

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2022年3月3日(03.03.2022)



(10) 国際公開番号

WO 2022/044121 A1

(51) 国際特許分類:

G06F 9/50 (2006.01)

(21) 国際出願番号: PCT/JP2020/032029

(22) 国際出願日: 2020年8月25日(25.08.2020)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).

(72) 発明者: 杉園 幸司(SUGISONO, Koji); 〒1808585 東京都武蔵野市緑町3丁目9-11 N T T 知的財産センタ内 Tokyo (JP). 河野 伸也(KAWANO,

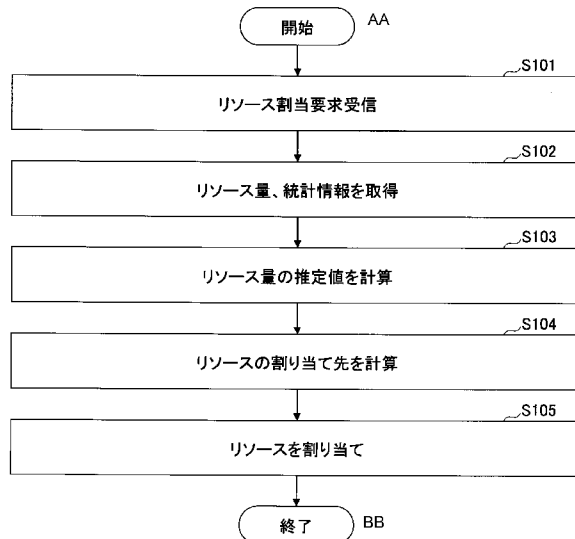
Shinya); 〒1808585 東京都武蔵野市緑町3丁目9-11 N T T 知的財産センタ内 Tokyo (JP). 加納 浩輝(KANO, Hiroki); 〒1808585 東京都武蔵野市緑町3丁目9-11 N T T 知的財産センタ内 Tokyo (JP). 木村 明寛(KIMURA, Akihiro); 〒1808585 東京都武蔵野市緑町3丁目9-11 N T T 知的財産センタ内 Tokyo (JP). 宮本 克真(MIYAMOTO, Katsuma); 〒1808585 東京都武蔵野市緑町3丁目9-11 N T T 知的財産センタ内 Tokyo (JP).

(74) 代理人: 伊東 忠重, 外(ITO, Tadashige et al.); 〒1000005 東京都千代田区丸の内二丁目1番1号 丸の内 M Y P L A Z A (明治安田生命ビル) 16階 Tokyo (JP).

(54) Title: RESOURCE QUANTITY ESTIMATION DEVICE, RESOURCE QUANTITY ESTIMATION METHOD, AND PROGRAM

(54) 発明の名称: リソース量推定装置、リソース量推定方法及びプログラム

[図5]



S101 Receive resource allocation request
S102 Acquire resource quantity and statistical information
S103 Calculate estimated value for resource quantity
S104 Calculate resource allocation destination
S105 Allocate resource
AA Start
BB End

(57) Abstract: This resource quantity estimation device increases the likelihood that resources will be allocated appropriately at the time of resource allocation by including: an acquisition unit that, in response to a resource allocation request, acquires a measured value for the resource quantity of the resource, a measurement time for the measured value, and statistical information relating to the resource; and a calculation unit that inputs, to a model for estimating the future quantity of the resource on the basis of the statistical information, the measured value acquired by the acquisition unit

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一 国際調査報告(条約第21条(3))

and the elapsed time from the measurement time until an expected time of allocation of the resource, and calculates an estimated value for the resource quantity at the expected time of allocation.

(57) 要約: リソース量推定装置は、リソースの割り当て要求に応じ、前記リソースのリソース量の計測値、前記計測値の計測時刻、及び前記リソースに関する統計情報を取得する取得部と、将来の前記リソース量を前記統計情報に基づいて推定するモデルに、前記取得部が取得した計測値と、前記計測時刻から前記リソースの割り当て予定時刻までの経過時間とを入力して、前記割り当て予定時刻における前記リソース量の推定値を計算する計算部と、を有することで、リソースの割り当て時点に適したリソースの割り当てが実施される可能性を高める。

明 細 書

発明の名称：

リソース量推定装置、リソース量推定方法及びプログラム

技術分野

[0001] 本発明は、リソース量推定装置、リソース量推定方法及びプログラムに関する。

背景技術

[0002] 仮想マシン技術やSDN (Software Defined Network) 技術の発達に伴い、サービス利用者が利用したいときにオンデマンドにデータ処理機能を仮想マシンやコンテナの形で提供し、機能間の通信経路を動的に設定することで、利用者にネットワークサービスを提供できるようになった。このような状況を受けて、従来、利用の多寡にかかわらず期間単位の定額料金を支払っていた低利用率ユーザがサービス利用コストを下げるためのサービスとして、リソース利用期間にのみ課金されるサーバレスコンピューティングモデルが登場した。

[0003] サーバレスコンピューティングモデルでは、オンデマンドにリソースが割り当てられ、処理終了後リソースが開放される。このため、バースト的にリソース割り当てや仮想マシン間の通信が発生する、という特徴を有する。

[0004] 利用料金を低減するための要求条件として、処理時間を短縮するということが挙げられる。処理時間は、計算機能及び通信機能が利用する、サーバやリンクなどの物理演算リソースの利用率により変化する。すいている装置に計算機能及び通信機能を割り当てることで、処理時間を短縮することができる。リソース利用率の低い装置に計算／通信機能を割り当てるため、サーバレスコンピューティング処理に必要とするCPUや想定通信データ量などの要求条件とサービス用物理演算リソースとを勘案しながら、すいている装置や通信経路を導き出し、サーバレスコンピューティング処理用計算／通信機能を割り当てる技術が長年検討されてきた。

先行技術文献

非特許文献

- [0005] 非特許文献1: A.Fischer et al., "Virtual Network Embedding: A Survey,"
IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL.15, NO.4, pp.1888-1906,
FOURTH QUARTER 2013.

発明の概要

発明が解決しようとする課題

- [0006] 従来技術は、計算機能及び通信機能の割り当て箇所を計算する際、サーバやネットワークスイッチなどからデータ量やCPU利用率などのリソース情報を、割り当て箇所計算機能に集約し、集めたリソース情報をアルゴリズムなどに適用することで割り当て箇所を計算していた。
- [0007] ここで、考慮すべき点は、使用しているリソース情報の鮮度、すなわち、リソース情報の計測時刻から割り当て箇所の計算時点までの経過時間の大きさである。
- [0008] 通常、物理リソースから割り当て箇所計算機能に測定データが届くまで、データ転送遅延や処理遅延等の影響からタイムラグが発生する。このため、割り当て箇所計算機能が計算時点で参考にするリソース情報と、当該リソース情報を送信したサーバやスイッチが当該計算時点で観測している統計情報には誤差が存在する。
- [0009] このため、割り当て箇所計算機能が、得られたリソース情報をもとに最適な割り当て箇所を計算したとしても、実際に経路を割り当てる際には状況が変化し、最適解ではなくなってしまう可能性が有る。
- [0010] 本発明は、上記の点に鑑みてなされたものであって、リソースの割り当て時点に適したリソースの割り当てが実施される可能性を高めることを目的とする。

課題を解決するための手段

- [0011] そこで上記課題を解決するため、リソース量推定装置は、リソースの割り

当て要求に応じ、前記リソースのリソース量の計測値、前記計測値の計測時刻、及び前記リソースに関する統計情報を取得する取得部と、将来の前記リソース量を前記統計情報に基づいて推定するモデルに、前記取得部が取得した計測値と、前記計測時刻から前記リソースの割り当て予定時刻までの経過時間とを入力して、前記割り当て予定時刻における前記リソース量の推定値を計算する計算部と、を有する。

発明の効果

[0012] リソースの割り当て時点に適したリソースの割り当てが実施される可能性を高めることができる。

図面の簡単な説明

[0013] [図1]本発明の実施の形態におけるシステム構成例を示す図である。

[図2]第1の実施の形態におけるリソース割当装置10のハードウェア構成例を示す図である。

[図3]第1の実施の形態におけるリソース割当装置10の機能構成例を示す図である。

[図4]第1の実施の形態におけるサービス実行装置20の機能構成例を示す図である。

[図5]第1の実施の形態におけるリソース割当装置10が実行する処理手順の一例を説明するためのフローチャートである。

[図6]100msec後の通信リソース量の確率分布を示す図である。

[図7]通信リソース量の補正の具体例を示す図である。

[図8]計測時刻ごとの計測値とt秒後の計測値とのペアの生成を説明するための図である。

[図9]離散化及び遷移パターンの生成を説明するための図である。

[図10]遷移パターンの発生確率を説明するための図である。

発明を実施するための形態

[0014] 以下、図面に基づいて本発明の実施の形態を説明する。図1は、本発明の実施の形態におけるシステム構成例を示す図である。図1に示されるように

、リソース割当装置１０は、ネットワークを介して１以上のクライアント装置３０に接続される。リソース割当装置１０は、また、サービス用プラットフォームＰ１を構成する複数のサービス実行装置２０とネットワークを介して接続される。なお、複数のサービス実行装置２０は、相互にネットワークを介して接続される。以下、２対のサービス実行装置２０間のネットワーク区間を「リンク」という。

[0015] サービス用プラットフォームＰ１は、仮想マシン技術やＳＤＮ（Software Defined Network）技術等を用いて、所定のサービス（例えば、ネットワークサービス）を提供するコンピュータシステムである。各サービス実行装置２０は、サービスに関する処理を実行するソフトウェア（本実施の形態では仮想マシン（ＶＭ（Virtual Machine）））が稼働するコンピュータである。

[0016] クライアント装置３０は、サービスに必要なシステムリソース（以下、単に「リソース」という。）の割当をリソース割当装置１０に要求するコンピュータである。

[0017] リソース割当装置１０は、クライアント装置３０からの要求（以下、「リソース割当要求」という。）に応じたリソースの割当を、サービス用プラットフォームＰ１に対して実行する１以上のコンピュータである。

[0018] 図２は、第１の実施の形態におけるリソース割当装置１０のハードウェア構成例を示す図である。図２のリソース割当装置１０は、それぞれバスＢで相互に接続されているドライブ装置１００、補助記憶装置１０２、メモリ装置１０３、プロセッサ１０４、及びインタフェース装置１０５等を有する。

[0019] リソース割当装置１０での処理を実現するプログラムは、ＣＤ－ＲＯＭ等の記録媒体１０１によって提供される。プログラムを記憶した記録媒体１０１がドライブ装置１００にセットされると、プログラムが記録媒体１０１からドライブ装置１００を介して補助記憶装置１０２にインストールされる。但し、プログラムのインストールは必ずしも記録媒体１０１より行う必要はなく、ネットワークを介して他のコンピュータよりダウンロードするようにしてもよい。補助記憶装置１０２は、インストールされたプログラムを格納

すると共に、必要なファイルやデータ等を格納する。

[0020] メモリ装置 103 は、プログラムの起動指示があった場合に、補助記憶装置 102 からプログラムを読み出して格納する。プロセッサ 104 は、CPU 若しくは GPU (Graphics Processing Unit)、又は CPU 及び GPU であり、メモリ装置 103 に格納されたプログラムに従ってリソース割当装置 10 に係る機能を実行する。インタフェース装置 105 は、ネットワークに接続するためのインタフェースとして用いられる。

[0021] なお、クライアント装置 30 及びサービス実行装置 20 等も図 2 に示されるようなハードウェア構成を有してもよい。

[0022] 図 3 は、第 1 の実施の形態におけるリソース割当装置 10 の機能構成例を示す図である。図 3 において、リソース割当装置 10 は、入力受付部 11、計測情報取得部 12、推定値計算部 13 及びリソース割当部 14 等を有する。これら各部は、リソース割当装置 10 にインストールされた 1 以上のプログラムが、プロセッサ 104 に実行させる処理により実現される。

[0023] 図 4 は、第 1 の実施の形態におけるサービス実行装置 20 の機能構成例を示す図である。図 4 において、サービス実行装置 20 は、リソース計測部 21、時刻計測部 22、計測情報通知部 23 及びサービス実行部 24 等を有する。これら各部は、サービス実行装置 20 にインストールされた 1 以上のプログラムが、サービス実行装置 20 のプロセッサに実行させる処理により実現される。

[0024] 以下、リソース割当装置 10 が実行する処理手順について説明する。図 5 は、第 1 の実施の形態におけるリソース割当装置 10 が実行する処理手順の一例を説明するためのフローチャートである。

[0025] ステップ S101 において、入力受付部 11 は、クライアント装置 30 からのリソース割当要求を受信する。リソース割当要求としては、例えば、割当が必要な VM に関する情報（以下、「VM 情報」という。）と、各 VM 間の通信に必要な帯域（例えば、10Mbps）等が含まれる。VM 情報には、例えば、必要な VM ごとに、当該 VM の実行処理のタイプ（すなわち、機

能)、各VMの実行順序等が含まれる。

[0026] 続いて、計測情報取得部12は、各サービス実行装置20や、サービス実行装置20間を接続する各リンクから、リソース量やリソースの稼働状況又は利用状況に関する統計情報等を取得する(S102)。各サービス実行装置20からは、例えば、プロセッサの利用率等、VMの演算処理に関連するリソース量(以下、「演算リソース量」という。)及び統計情報(以下、「装置統計情報」という。)が取得される。各リンクからは、例えば、空き帯域等、VM間の通信に関連するリソース量(以下、「通信リソース量」という。)及び統計情報(以下、「リンク統計情報」という。)が取得される。装置統計情報及びリンク統計情報の一例としては、平均リソース割り当て発生レート及び平均リソース割り当て時間等が挙げられる。

[0027] サービス実行装置20に関して、平均リソース割り当て発生レートとは、単位時間当たりのVMの割り当て回数をいい、平均リソース割り当て時間とは、1回の割り当てに係るVMの稼働時間の平均をいう。稼働時間とは、VMが割り当てられてから当該VMが解放されるまでの期間をいう。

[0028] リンクに関して、平均リソース割り当て発生レートとは、単位時間当たりのフロー(flow)の発生回数をいい、平均リソース割り当て時間とは、1フローの継続時間の平均をいう。なお、フローとは、同一の通信相手間の一連の通信をいう。

[0029] なお、本実施の形態において、各リソースの平均リソース割り当て時間、平均リソース割り当て発生レート、平均リソース利用量、バーストラヒック発生過程は、各リソース割当要求に対して同じ値であるとする。

[0030] 演算リソース量及び装置統計情報は、各サービス実行装置20のリソース計測部21によって計測される。また、演算リソース量及び装置統計情報の計測時点の時刻(計測時刻)が各サービス実行装置20の時刻計測部22によって計測される

通信リソース量及びリンク統計情報、並びにそれらの計測時刻は、例えば、リンクを構成するスイッチ等のネットワーク装置によって計測される。

[0031] 例えば、計測情報取得部 12 は、各サービス実行装置 20 及び各スイッチに対して、リソース量及び統計情報の送信を要求する。各サービス実行装置 20 のリソース計測部 21 は、当該要求を受信すると、その時点の演算リソース量を計測し、時刻計測部 22 は、その時点の時刻を計測する。各サービス実行装置 20 の計測情報通知部 23 は、当該演算リソース量及び当該演算リソース量の計測時刻と、当該要求とは非同期な（例えば、定期的に）なりリソース量等の計測から得られる装置統計情報とを計測情報取得部 12 に送信する。また、各ネットワーク装置は、当該要求を受信すると、その時点の通信リソース量及び計測時刻を計測する。各ネットワーク装置は、当該通信リソース量及び当該通信リソース量の計測時刻と、当該要求とは非同期な（例えば、定期的に）な計測から得られるリンク統計情報とを計測情報取得部 12 に送信する。

[0032] 続いて、推定値計算部 13 は、各リソース量（各演算リソース量及び各通信リソース量）について、将来のリソース量の推定値を、ステップ S102 において取得された各統計情報に基づいて推定するモデル（以下、「推定モデル」という。）に、ステップ S102 において取得されたリソース量の計測値及び、及び当該リソース量の計測時刻からリソースの割り当て完了予定時刻までの経過時間を入力して計算する（S103）。第 1 の実施の形態において、推定モデルは、 $M/M/100/100$ でモデル化される待ち行列理論が用いられて構築される。ここでは、説明を単純化するため、通信リソース量及びリンク統計情報に基づいて、リソース割り当て完了予定時刻の通信リソース量を推定する例について説明する。なお、 $M/M/100/100$ は、割り当て可能なリソース量が 100 であり、割り当て可能なリソース量の最大値が 100 であるモデルを意味する。例えば、1 フロー当たりの帯域が 1 Mbps であり、割り当て可能な帯域の最大値が 100 Mbps である場合が、 $M/M/100/100$ に該当する。

[0033] ここで、ステップ S102 において取得された通信リソース量が B (Mbps)、平均リソース割り当て発生レートが λ 、平均リソース割り当て時間

が μ である場合、 t 秒後の帯域 i ごとの確率は、以下の状態方程式に基づいて算出可能である。なお、 t の値は、リソース量の計測時刻からリソース割当完了予定時刻まで経過時間である。当該リソース割当完了予定時刻は、例えば、過去の実績等に基づいて、推定値計算部13が計算すればよい。

[0034] [数1]

$$\frac{d}{dt}P(t) = QP(t) \quad (1)$$

$$P(t) = (P(0, t), P(1, t), \dots, P(100, t))^T$$

$$Q = \begin{pmatrix} -\lambda & \mu & 0 & \dots & 0 \\ \lambda & -(\lambda + \mu) & 2\mu & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda & -(\lambda + 99\mu) & 100\mu \\ & & & 0 & \lambda & -100\mu \end{pmatrix}$$

但し、 $P(i, t)$ 時刻 t の通信リソース量が i Mbpsである確率

$$P(B, 0) = 1$$

$$P(i, 0) = 0 \quad (i \neq B)$$

Q ：状態遷移行列

したがって、ステップS103において、推定値計算部13は、リンクごとに、 $0 \sim 100$ Mbpsの1 Mbps刻みの帯域 i ごとに、式(1)の t 、 λ 、 μ に対して、 t_j 、 λ_j 、 μ_j を代入することで、 $P_j(t_j)$ を計算する。なお、 t_j は、リンク j の通信リソースの計測時刻からリソース割当完了予定時刻までの経過時間である。 λ_j は、リンク j の平均リソース割り当て発生レートである。 μ_j は、リンク j の平均リソース割り当て時間である。 $P_j(t_j)$ は、リンク j の t_j 秒後における各帯域の確率分布である。例えば、 $t_j = 100$ msecとした場合、式(1)は、図6に示されるような確率分布を示す。

[0035] すなわち、 $t_j = 100$ msecであるリンク j について、推定値計算部1

3は、式(1)について、 $P_j(0, 1) = \{P_j(0, 0.1), \dots, P_j(100, 0.1)\}$ を計算することで、計測時刻から100 msec後の帯域の確率分布を得る。

[0036] 続いて、推定値計算部13は、各リンクjについて、当該リンクjの帯域の t_j 秒後の確率分布に基づいて、 t_j 秒後の帯域の推定値 c_j を計算する。以下の式(2)に示される通り、確率分布の期待値(平均値)が推定値 c_j として算出される。

[0037] [数2]

$$c_j = \sum i P_j(i, t_j) \quad (2)$$

その結果、経路計算に用いられる各通信リソース量が、例えば、図7に示されるように補正される。図7では、リンクL1の計測時の通信リソース量(計測値)が60 Mbpsであったが、リソース割当完了予定時刻での通信リソース量(推定値)が80 Mbpsに補正されたことが示されている。また、リンクL2の計測時の通信リソース量(計測値)が40 Mbpsであったが、リソース割当完了予定時刻での通信リソース量(推定値)が30 Mbpsに推定されたことが示されている。更に、リンクL3計測時の通信リソース量(計測値)が80 Mbpsであったが、リソース割当完了予定時刻での通信リソース量(推定値)が50 Mbpsに推定されたことが示されている。

[0038] なお、上記では、通信リソース量について述べたが、各演算リソース量についても同様の考え方で補正する(リソース割当完了予定時刻での推定値を算出する)ことができる。

[0039] 続いて、リソース割当部14は、各リソース量について算出された推定値に基づいて、リソースの割り当て先とするサービス実行装置20や、VM間の通信経路を割り当てるリンク等を計算する(S104)。

[0040] 続いて、リソース割当部14は、ステップS104の計算結果に基づき、

リソース割当要求に応じたリソースの割り当てを実行する（S 1 0 5）。

[0041] なお、ステップS 1 0 4 及びS 1 0 5、公知技術に基づいて行われればよい。

[0042] 次に、第2の実施の形態について説明する。第2の実施の形態では第1の実施の形態と異なる点について説明する。第2の実施の形態において特に言及されない点については、第1の実施の形態と同様でもよい。

[0043] 第2の実施の形態では、図5のステップS 1 0 2 及びS 1 0 3の内容が第1の実施の形態と異なる。

[0044] ステップS 1 0 2において、計測情報取得部12は、各サービス実行装置20及び各リンクから、それぞれのリソース量（演算リソース量、通信リソース量）の時系列データを取得する。当該時系列データは、例えば、定期的に計測されたリソース量の履歴である。

[0045] 続くステップS 1 0 3において、推定値計算部13は、各リソース量について、時系列データに基づく統計処理に基づく推定モデルを用いてリソース割当完了時刻（計測時刻から t 秒後）の推定値を計算する。

[0046] 当該統計処理の詳細について説明する。ここでは、演算リソース量（CPU利用率）についての推定値の計算方法を説明する。

[0047] まず、推定値計算部13は、時系列データの各計測値の計測時刻ごとに、当該計測時刻の計測値 r_1 と、当該計測時刻から t 秒後の計測値 r_2 とを取得して、 (r_1, r_2) の計測値のペアを生成する。

[0048] 図8は、計測時刻ごとの計測値と t 秒後の計測値とのペアの生成を説明するための図である。図8の数直線上の各垂直線は、計測時刻を示す。なお、 t 秒後に完全に一致するタイミングで計測が行われていない場合には、 t 秒後に最も近い計測時刻の計測値が取得されればよい。このように取得される各ペアは、 t 秒間における計測値の遷移を示す。

[0049] 続いて、推定値計算部13は、各ペアの計測値 r_1 及び r_2 を離散化する。具体的には、予め設定された各離散区間のうち同一の離散区間に含まれる計測値が同一の値になるように離散化される。例えば、小数点以下の四捨五

入によって離散化が行われてもよい。推定値計算部 13 は、離散化されたペア群に基づいて、 r_1 から r_2 への遷移パターンを示すグラフを生成する。

[0050] 図 9 は、離散化及び遷移パターンの生成を説明するための図である。図 9 には、(10.1%, 20.3%) 及び (10.3%, 15.1%) 等のペアが示されている。これらの各計測値が、例えば、小数点以下の四捨五入によって離散化されると、各ペアは、(10%, 20%)、(10%, 15%) となる。この 2 つのペアによれば、10% については、 t 秒間において 20% 又は 15% への 2 つの遷移パターンが存在することが分かる。この遷移パターンを示すグラフが図 9 の右端の図である。当該グラフでは、共通の r_1 (遷移元) は、同一のノード (遷移元ノード) に統合され、共通の r_2 (遷移先) は、同一のノード (遷移先ノード) に統合される。以下、当該グラフを「遷移グラフ」という。

[0051] 続いて、推定値計算部 13 は、遷移元ノードを同じくする遷移パターンごとに、各遷移パターンの発生回数をカウントし、当該発生回数を当該遷移元ノードに係る遷移パターンの総数で除することで、発生確率 (各計測値からの遷移確率) を算出する。

[0052] 図 10 は、遷移パターンの発生確率を説明するための図である。図 10 には、10% からは 20% 又は 15% へ遷移する可能性があり、それぞれの遷移確率は 40%、60% であることが示されている。すなわち、この段階では、遷移グラフの各枝に当該枝に対応する遷移パターンの発生確率 (遷移確率) が付与される。この状態の遷移グラフが、第 2 の実施の形態において推定値の計算に用いられる推定モデルとなる。

[0053] 続いて、推定値計算部 13 は、各演算リソースについて、当該演算リソースの計測値 (ステップ S102 で取得された値) を遷移元ノードとする各遷移パターンの遷移確率を重みとする、当該各遷移パターンの遷移先ノードの値 (リソース量) の加重平均を、割当完了予定時刻の当該リソース量の推定値として計算する。

[0054] 図 5 のステップ S104 以降は、第 1 の実施の形態と同様でよい。

- [0055] なお、遷移グラフは、全てのサービス実行装置 20 からの時系列データに基づいて 1 つ生成されてもよいし、サービス実行装置 20 ごとに生成されてもよい。後者の場合、或る演算リソースのリソース量の推定値の計算には、当該演算リソースに係るサービス実行装置 20 の時系列データに基づく遷移グラフが用いられればよい。
- [0056] また、上記では、便宜上、演算リソースについて述べたが、推定値計算部 13 は、各通信リソースについても同様に、計測時刻から t 秒後のリソース量の推定値を計算する。
- [0057] 次に、第 3 の実施の形態について説明する。第 3 の実施の形態では第 1 の実施の形態と異なる点について説明する。第 3 の実施の形態において特に言及されない点については、第 1 の実施の形態と同様でもよい。
- [0058] 第 3 の実施の形態では、推定モデルとして機械学習に基づくモデルが用いられる例について説明する。なお、第 3 の実施の形態においては、便宜上、演算リソースを対象として説明する。
- [0059] 第 3 の実施の形態では、図 5 の処理手順が実行される前に推定モデルの学習が以下の (1) ~ (3) の手順で実行される。推定モデルは、ニューラルネットワークを用いたものでもよいし、他のモデルを用いたものでもよい。
- [0060] (1) 学習データとして、各サービス実行装置 20 のリソース計測部 21 が定期的な計測時刻において記録した、当該計測時刻、演算リソース量、当該リソース量に係るリソースを利用するソフトウェア (VM) に関する統計情報 (例えば、稼働 VM 数、各 VM の実行処理のタイプ) を各サービス実行装置 20 から計測情報取得部 12 が取得する。すなわち、各計測時点の計測時刻、リソース量及び統計情報を含む時系列データが取得される。
- [0061] (2) 推定値計算部 13 は、当該時系列データを学習データとして、計測時刻、リソース量及び統計情報の遷移関係を推定モデルに学習させる。例えば、第 2 の実施の形態において説明したような遷移パターンが推定モデルによって学習される。この際、実行処理のタイプは、例えば、各タイプを示す数値に置換されればよい。なお、当該推定モデルは、時刻、リソース量及び

統計情報を入力とし、当該時刻のリソース量の推定値を出力する関数である。

[0062] (3) 推定モデルによる出力を技術者が識別したうえで、有意と判断した出力を再度推定モデルに入力して、推定モデルが再学習されてもよい。

[0063] なお、推定モデルは、全てのサービス実行装置 20 に対して共通なものが学習されてよいし、サービス実行装置 20 ごとに学習が行われてもよい。

[0064] 第 3 の実施の形態では、推定モデルが学習済みである状態において、図 5 の処理手順が実行される。

[0065] ステップ S 102 において、計測情報取得部 12 は、各サービス実行装置 20 から現時点の演算リソース、当該演算リソースの計測時刻及び現時点のサービス実行装置 20 の統計情報（例えば、稼働 VM 数、各 VM の実行処理のタイプ）を取得する。

[0066] ステップ S 103 において、推定値計算部 13 は、サービス実行装置 20 ごとに、当該サービス実行装置 20 から取得された演算リソース、当該演算リソースの計測時刻から t 秒後の時刻（すなわち、リソース割当完了予定時刻）、及び現時点のサービス実行装置 20 の統計情報を学習済みの AI システムへ入力して、リソース割当完了予定時刻の当該演算リソースの推定値を計算する。なお、推定モデルがサービス実行装置 20 ごとに構築されている場合、推定値計算部 13 は、サービス実行装置 20 ごとに当該サービス実行装置 20 に関して学習済みの推定モデルを利用する。

[0067] 図 5 のステップ S 104 以降は、第 1 の実施の形態と同様でよい。

[0068] なお、上記では、便宜上、演算リソースについて述べたが、推定値計算部 13 は、各通信リソースについても同様に計測時から t 秒後のリソース量の推定値を計算する。

[0069] 上述したように、上記各実施の形態によれば、リソースに関する統計情報に基づくモデルに基づいて、計測時点のリソース量がリソース割当完了予定時刻における推定値に補正される。したがって、リソース量の計測時点とリソースの割り当て時点との間にタイムラグが存在する場合であっても、リソ

ースの割り当て時点に適したリソースの割り当てが実施される可能性を高めることができる。その結果、例えば、VMの処理時間やVM通信時間が低減され、サービス利用コストの低減を期待することができる。

[0070] なお、上記各実施の形態において、リソース割当装置10は、リソース量推定装置の一例である。

[0071] 以上、本発明の実施の形態について詳述したが、本発明は斯かる特定の実施形態に限定されるものではなく、請求の範囲に記載された本発明の要旨の範囲内において、種々の変形・変更が可能である。

符号の説明

[0072]	10	リソース割当装置
	11	入力受付部
	12	計測情報取得部
	13	推定値計算部
	14	リソース割当部
	20	サービス実行装置
	21	リソース計測部
	22	時刻計測部
	23	計測情報通知部
	24	サービス実行部
	30	クライアント装置
	100	ドライブ装置
	101	記録媒体
	102	補助記憶装置
	103	メモリ装置
	104	プロセッサ
	105	インタフェース装置
	B	バス
	P1	サービス用プラットフォーム

請求の範囲

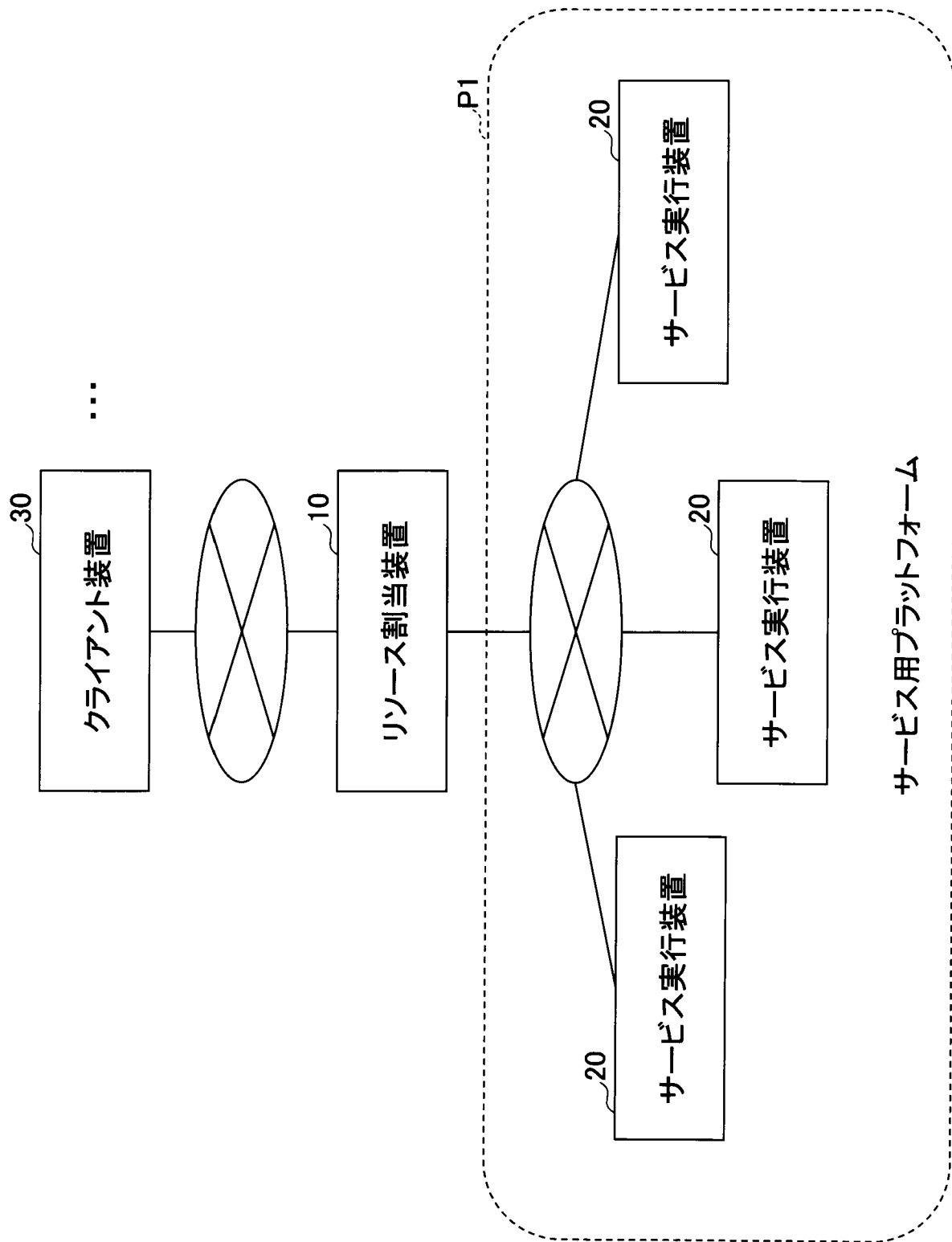
- [請求項1] リソースの割り当て要求に応じ、前記リソースのリソース量の計測値、前記計測値の計測時刻、及び前記リソースに関する統計情報を取得する取得部と、
- 将来の前記リソース量を前記統計情報に基づいて推定するモデルに、前記取得部が取得した計測値と、前記計測時刻から前記リソースの割り当て予定時刻までの経過時間とを入力して、前記割り当て予定時刻における前記リソース量の推定値を計算する計算部と、
- を有することを特徴とするリソース量推定装置。
- [請求項2] 前記モデルは、待ち行列理論、統計処理、又は機械学習を用いたモデルである。
- ことを特徴とする請求項1記載のリソース量推定装置。
- [請求項3] 前記統計情報は、前記リソースに関する平均リソース割り当て発生レート及び平均リソース割り当て時間、又は前記リソース量の時系列データを含む、
- ことを特徴とする請求項1又は2記載のリソース量推定装置。
- [請求項4] リソースの割り当て要求に応じ、前記リソースのリソース量の計測値、前記計測値の計測時刻、及び前記リソースに関する統計情報を取得する取得手順と、
- 将来の前記リソース量を前記統計情報に基づいて推定するモデルに、前記取得手順が取得した計測値と、前記計測時刻から前記リソースの割り当て予定時刻までの経過時間とを入力して、前記割り当て予定時刻における前記リソース量の推定値を計算する計算手順と、
- をコンピュータが実行することを特徴とするリソース量推定方法。
- [請求項5] 前記モデルは、待ち行列理論、統計処理、又は機械学習を用いたモデルである。
- ことを特徴とする請求項4記載のリソース量推定方法。
- [請求項6] 前記統計情報は、前記リソースに関する平均リソース割り当て発生

レート及び平均リソース割り当て時間、又は前記リソース量の時系列データを含む、

ことを特徴とする請求項4又は5記載のリソース量推定方法。

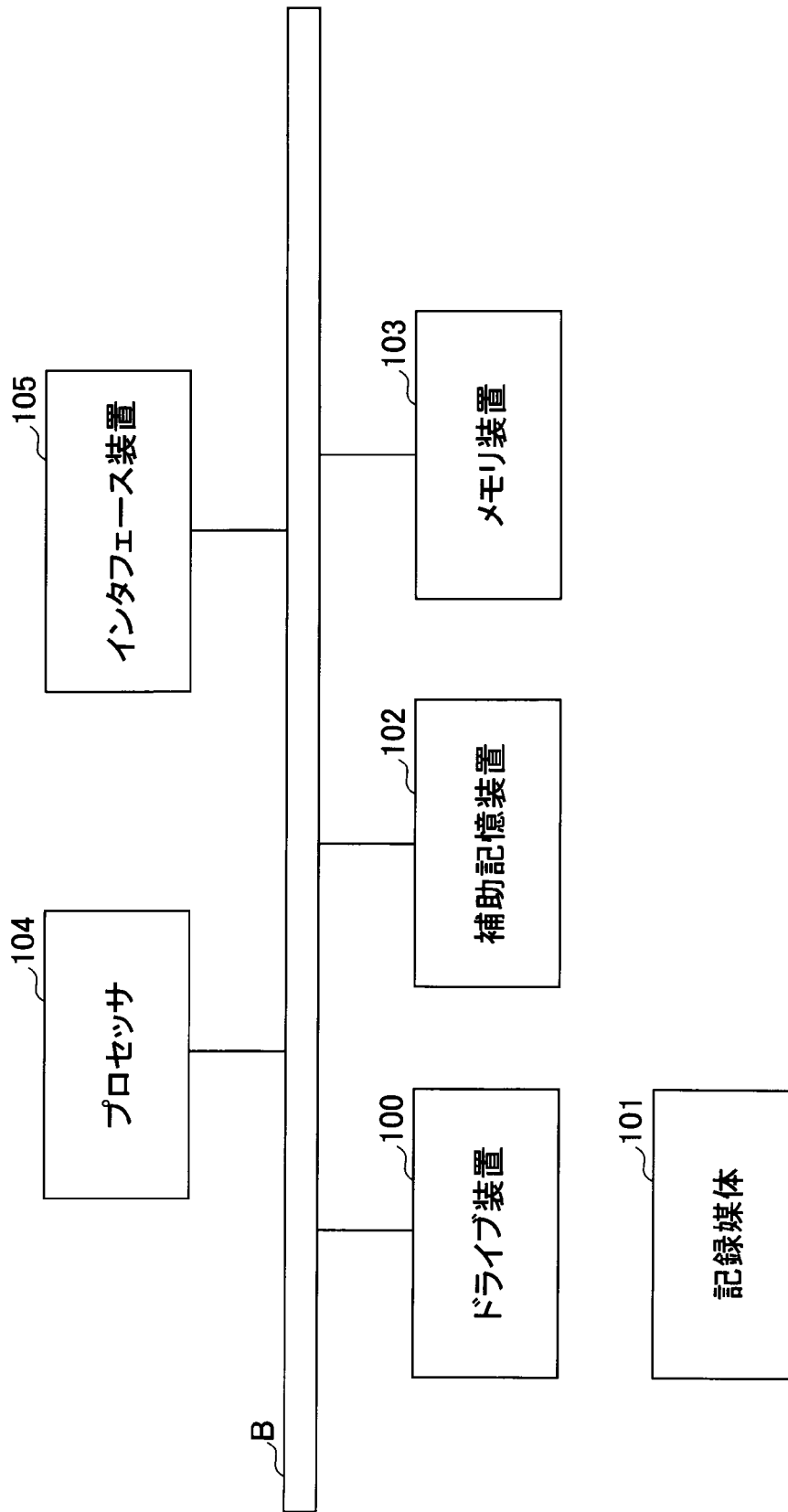
[請求項7] 請求項4乃至6いずれか一項記載のリソース量推定方法をコンピュータに実行させることを特徴とするプログラム。

[図1]

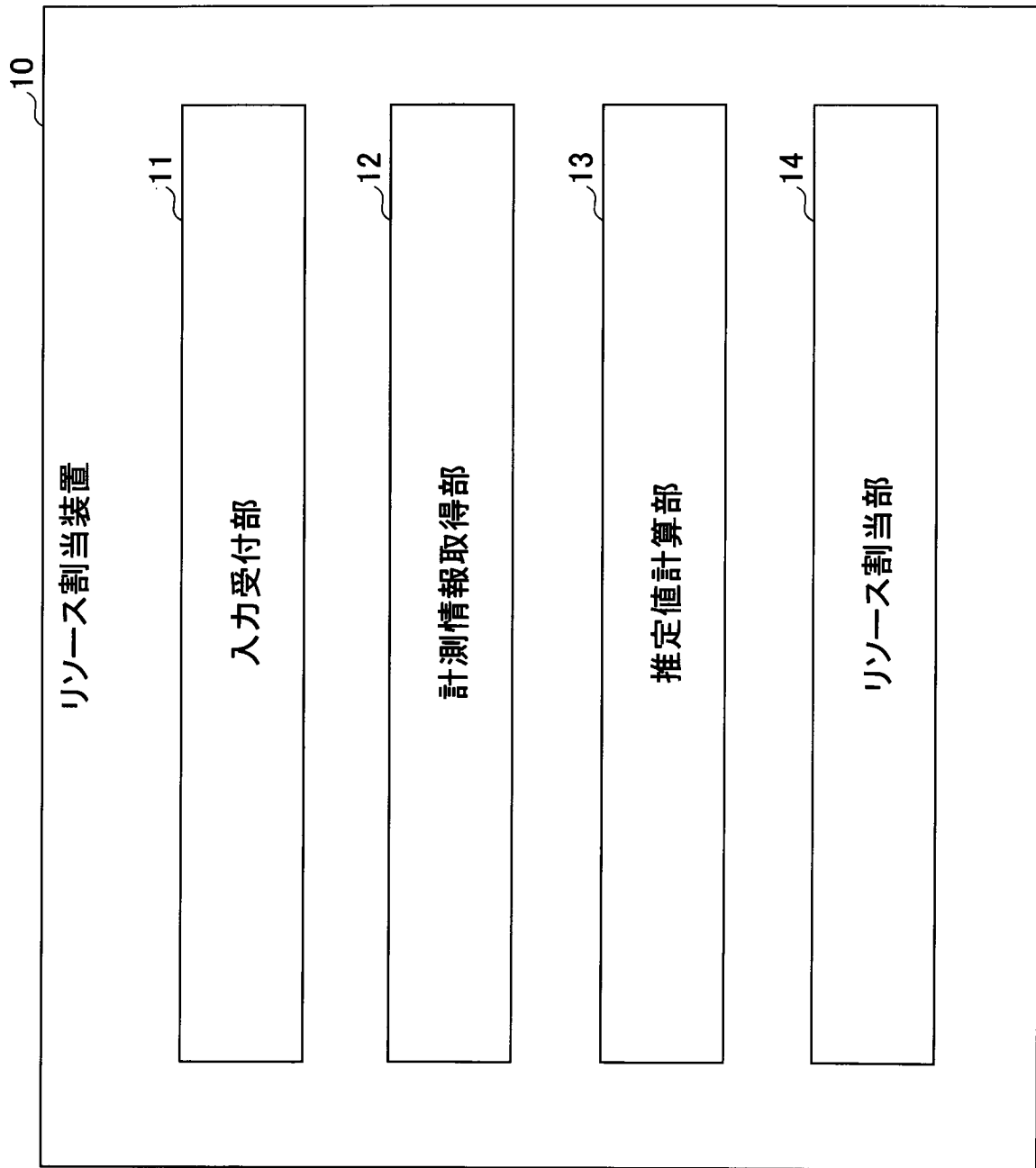


[図2]

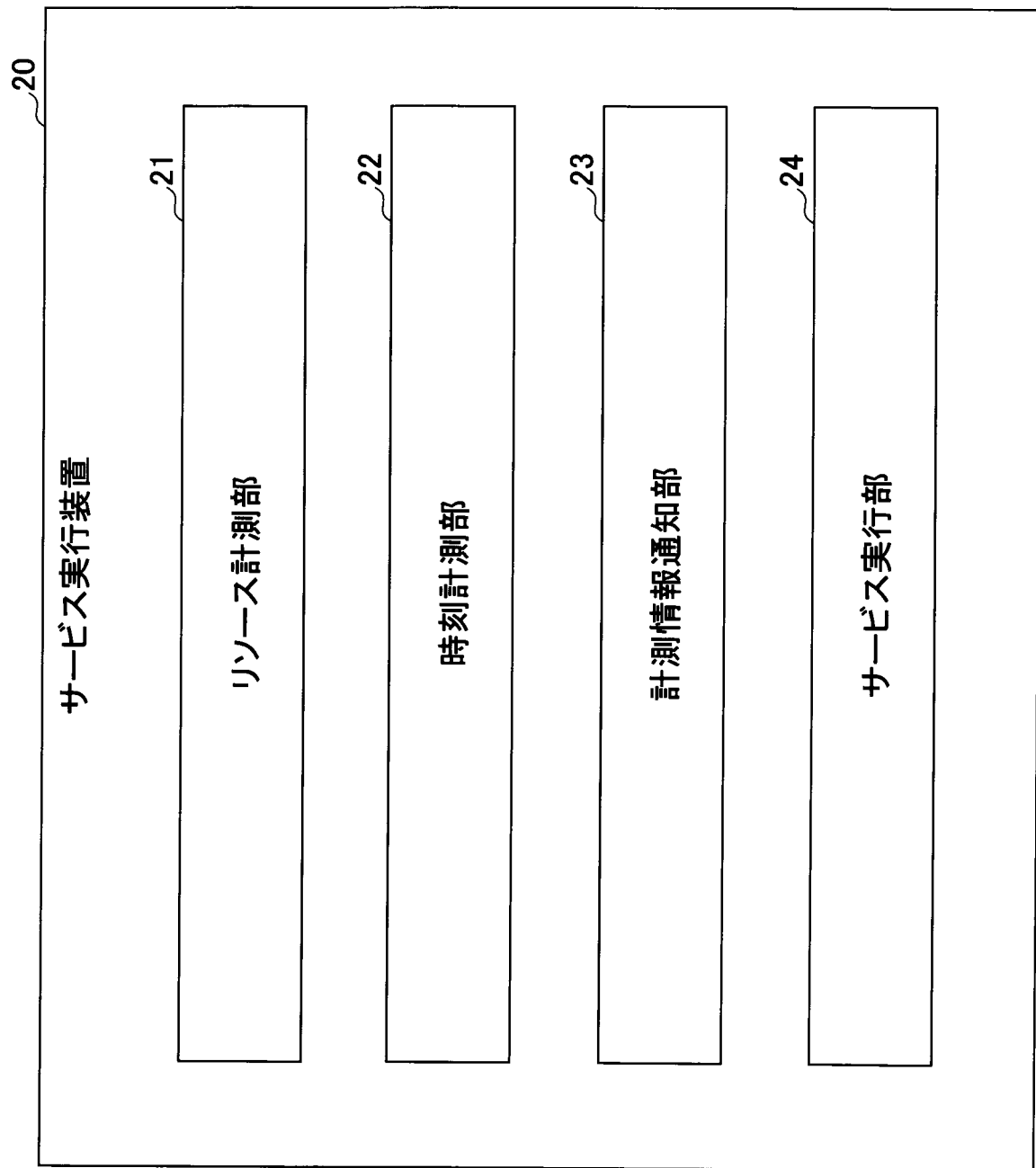
10



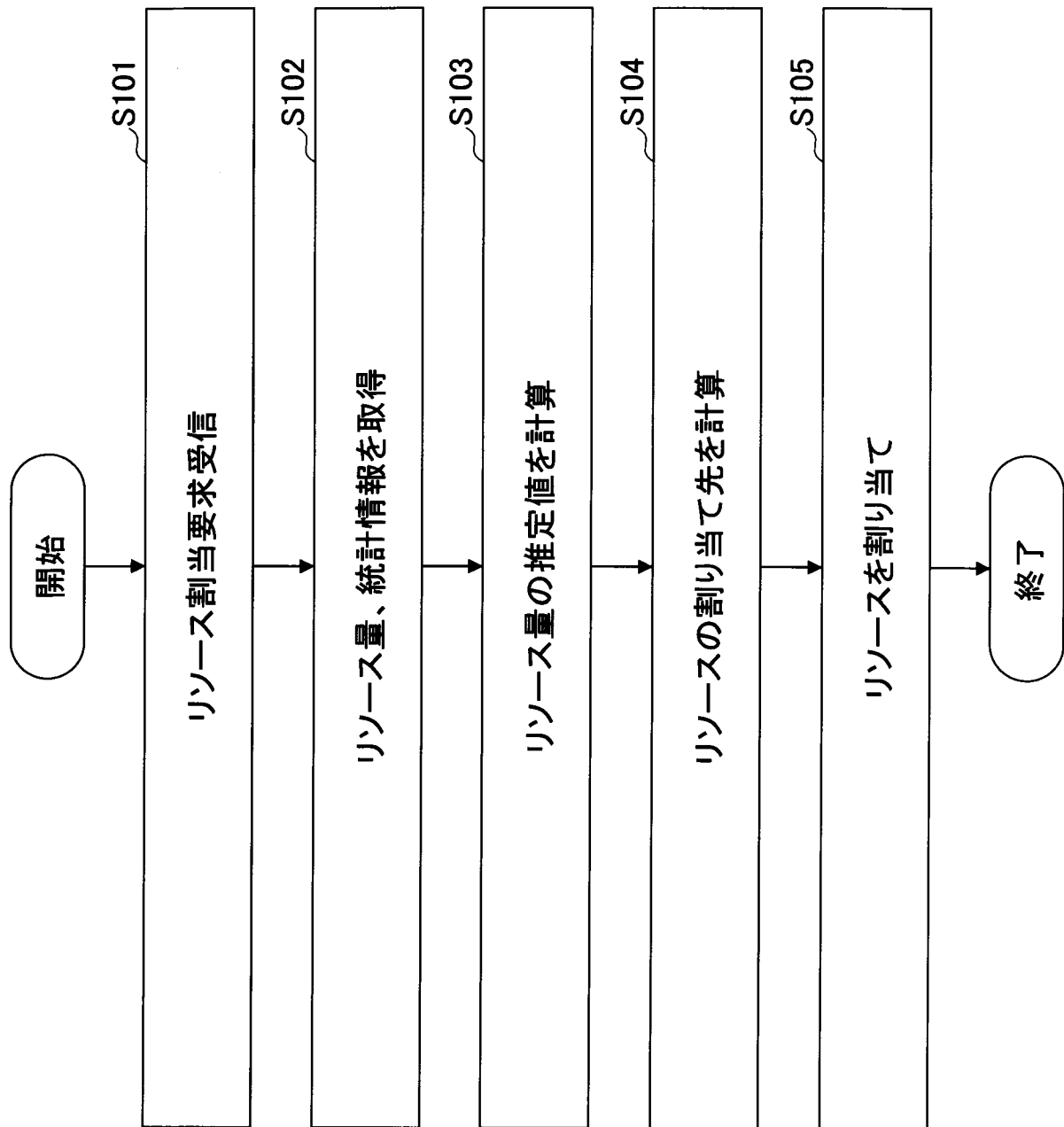
[図3]



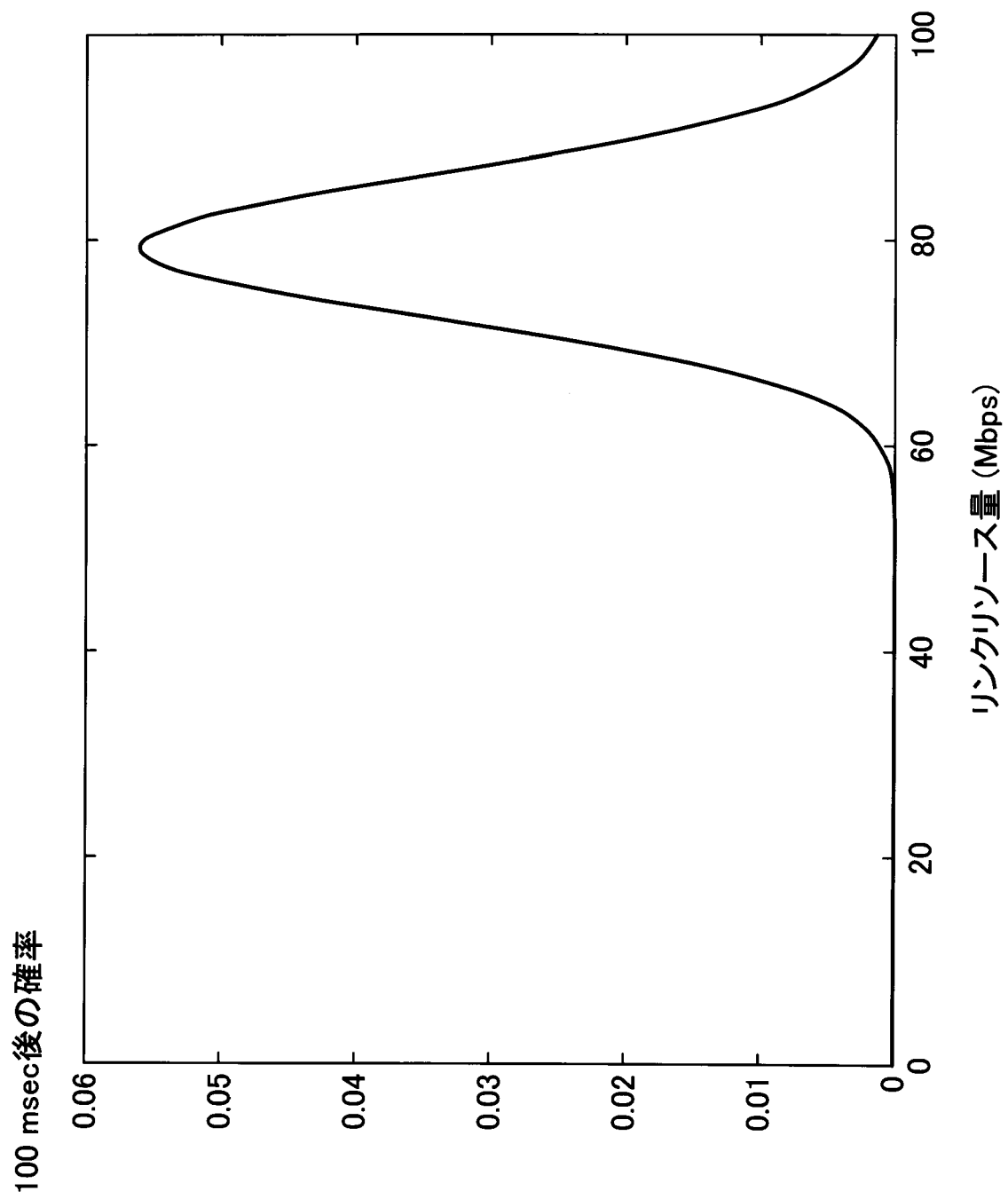
[図4]



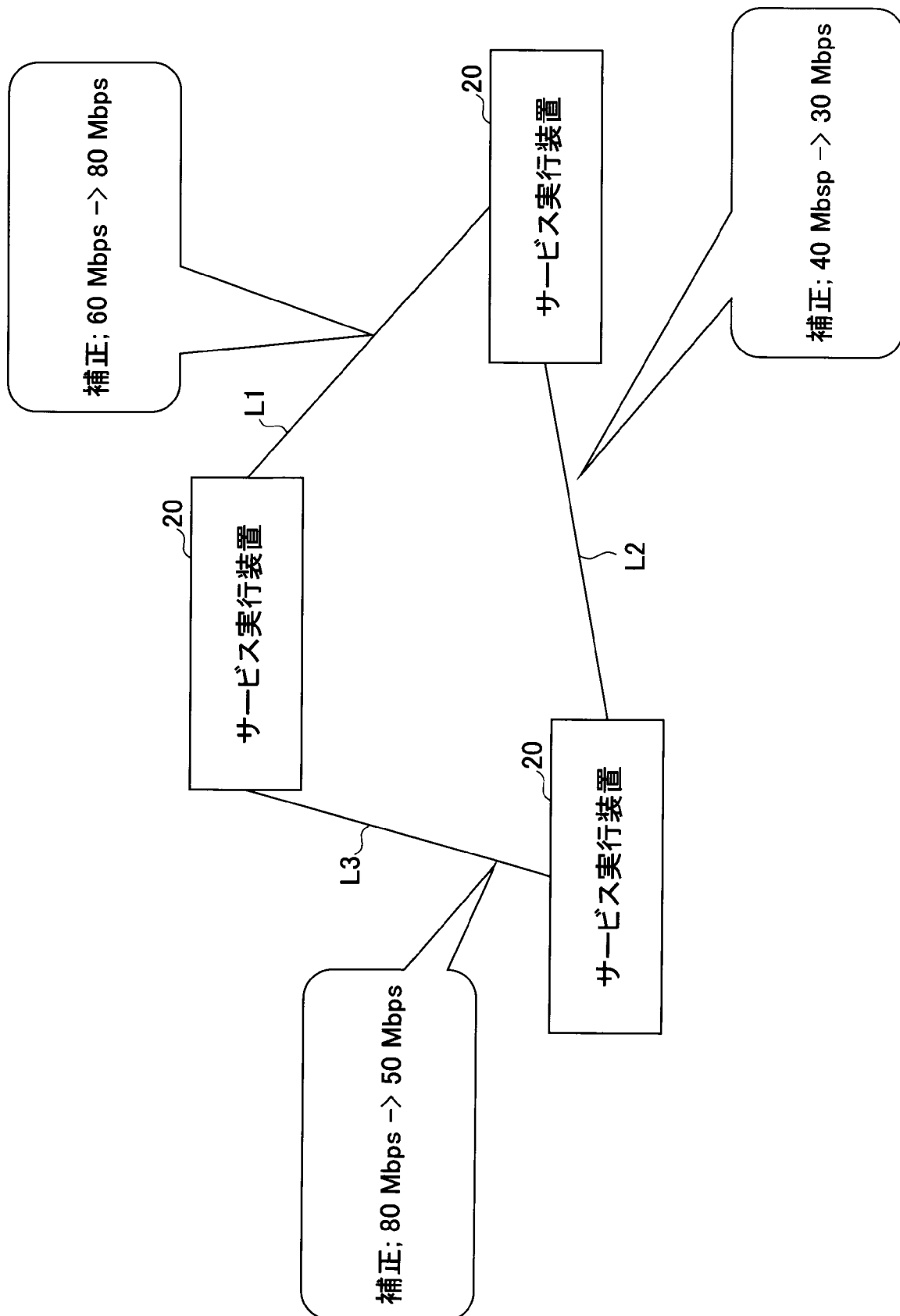
[図5]



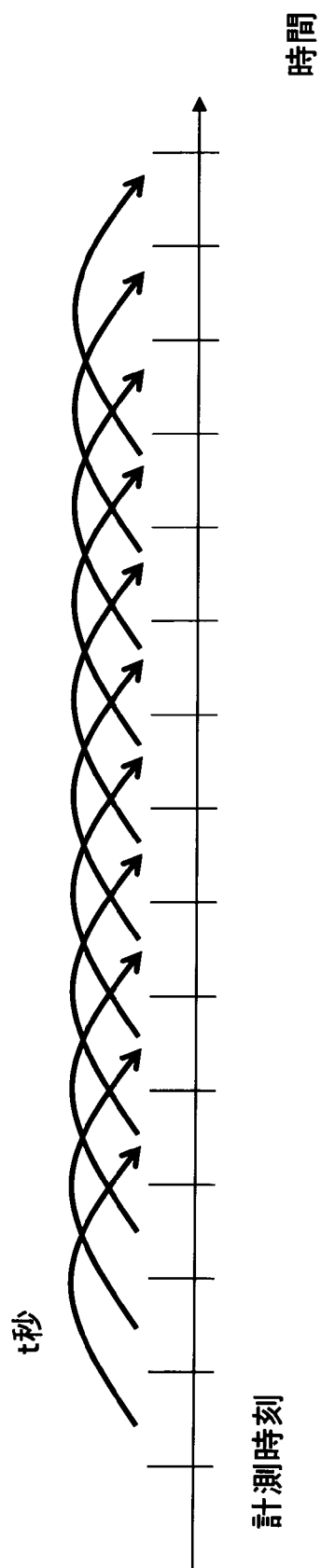
[図6]



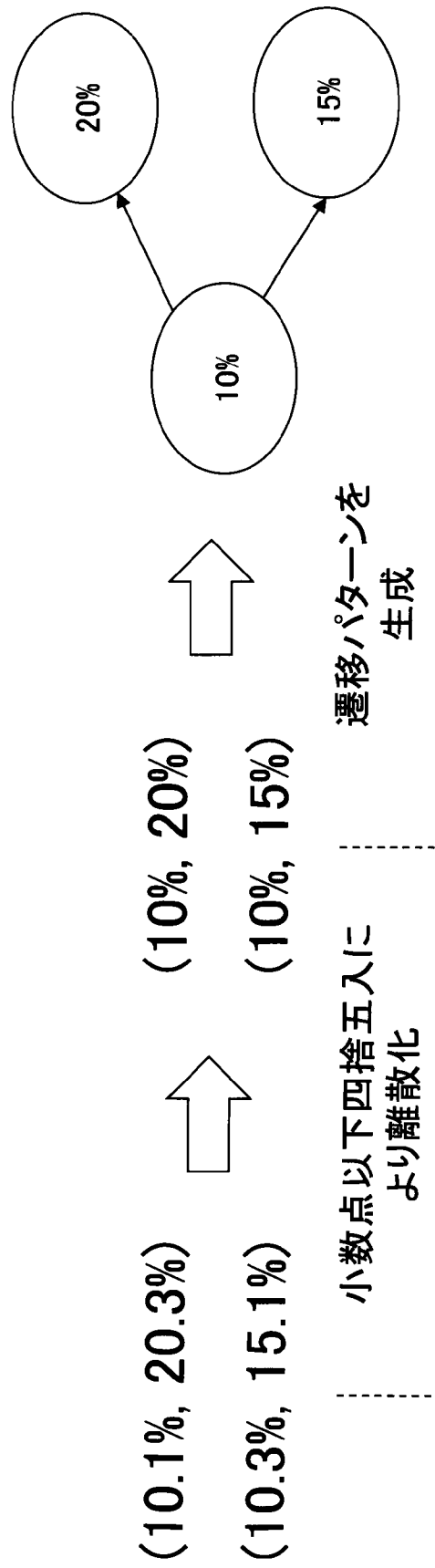
[図7]



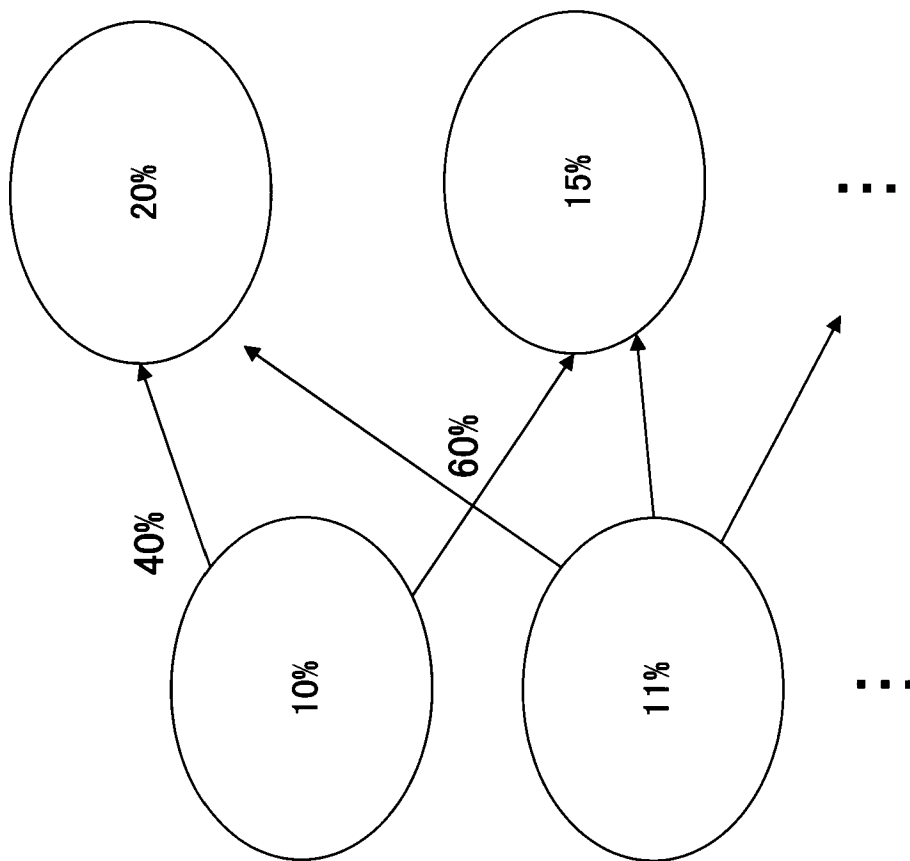
[図8]



[図9]



[図10]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/032029

A. CLASSIFICATION OF SUBJECT MATTER

G06F 9/50 (2006.01) i

FI: G06F9/50 120Z

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F9/50

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996

Published unexamined utility model applications of Japan 1971-2020

Registered utility model specifications of Japan 1996-2020

Published registered utility model applications of Japan 1994-2020

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2018-26050 A (FUJITSU LTD.) 15 February 2018 (2018-02-15) abstract	1-7
A	JP 2015-162701 A (KYOCERA CORP.) 07 September 2015 (2015-09-07) abstract	1-7



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"I" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
09 November 2020 (09.11.2020)Date of mailing of the international search report
17 November 2020 (17.11.2020)Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/JP2020/032029

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
JP 2018-26050 A	15 Feb. 2018	US 2018/0046505 A1 abstract	
JP 2015-162701 A	07 Sep. 2015	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06F 9/50(2006.01)i FI: G06F9/50 120Z														
B. 調査を行った分野														
調査を行った最小限資料（国際特許分類（IPC）） G06F9/50														
最小限資料以外の資料で調査を行った分野に含まれるもの <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922 - 1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971 - 2020年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996 - 2020年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994 - 2020年</td> </tr> </table>			日本国実用新案公報	1922 - 1996年	日本国公開実用新案公報	1971 - 2020年	日本国実用新案登録公報	1996 - 2020年	日本国登録実用新案公報	1994 - 2020年				
日本国実用新案公報	1922 - 1996年													
日本国公開実用新案公報	1971 - 2020年													
日本国実用新案登録公報	1996 - 2020年													
日本国登録実用新案公報	1994 - 2020年													
国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）														
C. 関連すると認められる文献														
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号												
A	JP 2018-26050 A（富士通株式会社）15.02.2018（2018 - 02 - 15） 【要約】	1-7												
A	JP 2015-162701 A（京セラ株式会社）07.09.2015（2015 - 09 - 07） 【要約】	1-7												
☐ C欄の続きにも文献が列挙されている。 ☒ パテントファミリーに関する別紙を参照。														
<table border="0"> <tr> <td>* 引用文献のカテゴリー</td> <td>“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの</td> </tr> <tr> <td>“A” 特に関連のある文献ではなく、一般的技術水準を示すもの</td> <td>“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの</td> </tr> <tr> <td>“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの</td> <td>“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの</td> </tr> <tr> <td>“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）</td> <td>“&” 同一パテントファミリー文献</td> </tr> <tr> <td>“O” 口頭による開示、使用、展示等に言及する文献</td> <td></td> </tr> <tr> <td>“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献</td> <td></td> </tr> </table>			* 引用文献のカテゴリー	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの	“A” 特に関連のある文献ではなく、一般的技術水準を示すもの	“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの	“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの	“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）	“&” 同一パテントファミリー文献	“O” 口頭による開示、使用、展示等に言及する文献		“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献	
* 引用文献のカテゴリー	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの													
“A” 特に関連のある文献ではなく、一般的技術水準を示すもの	“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの													
“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの													
“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）	“&” 同一パテントファミリー文献													
“O” 口頭による開示、使用、展示等に言及する文献														
“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献														
国際調査を完了した日 09.11.2020	国際調査報告の発送日 17.11.2020													
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 田中 幸雄 5B 9191 電話番号 03-3581-1101 内線 3545													

国際調査報告
パテントファミリーに関する情報

国際出願番号

PCT/JP2020/032029

引用文献			公表日	パテントファミリー文献	公表日
JP	2018-26050	A	15.02.2018	US 2018/0046505 A1 ABSTRACT	
JP	2015-162701	A	07.09.2015	(ファミリーなし)	