

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7181663号
(P7181663)

(45)発行日 令和4年12月1日(2022.12.1)

(24)登録日 令和4年11月22日(2022.11.22)

(51)国際特許分類

G 0 6 F 9/50 (2006.01)

F I

G 0 6 F

9/50

1 5 0 Z

請求項の数 5 (全20頁)

(21)出願番号	特願2019-3225(P2019-3225)	(73)特許権者	000005223 富士通株式会社 神奈川県川崎市中原区上小田中4丁目1 番1号
(22)出願日	平成31年1月11日(2019.1.11)	(74)代理人	100121083 弁理士 青木 宏義
(65)公開番号	特開2020-113032(P2020-113032 A)	(74)代理人	100138391 弁理士 天田 昌行
(43)公開日	令和2年7月27日(2020.7.27)	(74)代理人	100074099 弁理士 大菅 義之
審査請求日	令和3年10月7日(2021.10.7)	(74)代理人	100133570 弁理士 德 永 民雄
		(72)発明者	堀井 基史 神奈川県川崎市中原区上小田中4丁目1 番1号 富士通株式会社内

最終頁に続く

(54)【発明の名称】 通信装置、通信プログラム、および分散処理方法

(57)【特許請求の範囲】

【請求項1】

実行順序が予め指定された複数の部分処理を含む目的処理を実行する複数の通信装置の中の1つの通信装置において使用される通信プログラムであって、

前記複数の通信装置は、それぞれ、前記実行順序に従って前記複数の部分処理を処理するよう構成されており、

前記複数の通信装置は、それぞれ、前記複数の部分処理の状態を表す情報を記録する証跡を備えており、

前記複数の通信装置が備える証跡の内容は互いに同期しており、

当該通信装置の証跡を参照して、前記複数の部分処理のうちから、最新に完了した部分処理の次の部分処理である未完了部分処理を選択し、

前記未完了部分処理について当該通信装置の証跡に記録されている、前記複数の通信装置のうちの異なる通信装置により得られている同じ実行結果の数が所定の目標数に達しているかを判定し、

前記同じ実行結果の数が前記目標数に達していないときに、前記未完了部分処理を実行し、前記未完了部分処理の実行結果を当該通信装置の証跡に記録する処理をプロセッサに実行させる通信プログラム。

【請求項2】

前記未完了部分処理の実行結果は、他の各通信装置に送信され、

前記未完了部分処理の実行結果について前記複数の通信装置による合意が形成された後

に、前記未完了部分処理について前記複数の通信装置により合意が形成された実行結果を当該通信装置が備える証跡に記録する処理を前記プロセッサにさらに実行させることを特徴とする請求項1に記載の通信プログラム。

【請求項3】

同じ実行結果の数が目標数よりも少ない部分処理が存在しないときは、前記目的処理の実行結果を出力する

処理をプロセッサにさらに実行させる請求項1に記載の通信プログラム。

【請求項4】

複数の通信装置を用いて実行順序が予め指定された複数の部分処理を含む目的処理を実行する分散処理システムにおいて使用される、前記複数の通信装置の中の1つの通信装置であって、

前記複数の通信装置は、それぞれ、前記実行順序に従って前記複数の部分処理を処理するように構成されており、

前記複数の通信装置は、それぞれ、前記複数の部分処理の状態を表す情報を記録する証跡を備えており、

前記複数の通信装置が備える証跡の内容は互いに同期しており、

当該通信装置の証跡を参照して、前記複数の部分処理のうちから、最新に完了した部分処理の次の部分処理である未完了部分処理を選択する選択部と、

前記未完了部分処理について当該通信装置の証跡に記録されている、前記複数の通信装置のうちの異なる通信装置により得られている同じ実行結果の数が所定の目標数に達しているかを判定する判定部と、

前記同じ実行結果の数が前記目標数に達していないときに、前記選択部により選択された未完了部分処理を実行する実行部と、

前記実行部により得られる実行結果を当該通信装置の証跡に記録する証跡管理部と、備える通信装置。

【請求項5】

複数の通信装置を含む分散処理システムにおいて実行順序が予め指定された複数の部分処理を含む目的処理を実行する分散処理方法であって、

前記複数の通信装置は、それぞれ、前記実行順序に従って前記複数の部分処理を処理するように構成されており、

前記複数の通信装置は、それぞれ、前記複数の部分処理の状態を表す情報を記録する証跡を備えており、

前記複数の通信装置が備える証跡の内容は互いに同期しており、

各通信装置は、

当該通信装置の証跡を参照して、前記複数の部分処理のうちから、最新に完了した部分処理の次の部分処理である未完了部分処理を選択し、

前記未完了部分処理について当該通信装置の証跡に記録されている、前記複数の通信装置のうちの異なる通信装置により得られている同じ実行結果の数が所定の目標数に達しているかを判定し、

前記同じ実行結果の数が前記目標数に達していないときに、前記未完了部分処理を実行し、

前記未完了部分処理の実行結果を当該通信装置の証跡に記録することを特徴とする分散処理方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、通信装置、通信プログラム、および分散処理方法に係わる。

【背景技術】

【0002】

大型のコンピュータを用いることなく大容量の計算を実行するための技術の1つとして

10

20

30

40

50

分散コンピューティングが提案されている。分散コンピューティングにおいては、プログラムが複数のプログラム部分に分割され、複数のプログラム部分が複数のコンピュータ上で実行される。このとき、複数のコンピュータは、ネットワークを介して互いに通信を行いながら、全体として処理を進行させる。

【0003】

分散コンピューティングの一例として、ボランティアコンピューティングが実用化されている。ボランティアコンピューティングにおいては、分散して存在する余剰計算資源を活用して処理が実行される。例えば、地球外知的生命体探索のためにボランティアコンピューティングが利用されることがある。

【0004】

図1は、従来の分散処理システムの一例を示す。この例では、分散処理システムは、制御ノードコンピュータおよび複数の実行ノードコンピュータを備える。なお、分散処理システムは、分散コンピューティングにより実現される。

10

【0005】

各実行ノードコンピュータは、分散処理システムに余剰計算資源を提供する。制御ノードコンピュータは、ユーザから与えられたアプリケーションプログラムを複数のプログラム部分に分割して実行ノードコンピュータに分配する。このとき、制御ノードコンピュータは、どの実行ノードコンピュータにどのプログラム部分を実行させるのかを決定する。各実行ノードコンピュータは、与えられたプログラム部分を実行し、その計算結果を制御ノードコンピュータに送信する。そして、制御ノードコンピュータは、各実行ノードコンピュータによる計算結果を収集して検証する。

20

【0006】

分散コンピューティングまたはボランティアコンピューティングのプラットフォームの1つとして、BOINC (Berkeley Open Infrastructure for Network Computing) が知られている。また、特許文献1～2に関連技術が記載されている。

【先行技術文献】

【特許文献】

【0007】

【文献】WO 2018 / 142948号

特開2015-146165号公報

30

【発明の概要】

【発明が解決しようとする課題】

【0008】

従来の分散処理システムにおいては、図1を参照しながら説明したように、制御ノードコンピュータが複数の実行ノードコンピュータの処理を制御する。したがって、制御ノードコンピュータに障害が発生すると、アプリケーションの実行を継続できなくなる。すなわち、従来の分散処理システムの信頼性（特に、耐障害性）は低い。また、各実行ノードコンピュータの計算結果は、制御ノードコンピュータにより検証される。換言すれば、制御ノードコンピュータは、計算結果を改ざんできる。したがって、この観点においても、従来の分散処理システムの信頼性は低い。

40

【0009】

本発明の1つの側面に係わる目的は、分散処理システムの信頼性を向上させることである。

【課題を解決するための手段】

【0010】

本発明の1つの態様の通信プログラムは、複数の部分処理を含む目的処理を実行する複数の通信装置の中の1つの通信装置において使用される。この通信プログラムにより提供される方法は、前記複数の部分処理の状態を表す証跡を参照して、前記複数の部分処理のうちから、同じ実行結果の数が目標数よりも少ない未完了部分処理を選択し、前記未完了部分処理を実行し、前記未完了部分処理の実行結果を前記証跡に記録する。

50

【発明の効果】

【0011】

上述の態様によれば、分散処理システムの信頼性が向上する。

【図面の簡単な説明】

【0012】

【図1】従来の分散処理システムの一例を示す図である。

【図2】本発明の実施形態に係わる分散処理システムの一例を示す図である。

【図3】実行ノードコンピュータの処理の一例を示すフローチャートである。

【図4】分散処理システムによる処理の流れの一例を示す図（その1）である。

【図5】分散処理システムによる処理の流れの一例を示す図（その2）である。

【図6】分散処理システムによる処理の流れの一例を示す図（その3）である。

【図7】分散処理システムの構成および機能の一例を示す図である。

【図8】処理一覧表の一例を示す図である。

【図9】実行ノード一覧表の一例を示す図である。

【図10】証跡の一例を示す図である。

【図11】開始フェーズのシーケンスの一例を示す図である。

【図12】実行フェーズのシーケンスの一例を示す図である。

【図13】合意フェーズのシーケンスの一例を示す図である。

【図14】終了フェーズのシーケンスの一例を示す図である。

【図15】証跡を更新する処理の一例を示すフローチャートである。

【図16】通信装置として動作するコンピュータのハードウェア構成の一例を示す図である。

【図17】分散処理システムの構成例を示す図である。

【発明を実施するための形態】

【0013】

図2は、本発明の実施形態に係わる分散処理システムの一例を示す。本発明の実施形態に係わる分散処理システム100は、複数の実行ノードコンピュータ1（1A～1D）を含む。実行ノードコンピュータ1A～1Dは、互いに通信可能に接続されている。実行ノードコンピュータ1A～1Dは、光ファイバリンクで接続されてもよいし、電気回線で接続されてもよいし、無線リンクで接続されてもよい。なお、各実行ノードコンピュータ1（1A～1D）は、本発明の実施形態に係わる通信装置の一例である。

【0014】

分散処理システム100に与えられるアプリケーションの処理は、複数の部分処理から構成される。例えば、分散処理システム100に与えられるアプリケーションの処理は、予め複数の部分処理に分割されている。或いは、分散処理システム100は、与えられたアプリケーションの処理を複数の部分処理に分割してもよい。

【0015】

各実行ノードコンピュータ1A～1Dは、複数の部分処理の中から実行すべき部分処理を自律的に選択し、選択した部分処理を実行する。そして、各実行ノードコンピュータ1A～1Dは、選択した部分処理の実行結果を証跡に記録する。証跡は、例えば、選択した部分処理を実行したノードを識別する情報、選択した部分処理を識別する情報、選択した部分処理についての実行結果、および選択した部分処理についての実行が終了した時刻を表すタイムスタンプを含む。また、証跡は、この実施例では、各ノード内に保存される。ただし、あるノードにおいて証跡が更新されると、更新された内容は、ネットワーク内の全ノードに通知される。したがって、証跡は、実質的に、実行ノードコンピュータ1A～1Dにより共有される。或いは、証跡は、ネットワーク内の全実行ノードコンピュータ1A～1Dが閲覧可能な領域に保存されるようにしてよい。

【0016】

実行ノードコンピュータ1A～1Dは、部分処理ごとに実行結果について合意を形成する。合意は、各実行ノードコンピュータ1A～1Dにより記録される証跡の内容（たとえ

10

20

30

40

50

ば、処理内容および処理順番)が互いに同じであることを確認することで実現される。この結果、各ノードに保存される証跡が互いに同じであることが保証される。

【0017】

複数の部分処理は、例えば、予め決められた順番で実行されるように指定されている。この場合、ある部分処理について実行ノードコンピュータ1 A ~ 1 Dにより合意が形成されると、次の部分処理が実行される。そして、すべての部分処理について合意が形成されると、分散処理システム100は、与えられたアプリケーションについての実行結果を出力する。

【0018】

図3は、実行ノードコンピュータ1の処理の一例を示すフローチャートである。このフローチャートの処理は、複数の部分処理を含むアプリケーションが分散処理システム100に与えられたときに実行される。

10

【0019】

S1において、実行ノードコンピュータ1は、自ノードに保存されている証跡を参照し、複数の部分処理の中から実行すべき部分処理を選択する。なお、各部分処理について、実行すべき順番が予め指定されているものとする。この場合、完了していない部分処理の中で、最も早く実行されるべき部分処理が選択される。

【0020】

S2において、実行ノードコンピュータ1は、自ノードに保存されている証跡を参照し、選択した部分処理についての実行結果の個数(現状の回答数)が目標数(必要回答数)に達しているか否かを判定する。このとき、同じ実行結果の個数が目標数に達しているか否かが判定される。なお、目標数(必要回答数)は、分散処理システム100において予め決められているものとする。

20

【0021】

S1で選択した部分処理についての現状の回答数が必要回答数に達していれば、実行ノードコンピュータ1は、選択した部分処理が完了しており、且つ、その実行結果について合意が形成されていると判定する。この場合、実行ノードコンピュータ1の処理はS1に戻る。すなわち、次の部分処理が選択される。一方、S1で選択した部分処理についての現状の回答数が必要回答数より少なければ、実行ノードコンピュータ1の処理はS3に進む。

30

【0022】

S3において、実行ノードコンピュータ1は、選択した部分処理の1つ前の部分処理の実行結果を利用して、選択した部分処理を実行する。なお、選択した部分処理の1つ前の部分処理の実行結果は、証跡に記録されている。

【0023】

S4において、実行ノードコンピュータ1は、選択した部分処理についての実行結果を証跡に記録する。ただし、実行ノードコンピュータ1は、実行結果について複数またはすべての実行ノードコンピュータ1による合意が得られた後に、その実行結果を証跡に記録することが好ましい。また、実行ノードコンピュータ1は、現状の回答数が必要回答数より少ないとときに限って、選択した部分処理についての実行結果を証跡に記録することが好ましい。更新された証跡は、分散処理システム100内の全実行ノードコンピュータ1により共有される。

40

【0024】

S5において、実行ノードコンピュータ1は、自ノードに保存されている証跡を参照し、すべての部分処理が完了したか否かを判定する。そして、完了していない部分処理が残つていれば、実行ノードコンピュータ1の処理はS1に戻る。すなわち、次の部分処理が選択される。一方、すべての部分処理が完了していれば、実行ノードコンピュータ1の処理は終了する。この後、分散処理システム100は、実行結果を出力する。

【0025】

図4～図6は、分散処理システム100による処理の流れの一例を示す。この実施例で

50

は、分散処理システム 100 は、実行ノードコンピュータ 1A ~ 1C を備える。分散処理システム 100 に与えられるアプリケーションは、部分処理 1、部分処理 2、部分処理 3 を含む。部分処理 1 が最初に実行され、部分処理 1 の次に部分処理 2 が実行され、部分処理 3 が最後に実行される、ことが予め決められている。なお、図 4 ~ 図 6 では、部分処理 1、部分処理 2、部分処理 3 は、それぞれ「処理 1」、「処理 2」、「処理 3」と表記されている。

【0026】

各実行ノードコンピュータ 1A ~ 1C に保存される証跡は、互いに同期しているものとする。すなわち、実行ノードコンピュータ 1A ~ 1C がそれぞれ参照する証跡の内容は互いに同じである。証跡は、与えられたアプリケーションに含まれる複数の部分処理の状態を表す。図 4 ~ 図 6 に示す例では、証跡は、部分処理のリストおよび各部分処理についての実行状態を表す情報を含む。部分処理のリストにおいては、部分処理 1、部分処理 2、部分処理 3 が順番に並べられている。各部分処理についての実行状態を表す情報は、部分処理を実行したノードを識別する情報および実行結果を含む。

10

【0027】

例えば、図 4 (a) に示す証跡は、部分処理 1 について、実行ノードコンピュータ 1A による実行結果が「100」であり、実行ノードコンピュータ 1C による実行結果も「100」である状態を表している。また、この証跡は、部分処理 2 について、実行ノードコンピュータ 1B による実行結果が「150」である状態を表している。

20

【0028】

以下、図 3 に示すフローチャートを参照しながら、図 4 (a) に示す状態の後の実行ノードコンピュータ 1 の処理を説明する。なお、この実施例では、上述した目標数（必要回答数）は「2」である。この場合、ある部分処理について 2 個以上の同じ実行結果が得られていれば、その実行結果について分散処理システム 100 内で合意が形成されていると判定される。即ち、ある部分処理について 2 個以上の同じ実行結果が得られていれば、その部分処理は正しく完了していると判定される。

30

【0029】

S1：各実行ノードコンピュータ 1 は、自ノードに保存されている証跡を参照し、部分処理 1 ~ 3 の中から実行すべき部分処理を選択する。ここで、部分処理 1 に対して、2 個の同じ実行結果が記録されている。よって、各実行ノードコンピュータ 1A ~ 1C は、証跡を参照することにより、部分処理 1 は既に完了していると判定する。よって、各実行ノードコンピュータ 1A ~ 1C は、部分処理 1 の次の部分処理（すなわち、部分処理 2）について証跡を参照する。

30

【0030】

例えば、実行ノードコンピュータ 1A は、図 4 (a) に示すように、部分処理 2 を選択する。同様に、実行ノードコンピュータ 1C も、部分処理 2 を選択する。なお、各実行ノードコンピュータ 1A ~ 1C は、自律的に動作するので、複数の実行ノードコンピュータ 1 が同時またはほぼ同時に同じ部分処理を選択することがある。

40

【0031】

S2：実行ノードコンピュータ 1A、1C は、図 4 (b) に示すように、それぞれ部分処理 2 についての同じ実行結果の個数（現状の回答数）が目標数（必要回答数）に達しているか否かを判定する。この例では、部分処理 2 に対して 1 個の実行結果が記録されている。すなわち、現状の回答数は、必要回答数よりも少ない。よって、実行ノードコンピュータ 1A、1C の処理は、それぞれ S3 に進む。

40

【0032】

S3：実行ノードコンピュータ 1A、1C は、図 5 (a) に示すように、それぞれ部分処理 2 を実行する。このとき、部分処理 2 の 1 つ前の部分処理（すなわち、部分処理 1）の実行結果を利用して、部分処理 2 が実行される。すなわち、実行ノードコンピュータ 1A、1C は、それぞれ「部分処理 1 の実行結果 = 100」を利用して部分処理 2 を実行する。

50

【 0 0 3 3 】

このように、実行ノードコンピュータ 1 A、1 C は、それぞれ部分処理 2 を実行する。ここで、実行ノードコンピュータ 1 C による部分処理 2 の実行を終了する前に、実行ノードコンピュータ 1 A による部分処理 2 の実行を終了したものとする。

【 0 0 3 4 】

この場合、実行ノードコンピュータ 1 A は、証跡を参照し、部分処理 2 の状態を確認する。このとき、部分処理 2 に対して 1 個の実行結果が記録されている。すなわち、現状の回答数は、必要回答数よりも少ない。よって、実行ノードコンピュータ 1 A は、部分処理 2 の実行結果を証跡に記録する。この例では、図 5 (b) に示すように、実行ノードコンピュータ 1 A により「部分処理 2 の実行結果 = 1 5 0 」が証跡に書き込まれる。

10

【 0 0 3 5 】

続いて、実行ノードコンピュータ 1 C は、部分処理 2 の実行が終了すると、証跡を参照し、部分処理 2 の状態を確認する。ところが、この時点では、部分処理 2 に対して 2 個の実行結果が記録されている。また、これら 2 個の実行結果は、互いに同じである。すなわち、部分処理 2 についての現状の回答数が必要回答数に達している。したがって、実行ノードコンピュータ 1 C は、部分処理 2 が正しく完了していると判定し、部分処理 2 の実行結果を証跡に記録しない。

【 0 0 3 6 】

この後、同様の手順で部分処理 3 が実行され、図 6 に示す証跡が作成される。そして、すべての部分処理 1 ~ 3 が完了すると、分散処理システム 100 は、最終的な実行結果を出力する。すなわち、「実行結果 = 2 0 0 」が出力される。

20

【 0 0 3 7 】

このように、分散処理システム 100 は、各実行ノードコンピュータ 1 を制御する制御ノードコンピュータなしで分散処理を実現する。よって、図 1 に示す構成（すなわち、各実行ノードコンピュータを制御する制御ノードコンピュータを必要とする構成）と比較すると、耐障害性が高くなる。例えば、分散処理システム 100 においては、複数の実行ノードコンピュータ 1 のうちの幾つかが故障しても、他の実行ノードコンピュータ 1 により処理を継続することが可能である。

【 0 0 3 8 】

また、各部分処理について複数の実行ノードコンピュータ 1 により複数の実行結果が生成され、所定数（目標数または必要回答数）以上の実行結果が一致したときに、その部分処理が完了する。したがって、1 または少数の実行ノードコンピュータ 1 が不正または改ざんを行うことは困難である。或いは、1 または少数の実行ノードコンピュータ 1 が悪意あるユーザに乗っ取られても、誤った実行結果が出力されることはない。

30

【 0 0 3 9 】**< 実施例 >**

図 7 は、分散処理システム 100 の構成および機能の一例を示す。この実施例では、分散処理システム 100 は、要求ノードコンピュータ 3 および複数の実行ノードコンピュータ 1 (1 A、1 B) を備える。なお、分散処理システム 100 は、要求ノードコンピュータ 3 を含まなくてもよい。すなわち、要求ノードコンピュータ 3 は、分散処理システム 100 の外部に設けられてもよい。また、実行ノードコンピュータ 1 の数は、特に限定されるものではない。

40

【 0 0 4 0 】

要求ノードコンピュータ 3 は、処理管理部 31、要求部 32、検証部 33、通信部 34 を備える。ただし、要求ノードコンピュータ 3 は、図 7 に示していない他の機能を備えていてもよい。

【 0 0 4 1 】

処理管理部 31 は、処理一覧表を用いて、実行ノードコンピュータ 1 に実行を要求すべきアプリケーションを管理する。なお、実行ノードコンピュータ 1 に実行を要求すべきアプリケーションは、例えば、ユーザから与えられる。

50

【 0 0 4 2 】

図 8 は、処理一覧表の一例を示す。この実施例では、処理一覧表には、アプリ名、部分処理リスト、処理内容コード、必要回答数などが登録される。アプリ名は、ユーザから受け付けたアプリケーションの名称（または、識別情報）を表す。部分処理リストは、ユーザから受け付けたアプリケーションに含まれる複数の部分処理の名称（または、識別別情報）を表す。この実施例では、動画像処理アプリケーションが、部分処理として、圧縮処理、音声加工処理、モザイク加工処理、字幕追加処理などを含んでいる。処理内容コードは、対応する部分処理の内容を記述する。必要回答数は、部分処理の完了までに必要な同じ実行結果の数を表す。例えば、圧縮処理の必要回答数は「3」である。この場合、互いに一致する3個の実行結果が得られるまで、異なる実行ノードコンピュータ1により圧縮処理が実行される。

10

【 0 0 4 3 】

なお、この実施例では、例えば、ユーザがアプリケーションを分割して複数の部分処理を定義する。この場合、各部分処理の必要回答数もそれぞれユーザにより定義される。

【 0 0 4 4 】

要求部32は、実行ノードコンピュータ1に、処理一覧表に登録されているアプリケーションの実行を要求する。ここで、要求部32は、実行ノード一覧表を備える。実行ノード一覧表には、図9に示すように、分散処理システム100内で動作する実行ノードコンピュータ1が登録されている。この実施例では、実行ノード一覧表に実行ノードコンピュータ1A、1B、1Cなどが登録されている。また、実行ノード一覧表には、各実行ノードコンピュータ1にアクセスするための情報も登録されている。

20

【 0 0 4 5 】

検証部33は、実行ノードコンピュータ1に実行を要求したアプリケーションが正しく完了したか否かを検証する。通信部34は、ネットワークを介して各実行ノードコンピュータ1と通信を行う。

【 0 0 4 6 】

実行ノードコンピュータ1は、証跡管理部11、合意形成部12、選択部13、実行部14、通信部15を備える。ただし、実行ノードコンピュータ1は、図7に示していない他の機能を備えていてもよい。

30

【 0 0 4 7 】

証跡管理部11は、実行ノードコンピュータ1が要求ノードコンピュータ3からアプリケーションの実行要求を受信したときに、そのアプリケーションに係わる情報を証跡に記録する。また、証跡管理部11は、実行ノードコンピュータ1による実行結果を証跡に記録する。

【 0 0 4 8 】

図10は、証跡の一例を示す。証跡は、この実施例では、実行対象アプリケーション毎に作成される。引数は、実行対象アプリケーションにより処理されるデータを指示する。要求ノード名は、実行ノードコンピュータ1に対してアプリケーションの実行を要求したノードを識別する。部分処理リスト、処理内容コード、必要回答数は、図8に示す処理一覧表および図10に示す証跡において実質的に同じなので、説明を省略する。

40

【 0 0 4 9 】

有効回答数は、対応する部分処理について得られている、互いに一致する実行結果（即ち、回答）の数を表す。状態は、対応する部分処理が完了しているか否かを表す。実行結果は、対応する部分処理についての実行ノードコンピュータ1による実行結果を表す。なお、実行結果は、情報量を削減するために、例えば、ハッシュ値で表される。実行ノードは、対応する部分処理を実行したノードを識別する。実行時刻は、対応する部分処理が実行された時刻を表す。

【 0 0 5 0 】

例えば、動画処理アプリケーションに含まれる圧縮処理は、実行ノードコンピュータ1A、1B、1Cにより実行されている。ここで、実行ノードコンピュータ1A、1B、1

50

Cによる実行結果は、互いに同じである。したがって、有効回答数は「3」である。この場合、有効回答数が必要回答数に達しているので、圧縮処理の状態は「完了」である。

【0051】

モザイク加工処理は、実行ノードコンピュータ1D、1B、1Aにより実行されている。ところが、実行ノードコンピュータ1D、1Aによる実行結果は互いに一致するが、実行ノードコンピュータ1Bによる実行結果は、他の2つの実行結果とは異なっている。したがって、有効回答数は「2」である。この場合、有効回答数は必要回答数より少ないので、モザイク加工処理の状態は「未完了」である。

【0052】

合意形成部12は、他のノードの合意形成部12と連携して、実行ノードコンピュータ1による実行結果についての合意の形成を試みる。なお、合意形成部12は、要求ノードコンピュータ3の要求部32と同様に、図9に示す実行ノード一覧表を備える。

10

【0053】

選択部13は、実行対象アプリケーションに含まれる部分処理のうちから、完了していない部分処理を選択する。すなわち、選択部13は、自ノードの証跡を参照し、複数の部分処理のうちから、同じ実行結果の数が必要回答数よりも少ない未完了部分処理を選択する。例えば、図10に示す例では、圧縮処理および音声加工処理は完了しているが、モザイク可能処理および字幕追加処理は未だ完了していない。よって、この場合、選択部13は、モザイク加工処理または字幕追加処理を選択する。ただし、各部分処理の実行順序が予め指定されているときは、選択部13は、完了していない部分処理のうちで、実行順番が最も早い部分処理を選択する。

20

【0054】

実行部14は、選択部13により選択された部分処理を実行する。尚、選択部13および実行部14は、一体的に動作してもよい。この場合、選択部13および実行部14は、完了していない部分処理を選択して実行する。通信部15は、ネットワークを介して、要求ノードコンピュータ3および他の実行ノードコンピュータ1と通信を行う。

【0055】

図11～図14は、分散処理システム100が与えられたアプリケーションを実行するときのシーケンスの一例を示す。このシーケンスは、図11に示す開始フェーズ、図12に示す実行フェーズ、図13に示す合意フェーズ、および図14に示す終了フェーズを含む。なお、このシーケンスが開始する前に、実行すべきアプリケーションがユーザから要求ノードコンピュータ3に与えられているものとする。この場合、このアプリケーションは、処理管理部31が管理する処理一覧表に登録される。また、以下の記載では、分散処理システム100に与えられるアプリケーションを「目的アプリケーション」または「目的処理」と呼ぶことがある。

30

【0056】

図11は、開始フェーズのシーケンスの一例を示す。開始フェーズは、例えば、ユーザから入力される指示に応じて、要求ノードコンピュータ3において起動される。

【0057】

S11において、要求部32は、処理管理部31に対して目的アプリケーションに係わる処理一覧表を要求する。この要求は、目的アプリケーションの名称を表す情報（又は、目的アプリケーションの識別情報）を含む。

40

【0058】

S12において、処理管理部31は、要求部32から受信する要求に応じて、対応するアプリケーションに係わる処理一覧表を要求部31に送信する。処理一覧表は、例えば、図8に示すように、部分処理リストを含む。また、処理一覧表は、各部分処理についての必要解答数を表す情報を含む。

【0059】

S13において、要求部32は、目的アプリケーションの実行を要求する実行要求を通信部34に渡す。この実行要求は、目的アプリケーションに係わる処理一覧表および引数

50

を含む。

【 0 0 6 0 】

S 1 4において、通信部 3 4は、要求部 3 1から受信した実行要求を、すべての実行ノードコンピュータ 1に送信する。ただし、この実行要求は、目的アプリケーションに係わる処理一覧表および引数に加えて、実行要求の送信元（ここでは、要求ノードコンピュータ 3）を識別する情報含む。

【 0 0 6 1 】

S 1 5～S 1 6は、要求ノードコンピュータ 3から実行要求を受信した各実行ノードコンピュータ 1により実行される。すなわち、S 1 5において、通信部 1 5は、要求ノードコンピュータ 3から受信した実行要求を証跡管理部 1 1に渡す。そうすると、証跡管理部 1 1は、受信した実行要求の内容を証跡に記録する。図 1 0に示す例では、証跡管理部 1 1は、受信した実行要求に基づいて、アプリケーション名、引数、要求ノード名、部分処理リスト、必要回答数などを証跡に記録する。

10

【 0 0 6 2 】

S 1 6において、通信部 1 5は、要求ノードコンピュータ 3から受信した実行要求に基づいて、選択部 1 3に実行指示を与える。そうすると、実行フェーズが起動される。

【 0 0 6 3 】

図 1 2は、実行フェーズのシーケンスの一例を示す。実行フェーズは、要求ノードコンピュータ 3から実行要求を受信した各実行ノードコンピュータ 1において起動される。具体的には、図 1 1に示すS 1 6の実行指示が選択部 1 3に与えられると実行フェーズが起動される。

20

【 0 0 6 4 】

実行フェーズは、目的アプリケーションの各部分処理に対して実行される。すなわち、S 2 1～S 2 8は繰り返し実行される。そして、すべての部分処理が完了すると、実行フェーズは終了する。具体的には以下の通りである。

【 0 0 6 5 】

S 2 1において、選択部 1 3は、未だ完了していない部分処理が残っている否かを証跡管理部 1 1に問い合わせる。このとき、選択部 1 3は、目的アプリケーションの名称を表す情報（又は、目的アプリケーションの識別情報）を証跡管理部 1 1に通知する。なお、以下の記載では、未だ完了していない部分処理を「未完了部分処理」と呼ぶことがある。

30

【 0 0 6 6 】

S 2 2において、証跡管理部 1 1は、証跡を参照し、未完了部分処理を検索する。図 1 0に示す例では、圧縮処理および音声加工処理は完了しているが、モザイク加工処理および字幕追加処理は未だ完了していない。そして、証跡管理部 1 1は、検索結果を選択部 1 1に通知する。ここで、未完了部分処理が残っているときは、証跡管理部 1 1は、すべての未完了部分処理を選択部 1 3に通知する。

【 0 0 6 7 】

S 2 3において、選択部 1 3は、証跡管理部 1 1から受信する検索結果に基づいて、未完了部分処理が残っているか否かを検出する。そして、未完了部分処理が残っているときは、S 2 4において、選択部 1 3は、通知された未完了部分処理のうちから実行すべき部分処理を選択する。このとき、選択部 1 3は、通知された未完了部分処理のうちで最も先に実行すべき部分処理を選択する。そして、選択部 1 3は、選択した部分処理についての実行指示を実行部 1 4に与える。この実行指示は、目的アプリケーションの名称および選択した部分処理の名称を含む。

40

【 0 0 6 8 】

S 2 5において、実行部 1 4は、選択部 1 1により選択された部分処理を実行する。そして、S 2 6において、実行部 1 4は、合意形成部 1 2に対して、実行結果についての合意の形成を要求する。合意形成要求は、目的アプリケーションの名称、実行した部分処理の名称、実行結果を含む。

【 0 0 6 9 】

50

S 2 7において合意フェーズが実行される。なお、合意フェーズについては、後で図13を参照しながら説明する。ここでは、S 2 5の実行結果について、複数の実行ノードコンピュータ1により合意が形成されたものとする。

【0070】

S 2 8において、合意形成部12は、実行結果について合意が形成されたことを実行部14に通知する。S 2 9において、実行部14は、この通知を選択部13に転送する。このとき、実行部14は、S 2 5で得られている実行結果を選択部13に渡す。

【0071】

この後、実行フェーズのシーケンスはS 2 1に戻る。すなわち、実行ノードコンピュータ1は、未完了部分処理を1つずつ順番に選択して実行する。そして、すべての部分処理が完了すると、S 2 3において、選択部13は、通信部15に終了通知を与える。また、選択部13は、選択部13が各部分処理についての実行結果を通信部15に与える。

10

【0072】

なお、この実施例では、図4～図6を参照して説明したように、実行ノードコンピュータ1は、選択した部分処理の1つ前の部分処理の実行結果を利用して、選択した部分処理を実行する。したがって、最後の部分処理の実行結果は、すべての部分処理についての実行結果（即ち、目的アプリケーションの実行結果）に相当する。すなわち、S 2 3において、すべての部分処理が完了すると、目的アプリケーションの実行結果が選択部13から通信部15に与えられる。

20

【0073】

このように、実行フェーズにおいては、選択部13により選択された部分処理が実行部14により実行される。そして、すべての部分処理が完了すると、選択部13から通信部15に実行結果が与えられる。

【0074】

図13は、合意フェーズのシーケンスの一例を示す。合意フェーズは、図12に示す実行フェーズ中のS 2 7に相当する。すなわち、選択された部分処理の実行結果が合意形成部12に与えられると、合意フェーズが起動される。なお、以下の記載では、実行ノードコンピュータ1Aの実行結果について、複数または全ての実行ノードコンピュータ1により合意形成が行われるものとする。また、図13において、実行ノードコンピュータ1Bは、実行ノードコンピュータ1A以外の任意の実行ノードコンピュータ1を表す。

30

【0075】

S 3 1において、合意形成部12は、実行部14から受け取った合意形成要求を通信部15に転送する。S 3 2において、通信部15は、合意形成部12から受け取った合意形成要求を全実行ノードコンピュータ1（図13では、実行ノードコンピュータ1B）に送信する。このとき、合意形成要求は、合意形成要求の送信元（すなわち、実行ノードコンピュータ1A）を識別する情報を含む。

【0076】

S 3 3において、実行ノードコンピュータ1A、1Bは、所定の合意形成プロトコルに従って、実行ノードコンピュータ1Aの実行結果について合意を形成する。合意形成プロトコルとしては、例えば、P B F T (Practical Byzantine Fault Tolerance)、P o W (Proof of Work)、またはP o S (Proof of Stake)が使用される。また、この実施例では、選択された部分処理を実行した実行ノード、実行結果、実行時刻などについて合意が形成される。そして、合意形成プロトコルによる合意内容を表す合意結果は、各実行ノードコンピュータ1に送信される。すなわち、各実行ノードコンピュータ1の通信部15は、他の実行ノードコンピュータ1から合意結果を受信する。

40

【0077】

S 3 4～S 3 6は、実行ノードコンピュータ1Bにおいて実行される。具体的には、S 3 4において、通信部15は、受信した合意結果を合意形成部12に与える。合意形成部12は、この合意内容を証跡管理部12に与える。そして、証跡管理部11は、各実行ノードコンピュータ1A、1Bにより合意が形成された内容を証跡に記録する。なお、証跡

50

を更新する方法については、後で図 15 を参照して説明する。

【0078】

S37において、合意結果が実行ノードコンピュータ1Aに送信される。そうすると、実行ノードコンピュータ1AにおいてS38～S40が実行される。ここで、S38～S40は、実質的にS34～S36と同じである。すなわち、実行ノードコンピュータ1Aにおいても、証跡管理部11は、各実行ノードコンピュータ1A、1Bにより合意が形成された内容を証跡に記録する。

【0079】

S41において、証跡管理部11は、証跡の更新が完了したことを表す通知を合意形成部12に与える。すなわち、S25の実行結果について複数または全ての実行ノードコンピュータ1による合意が形成されたことを表す通知が、証跡管理部11から合意形成部12に与えられる。この後、合意形成部12は、図12に示すS28を実行する。

10

【0080】

このように、合意フェーズにおいては、ある実行ノードコンピュータ（図12～図13では、実行ノードコンピュータ1A）による実行結果について、複数または全ての実行ノードコンピュータ1により合意が形成される。そして、合意が得られた実行結果は、各実行ノードコンピュータ1の証跡に記録される。したがって、すべての実行ノードコンピュータ1は、同じ証跡を保持することになる。

【0081】

図14は、終了フェーズのシーケンスの一例を示す。終了フェーズは、図12に示すS23において、すべての部分処理が完了したと判定されたときに実行される。

20

【0082】

S51において、各実行ノードコンピュータ1の通信部15は、図12のS23で選択部13から受け取った実行結果を要求ノードコンピュータ3に送信する。よって、要求ノードコンピュータ3の通信部34は、各実行ノードコンピュータ1から実行結果を受信する。

【0083】

S52において、通信部34は、各実行ノードコンピュータ1から受信した実行結果を検証部33に渡す。そうすると、S53において、検証部33は、各実行ノードコンピュータ1から受信した実行結果の正当性を確認するために、各実行ノードコンピュータ1の証跡を収集する。すなわち、検証部33は、証跡の送信を要求する証跡要求を生成して要求部32に渡す。この証跡要求は、目的アプリケーションの名称および目的アプリケーションを構成する各部分処理の名称を含む。

30

【0084】

S54～S56において、証跡要求は、要求ノードコンピュータ3の通信部34を介して送信される。そうすると、各実行ノードコンピュータ1の通信部15は、受信した証跡要求を証跡管理部11に渡す。したがって、各実行ノードコンピュータ1の証跡管理部11は、それぞれ証跡要求を受信する。そうすると、S57において、証跡管理部11は、証跡要求により指定された目的アプリケーションに係わる証跡データを取得する。このとき、証跡管理部11は、自ノード内の所定の記憶領域に保存されている証跡データを取得する。そして、証跡管理部11は、取得した証跡データを要求ノードコンピュータ3に送信する。

40

【0085】

S58～S60において、証跡データは、通信部34および要求部31を介して、検証部33に与えられる。そうすると、S61において、検証部33は、各実行ノードコンピュータ1から収集した証跡データを互いに比較する。この結果、すべての証跡データが互いに一致すれば、その証跡データが正しいと判定される。或いは、収集した証跡データのうちの所定の割合以上の証跡データが互いに一致したときは、その証跡データを正しいと判定してもよい。いずれにしても、正しい証跡データが得られたときは、検証部33は、S62において、実行ノードコンピュータ1から受信した実行結果を要求部32に渡す。

50

そして、要求部 3 2 は、実行結果を出力する。

【 0 0 8 6 】

この後、要求ノードコンピュータ 3 は、受信した証跡データを参照し、有効な実行結果を提供した実行ノードコンピュータ 1 に対して報酬を与えてよい。すなわち、ある部分処理に対して必要回答数以上の同じ実行結果が提供されたときに、その実行結果を提供した各実行ノードコンピュータ 1 に報酬が与えられる。例えば、図 4 ~ 図 6 に示す実施例では、部分処理 1 に対して有効な実行結果を提供した実行ノードコンピュータ 1 A、1 C に対して報酬が与えられ、部分処理 2 に対して有効な実行結果を提供した実行ノードコンピュータ 1 A、1 B に対して報酬が与えられ、部分処理 3 に対して有効な実行結果を提供した実行ノードコンピュータ 1 A、1 C に対して報酬が与えられる。

10

【 0 0 8 7 】

図 1 5 は、証跡を更新する処理の一例を示すフローチャートである。このフローチャートの処理は、図 1 3 に示す S 3 6 または S 4 0 に相当する。すなわち、このフローチャートの処理は、実行結果について合意形成が行われたときに、各実行ノードコンピュータ 1 において証跡管理部 1 1 により実行される。このとき、証跡管理部 1 1 は、自ノード内に保存されている証跡データを参照する。

【 0 0 8 8 】

S 7 1 において、証跡管理部 1 1 は、各実行結果について、同じ実行結果の個数をそれぞれカウントする。例えば、各実行結果のハッシュ値が計算されている場合は、同じハッシュ値を有する実行結果の個数がそれぞれカウントされる。

20

【 0 0 8 9 】

S 7 2 において、証跡管理部 1 1 は、同じ実行結果の個数が必要回答数以上となっている実行結果があるか否かを判定する。なお、必要回答数は、部分処理ごとに予め指定されている。

【 0 0 9 0 】

同じ実行結果の個数が必要回答数以上となっている実行結果が存在するときは、認証管理部 1 1 は、S 7 3 において、最も早く必要回答数以上の同じ実行結果が集まった実行結果を特定する。そして、認証管理部 1 1 は、特定した実行結果を証跡に記録する。S 7 4 において、証跡管理部 1 1 は、S 7 3 で特定された実行結果の個数を「有効回答数」として記録する。S 7 5 において、証跡管理部 1 1 は「状態：完了」を記録する。

30

【 0 0 9 1 】

例えば、図 1 0 に示す実施例において、実行ノードコンピュータ 1 により圧縮処理が実行されているものとする。この場合、ハッシュ値が「7d97...」である実行結果が、3 個得られている。また、必要回答数は 3 である。よって、S 7 2 の判定結果は「Yes」であり、S 7 3 ~ S 7 5 が実行される。すなわち、圧縮処理の実行結果が確定する。また、「有効回答数：3」および「状態：完了」が記録される。

【 0 0 9 2 】

一方、同じ実行結果の個数が必要回答数以上となっている実行結果が存在しないときは、認証管理部 1 1 は、S 7 6 において、すべての実行結果を証跡に記録する。S 7 7 において、証跡管理部 1 1 は、同じ実行結果の個数の最大値を「有効回答数」として記録する。S 7 8 において、証跡管理部 1 1 は、「状態：未完了」を記録する。

40

【 0 0 9 3 】

例えば、図 1 0 に示す実施例において、実行ノードコンピュータ 1 によりモザイク加工処理が実行されているものとする。この場合、ハッシュ値が「dead...」である実行結果が 2 個であり、ハッシュ値が「beaf...」である実行結果が 1 個である。また、必要回答数は 3 である。したがって、S 7 2 の判定結果は「No」であり、S 7 6 ~ S 7 8 が実行される。すなわち、モザイク加工処理の実行結果は未だ確定しない。また、「有効回答数：2」および「状態：未完了」が記録される。

【 0 0 9 4 】

図 1 6 は、各ノードに実装される通信装置として動作するコンピュータのハードウェア

50

構成の一例を示す。コンピュータ 200 は、プロセッサ 201、メモリ 202、記憶装置 203、I/O デバイス 204、記録媒体デバイス 205、通信インターフェース 206 を備える。なお、コンピュータ 200 は、実行ノードコンピュータ 1 に相当する。

【0095】

プロセッサ 201 は、記憶装置 203 に格納されている通信プログラムを実行することにより、実行ノードコンピュータ 1 の機能を提供することができる。すなわち、プロセッサ 201 は、図 3 および図 15 に示すフローチャートの処理、および図 11～図 14 に示すシーケンス中の実行ノードコンピュータ 1 の処理を記述した通信プログラムを実行することにより、証跡管理部 11、合意形成部 12、選択部 13、実行部 14、通信部 15 の機能を提供する。

10

【0096】

メモリ 202 は、例えば半導体メモリであり、プロセッサ 201 の作業領域として使用される。記憶装置 203 は、コンピュータ 200 内に実装されていてもよいし、コンピュータ 200 に接続されてもよい。なお、証跡は、メモリ 202 または記憶装置 203 に保存される。I/O デバイス 204 は、ユーザまたはネットワーク管理者の指示を受け付ける。また、I/O デバイス 204 は、プロセッサ 201 による処理結果を出力する。記録媒体デバイス 205 は、可搬型記録媒体 207 に記録されている信号を読み取る。なお、上述した通信プログラムは、可搬型記録媒体 207 に記録されていてもよい。通信インターフェース 206 は、データ通信のためのインターフェースおよび制御情報を通信するためのインターフェースを含む。

20

【0097】

図 17 は、分散処理システム 100 の構成例を示す。なお、図 17 においては、2 台の実行ノードコンピュータが描かれているが、分散処理システム 100 はより多くの実行ノードコンピュータを備えていてもよい。

【0098】

各実行ノードコンピュータは、計算資源 51、計算資源制御部 52、アプリケーション実行制御部 53、分散台帳制御部 54、分散台帳 55 を備える。計算資源 51、計算資源制御部 52、およびアプリケーション実行制御部 53 は、図 7 に示す選択部 13 および実行部 14 に対応する。分散台帳制御部 54 は、証跡管理部 11 および合意形成部 12 に対応する。分散台帳制御部 54 は、HLF (Hyperledger Fabric) で実現してもよい。分散台帳 55 は、証跡管理部 11 により管理される証跡に対応する。なお、各ノードにおいて保存される分散台帳 55 の内容は、互いに一致している。また、分散台帳 55 は、例えば、ブロックチェーン技術を利用して実現してもよい。

30

【符号の説明】

【0099】

1 (1A～1D) 実行ノードコンピュータ

3 要求ノードコンピュータ

1 1 証跡管理部

1 2 合意形成部

1 3 選択部

1 4 実行部

1 5 通信部

2 0 1 プロセッサ

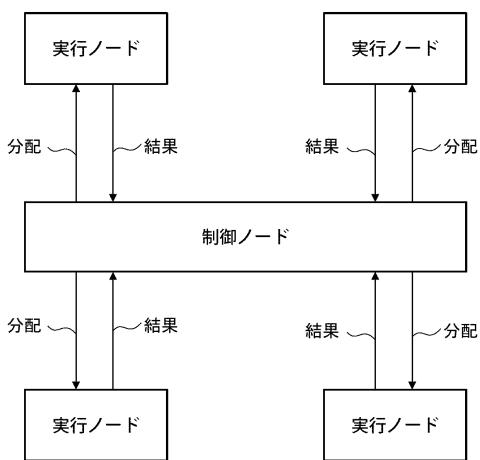
40

50

【図面】

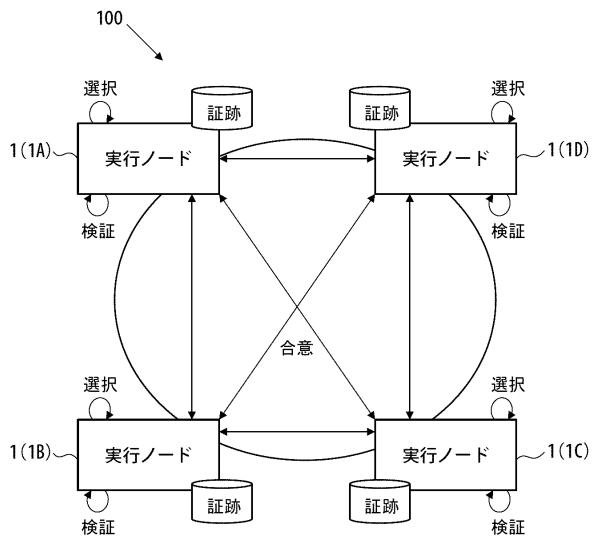
【図1】

従来の分散処理システムの一例を示す図



【図2】

本発明の実施形態に係わる分散処理システムの一例を示す図

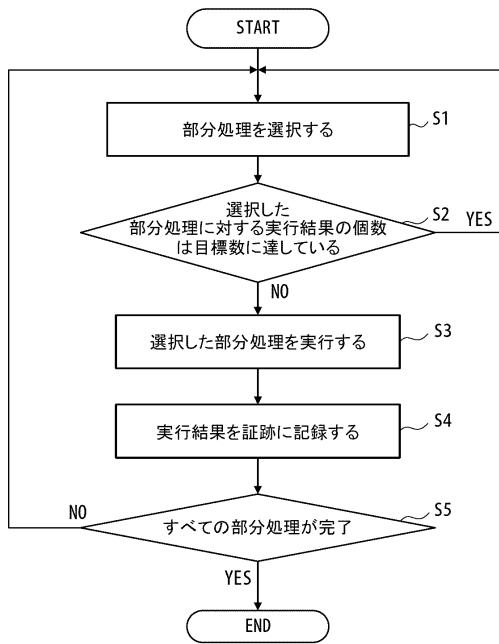


10

20

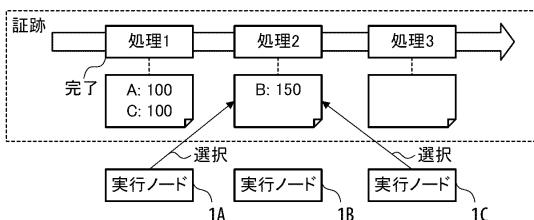
【図3】

実行ノードコンピュータの処理の一例を示すフローチャート



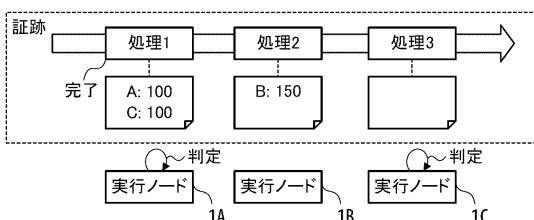
【図4】

分散処理システムによる処理の流れの一例を示す図(その1)



30

(a)



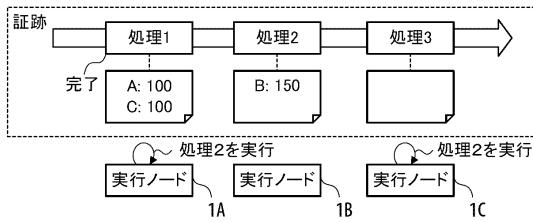
40

(b)

50

【図 5】

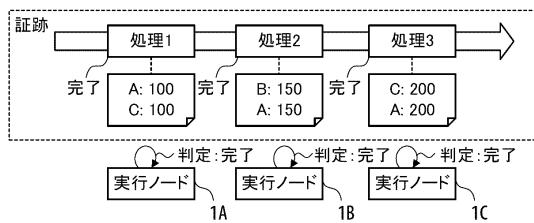
分散処理システムによる処理の流れの一例を示す図(その2)



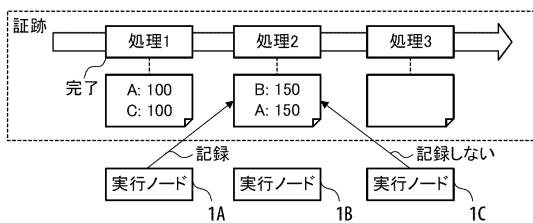
(a)

【図 6】

分散処理システムによる処理の流れの一例を示す図(その3)



10

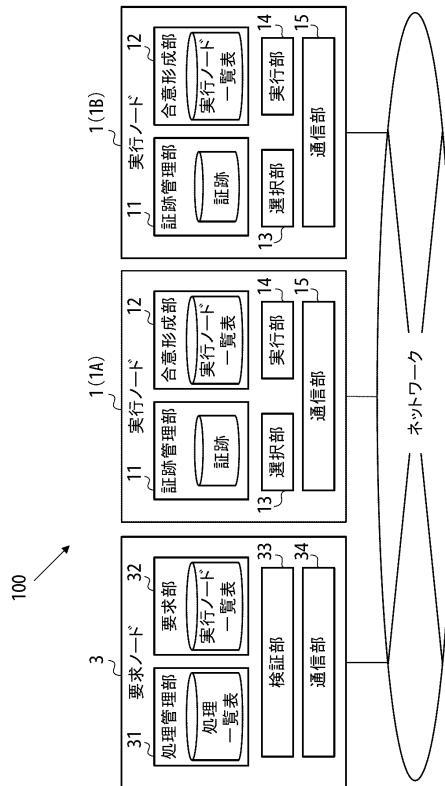


(b)

20

【図 7】

分散処理システムの構成および機能の一例を示す図



【図 8】

処理一覧表の一例を示す図

アプリ名	処理リスト	処理内容コード	必要回答数
動画処理アプリ	圧縮	(コードで記述された処理内容)	3
	音声加工	(同上)	2
	モザイク加工	(同上)	3
	字幕追加	(同上)	2
...
...

30

40

50

【図 9】

実行ノード一覧表の一例を示す図

ノード名	接続先
1A	https://node1a.com/api-gw
1B	https://node1b.com/api-gw
1C	https://node1c.com/api-gw
...	...

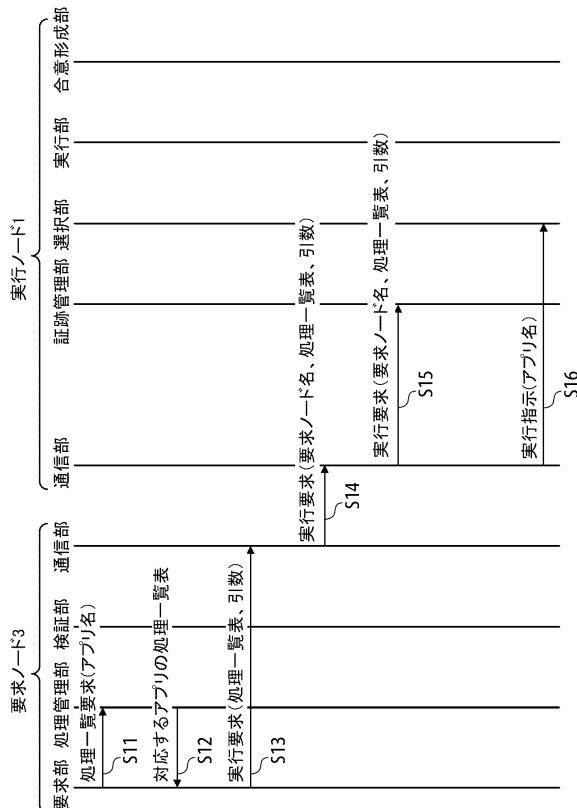
【図 10】

証跡の一例を示す図

アプリ名	引数	要求ノード	処理リスト	処理内容	必要回数	有効回答数	状態	実行結果(ハッシュ値)	実行ノード	実行時刻
動画アプリ	(動画データ)	ノード3	圧縮	(コードで記述された処理内容)	3	3	完了	7497...	1A	2018/11/5 15:22:35
								7397...	1B	2018/11/5 15:22:36
								7397...	1C	2018/11/5 15:22:39
								6042...	1B	2018/11/5 15:22:42
								6042...	1D	2018/11/5 15:22:45
								dead...	1D	2018/11/5 15:23:01
								beaf...	1B	2018/11/5 15:23:04
								dead...	1A	2018/11/5 15:23:05
							
							
							

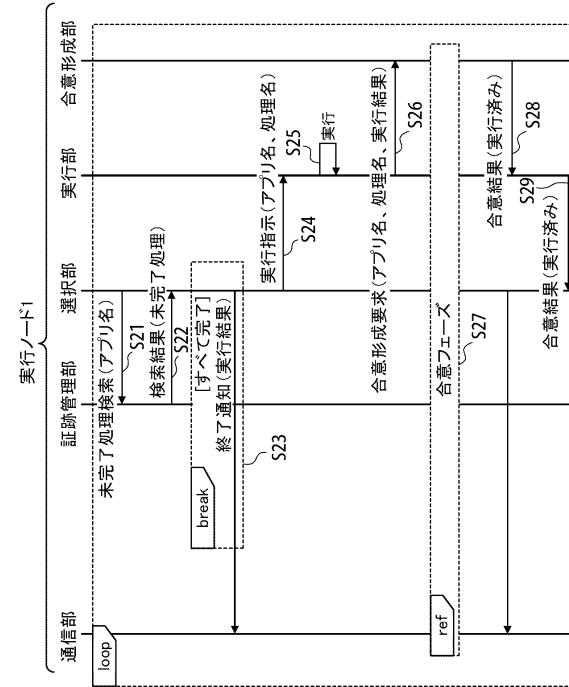
【図 11】

開始フェーズのシーケンスの一例を示す図



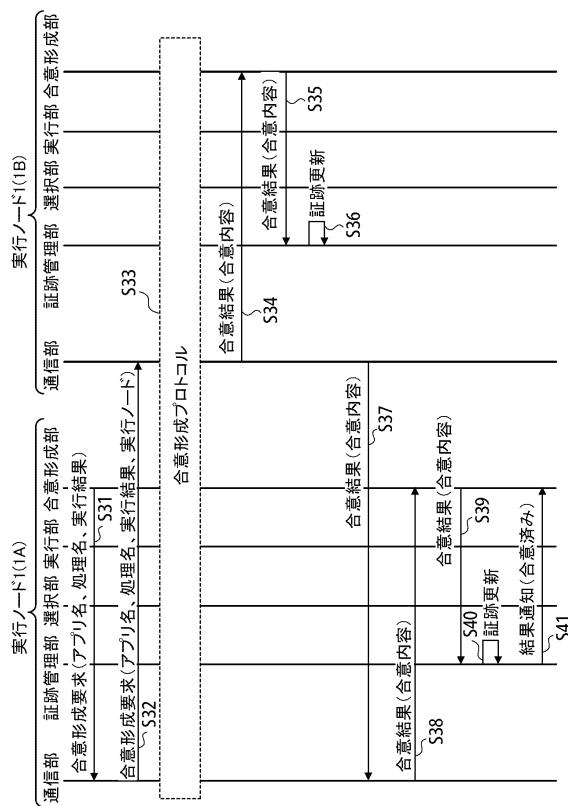
【図 12】

実行フェーズのシーケンスの一例を示す図



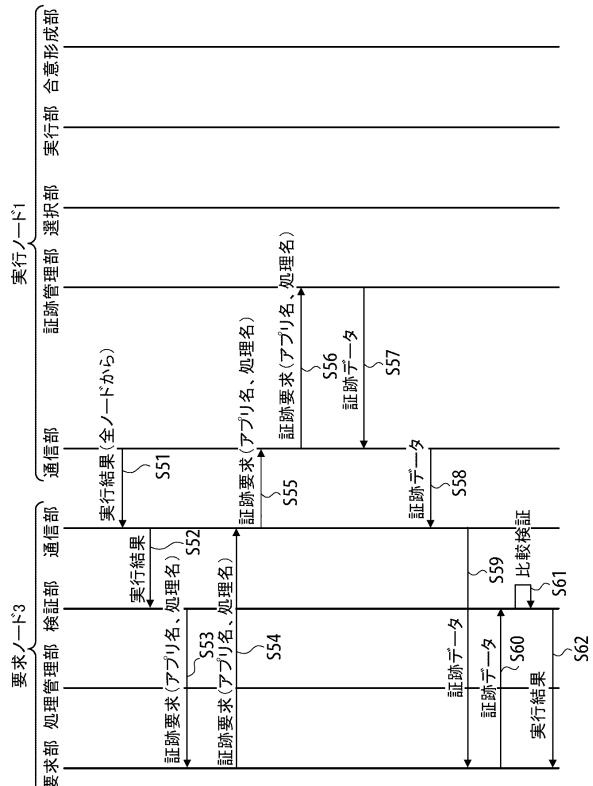
【図13】

合意フェーズのシーケンスの一例を示す図



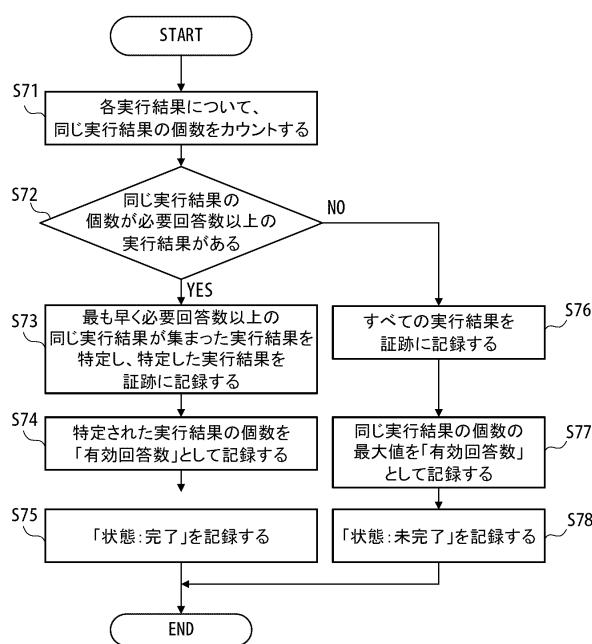
【図14】

終了フェーズのシーケンスの一例を示す図



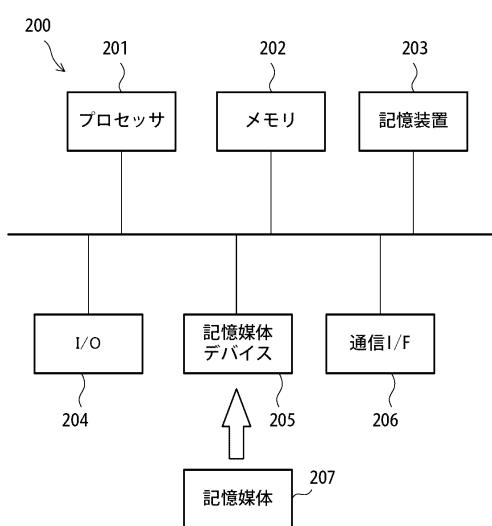
【図15】

証跡を更新する処理の一例を示すフローチャート

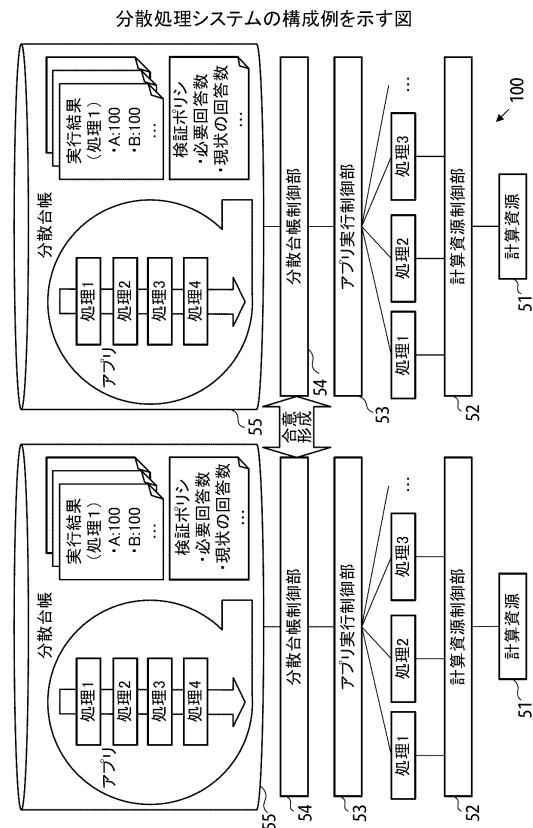


【図16】

通信装置として動作するコンピュータの ハードウェア構成の一例を示す図



【図 17】



10

20

30

40

50

フロントページの続き

審査官 三坂 敏夫

(56)参考文献 特開2005-242986 (JP, A)

特開平07-114520 (JP, A)

国際公開第2012/056487 (WO, A1)

特開2018-109878 (JP, A)

FABRE,Jean-Charles et al. , Saturation: reduced idleness for improved fault-tolerance , [1988] The Eighteenth International Symposium on Fault-Tolerant Computing. Digest of Papers , 米国 , IEEE , 1988年06月30日 , pages:200-205

(58)調査した分野 (Int.Cl. , D B名)

G 0 6 F 1 1 / 1 6 - 1 1 / 2 0

G 0 6 F 9 / 5 0