする。計算機システムを構成する計算機 10 の台数は限定事項ではない。

【0015】

複数の計算機 10 の何れかが、入力装置と表示装置を含んでもよいし、計算機システムが、入力装置と表示装置を含む管理計算機を含んでもよい。表示装置は、入力のための画面を表示し、入力装置は、ユーザからの入力を受け付ける。

40

【0016】

図 2 は、計算機 A 及び B のソフトウェア構成を示す。

【0017】

このソフトウェアは、記憶装置 12 内で記憶され、演算装置 11 によって実行される。このソフトウェアは、OS 31 と、システム状態再現部 41 と、AP 管理ミドル 51 と、少なくとも 1 つのアプリ (アプリケーション) 61 とを含む。AP 管理ミドル 51 は、アプリ 61 の実行に必要な機能を提供するミドルウェアである。計算機 A と B で実行されるアプリ 61 は、AP 管理ミドル 51 によって実行を管理される。アプリ 61 は、周期的に

50

動作する。

【0018】

A P管理ミドル51は、通信処理部51Aと、A P制御部51Bと、OS31からのタイマーイベントやシグナルイベントを処理するイベント管理部51Cとを含む。通信処理部51Aは、OS31の機能を利用して、アプリ61に必要なデータを通信装置13または通信装置14を通じて送受信を行う。A P制御部51Bは、アプリ61の順序制御（シーケンス制御）を行う。イベント管理部51Cは、OS31の機能を利用して、タイマーイベントについてアプリ61からの設定や、OS31からのタイマーイベントの受信や、複数のアプリ61間のシグナルの送受信処理を行う。

【0019】

A P管理ミドル51において、通信装置13または通信装置14からアプリ61へ入力される送受信データと、OS31からのタイマーイベント及びシグナルイベントを含む内部イベント入力データとが、管理される。また、通信処理部51AとA P制御部51Bとイベント管理部51Cとは、実行のログを記録する。ログ記録の処理方法は、限定されないが、例えば標準的なOS31のトレース機能を用いることで実現できる。

【0020】

アプリ61の実行状態再現処理を行うシステム状態再現部41は、通信管理部42と、A Pログ制御部43と、Snapshot制御部44と、状態再現制御部45と、Snapshot制御テーブル46と、ログ情報47と、状態再現同期制御テーブル49とを含む。

【0021】

通信管理部42は、計算機A、計算機Bと計算機C間でのシステム状態再現に必要なデータ送受信と、A P管理ミドル51の通信処理部51Aからの外部通信ログとを処理する。A Pログ制御部43は、A P管理ミドル51からA Pログを取得する。Snapshot制御部44は、ある特定の時刻の、計算機Aと計算機Bと計算機Cとの夫々のアプリ61の状態を再現するために必要なシステム状態をコピーするSnapshot作成を行う。状態再現制御部45は、Snapshot制御部44により作成されたSnapshotと、A P管理ミドル51のログ情報47をもとに、ある特定の時刻の計算機Aと計算機Bと計算機Cの夫々のアプリ61の状態を再現する。本実施例では、計算機システム内のアプリ61の状態をシステム状態と呼ぶ。Snapshot制御テーブル46は、Snapshot作成とアプリ61の状態の再現とに必要なパラメータを格納する。ログ情報47は、通信処理部51Aにより作成される外部通信ログと、A Pログ制御部43により作成されアプリ61の実行結果を示すA Pログと、イベント管理部51Cにより作成されるイベントログとを格納する。状態再現同期制御テーブル49は、計算機Aと計算機B、計算機Cの間でSnapshot作成時刻と再現時刻の同期をとるための情報を格納する。

【0022】

計算機Aと計算機Bの夫々におけるアプリ61は、それぞれの制御周期で動作し、制御処理を行う。計算機A及び計算機Bは、制御対象機器に接続されている。制御対象機器は例えば、工場や鉄道における機器である。制御処理は、制御対象機器を制御する。制御処理の実行は、制御周期の一部の時間で終了する。そのため、制御周期のうち、制御処理の実行状態の残りの期間において、アプリ61は、アイドル状態である。例えば、制御周期が1秒であり、制御周期の最初の0.5秒間においてアプリ61は実行状態であり、残りの0.5秒間においてアプリ61はアイドル状態である。制御処理を行うアプリ61を制御アプリと呼ぶことがある。A P管理ミドル51は、制御処理を制御する。

【0023】

図3は、計算機Cのソフトウェア構成を示す。

【0024】

このソフトウェアは、OS31と、システム状態再現部41と、仮想化基盤71と、仮想計算機81とを含む。計算機Cのソフトウェアは、計算機Aと計算機BのA P管理ミドル51の代わりに仮想化基盤71を含み、計算機Aと計算機Bのアプリ61の代わりに仮

10

20

30

40

50

想計算機 8 1 を含む。仮想化基盤 7 1 は、仮想計算機 8 1 を実行する。仮想計算機 8 1 は、OS 8 2 と、少なくとも 1 つのアプリ 6 1 とを実行する。仮想化基盤 7 1 は、A P ログを記録する。この A P ログは、計算機 A、B のようにアプリ 6 1 の実行結果を示す A P ログではなく、仮想計算機 8 1 の実行結果を示す A P ログである。仮想化基盤 7 1 は、計算機と計算機 B の A P 管理ミドル 5 1 と同様、仮想計算機 8 1 の A P ログを記録する。仮想化基盤 7 1 は、一つの仮想計算機 8 1 を一つのアプリとして運用する。仮想計算機 8 1 が実行されると仮想計算機 8 1 内の OS 8 2 が仮想計算機 8 1 内のアプリ 6 1 を実行する。仮想計算機 8 1 の A P ログを記録することで、アプリ 6 1 の実行結果を示す A P ログを記録することが可能である。

【0025】

計算機 C のシステム状態再現部 4 1 において、通信管理部 4 2 と A P ログ制御部 4 3 により受信されるアプリ 6 1 の A P ログは、仮想化基盤 7 1 からの A P ログであり、ログ内容は、仮想計算機 8 1 の A P ログである。計算機 C のアプリ 6 1 の状態の S n a p s h o t を作成し、その状態を再現するには、仮想計算機 8 1 の S n a p s h o t を作成し、その状態を再現すればよい。このように計算機 C の仮想計算機 8 1 と、計算機 A と計算機 C のアプリ 6 1 とは、計算機 A と計算機 B と計算機 C のシステム状態再現部 4 1 による再現対象のアプリとして扱うことができる。以下の実施例においては、S n a p s h o t 作成とシステム状態再現におけるアプリ 6 1 と A P 管理ミドル 5 1 の処理の内容について述べるが、計算機 C においては仮想計算機 8 1 と仮想化基盤 7 1 の処理に置き換えればよい。

【0026】

計算機 A 及び B においては、アプリ 6 1 を実行部と呼び、計算機 C においては、仮想計算機 8 1 を実行部と呼ぶ。計算機 A 及び B においては、A P 管理ミドル 5 1 を実行制御部と呼び、計算機 C においては、仮想化基盤 7 1 を実行制御部と呼ぶ。

【0027】

本実施例では、システムの中に、A P 管理ミドル 5 1 と、A P 管理ミドル 5 1 によって実行管理されるアプリ 6 1 とを含む計算機が、少なくとも 1 台あることを前提とする。

【0028】

A P 管理ミドル 5 1 または仮想化基盤 7 1 が、外部通信と、アプリ 6 1 の実行結果と、OS 3 1 から発行されるイベントとのログを記録し、システム状態再現部 4 1 が、再現を制御することで、OS 3 1 を変更することなく、複数の計算機の状態を再現することができる。これにより、計算機 A、B のような物理計算機上のアプリ 6 1 であっても、計算機 C のような仮想計算機上のアプリ 6 1 であっても、同様に扱うことができ、物理計算機と仮想計算機が混在していても、システム状態を再現することができる。

【0029】

図 4 は、S n a p s h o t 制御テーブル 4 6 を示す。

【0030】

本テーブルは、システム状態を再現する目標の再現時刻からシステムの最後の実行時刻までの時間を示す戻り時間 R T N と、特定時刻で作成された S n a p s h o t から再現時刻までの再現にかかる時間の許容時間 T s と、C o o r d i n a t o r の役割の計算機をあらわす C o o r d i n a t o r 情報と、計算機 A と計算機 B と計算機 C が再現処理状態か否かをあらわす動作モードとを含む。C o o r d i n a t o r は、計算機 A と計算機 B と計算機 C の間で、S n a p s h o t 作成の同期とシステム再現の同期とを調停する。本実施例において、復元に用いられる復元 S n a p s h o t の状態から再現時刻の状態までログに基づいて再現する時間は、復元 S n a p s h o t の作成時刻から再現時刻までの時間に等しいとする。

【0031】

戻り時間 R T N は、ミリ秒単位でユーザから設定される。許容時間 T s は、分単位でユーザから設定される。その設定方法は、コマンド入力でも設定ファイルによるものでもよい。C o o r d i n a t o r 情報は、「1」であれば、C o o r d i n a t o r の役割の計算機であることを示し、「0」であれば C o o r d i n a t o r の役割をしない計算機

10

20

30

40

50

であることを示す。動作モードは、「L」であれば、通信管理部42とAPログ制御部43がログテーブル48にログを格納するログモードであることを示す。「R」であれば、計算機システムがある特定時刻の状態を再現する状態再現モードであることを示す。Coordinatorは、通信管理部42の初期処理により設定される。動作モードは通信管理部42と状態再現制御部45によって設定される。

【0032】

Coordinatorである計算機は、表示装置と入力装置に接続されていてもよいし、表示装置と入力装置を含む端末に、ネットワークを介して接続されていてもよい。Coordinatorである計算機は、許容時間を受け付ける画面（第1入力画面）を表示装置に表示させ、許容時間が入力装置に入力される場合、入力された許容時間を取得し、Snapshot制御テーブル46へ設定してもよい。また、Coordinatorである計算機は、再現要求を受け付ける画面（第2入力画面）を表示装置に表示させ、再現要求が入力装置に入力される場合、入力された再現要求を取得し、後述の再現同期開始処理を実行してもよい。再現要求は、戻り時間及び再現時刻の何れかを含む。ユーザが許容時間を入力することにより、再現に要する時間を制限することができる。ユーザが戻り時間又は再現時刻を入力することにより、任意の時刻の再現を行うことができる。

【0033】

図5は、ログ情報47を示す。

【0034】

ログ情報47は、ログ管理テーブル47Aと、外部通信ログテーブル48Aと、APログテーブル48Bと、イベントログテーブル48Cとを含む。

【0035】

ログ管理テーブル47Aは、アプリ61に関わるログを格納する。ログは、外部通信ログと、APログと、イベントログとを含む。ログ管理テーブル47Aは、ログ毎のエントリを含む。各エントリは、ログ識別子と、ログ種類と、ログ内容と、照合状態とを含む。ログ識別子は、ログを識別するためのものである。ログ種類は、通信管理部42とAP管理ミドル51の通信処理部51Aからの外部通信ログを表すCOMMと、AP管理ミドル51のAP制御部51BからのAPログを表すAPと、AP管理ミドル51のイベント管理部51Cからのイベントログを表すEVTとの何れか一つを示す。ログ内容は、当該種類のテーブルの対応するエントリの内容を示す。ログ種類がCOMMであれば、ログ内容は、外部通信ログテーブル48Aの対応するエントリの内容を示す。ログ種類がAPであれば、ログ内容は、APログテーブル48Bの対応するエントリの内容を示す。ログ種類がEVTであれば、ログ内容は、イベントログテーブル48Cの対応するエントリの内容を示す。照合状態は、Snapshot制御テーブルの動作モードが「R」の状態再現モードである時、特定時刻のアプリ61の再現の照合結果を示す。照合処理は後述する。照合が未実施である場合、照合状態は「0」、照合結果が一致である場合、照合状態は「OK」、照合結果が不一致である場合、照合状態は「NG」となる。照合状態は、照合結果によってシステムの再現処理の継続可否を決定し、システム再現処理の健全性を確保するために用いられる。

【0036】

外部通信ログテーブル48Aは、通信管理部42によって作成され、計算機間の通信パケットの入力と出力である外部通信のログを示す。外部通信ログテーブル48Aは、外部通信毎のエントリを含む。各エントリは、タイムスタンプと、ログ内容と、通信内容保存先とを含む。タイムスタンプは、通信が行われた時刻が格納される。ログ内容は、外部通信ログフォーマットにしたがって格納される。詳細は後述する。通信内容保存先は、計算機10間の通信パケットの内容の保存先のディスク上のファイル名や、メモリ上の番地等を示す。

【0037】

APログテーブル48Bは、AP管理ミドル51のAP制御部51Bから送られ、アプリ間の通信のログを示す。APログテーブル48Bは、IPC（プロセス間通信）毎のエ

10

20

30

40

50

ントリを含む。各エントリは、タイムスタンプと、タスク識別子と、IPC識別子と、開始終了情報とを含む。タイムスタンプは該当ログの発生時刻を表す。タスク識別子は、アプリ61を識別するための情報であり、例えばOS31のプロセス識別子のPIDである。IPC識別子は、複数のアプリ61間で順序制御を行うために、AP管理ミドル51により作成され管理されるIPCの情報である。アプリ61が、処理終了時に次に実行が必要な別のアプリ61を、AP管理ミドル51により提供されるIPCを通じて起し実行させることで、複数のアプリ61の順序制御が実現される。開始終了情報は、制御処理の初めに処理するアプリに対して「S」を示し、制御処理の終了時に処理するアプリに対して「E」を示し、それ以外のアプリに対して「0」を示す。

【0038】

10

イベントログテーブル47Cは、AP管理ミドル51のイベント管理部51Cから送られ、イベントのログを示す。イベントログテーブル47Cは、イベント毎のエントリを含む。各エントリは、タイムスタンプと、タスク識別子と、イベント識別子と、イベント内容とを含む。タイムスタンプは、該当ログの発生時刻を表す。タスク識別子は、イベント送受信対象のアプリ61を識別するための情報である。イベント識別子は、発生イベントを識別するための情報である。イベント内容は、イベント種類がタイマーである場合に「TRi」を示し、イベント種類がシグナルである場合に「SIGj」を示す。ここでi、jは、イベント種類毎の発生順を表す整数である。

【0039】

図6は、外部通信ログフォーマットで表されたログ内容を示す。

20

【0040】

前述の外部通信ログテーブルのログ内容48AAの各エントリは、予め設定された外部通信フォーマットで表され、計算機識別子と、ロジックカウンタと、入出力識別子と、通信相手計算機識別子とを含む。計算機識別子は、対応するエントリに示された動作を実行する計算機を識別する。ロジックカウンタは、当該エントリが発生した計算機識別子と計算機Aと計算機Bと計算機Cのそれぞれにより管理される。入出力識別子は、当該エントリが入力か出力かを識別する。通信相手計算機識別子は、当該エントリの通信相手計算機を識別する。

【0041】

この例では、計算機識別子は、計算機Aであれば「a」、計算機Bは「b」、計算機Cは「c」を示す。ロジカルカウンタは、外部通信が発生するたびに1つ加算され、通信発生を順序を保証する。入出力識別子には、通信パケットの入力（受信）であれば「E」、通信パケットの出力（送信）であれば「P」が格納される。通信相手計算機識別子には、通信パケットの送信元または送信先の計算機識別子が格納される。通信パケットが後述の再現同期チェックメッセージであれば、通信相手計算機識別子には「x」が格納される。これらの情報によって、通信の因果性確認が可能である。

30

【0042】

図7は、状態再現同期制御テーブル49を示す。

【0043】

状態再現同期制御テーブル49は、計算機Aと計算機Bと計算機Cの間でのシステム状態の再現の同期のために必要な情報を格納する。状態再現同期制御テーブル49は、再現時刻TRと、復元Snapshot識別子と、リストレコード数と、同期リストとを含む。

40

【0044】

再現時刻TRは、最後の実行時刻から戻り時間RTNを減じた時刻である。復元Snapshot識別子は、Snapshot制御部44により作成されたSnapshotの中で再現時刻の直前のSnapshotである復元Snapshotの識別子である。リストレコード数は、計算機A、B、Cで発生した外部通信のログの数が格納される。

【0045】

同期リストは、復元Snapshotの作成時刻から再現時刻までの期間に、計算機A

50

、B、Cで発生した外部通信ログのエントリを含む。各エントリは、タイムスタンプと、ログ内容と、再現状態とを含む。計算機A、B、Cの夫々がSnapshotからアプリ61の状態の再現処理を行う際、同期リストのログに現れる通信の再現をするたびに全計算機の再現処理の歩調を合わせる処理を行う。それによって、全計算機の再現処理の同期が可能となる。再現状態は、状態再現制御部45によって設定される。再現状態は、対応するログが再現済みのログの場合に「1」を示し、それ以外の場合に「0」を示す。詳細処理は後述する。

【0046】

以下、各計算機10のAP管理ミドル51の動作について説明する。

【0047】

図8は、通信管理処理を示す。

【0048】

各計算機10の通信管理部42は、通信管理処理を行う。

【0049】

ステップ101：通信管理部42は、ユーザ入力から、戻り時間RTNと許容時間Tsの初期設定値を取得する。通信管理部42は更に、外部通信における送受信の順序性保証に利用するロジカルカウンタを0に設定する。

【0050】

ステップ102：通信管理部42は、Snapshot制御テーブル46へ初期設定値を設定する。

【0051】

ステップ103：通信管理部42は、Snapshot制御テーブル46の動作モードヘログモード「L」を設定する。

【0052】

ステップ104：計算機Aにおいて、アプリ61は、制御周期で動作し、AP管理ミドル51によって制御される。計算機Bにおいて、アプリ61は、別の制御周期で動作し、AP管理ミドル51によって制御される。Snapshot作成の同期とシステム再現の同期をするCoordinatorは、制御周期を持つアプリ61を実行する計算機の中から選定される。そのために、通信管理部42は、AP管理ミドル51で管理されているアプリ61の制御周期を、互いに送信しあって、計算機Aの制御周期と計算機Bの制御周期との共通制御周期LCMを算出する。例えば、通信管理部42は、計算機Aの制御周期と計算機Bの制御周期との最小公倍数を、共通制御周期LCMとして算出する。

【0053】

ステップ105：計算機Aの制御周期と計算機Bの制御周期との中の最小の制御周期を持つ計算機がCoordinatorとなる。例えば、計算機Aの制御周期が1秒であり、計算機Bの制御周期が3秒である場合、計算機AがCoordinatorとなり、計算機Bと制御周期アプリを持たない計算機Cが非Coordinatorとなる。通信管理部42は、Snapshot制御テーブル46のCoordinator情報へ、ローカルの計算機がCoordinatorである場合に「1」を、それ以外である場合に「0」を格納する。

【0054】

ステップ106：通信管理部42は、後述の同期メッセージ受信処理のためのプロセス（同期メッセージ受信処理用プロセス）を生成する。

【0055】

ステップ107：通信管理部42は、後述のSnapshot同期処理（Coordinator）又は後述のSnapshot同期受信処理（非Coordinator）のためのプロセス（Snapshot同期処理用プロセス）を生成する。

【0056】

ステップ108：通信管理部42は、後述の再現同期開始処理（Coordinator）又は後述の再現同期開始受信処理（非Coordinator）のためのプロセス（

10

20

30

40

50

再現同期開始処理用プロセス)を生成し、このフローを終了する。

【0057】

以上の通信管理処理によれば、各計算機10は、再現に必要な、定期的なSnapshotの作成と、ログの記録とを開始することができる。また、複数の計算機は、その中の一つの計算機を同期の基準となるCoordinatorに設定することができる。Coordinatorは、共通制御周期毎に、制御処理の終了時を再現同期ポイントとすることで、複数の制御アプリのアイドル状態で再現を同期させることができる。Coordinatorは、Snapshot周期毎に、制御処理の終了時にSnapshotを作成することで、複数の制御アプリに影響を与えることなく、Snapshotを作成することができる。

10

【0058】

図9は、同期メッセージ受信処理の動作を示す。

【0059】

各計算機10の通信管理部42は、前述の通信管理処理のステップ106で生成されたプロセスを用いて、同期メッセージ受信処理を行う。

【0060】

ステップ201：通信管理部42は、他計算機または自計算機からのメッセージを受信したか判定する。メッセージを受信した(Y)場合、通信管理部42は、次のステップを処理する。そうでない(N)場合、通信管理部42は、ステップ201へ戻る。

【0061】

ステップ202：通信管理部42は、受信したメッセージがSnapshot同期メッセージであるかを判定する。

20

【0062】

ステップ203：通信管理部42は、受信したメッセージがSnapshot同期メッセージであれば(Y)、後述のSnapshot同期処理を行い、ステップ201へ戻る。

【0063】

ステップ211：通信管理部42は、受信したメッセージが再現開始メッセージか判定する。

【0064】

ステップ221：通信管理部42は、受信したメッセージが再現開始メッセージであれば(Y)、後述の再現同期受信処理を行い、ステップ201へ戻る。

30

【0065】

ステップ231：通信管理部42は、受信したメッセージが再現開始メッセージでなければ(N)、後述の外部通信ログ処理を行い、ステップ201へ戻る。

【0066】

以上の同期メッセージ受信処理によれば、計算機は、受信したメッセージに応じて、Snapshot作成の同期、再現の開始、外部通信ログの作成の何れかを行うことができる。

【0067】

図10は、Snapshot同期処理を示す。

40

【0068】

Coordinatorの通信管理部42は、前述の通信管理処理のステップ107で生成されたプロセスを用いて、Snapshot同期処理を行う。Snapshot同期処理は、前述のステップ105で決定したCoordinatorとなった計算機Aにより行われる。非Coordinatorである計算機B及びCは、本処理は行わず、後述のSnapshot同期受信処理を行う。

【0069】

ステップ301：通信管理部42は、AP管理ミドル51からのAPログテーブル48Bの開始終了情報を参照することで、アプリ61の制御処理終了を待つ。アプリ61の制御処理終了時点は、アプリ61の状態が実行状態でなく、Snapshotの実行に必要な

50

な計算機資源の状態を決定できるため、システム全体の S n a p s h o t 同期ポイントとして適切である。

【0070】

ステップ302：通信管理部42は、前述のステップ104で算出した共通周期LCMが経過したか判定する。経過していなければ（N）、通信管理部42は、ステップ301に戻り次の制御処理終了を待つ。

【0071】

ステップ303：共通周期LCMが経過した場合（Y）、通信管理部42は、計算機B及びCに対して再現同期チェックメッセージを送信する。再現同期チェックメッセージの送信の時点で、制御アプリは、アイドル状態であり、状態が確定している。この時点を、再現同期ポイントと呼ぶ。計算機B及びCは、計算機Aにより本メッセージの送信の再現に従って、再現の歩調を合わせる。

【0072】

ステップ304：ステップ303の再現同期チェックメッセージは、計算機をまたがる外部通信であるため、通信管理部42は、外部通信を記録する外部通信ログ処理を行う。外部通信ログ処理については後述する。

【0073】

ステップ305：通信管理部42は、S n a p s h o t 制御テーブル46の許容時間TsをS n a p s h o t 周期とし、前回のS n a p s h o t の実行からS n a p s h o t 周期が経過したか否かを判定する。なお、通信管理部42は、許容時間から予め設定されたマージンを減ずることで、S n a p s h o t 周期を算出してもよい。経過していなければ（N）、通信管理部42は、ステップ301に戻り次の制御周期の終了を待つ。

【0074】

ステップ306：S n a p s h o t 周期が経過した場合（Y）、通信管理部42は、計算機BとCにS n a p s h o t の作成を要求するためのS n a p s h o t 同期メッセージを送信し、計算機BとCからの合意メッセージであるA C Kを待つ。

【0075】

ステップ307：通信管理部42は、自計算機のS n a p s h o t の作成を要求するためのS n a p s h o t 作成要求をS n a p s h o t 制御部44へ送信し、ステップ301に戻り、次の制御処理終了を待つ。

【0076】

以上のS n a p s h o t 同期処理によれば、C o o r d i n a t o r は、共通周期毎の制御処理の終了時を、再現同期ポイントとして各計算機10の外部通信ログに記録することができる。これにより、各計算機は、再現時に、再現同期ポイントを参照して同期することができる。更に、C o o r d i n a t o r は、制御処理の終了後に、S n a p s h o t 周期で各計算機10にS n a p s h o t を作成させることができる。

【0077】

図11は、S n a p s h o t 同期受信処理を示す。

【0078】

非C o o r d i n a t o r の通信管理部42は、前述の通信管理処理のステップ107で生成されたプロセスを用いて、C o o r d i n a t o r から送信されたS n a p s h o t 同期メッセージの受信に応じて、S n a p s h o t 同期受信処理を行う。

【0079】

ステップ401：通信管理部42は、実行制御部の処理終了を待つ。ここで通信管理部42は、A P ログテーブル48Bの開始終了情報を参照することで処理終了を判断する。各計算機10は、アプリ61の終了状態において同期的にS n a p s h o t を作成する。アプリ61の終了状態は、アプリ61の実行時に使用される計算機資源の状態が安定している状態である。例えば、アプリ61の終了状態は、通信ソケットにメッセージを書き終えた状態や、アプリ間制御メッセージをメッセージキューで送信し終えた状態や、アプリ61の処理で使用した計算機資源を解放した状態である。

【0080】

ステップ402：通信管理部42は、受信したSnapshot同期メッセージに対し、Snapshot作成同期の合意を表すACKをCoordinatorへ送信する。

【0081】

ステップ403：通信管理部42は、自計算機のSnapshotを作成するためのSnapshot作成要求をSnapshot制御部44へ送信し、このフローを終了する。

【0082】

以上のSnapshot同期受信処理によれば、非Coordinatorは、CoordinatorからのSnapshot同期メッセージに応じて、Snapshotを作成することで、Coordinatorに同期してSnapshotを作成することができる。

10

【0083】

図12は、Snapshot作成処理を示す。

【0084】

各計算機10のSnapshot制御部44は、前述のSnapshot同期処理のステップ307またはSnapshot同期受信処理のステップ403で送信されたSnapshot作成要求に応じて、Snapshotを作成するSnapshot作成処理を行う。

【0085】

ステップ501：Snapshot制御部44は、Snapshot作成要求の受信を待つ。

20

【0086】

ステップ502：Snapshot制御部44は、Snapshot作成要求を受信すると、実行部の状態をコピーする。例えば、Snapshot制御部44は、OS31のforkシステムコールを使用して、実行部の状態をコピーする。本実施例は、実行部の状態のコピー手法について限定しない。本実施例は、forkによるコピーを行う場合について、以降のステップを述べる。

【0087】

ステップ503：forkによって、実行部のプロセスのコピーがOS31によって生成される。forkによって生成されたプロセス（コピープロセス）は、もとの実行部のプロセスの識別子と異なる識別子を持ち、OS31からの別のプロセスとして扱われる。本ステップで、Snapshot制御部44は、コピープロセスを別のCPUコア（管理用コア）へ移動させる。

30

【0088】

ステップ504：Snapshot制御部44は、ステップ503で移動したコピープロセスを停止する。

【0089】

ステップ505：Snapshot制御部44は、前回のSnapshot作成処理で作成されて停止しているコピープロセスが存在するか判定する。なければ（N）、Snapshot制御部44は、ステップ501に戻る。

40

【0090】

ステップ506：Snapshot制御部44は、存在すると判定されたコピープロセスをファイルに書き出し、ファイル名として「Snapshot_Ts+n」を与える。Tsは、Snapshot周期である。nは、Snapshot作成処理実施回数を表す整数である。

【0091】

ステップ507：Snapshot制御部44は、最初のSnapshotファイルである「Snapshot0」があるか判定する。

【0092】

50

ステップ508:「Snapshot 0」があれば(Y)、Snapshot 制御部 44は、「Snapshot 0」と「Snapshot__Ts+n」の差分である「 $\Delta Ts+n$ 」を作成する。Snapshot 制御部 44は、差分と共に、そのSnapshotの識別子、Snapshotの作成時刻を格納する。

【0093】

ステップ509:Snapshot 制御部 44は、ステップ506で作成した「Snapshot__Ts+n」を削除し、ステップ501へ戻る。

【0094】

ステップ511:「Snapshot 0」がなければ(N)、Snapshot 制御部 44は、ステップ506で作成した「Snapshot__Ts+n」を「Snapshot 0」としてコピーし、ステップ509に進む。

10

【0095】

以上のSnapshot作成処理によれば、各計算機10は、実行部のプロセスのコピーをSnapshotとして作成することができる。また、各計算機10は、最初のSnapshotと最新のSnapshotの差分 $\Delta Ts+n$ を作成することにより、Snapshotに必要な容量を抑えることができる。

【0096】

図13は、外部通信ログ処理を示す。

【0097】

通信管理部42は、前述の同期メッセージ受信処理のステップ231、前述のSnapshot同期処理のステップ304、または後述のAPログ制御処理のステップ705において、外部通信ログ処理を行う。外部通信ログは、システム状態再現において同期ポイントとして使用される。

20

【0098】

ステップ601:通信管理部42は、タイムスタンプを取得する。

【0099】

ステップ602:通信管理部42は、前述のステップ101で初期化されたロジカルカウンタに1を加算する。

【0100】

ステップ603:通信管理部42は、外部通信の通信パケットから、その外部通信の送信元又は送信先を示す通信相手計算機情報を取得する。

30

【0101】

ステップ604:通信管理部42は、各ステップで取得した情報をもとに、外部通信ログフォーマットにしたがって外部通信ログを作成する。

【0102】

ステップ605:通信管理部42は、外部通信ログテーブル48Aへ外部通信ログを格納する。

【0103】

ステップ606:通信管理部42は、当該通信パケットが、Snapshot同期処理のステップ303で送信された再現同期チェックメッセージであるかを判定する。

40

【0104】

ステップ607:当該通信パケットが再現同期チェックメッセージであれば(Y)、通信管理部42は、外部通信ログの通信相手計算機識別子に対して「x」を設定し、当該通信パケットの保存を行わず、外部通信ログテーブル48Aの通信内容保存先に対してNULLを設定する。

【0105】

ステップ608:通信管理部42は、ログ管理テーブル47Aにエントリを作成し、そのエントリのログ内容に、ステップ605で格納した外部通信ログを格納し、そのエントリの照合状態に対して「0」を設定し、このフローを終了する。

【0106】

50

ステップ 6 1 1 : 再現同期チェックメッセージでなければ (N)、通信管理部 4 2 は、通信パケットの保存を行い、その保存先を外部通信ログテーブル 4 8 A の通信内容保存先に設定する。

【0 1 0 7】

以上の外部通信ログ処理によれば、各計算機 1 0 は、外部通信の度に外部通信ログを記録することができる。これにより、各計算機は、再現時に外部通信ログに基づいて同期することができる。

【0 1 0 8】

図 1 4 は、A P ログ制御処理を示す。

【0 1 0 9】

10

各計算機 1 0 の A P ログ制御部 4 3 は、A P ログ制御処理を行う。

【0 1 1 0】

ステップ 7 0 1 : A P ログ制御部 4 3 は、実行制御部からの、実行部の動作を表すログを待つ。

【0 1 1 1】

ステップ 7 0 2 : A P ログ制御部 4 3 は、S n a p s h o t 制御テーブル 4 6 の動作モードを読み取る。

【0 1 1 2】

ステップ 7 0 3 : A P ログ制御部 4 3 は、動作モードがログモード「L」であるか判定する。

20

【0 1 1 3】

ステップ 7 0 4 : 動作モードがログモード「L」であれば (Y)、A P ログ制御部 4 3 は、受信したログが、実行制御部から実行部へ受け渡す外部通信ログであるか判定する。

【0 1 1 4】

ステップ 7 0 5 : 受信したログが外部通信ログであれば (Y)、A P ログ制御部 4 3 は、前述の外部通信ログ処理を行い、ステップ 7 0 1 へ戻る。

【0 1 1 5】

ステップ 7 1 1 : 受信したログが外部通信ログでなければ (N)、A P ログ制御部 4 3 は、受信したログが、実行部の実行結果を示す A P ログであるか判定する。

【0 1 1 6】

30

ステップ 7 1 2 : 受信したログが A P ログであれば (Y)、A P ログ制御部 4 3 は、A P ログを A P ログテーブル 4 8 B へ格納し、ログ管理テーブル 4 7 A にエントリを作成し、そのエントリのログ内容に当該 A P ログを格納し、そのエントリの照合状態を「0」に設定し、ステップ 7 0 1 へ戻る。

【0 1 1 7】

ステップ 7 2 1 : 受信したログが A P ログでなければ (N)、A P ログ制御部 4 3 は、受信したログがイベントログであると判定し、そのイベントログをイベントログテーブル 4 8 C へ格納し、ログ管理テーブル 4 7 A にエントリを作成し、そのエントリのログ内容に当該イベントログを格納し、そのエントリの照合状態を「0」に設定し、ステップ 7 0 1 へ戻る。

40

【0 1 1 8】

ステップ 7 3 1 : 動作モードがログモード「L」でなければ (N)、即ち、動作モードが状態再現モードであれば、A P ログ制御部 4 3 は、後述の再現照合処理を行う。

【0 1 1 9】

以上の A P ログ制御処理によれば、各計算機 1 0 は、実行制御部により作成されたログを分類して格納することができる。

【0 1 2 0】

図 1 5 は、再現同期開始処理を示す。

【0 1 2 1】

C o o r d i n a t o r の通信管理部 4 2 は、前述の通信管理処理のステップ 1 0 8 で

50

生成されたプロセスを用いて、再現同期開始処理を行う。

【0122】

ステップ801：通信管理部42は、ユーザからの再現要求を待つ。ユーザからの再現要求はコマンドでも設定ファイルも形でもよい。再現要求は、ユーザから入力される情報として、再現時刻までの戻り時間RTNを含む。再現要求は、戻り時間RTNの代わりに、再現時刻を含んでもよい。

【0123】

ステップ802：再現要求から再現時刻までの戻り時間RTNを読み取り、戻り時間RTNと現在時刻に基づいて、状態再現同期制御テーブル49の再現時刻を設定する。

【0124】

ステップ803：通信管理部42は、Snapshot作成処理で作成されたSnapshotの中から、再現時刻直前のSnapshotの識別子を選択し、選択された識別子を状態再現同期制御テーブル49の復元Snapshot識別子に設定する。通信管理部42は、復元Snapshotを示す情報を、表示装置に表示させてもよい。

【0125】

ステップ804：通信管理部42は、復元Snapshotの作成時刻から再現時刻までの外部通信ログを、外部通信ログテーブル48Aから読み取り、読み取られた外部通信ログを状態再現同期制御テーブル49の同期リストに設定し、同期リストのエントリ数をリストレコード数へ設定する。

【0126】

ステップ805：通信管理部42は、再現開始メッセージを計算機B及びCへ送信する。

【0127】

ステップ806：通信管理部42は、再現開始合意を示すACKの受信を待つ。

【0128】

ステップ807：通信管理部42は、ACKとともに計算機B及びCから送られた状態再現同期制御テーブル49の同期リストをマージし、各ログ内容のロジカルカウンタと入出力識別子と通信相手計算機識別子から外部通信発生因果性を判定し、因果性に基づいて、マージされた同期リストを発生順に並べる。例えば、通信管理部42は、状態再現同期制御テーブル49において、計算機Aが計算機B及びCに対して再現同期チェックメッセージを送信したログ「a1P(x)」が最初のログである。その時、計算機Aのロジカルカウンタは1となる。計算機BとCは、計算機Aからの再現同期チェックメッセージの受信を示すログとして、「b1E(x)」と「c1E(x)」を外部通信ログテーブル48Aに格納する。次に、計算機Cは計算機Bに対してデータを送信し、計算機Bはそのデータを受信し、処理を行った上で、計算機Aへ制御データを送信したことが、ログから読み取れる。計算機Cの外部通信は計算機Bへのデータ送信が最後なので、計算機Cのロジカルカウンタは2である。計算機Bは計算機Cからの受信と計算機Aへの送信を行ったので、計算機Bのロジカルカウンタは3となる。計算機Aは計算機Bからの制御データの受信を行ったので、計算機Aのロジカルカウンタは2となる。

【0129】

ステップ808：通信管理部42は、ステップ807でマージされ並べ替えられた同期リストを状態再現同期制御テーブル49に再設定し、その同期リストと、同期リストの合意の要求とを示す、同期リスト合意要求メッセージとを、計算機BとCへ送信する。

【0130】

ステップ809：通信管理部42は、計算機BとCからの同期リストの合意を示すACK受信を待つ。

【0131】

ステップ810：通信管理部42は、状態再現制御部45へ状態再現制御要求を送信し、ステップ801へ戻る。

【0132】

以上の再現同期開始処理によれば、C o o r d i n a t o r は、再現開始メッセージを非C o o r d i n a t o r へ送信することで、全ての計算機の、復元S n a p s h o t から再現時刻までの外部通信ログを取得し、外部通信ログの内容を実行順に並べ替え、並べ替えた内容を同期リストとして非C o o r d i n a t o r へ送信することができる。これにより、各計算機は、同じ同期リストに基づいて、再現を同期することができる。

【0133】

図16は、再現同期開始受信処理を示す。

【0134】

非C o o r d i n a t o r の通信管理部42は、前述の通信管理処理のステップ108で生成されたプロセスを用いて、再現開始メッセージの受信に応じて、再現同期開始受信処理を行う。

10

【0135】

ステップ901：通信管理部42は、受信した状態再現同期制御テーブル49から、再現時刻及び復元S n a p s h o t 識別子を読み取る。

【0136】

ステップ902：通信管理部42は、復元S n a p s h o t の作成時刻から再現時刻までの外部通信ログを、外部通信ログテーブル48Aから読取り、状態再現同期制御テーブル49の同期リストとして設定し、その同期リストのエントリ数をリストレコード数に設定する。

【0137】

20

ステップ903：通信管理部42は、状態再現同期制御テーブル49の同期リストと再現開始メッセージに対する合意とを示すA C Kを、C o o r d i n a t o r である計算機Aへ送信する。

【0138】

ステップ904：通信管理部42は、計算機Aからの同期リスト合意要求メッセージを待つ。

【0139】

ステップ905：通信管理部42は、受信された同期リスト合意要求メッセージに含まれている同期リストを、状態再現同期制御テーブル49に再設定する。

【0140】

30

ステップ906：通信管理部42は、計算機Aへ同期リストの合意を示すA C Kを送信する。

【0141】

ステップ907：通信管理部42は、状態再現制御部45へ状態再現制御要求を送信し、このフローを終了する。

【0142】

以上の再現同期開始受信処理によれば、非C o o r d i n a t o r は、再現に必要な外部通信ログをC o o r d i n a t o r へ送信することができる。

【0143】

図17は、状態再現制御処理を示す。

40

【0144】

各計算機10の状態再現制御部45は、状態再現制御処理を行う。

【0145】

ステップ1001：状態再現制御部45は、前述の再現同期開始処理のS810または前述の再現同期開始受信処理のS907で発行される、状態再現制御要求を待つ。

【0146】

ステップ1002：状態再現制御部45は、状態再現同期制御テーブル49から再現時刻と復元S n a p s h o t 識別子を読み取る。

【0147】

ステップ1003：状態再現制御部45は、復元S n a p s h o t の差分 $\Delta T s + n$ を

50

S n a p s h o t 0に適用することで復元S n a p s h o tを作成し、復元S n a p s h o tをメモリにロードすることで、実行部を復元する。

【0148】

ステップ1004：状態再現制御部45は、外部通信ログテーブル48AとA Pログテーブル48Bとイベントログテーブル48Cとを基に、実行部の実行に必要な計算機資源の再構築を、実行制御部に要求する。これに応じて、実行制御部は、計算機資源を再構築する。本実施例では、実行制御部は、システム状態の再現のために必要な計算機資源構築機能を含む。例えば、計算機資源構築機能は、c r i uの機能を含み、アプリ61と仮想計算機81で使用されている通信ソケットやアプリ間順序制御用のI P Cの資源及びタイマーイベント、シグナルなどが再構築可能である。

10

【0149】

ステップ1005：状態再現制御部45は、後述の状態再現同期受信処理のためのプロセスを作成する。状態再現同期受信処理は、各計算機10がシステム状態の再現の同期に必要な外部通信ログの再現のために、他計算機に対して再現同期要求メッセージを送信する。その詳細は、後述する。

【0150】

ステップ1006：状態再現制御部45は、S n a p s h o t 制御テーブル46の動作モードに状態再現モード「R」を設定する。

【0151】

ステップ1007：状態再現制御部45は、実行部の処理を開始する。ステップ1003で復元されメモリにロードされたS n a p s h o t は、O S 31からみて処理停止中の一つのプロセスに見える。処理停止中のプロセスの開始処理は、例えば、c g r o u p s の機能を用いることができる。本実施例は、処理停止中のプロセスの開始方法について限定することはない。

20

【0152】

ステップ1008：状態再現制御部45は、状態再現同期処理を行う。その詳細は後述する。

【0153】

その後、状態再現制御部45は、ステップ1001に戻る。

【0154】

以上の状態再現制御処理によれば、各計算機10は、再現時刻の直前の復元S n a p s h o t を用いて実行部の状態を再現し、実行制御部の処理を開始することで、実行部の処理を開始することができる。

30

【0155】

図18は、再現照合処理を示す。

【0156】

各計算機10のA Pログ制御部43は、前述のA Pログ制御処理のステップ731において、再現照合処理を行う。即ち、状態再現モードにおいて、A Pログ制御部43が実行制御部からログを受信した場合、A Pログ制御部43は、再現照合処理を行う。

【0157】

ステップ1101：状態再現制御処理のステップ1007で処理を開始した実行部が、実行制御部により提供される機能を実行すると、A Pログ制御部43は、実行により作成されるログを、実行制御部から受信する。A Pログ制御部43は、受信されたログに対応する、ログ管理テーブル47A内のログを選択する。

40

【0158】

ステップ1102：A Pログ制御部43は、ログ管理テーブル47Aの中で、ステップ1101で選択されたログの直前のログの照合状態が「OK」であるかを判定する。照合状態が「OK」であれば、A Pログ制御部43は、システム状態が正常に再現されていると判断する。照合状態が「0」のままの場合、A Pログ制御部43は、再現順序がログと異なっているため、再現処理エラーと判断し、エラー処理を行う。エラー処理は、すべて

50

の計算機に対して再現中止メッセージを送信する。エラー処理は更に、再現のリトライやユーザへの通知等を行う。本実施例では、A P管理ミドル5 1がO S 3 1のイベントを格納して処理することから、再現の順序が入れ替わる可能性がある。そのため、各計算機1 0は、システム状態が正常に再現されているか否かを判定する。

【0 1 5 9】

ステップ1 1 0 3：ステップ1 1 0 1で選択されたログの直前のログが外部通信ログ（C O M M）であれば、A Pログ制御部4 3は、状態再現制御部4 5へ再現継続要求を送信する。再現継続要求は、選択されたログの内容を含む。ここで、直前のログが外部通信ログであり正常に処理されているので、状態再現制御部4 5は、後述の状態再現同期処理において、再現を進めることを示す再現継続要求メッセージを他計算機へ送信する。

10

【0 1 6 0】

ステップ1 1 0 4：A Pログ制御部4 3は、ステップ1 1 0 1で参照したログの直後のログが外部通信ログであるか判定する。直後のログが外部通信ログである場合、A Pログ制御部4 3は、再現処理の同期ポイントであるため、次のステップ1 1 0 5で、その外部通信ログの再現の直前に全計算機の状態再現の歩調をあわせる処理を行う。直後のログが外部通信ログでない場合、A Pログ制御部4 3は、ステップ1 1 0 9へ進む。

【0 1 6 1】

ステップ1 1 0 5：A Pログ制御部4 3は、状態再現同期制御テーブル4 9の同期リストにおいて、当該外部通信ログを検索し、その直前のログの再現状態を判定する。A Pログ制御部4 3は、その再現状態が「1」になるまで処理を待つ。A Pログ制御部4 3は、全計算機の同期リストに従って外部通信ログの順序性を確保しながら再現する必要があるため、直前のログの再現状況を判定することで、当該外部通信ログの再現を待つ。

20

【0 1 6 2】

ステップ1 1 0 6：A Pログ制御部4 3は、ステップ1 1 0 1で判定したログの直後のログが、外部通信ログでありかつ再現同期チェックメッセージを示すか判定する。

【0 1 6 3】

ステップ1 1 0 7：A Pログ制御部4 3は、S n a p s h o t制御テーブル4 6のC o o r d i n a t o r情報を読み取り、自計算機がC o o r d i n a t o rであるか判定する。前述のS n a p s h o t同期処理において再現同期チェックメッセージの送信元はC o o r d i n a t o rであるため、C o o r d i n a t o rが再現同期チェックメッセージの送信を再現する。C o o r d i n a t o rが計算機AまたはBである場合、A P管理ミドル5 1が再現同期チェックメッセージの送信を再現する。再現同期チェックメッセージは、外部通信ログに含まれており、共通周期毎の制御処理の終了時を示す再現同期ポイントを示す。C o o r d i n a t o rが計算機Cである場合、システム状態再現部4 1が再現同期チェックメッセージの送信を再現する。自計算機が非C o o r d i n a t o rである場合、A Pログ制御部4 3は、ステップ1 1 0 9へ進む。

30

【0 1 6 4】

ステップ1 1 0 8：A Pログ制御部4 3は、状態再現制御部4 5へ再現同期要求を送信する。再現同期要求は、選択されたログの内容を含む。

【0 1 6 5】

ステップ1 1 0 9：A Pログ制御部4 3は、ログ管理テーブル4 7 Aにおいて、ステップ1 1 0 1で選択されたログの照合状態へ「O K」を設定する。これは、実行制御部から送られてきたログがログ管理テーブル4 7 Aに存在し、その再現順序も正しいと判定されるためである。その後、A Pログ制御部4 3は、このフローを終了し、A Pログ制御処理へ戻る。

40

【0 1 6 6】

以上の再現照合処理によれば、各計算機1 0は、状態再現モードにおいて、A Pログ制御部4 3が再現により発生したログを実行制御部から受信した場合、ログ情報のうち受信されたログに対応するログを選択し、選択されたログの直前のログが照合済の外部通信ログであれば、選択されたログを再現済にするための再現継続要求メッセージを他計算機へ

50

送信することができる。更に、各計算機は、選択されたログの直後のログが外部通信ログであれば、選択されたログの直前のログの再現を待って、選択されたログの再現を同期するための再現同期要求メッセージを他計算機へ送信することができる。

【0167】

図19は、状態再現同期処理を示す。

【0168】

各計算機10の状態再現制御部45は、前述の状態再現制御処理のステップ1008において状態再現同期処理を行う。

【0169】

ステップ1201：状態再現制御部45は、要求を待つ。

10

【0170】

ステップ1202：状態再現制御部45は、受信された要求が、前述の再現照合処理のステップ1108で送信される再現同期要求であるか判定する。

【0171】

ステップ1203：要求が再現同期要求であれば（Y）、状態再現制御部45は、実行制御部の処理を停止する。実行制御部の処理停止方法は、システム状態再現部41との間にインタフェースを設けることで、実施する。インタフェースは、例えば、Unix Domain Socketなどで容易に実装可能である。

【0172】

ステップ1204：状態再現制御部45は、他計算機に対して再現同期要求メッセージを送信する。

20

【0173】

ステップ1205：状態再現制御部45は、他計算機から、再現同期要求メッセージに対する合意を示すACKの受信を待つ。

【0174】

ステップ1206：状態再現制御部45は、再現同期要求に含まれているログを状態再現同期制御テーブル49から探し、そのログの再現状態を「1」に設定する。

【0175】

ステップ1207：状態再現制御部45は、AP管理ミドル51または仮想化基盤71の処理を開始し、アプリ61と仮想計算機81の実行を継続し、再現時刻までの再現を継続する。その後、状態再現制御部45は、ステップ1201へ戻る。

30

【0176】

ステップ1211：要求が再現同期要求でなければ（N）、状態再現制御部45は、再現継続要求メッセージを他計算機へ送信する。その後、状態再現制御部45は、ステップ1201へ戻る。

【0177】

以上の状態再現同期処理によれば、各計算機10は、再現同期要求に応じて、実行制御部の処理を停止させ、再現同期要求メッセージを他計算機へ送信し、実行制御部の処理を再開させることができる。これにより、各計算機10は、他計算機の再現を待って、他計算機と同期することができる。また、各計算機10は、再現継続要求に応じて再現継続要求メッセージを他計算機へ送信することができる。

40

【0178】

図20は、状態再現同期受信処理を示す。

【0179】

状態再現制御部45は、前述の状態再現制御処理のステップ1005で生成されたプロセスを用いて、状態再現同期受信処理を行う。

【0180】

ステップ1301：状態再現制御部45は、他計算機からの要求メッセージを待つ。要求メッセージは、再現同期要求メッセージと再現継続要求メッセージの何れかである。

【0181】

50

ステップ 1 3 0 2 : 状態再現制御部 4 5 は、受信された要求メッセージが再現同期要求メッセージであるか判定する。

【0 1 8 2】

ステップ 1 3 0 3 : 要求メッセージが再現同期要求メッセージであれば (Y)、状態再現制御部 4 5 は、再現同期要求メッセージから要求元の計算機識別子を取得し、外部通信ログテーブル 4 8 A において、要求元の計算機識別子を持つログを探す。

【0 1 8 3】

ステップ 1 3 0 4 : 要求元の計算機識別子を持つログが存在し、かつログ管理テーブル 4 7 A において、その直前のログの照合状態が「0」であるかを判定する。

【0 1 8 4】

ステップ 1 3 0 5 : 直前のログの照合状態が「0」であれば (Y)、状態再現制御部 4 5 は、直前のログの照合状態が「OK」になるまで待つことで、再現順序を守る。

【0 1 8 5】

ステップ 1 3 0 6 : 状態再現制御部 4 5 は、実行制御部の処理を停止することで、再現の歩調を他計算機に合わせる。

【0 1 8 6】

ステップ 1 3 0 7 : 状態再現制御部 4 5 は、状態再現同期制御テーブル 4 9 において、再現同期要求メッセージに含まれているログを探し、そのログの再現状態を「1」に設定する。

【0 1 8 7】

ステップ 1 3 0 8 : 状態再現制御部 4 5 は、再現同期要求メッセージに対する合意を示す ACK を、再現同期要求メッセージの送信元へ送信する。

【0 1 8 8】

ステップ 1 3 0 9 : 状態再現制御部 4 5 は、実行制御部の処理を再開する。その後、状態再現制御部 4 5 は、ステップ 1 3 0 1 へ戻る。

【0 1 8 9】

ステップ 1 3 1 1 : 要求メッセージが再現同期要求メッセージでなければ (N)、状態再現制御部 4 5 は、要求メッセージが再現継続要求メッセージであるか判定する。全計算機は、再現継続要求メッセージを用いて、再現の正しさを確認し合いながら、再現していくことで、再現の健全性を確保する。

【0 1 9 0】

ステップ 1 3 1 2 : 要求メッセージが再現継続要求メッセージであれば (Y)、状態再現制御部 4 5 は、状態再現同期制御テーブル 4 9 において、再現継続要求メッセージに含まれているログを探し、そのログの再現状態を「1」に設定する。その後、状態再現制御部 4 5 は、ステップ 1 3 0 1 へ戻る。

【0 1 9 1】

ステップ 1 3 1 3 : 要求メッセージが再現継続要求メッセージでなければ (N)、状態再現制御部 4 5 は、再現処理エラーと判断し、エラー処理を行う。

【0 1 9 2】

以上の状態再現同期受信処理によれば、各計算機 1 0 は、他計算機からの再現同期要求メッセージに応じて、要求されたログ以前の照合を待って、ACK を返すことができる。また、各計算機 1 0 は、他計算機からの再現継続要求メッセージに応じて、要求されたログ以前の照合を待って、ACK を返すことができる。

【0 1 9 3】

計算機システムは、再現時刻の直前の Snapshot とログを用いることで、複数の計算機の状態を迅速に再現することができる。許容時間に基づいて Snapshot 周期を決定することで、復元 Snapshot による実行部の復元から、再現時刻までの、再現に要する時間を許容時間以下に抑えることができる。Snapshot 同期処理と、Snapshot 同期受信処理と、Snapshot 作成処理とによれば、計算機システムの性能の低下を防ぐことができるとともに、複数の計算機の Snapshot の作成を同

10

20

30

40

50

期させることができる。再現同期開始処理と、再現同期開始受信処理と、状態再現制御処理と、再現照合処理と、状態再現同期処理と、状態再現同期受信処理とによれば、複数の計算機の間通信のログを用いて、複数の計算機の再現の歩調を合わせると共に、通信の発生順序を保証することができる。

【0194】

計算処理は、アプリ61及び仮想計算機81の処理等に対応する。第1計算機は、Coordinatorの計算機等に対応する。第1計算処理は、制御アプリの処理等に対応する。作成同期処理は、Snapshot同期処理等に対応する。作成処理は、Snapshot作成処理等に対応する。記録処理は、外部通信ログ処理、APログ制御処理等に対応する。再現開始処理は、再現同期開始処理等に対応する。特定時刻は、再現時刻等に対応する。再現処理は、状態再現制御処理、再現照合処理、状態再現同期処理、状態再現同期受信処理等に対応する。作成条件は、ステップ301、302、305等の条件に対応する。制御計算機は、制御アプリを実行する計算機等に対応する。選択条件は、ステップ105の条件等に対応する。作成情報は、Snapshot同期メッセージ等に対応する。復元スナップショット情報は、再現開始メッセージ、復元Snapshot識別子等に対応する。順序情報は、同期リスト、同期リスト合意要求メッセージ等に対応する。再現同期情報は、再現同期要求メッセージ、再現継続要求メッセージ等に対応する。同期情報は、再現同期チェックメッセージ等に対応する。ミドルウェアは、システム状態再現部41等に対応する。管理プログラムは、AP管理ミドル51、仮想化基盤71等に対応する。

10

20

【0195】

以上、本発明の実施形態を説明したが、これは本発明の説明のための例示であって、本発明の範囲を上記構成に限定する趣旨ではない。本発明は、他の種々の形態でも実施する事が可能である。

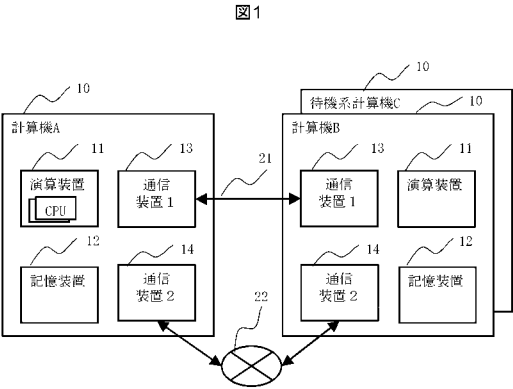
【符号の説明】

【0196】

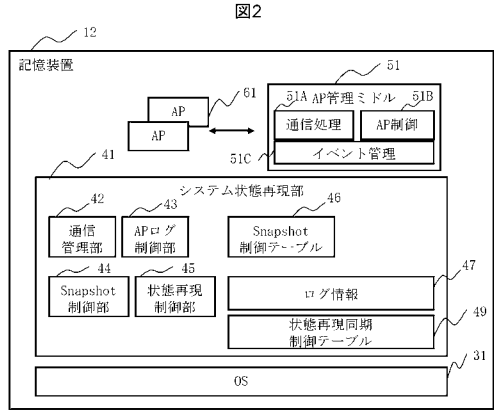
10…計算機、 11…演算装置、 12…記憶装置、 13…通信装置、 14…通信装置、 21…内部ネットワーク、 22…外部ネットワーク、 41…システム状態再現部、 51…AP管理ミドル、 61…アプリ、 71…仮想化基盤、 81…仮想計算機

30

【図 1】



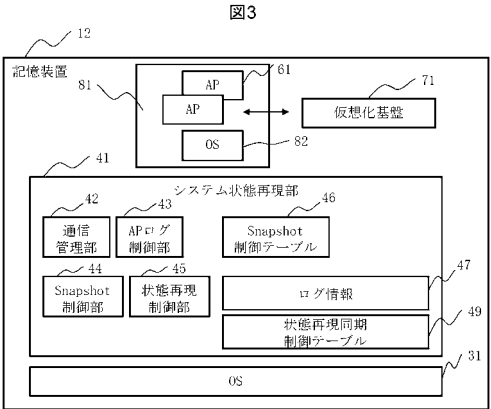
【図 2】



【図 5】



【図 3】



【図 4】



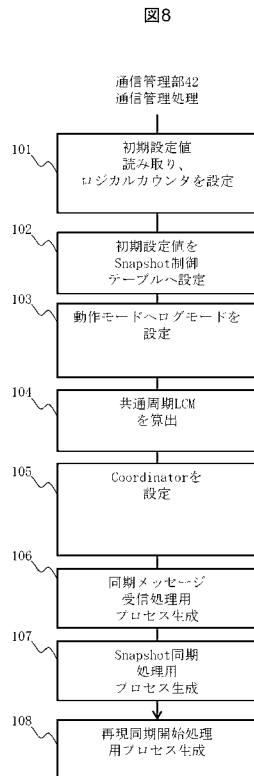
【図 6】



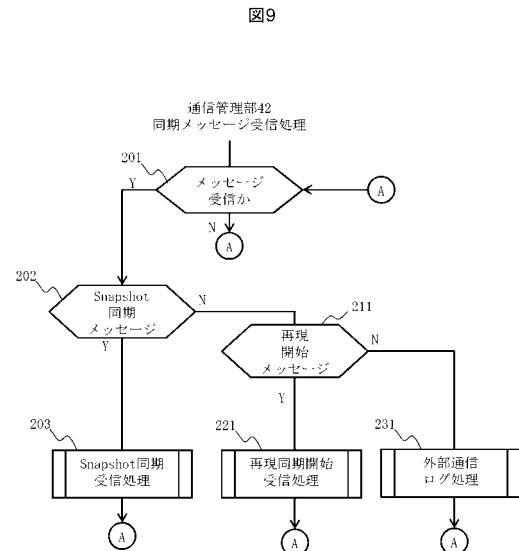
【図 7】



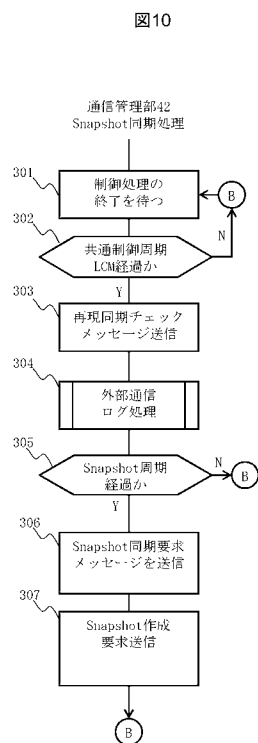
【図 8】



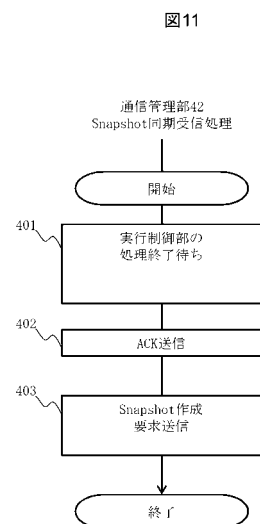
【図 9】



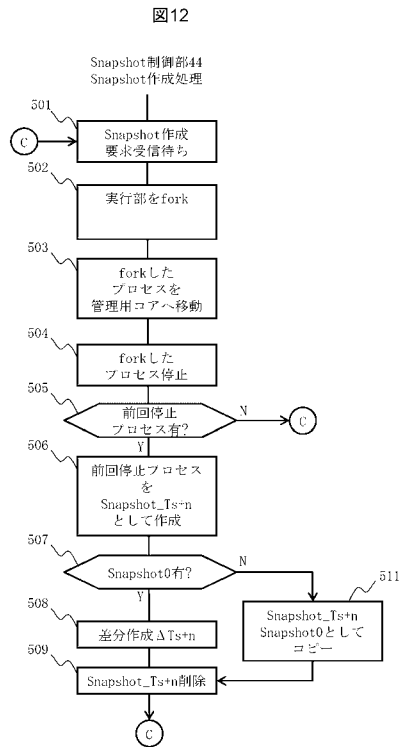
【図 10】



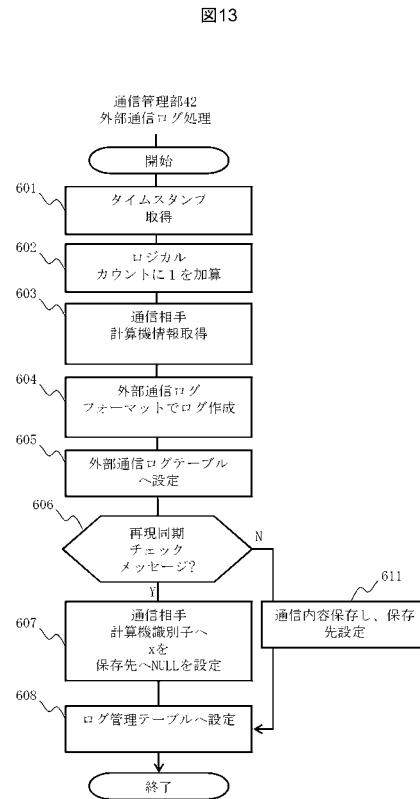
【図 11】



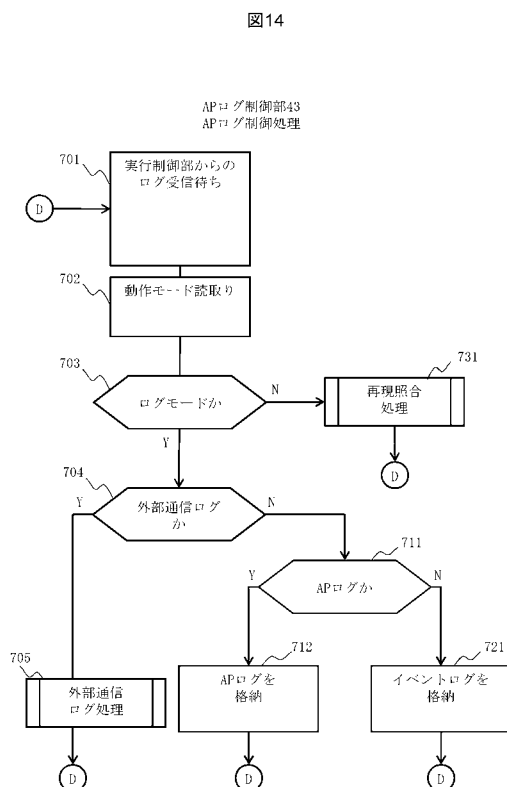
【図 1 2】



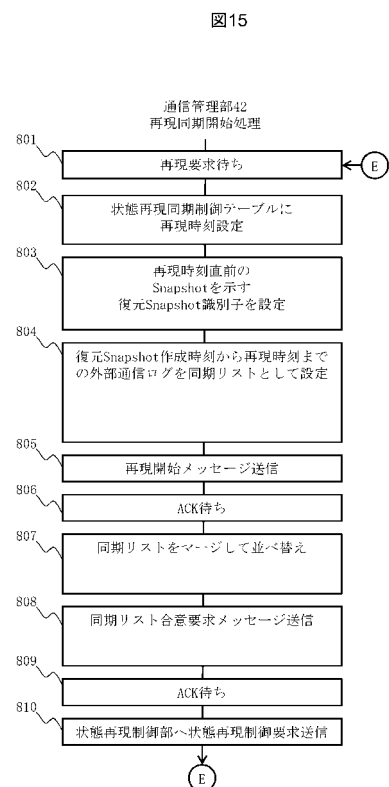
【図 1 3】



【図 1 4】



【図 1 5】



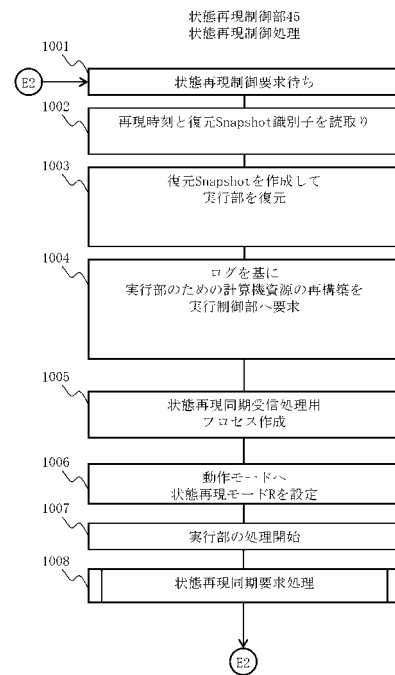
【図 16】

図16



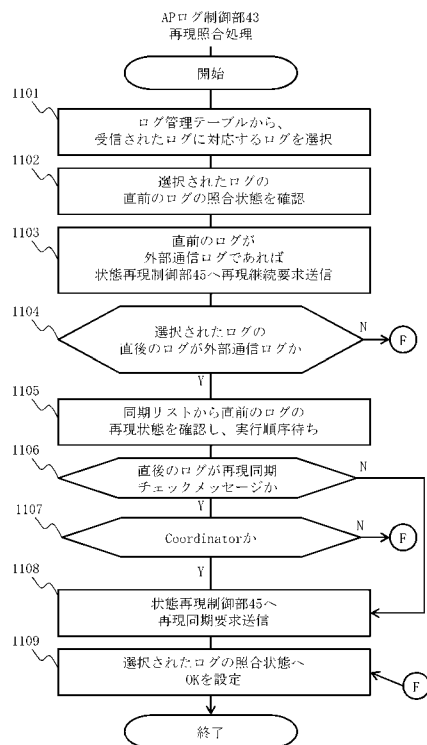
【図 17】

図17



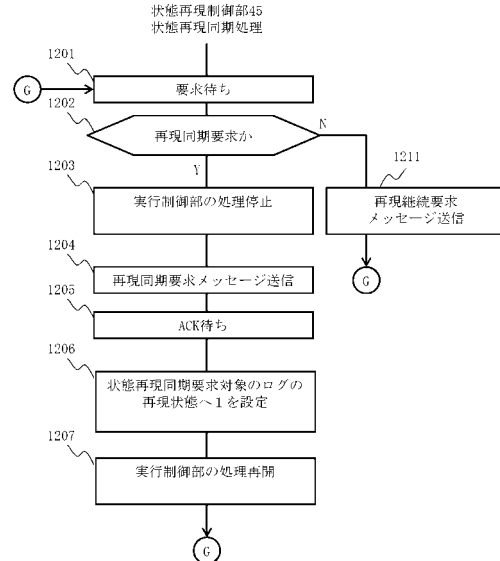
【図 18】

図18



【図 19】

図19



【図 20】

図20

