

(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
2014 年 7 月 3 日(03.07.2014)



(10) 国際公開番号
WO 2014/102996 A1

- (51) 国際特許分類:
H04L 12/701 (2013.01) **G06F 15/173** (2006.01)
- (21) 国際出願番号: PCT/JP2012/084019
- (22) 国際出願日: 2012 年 12 月 28 日(28.12.2012)
- (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: 株式会社日立製作所(HITACHI, LTD.)
[JP/JP]; 〒1008280 東京都千代田区丸の内一丁目
6 番 6 号 Tokyo (JP).
- (72) 発明者: 林 真人(HAYASHI Masato); 〒1858601 東京
都国分寺市東恋ヶ窪一丁目 2 8 0 番地 株式
会社日立製作所 中央研究所内 Tokyo (JP). 加藤
猛(KATO Takeshi); 〒1858601 東京都国分寺市東
恋ヶ窪一丁目 2 8 0 番地 株式会社日立製作所
中央研究所内 Tokyo (JP).
- (74) 代理人: 井上 学, 外(INOUE Manabu et al.); 〒
1008220 東京都千代田区丸の内一丁目 6 番 1 号
株式会社日立製作所内 Tokyo (JP).

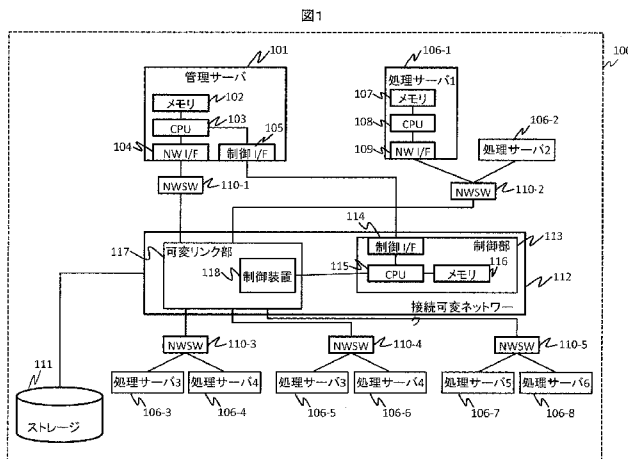
- (81) 指定国 (表示のない限り、全ての種類の国内保
護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA,
BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN,
CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES,
FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN,
IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR,
LS, LT, LU, LY, MA, MD, 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, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, 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, 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, ML, MR,
NE, SN, TD, TG).

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

(54) Title: INFORMATION PROCESSING SYSTEM

(54) 発明の名称: 情報処理システム



- 101 Management server
102, 107, 116 Memory
105, 114 Control interface
106-1, 106-2, 106-3, 106-4, 106-5, 106-6, 106-7, 106-8 Processing server
111 Storage
112 Variable connection network
113 Control unit
117 Variable link unit
118 Control section

(57) Abstract: With this information processing system, processing loads are dispersed among multiple computation nodes, communication loads resulting from the processing loads are predicted, and a network topology among the computation nodes is determined on the basis of the prediction results. Consequently, even in the event of an imbalance in the volumes of communication between the computation nodes during parallel processing as is the case with graph processing, a network topology in accordance with the imbalance in the communication volumes can be constructed, and thus the processing speed can be increased.

(57) 要約: 本発明の情報処理システムでは、処理負荷を複数の計算ノードに分散配置し、処理負荷により生じる通信負荷を予測し、予測結果に基づき計算ノード間のネットワークトポロジを決定する。これにより、グラフ処理のように並列処理時に計算ノード間の通信量に偏りが生じる場合であっても、通信量の偏りに応じたネットワークトポロジを構成することが可能となり、ひいては処理の高速化を図ることができる。

明 細 書

発明の名称： 情報処理システム

技術分野

[0001] 本発明は、情報処理システムに関し、特にリソース間を接続するネットワークの制御に関する。

背景技術

[0002] 本技術分野の背景技術として特許文献１に開示されている技術がある。この公報には、アプリケーションフレームワークによって制御される少なくとも一台の回線交換機と、多数の計算ノードが接続されたネットワークを備え、ＣＰＵ負荷やメモリ使用率、ボトルネック等の監視結果に基づいて回線交換機を制御する発明が開示されている。

[0003] 特許文献２では、ストリーム処理グラフを複数のプロセッサからなる処理ノードの集合（特許文献２ではスーパーノードクラスタ（Super node cluster）と記載）に割当て、該集合間が相互に接続されるように光回線交換機を制御する技術が開示されている。

先行技術文献

特許文献

[0004] 特許文献１：米国特許第８１２５９８４号明細書

特許文献２：米国特許第８０３７２８４号明細書

発明の概要

発明が解決しようとする課題

[0005] 近年、ソーシャルネットワークにおける人間関係や企業間の取引関係など、繋がりのある大規模なデータの解析が注目されている。このようなデータは一般的にグラフとして表現される。グラフを解析する際には処理速度とデータサイズの要求に応じて複数のサーバを用いて並列分散処理を行う必要が生じる。上記のような自然発生的なグラフは、一般に、複雑な構造を有し疎密の偏りが大きいいため、並列分散処理を行う際にサーバ間の通信負荷の偏り

が生じ易い。通信負荷の高いサーバは、通信負荷のために処理時間が長くなり、全体の処理性能のボトルネックになりやすいという問題を生じる。

[0006] ここで、前述した特許文献 1 に記載のシステムでは、実行中の負荷監視結果に基づいてネットワークの構成を変えることができるが、プログラム実行中のトポロジ変更処理自体がオーバーヘッドとなる。また、プログラム実行開始時にどのようなネットワークトポロジを用いるべきかについては言及されていない。

[0007] 一方、前述した特許文献 2 に記載のシステムでは、ストリームグラフの処理ノードの割当て結果をもとに光回線交換機を用いてネットワークトポロジを変更しているが、単に全ての処理ノード間で、発生する通信量の総合計を上回る帯域を与えると記載されている。大規模なグラフ処理になるほど処理ノード間の通信量が大きく偏る場合が増え、全ての処理ノード間に通信量の総合計を上回る帯域を与えると無駄が多くなり、物量の観点から不利である。

[0008] 本発明は上記の課題に鑑みてなされたものであり、その目的は各処理ノード間で発生する通信量の予測値をもとに、処理ノード間の通信量の偏りに応じたネットワークトポロジを決定可能な情報処理システムを提供することにある。

課題を解決するための手段

[0009] 本発明の情報処理システムでは、処理負荷を複数の計算ノードに分散配置し、処理負荷により生じる通信負荷を予測し、予測結果に基づき計算ノード間のネットワークトポロジを決定することで、上述の課題を解決する。

発明の効果

[0010] 本発明によれば、グラフ処理のように並列処理時に計算ノード間の通信量に偏りが生じる場合であっても、通信量の偏りに応じたネットワークトポロジを構成することが可能となり、ひいては処理の高速化を図ることができる。

図面の簡単な説明

[0011] [図1]情報処理システムの構成を示すブロック図である。

[図2]情報処理システムの機能を示すブロック図である。

[図3]接続可変ネットワークの説明図である。

[図4]情報処理システム全体の動作を示すフローチャートである。

[図5]情報処理システムにおける負荷配置の例を示す図である。

[図6]情報処理システムにおける負荷配置情報の例を示す図である。

[図7]情報処理システムにおける負荷とサーバの対応関係の例を示す図である。

[図8]情報処理システムにおけるサーバとネットワークスイッチの対応関係の例を示す図である。

[図9]情報処理システムにおいて負荷配置情報から通信負荷を予測する手順を示すフローチャートである。

[図10]情報処理システムにおいて負荷配置情報から通信負荷を予測した結果の例を示す図である。

[図11]情報処理システムにおいてトポロジ決定の例を示した図である。

[図12]情報処理システムにおいて、処理サーバ間のネットワークトポロジの例を示す図である。

[図13]情報処理システムの入力となるグラフの例を示す図である。

発明を実施するための形態

[0012] 以下、実施例を図面を用いて説明する。

実施例 1

[0013] 図13に、本発明の情報処理システムが扱う処理負荷の例として、入力グラフの例を示す。頂点が丸印で、辺が頂点間を結ぶ矢印でグラフ構造が表現されている。各頂点には固有の通し番号であるIDが付けられている。また辺で繋がった頂点どうしを「隣接する」と呼ぶ。具体的なグラフ処理としては、頂点間の最短経路を算出する処理や、グラフ中の重要な頂点を算出する中心性計算など様々なものがある。これらの多くは辺に沿って情報のやり取りを行うトラバース処理を中心とする処理である。

- [0014] トラバース処理は辺に沿って頂点間でデータをやり取りする処理で、以下のステップを頂点毎に行う。まず、隣接する頂点から辺に沿ってデータを受信する。次に、受信したデータをもとに計算を行い頂点のデータを更新する。そして更新したデータを辺に沿って隣接する頂点に送信する。
- [0015] このようなグラフ処理を複数のサーバ上で並列に実行する場合、グラフを分割してその一部をサーバに割当てて。このとき、辺の両端の頂点異なるサーバに割当てられていると、トラバース処理中に発生する辺に沿ったデータの送受信はネットワークを介したサーバ間の通信処理として実現される。
- [0016] 図 1 に、本発明の実施例である情報処理システム 100 を示す。情報処理システム 100 は、管理サーバ 101 と、複数の処理サーバ 106-1~8 と、ストレージ 111 と、を備える。管理サーバ 101 と処理サーバ 106-1~8 は、夫々対応するネットワークスイッチ 110-1~110-5 に接続されている。ネットワークスイッチ 110-1~5 は、接続可変ネットワーク装置 112 を介して相互に接続されている。処理サーバ 106-1~8 は、並列処理を行う計算ノードとして働く。
- [0017] 管理サーバ 101 は、メモリ 102 と、中央処理装置（CPU）103 と、ネットワークインターフェース 104 と、接続可変ネットワーク制御インターフェース 105 と、を備える。管理サーバ 101 は、情報処理システム 100 内で、管理ノードとして働く。なお、管理サーバ 101 は、さらに計算ノードとして処理サーバ 106-1~8 と同様に働くことも可能である。
- [0018] 接続可変ネットワーク制御インターフェース 105 は、接続可変ネットワーク装置 112 に応じて専用の接続を用いるほか、例えば USB、シリアル接続など汎用の接続を用いることができる。処理サーバ 106-1~8 は同様のハードウェア構成を備え、夫々メモリ 107 と、CPU 108 とネットワークインターフェース 109 と、を備える。
- [0019] 接続可変ネットワーク装置 112 について図 4 を併用しながら説明する。接続可変ネットワーク装置 112 は、複数のポート 302 を備え、接続対象のネットワークスイッチとひとつまたは複数のリンク 301 を用いて接続さ

れる。リンク301には、例えば光ファイバや電気ケーブルを用いることができる。また、接続可変ネットワーク装置112は、制御部113と、可変リンク部117とを備える。制御部113は、制御インターフェース(I/F)114と、CPU115と、メモリ116と、を備える。可変リンク部117は、制御装置118と、複数の可変リンク303と、を備える。可変リンク部117は、制御装置118からの指示に基づいて任意のポート間に可変リンク303を張ることができる。図4の例ではポート1と7、ポート2と4、ポート8と10、ポート6と12が接続されている状態を示している。可変リンク303を制御することは接続可変ネットワーク装置112に接続されているネットワークスイッチ間の物理的な結線を変更することに相当する。これにより、固定された結線のネットワーク上でルーティングを制御する場合と比べて、より柔軟にネットワークスイッチ間の帯域を決めることが可能となる。接続可変ネットワーク装置112としては、例えばMEMS(Micro Electro Mechanical Systems)による光スイッチ等を用いることができる。

[0020] 図2に情報処理システム100の機能ブロック図を示す。管理サーバ101および処理サーバの各部と情報はそれぞれのメモリ上に格納される。管理サーバ101は、メモリ102上に、負荷配置部201と、予測通信量集計部202と、予測通信量通知部203と、通信量予測結果204と、負荷配置リスト205と、ネットワークスイッチ対応情報206と、を格納する。処理サーバ106-1~8は、メモリ107に、夫々通信量予測部207と、処理部208と、負荷配置リスト205と、ネットワークスイッチ対応情報206と、処理負荷情報209と、を格納する。ストレージ111は、負荷データ212と、ネットワークスイッチ情報206と、を格納する。

[0021] 図4に、情報処理システム100の動作フローを示す。ステップ401では、管理サーバ101が、ストレージ111上に格納されたネットワークスイッチ対応情報206を読み込んでメモリ102上に格納する。その後、管理サーバ101は、全ての処理サーバ106-1~8にスイッチ対応情報2

06を送信する。各処理サーバ106-1~8は、受信したネットワークスイッチ対応情報を夫々のメモリ107に格納する。本実施例では、各サーバは固定的にネットワークスイッチに接続されており、実行中にサーバとスイッチの対応関係が変化しないため、静的な構成情報として事前に準備されているものとした。図8にスイッチ対応情報の例を示す。各サーバの識別情報(ID)に対して、各サーバが接続されているネットワークスイッチのIDが格納されている。なお、対応関係に規則性があれば、対応情報を数式で表現してもかまわない。

[0022] ステップ402では、管理サーバ101の負荷配置部201が、ストレージ111上に格納された負荷データ212を読み込んで処理サーバ106-1~8のメモリ107上に格納する。分散処理を前提とするため、各処理サーバは夫々異なる負荷データの部分をメモリに格納し処理を行う。例えば、管理サーバ101の負荷配置部201は、ストレージ111から負荷データ212のデータサイズを取得する。その後、管理サーバ101の負荷配置部201は、負荷データ212のデータサイズを処理サーバ数（本実施例では8）で割って、処理サーバ1台分の読み込みサイズを決定する。その後、管理サーバ101の負荷配置部201は、各処理サーバに読み込む範囲を通知する。各処理サーバは、受け取った読み込み範囲に従いストレージ111から負荷データ212の一部を読み取り夫々のメモリ107に処理負荷情報209として格納する。

[0023] 図5にステップ402による負荷配置イメージを示す。処理サーバ1は頂点1~6、処理サーバ2は頂点7~12というように、各サーバに連続した番号の頂点が格納されている。図6は、各処理サーバ106-1~8に格納されている配置情報の例である。各リストには、各サーバに配置されている頂点と、該頂点に隣接する頂点の番号とが格納されている。例えば、サーバ1には頂点1~6が割り当てられており、頂点1には頂点2と頂点43が隣接していることを示す。

[0024] ステップ403では、ステップ402の結果をもとに各サーバ間の通信量

の予測を行う。ステップ403の詳細な手順を図9に示す。通信量予測処理は管理サーバ101からの指示に基づいて処理サーバ106-1~8が夫々独立に実行し、最終的に管理サーバ101の予測通信量集計部202が結果を集計する。

[0025] ステップ901では、管理サーバ101の予測通信量集計部202は、処理サーバの中からまだ予測処理を実行していないサーバを一つ選択する。例えば、管理サーバ101は、処理サーバ106-1を選択する。

[0026] ステップ902では、管理サーバ101は、ステップ901で選択した処理サーバに対して、通信量予測処理を開始するよう指示する。

[0027] ステップ903では、ステップ901で選択された処理サーバの通信量予測部207が、自身のメモリ上に格納されているネットワークスイッチ対応情報206を参照し、自身と接続しているネットワークスイッチのIDを取得する。図8のとおり、処理サーバ106-1は、ネットワークスイッチ1と接続している。また、該処理サーバの通信量予測部207は、取得したIDに対応するネットワークスイッチと、それ自身を含むすべてのネットワークスイッチとの間の通信量のカウンタを0に初期化する。

[0028] ステップ904では、ステップ901で選択された処理サーバの通信量予測部207が、該処理サーバに割り当てられている頂点のうち、まだ処理されていないものを選択する。選択の順序は任意であるが、例えばIDの小さい順から選択する。選択中の処理サーバ1は図6に示すとおり頂点1~6が割り当てられていて、初期状態ではどの頂点も処理されていないので、最初に頂点1が選択される。

[0029] ステップ905では、ステップ901で選択された処理サーバの通信量予測部207が、ステップ904で選択した頂点に隣接する頂点のうち、まだ処理されていないものを選択する。選択の順序は任意であるが、例えばIDの小さい順から選択する。図6のとおり、ステップ904で選択した頂点1には頂点2と頂点43が隣接している。初期状態ではいずれも処理されていないので頂点2を選択する。

- [0030] ステップ906では、ステップ901で選択された処理サーバの通信量予測部207が、負荷配置リスト205を参照することによりステップ905で選択した頂点の所属するサーバのIDを取得する。図7のとおり、ステップ905で選択した頂点2はサーバ1に所属していることが分かる。
- [0031] ステップ907では、ステップ901で選択された処理サーバの通信量予測部207が、ネットワークスイッチ対応情報206を参照することによりステップ906で取得したIDに対応するサーバが接続しているネットワークスイッチのIDを取得する。図9のとおり、サーバ1はネットワークスイッチ1に接続していることがわかる。
- [0032] ステップ908では、ステップ901で選択された処理サーバの通信量予測部207が、ステップ903で取得したIDに対応するネットワークスイッチとステップ907で取得したIDに対応するネットワークスイッチの間の通信量を加算する。ここではステップ907で取得したネットワークスイッチのIDは1なので、ネットワークスイッチ1に対する通信量に1を加算する。
- [0033] ステップ910では、ステップ901で選択された処理サーバの通信量予測部207が、ステップ904で選択された頂点に隣接する頂点のうち、未処理のものがあるかどうか判定する。未処理のものがある場合、ステップ905に戻り、該処理サーバは前回実行時と異なる隣接頂点を選択する。未処理のものが無い場合、ステップ911に進む。ここでは、上記のとおり選択中の頂点1に頂点2と頂点43が隣接しており、頂点43が未処理であるため判定の結果がyesとなり、ステップ905に戻り頂点43が選択される。
- [0034] ステップ911では、ステップ901で選択された処理サーバの通信量予測部207が、ステップ901で選択された処理サーバに割り当てられている頂点のうち、まだ処理されていないものがあるかどうか判定する。処理されていないものがある場合、ステップ904に戻り、該処理サーバの通信量予測部207は、前回実行時と異なる頂点を選択する。処理されていないも

のが無い場合、ステップ 9 1 2 に進む。ここでは、上記のとおり選択中のサーバ 1 には頂点 1 ～ 6 が割り当てられており、頂点 2 ～ 6 は未処理であるため、判定結果が y e s となり、ステップ 9 0 4 に戻り、異なる頂点が選択される。

[0035] ステップ 9 1 2 では、ステップ 9 0 1 で選択された処理サーバが、ステップ 9 0 3 ～ 9 1 1 で求めた結果を管理サーバ 1 0 1 に送信する。

[0036] ステップ 9 1 3 では、管理サーバ 1 0 1 の予測通信量集計部 2 0 2 が、通信量予測処理を行っていない処理サーバがあるかどうかを判定する。通信量予測処理を行っていない処理サーバがある場合、ステップ 9 0 1 に戻り、管理サーバ 1 0 1 は、前回実行時と異なる処理サーバを選択する。通信量予測処理を行っていない処理サーバが無い場合、管理サーバ 1 0 1 は通信量予測処理を終了する。

[0037] 図 1 0 に、ステップ 4 0 3 の出力結果である通信量予測結果を示す。この通信量予測結果は、ネットワークスイッチの組ごとに予測される通信量を格納した行列の形式で示されている。計算結果は通信量予測結果 2 0 4 として管理サーバ 1 0 1 のメモリ 1 0 2 に格納される。説明の便宜上、処理サーバを逐次を選択して通信量予測処理を行うようにしたが、各処理サーバで並列に通信量予測処理を行っても良い。

[0038] ステップ 4 0 4 では、管理サーバ 1 0 1 の予測通信量通知部 2 0 3 が、ステップ 4 0 3 で計算された通信量予測結果 2 0 4 を制御 I / F 1 0 5 を通じて接続可変ネットワーク装置 1 1 2 に送信する。

[0039] ステップ 4 0 5 では、接続可変ネットワーク装置 1 1 2 の CPU 1 1 5 で実行されるトポロジ最適化部 2 1 3 が、受信した通信量予測結果に応じたトポロジを決定する。トポロジの決定方法としては、例えば全体の通信量に対する特定のネットワークスイッチ間通信量の割合に応じてリンクを分配する方法がある。

[0040] 例えば、図 1 0 に示した通信量予測結果から、全てのネットワークスイッチの組の通信量を列挙すると図 1 1 のような結果が得られる。各組の通信量

を合計すると全体の通信量が得られる。図 11 の例では全体の通信量は 30 となる（図示せず）。こうして求めた全体の通信量に対して各ネットワークスイッチの組の通信量の割合を求めると、図 11 の“割合”列のような結果が得られる。この割合に従ってリンクを分配する。図 3 に示したように各ネットワークスイッチから 3 ポートに対してリンクを張っている場合、全部で 12 ポート分が可変リンクで利用可能である。可変リンク 1 本につき 2 ポートを使用するため、使用できる可変リンクの数は最大で 6 本となる。前記割合に沿って 6 本を分配すると図 11 の“リンク本数”列のようになる。リンク本数は整数値しか取れないため、小数点以下は切り捨てる。CPU 115 で実行されるトポロジ最適化部 213 は、このようにして得られたトポロジ最適化結果を制御装置 118 に送り、最適結果を反映するように可変リンク 303 を変更する。図 12 は、この結果を反映したトポロジの例である。特定のネットワークスイッチ間でポート 8 と 11 及びポート 9 と 10 という 2 本の可変リンクが張られており、当該スイッチ間でより多くの帯域を使用する事が可能になっている。

[0041] ステップ 406 では、管理サーバ 101 が、各処理サーバ 106-1~8 に処理開始の準備が整ったことを通知する。通知を受けた各処理サーバ 106-1~8 は、夫々の持つ処理部 208 の実行を開始する。

[0042] 一連の手順により、グラフデータの不均一性に起因する通信量の偏りを反映したネットワークトポロジを構成することができる。これにより、サーバ間の通信で発生していたボトルネックを解消することができるため処理を高速に実行することが可能になる。

符号の説明

[0043] 100：情報処理システム、101：管理サーバ、102：メモリ、103：CPU、104：ネットワークインターフェース、105：制御インターフェース、106-1~8：処理サーバ、107：メモリ、108：CPU、109：ネットワークインターフェース、110-1~5：ネットワークスイッチ、111：ストレージ、112：接続可変ネットワーク装置、1

1 3 : 制御部、1 1 4 : 制御インターフェース、1 1 5 : CPU、1 1 7 : 可変リンク部、1 1 8 : 制御装置、2 0 1 : 負荷配置部、2 0 2 : 予測通信
量集計部、2 0 3 : 予測通信量通知部、2 0 4 : 通信量予測結果、2 0 5 :
負荷配置リスト、2 0 6 : ネットワークスイッチ対応情報、2 0 7 : 通信量
予測部、2 0 8 : 処理部、2 0 9 : 処理負荷情報、2 1 2 : 負荷データ、2
1 3 : トポロジ最適化部、3 0 1 : リンク、3 0 2 : ポート、3 0 3 : 可変
リンク、1 3 0 1 : グラフの頂点、1 3 0 2 : グラフの辺。

請求の範囲

- [請求項1] 複数の計算ノードと、
前記複数の計算ノード間の接続を可変とするネットワーク装置と、
を有し、
処理負荷を前記複数の計算ノードに分散して配置し、
前記処理負荷によって生じる前記複数のノード間の通信量を予測し、
予測結果に基づいて前記接続を決定することを特徴とする情報処理システム。
- [請求項2] 請求項1に記載の情報処理システムにおいて、
前記処理負荷はグラフ構造を有しており、
前記処理負荷を分散して配置する際に、グラフ頂点を各計算ノードに配置し、
前記通信量を予測する際に、互いに異なる計算ノードに配置されたグラフ頂点を接続する辺の数に基づいて前記通信量を予測することを特徴とする情報処理システム。
- [請求項3] 請求項1に記載の情報処理システムにおいて、
管理ノードを有し、
前記管理ノードが、
前記処理負荷を前記複数の計算ノードに分散して配置し、
前記予測を各計算ノードに指示することを特徴とする情報処理システム。
- [請求項4] 請求項1に記載の情報処理システムにおいて、
前記複数の計算ノードは、グループに分かれてグループ毎にスイッチに接続され、
各スイッチが前記ネットワーク装置の各ポートに接続されていることを特徴とする情報処理システム。
- [請求項5] 請求項4に記載の情報処理システムにおいて、

前記予測結果に基づいて前記接続を決定する際に、

前記予測結果から各スイッチ間の通信量を予測して前記接続を決定することを特徴とする情報処理システム。

[請求項6]

請求項4に記載の情報処理システムにおいて、

各スイッチは、それぞれ複数の前記ポートに接続されていることを特徴とする情報処理システム。

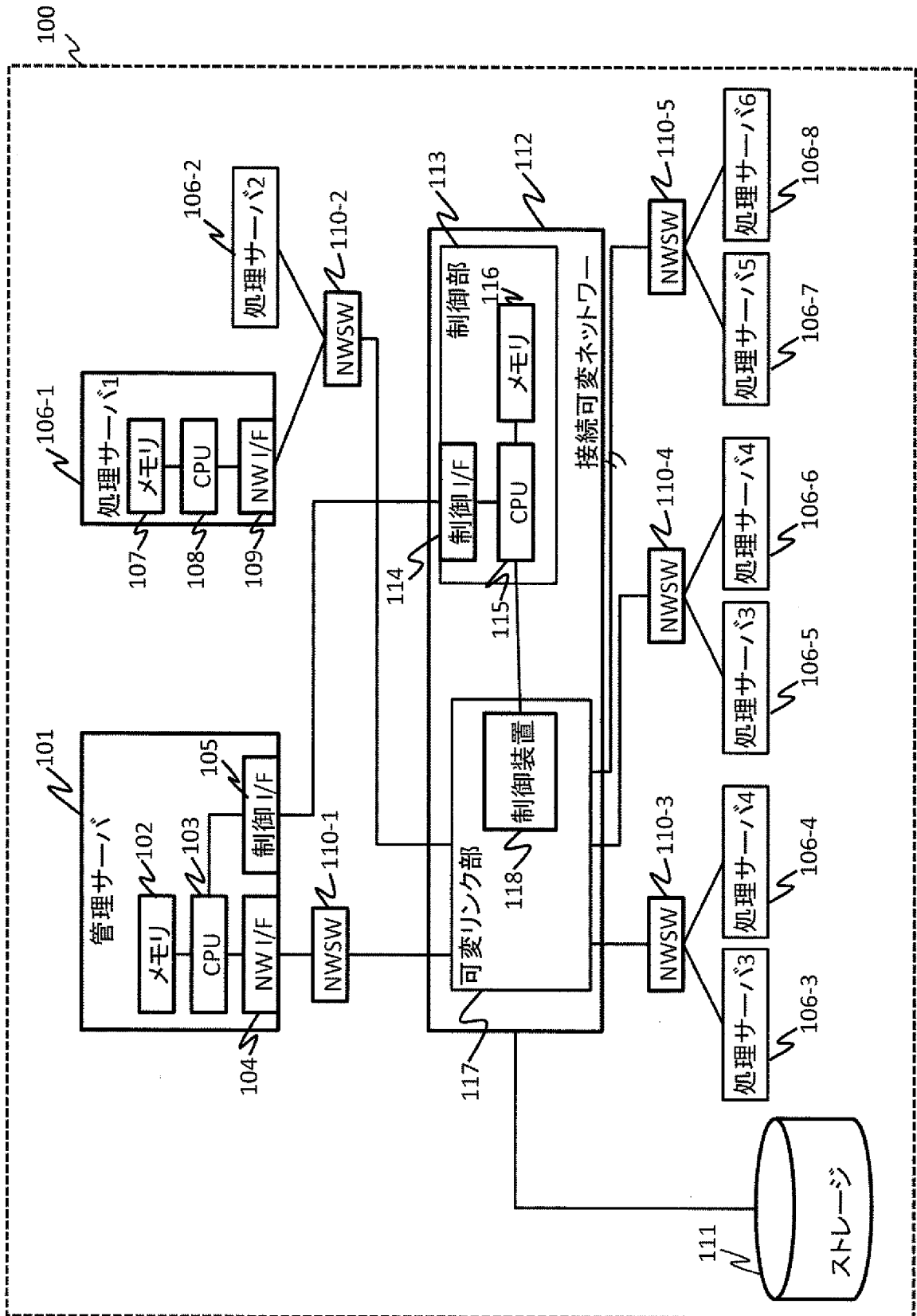
[請求項7]

請求項1に記載の情報処理システムにおいて、

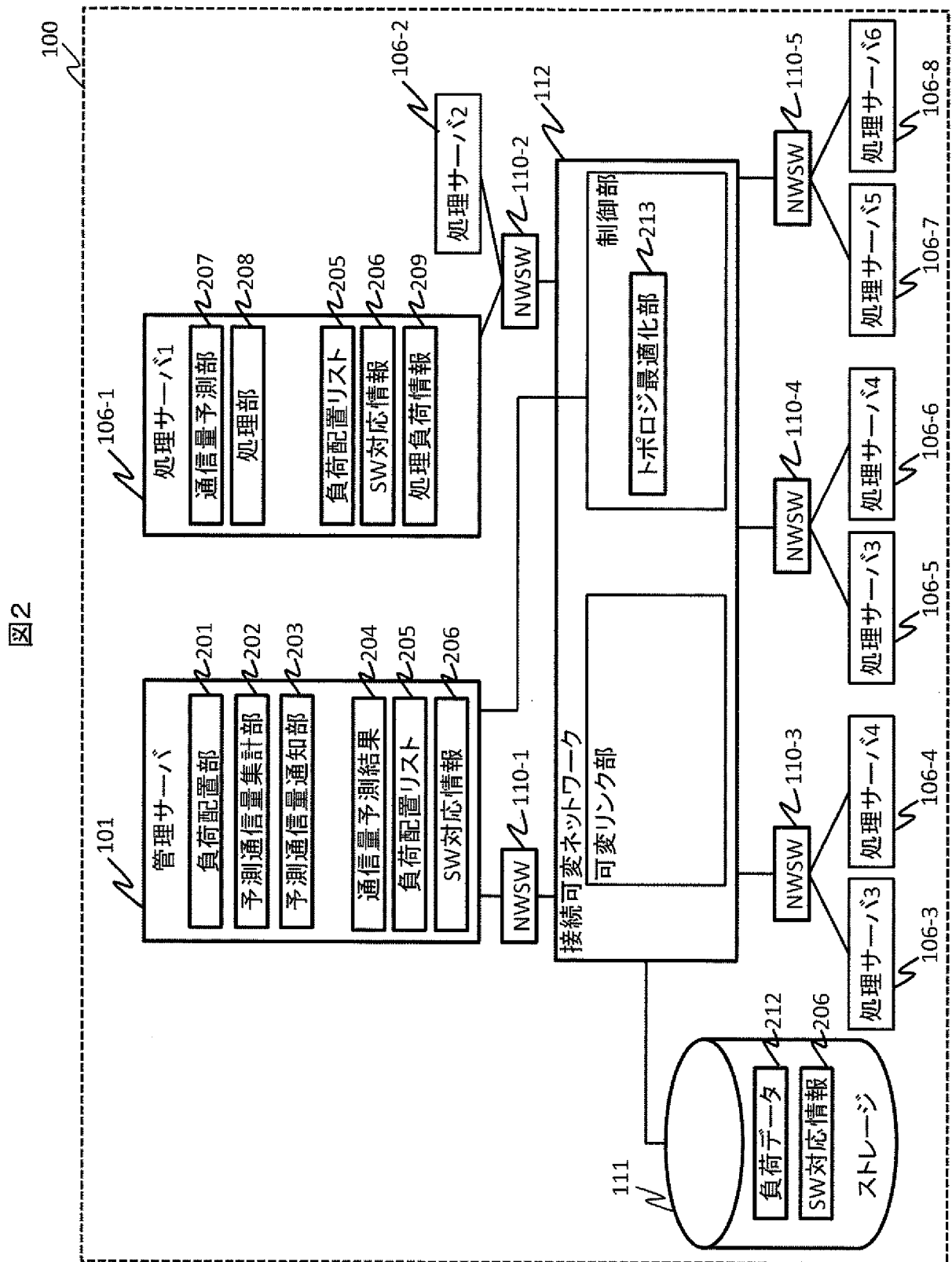
前記ネットワーク装置は、光スイッチであることを特徴とする情報処理システム。

[図1]

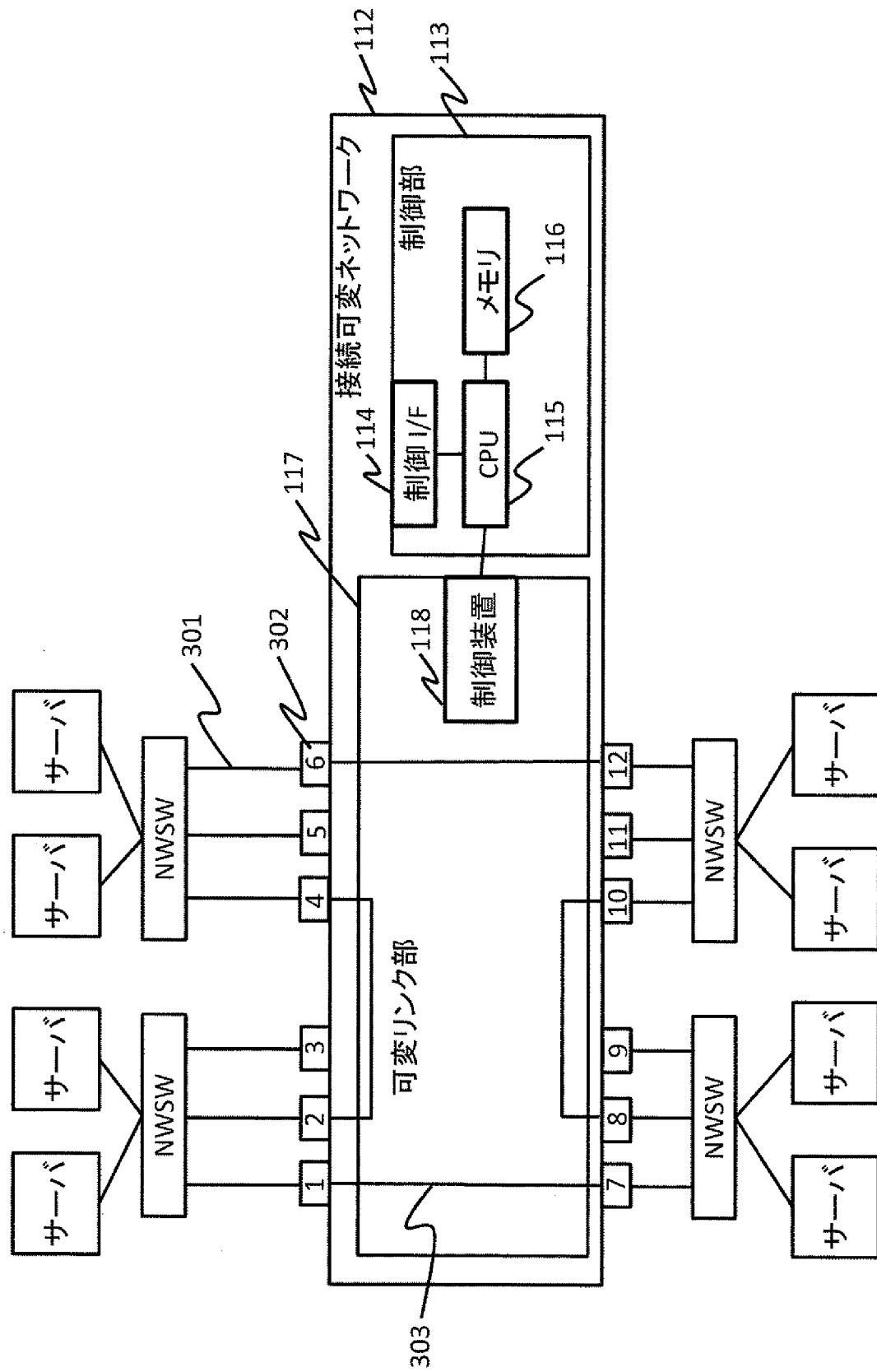
図1



[図2]

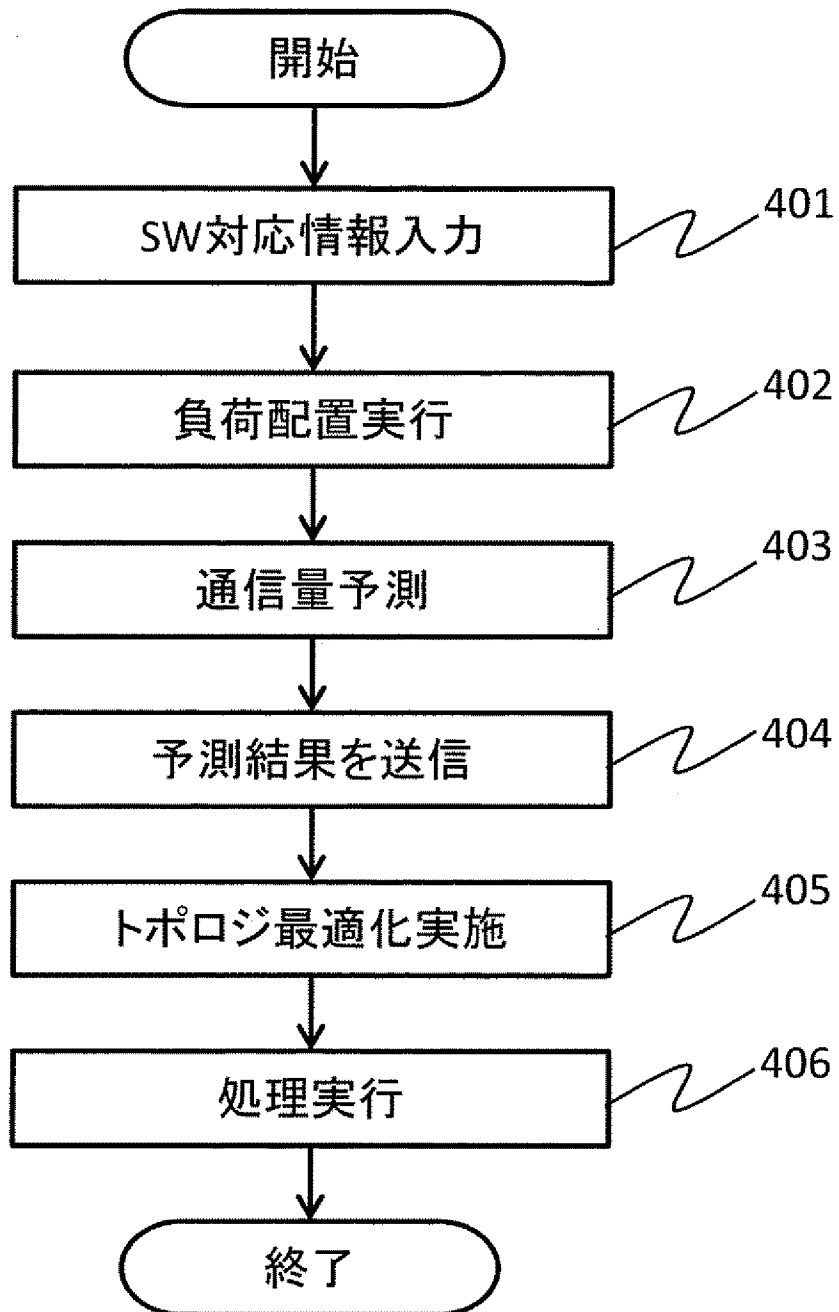


[図3]

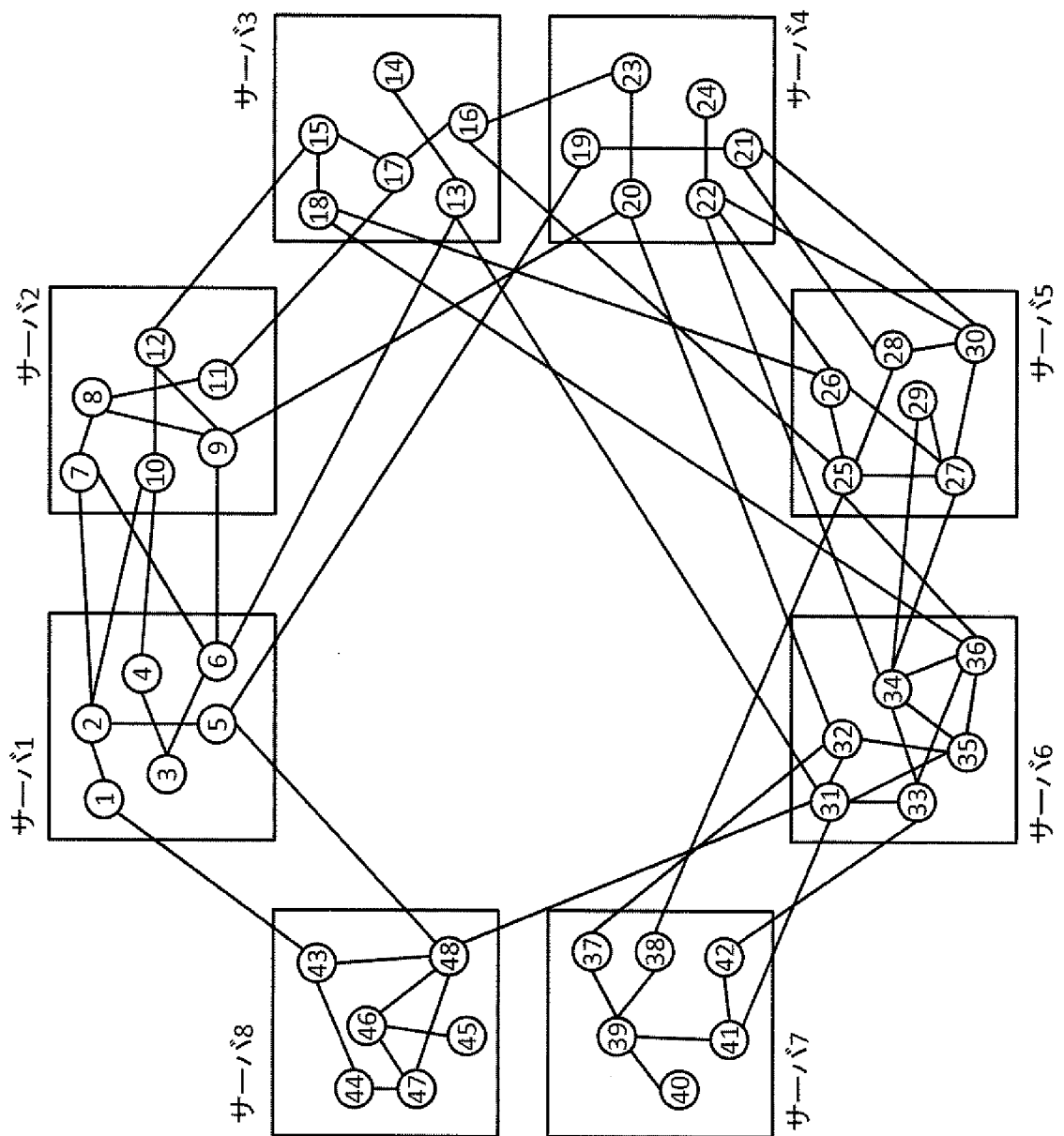


[図4]

図4



[図5]



[図6]

図6

サーハバ別隣接リスト

サーハバ1		サーハバ3		サーハバ5		サーハバ7	
頂点	隣接頂点	頂点	隣接頂点	頂点	隣接頂点	頂点	隣接頂点
1	2,43	13	6,14,31	25	16,26,27,36,38	37	32,39
2	1,5,7,10	14	13	26	18,22,25,27	38	25,39
3	4,6	15	12,17,18	27	25,26,29,30,34	39	37,38,40
4	3,10	16	17,23,25	28	21,25,30	40	39
5	2,19,48	17	11,15,16	29	27,34	41	31,39,42
6	3,7,9,13	18	15,26,36	30	21,22,28,27	42	33,41

サーハバ2		サーハバ4		サーハバ6		サーハバ8	
頂点	隣接頂点	頂点	隣接頂点	頂点	隣接頂点	頂点	隣接頂点
7	2,6,8	19	5,21	31	13,32,33,35,41,48	43	1,44,48
8	7,9,11	20	9,23,32	32	20,31,35,37	44	43,47
9	6,8,12,20	21	19,28,30	33	22,31,36,42	45	46
10	2,4,12	22	24,26,30,34	34	22,27,29,33,35,36	46	45,47,48
11	8,17	23	16,20	35	31,32,34,36	47	44,46,48
12	9,10,15	24	22	36	18,25,33,35	48	5,31,43,46,47

[図7]

図7

配置リスト(各サーバ共通)

頂点番号	割当サーバ
1～6	1
7～12	2
13～18	3
19～24	4
25～30	5
31～36	6
37～42	7
43～48	8

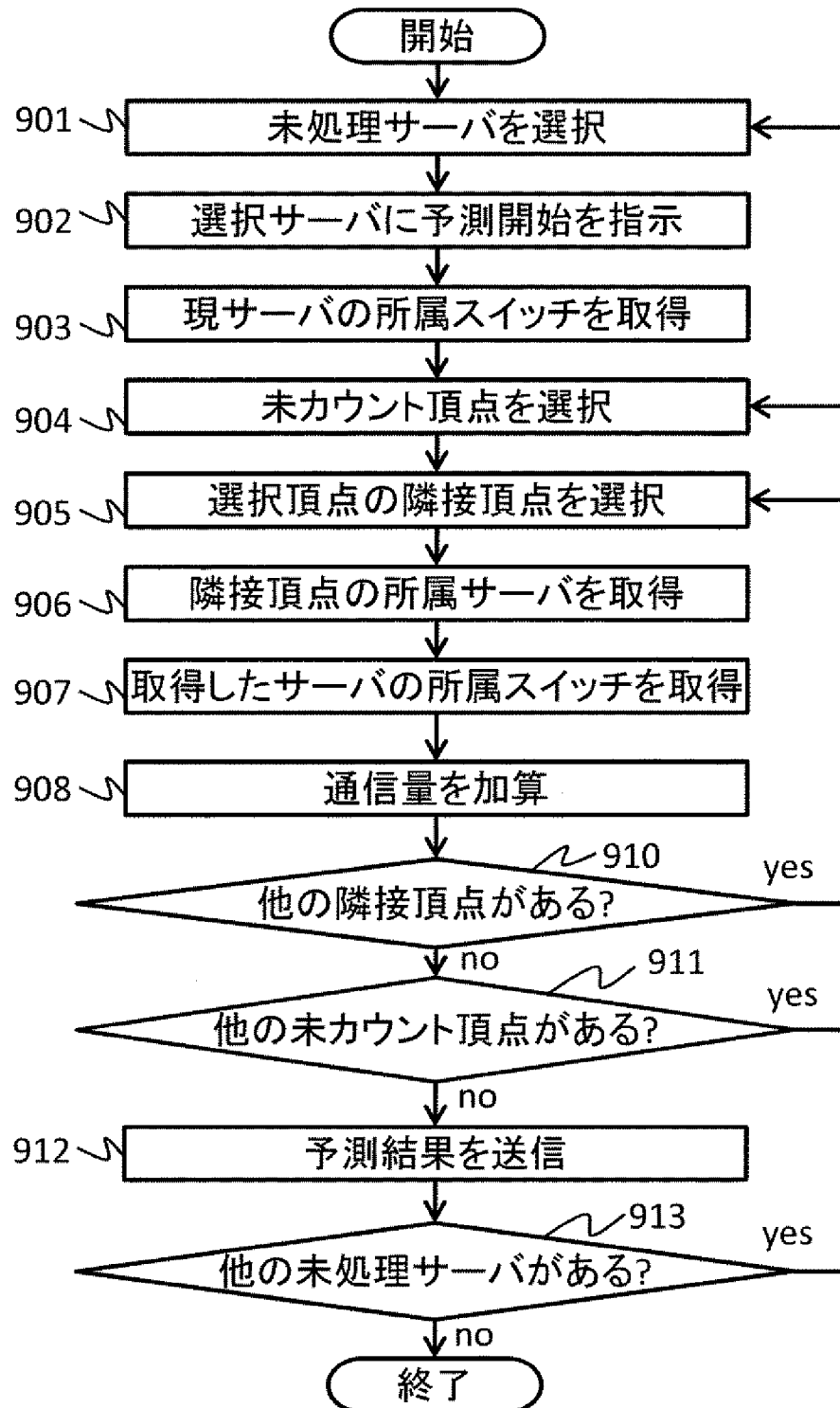
[図8]

図8

サーバ番号	SW番号
1	1
2	1
3	2
4	2
5	3
6	3
7	4
8	4

[図9]

図9



[図10]

図10

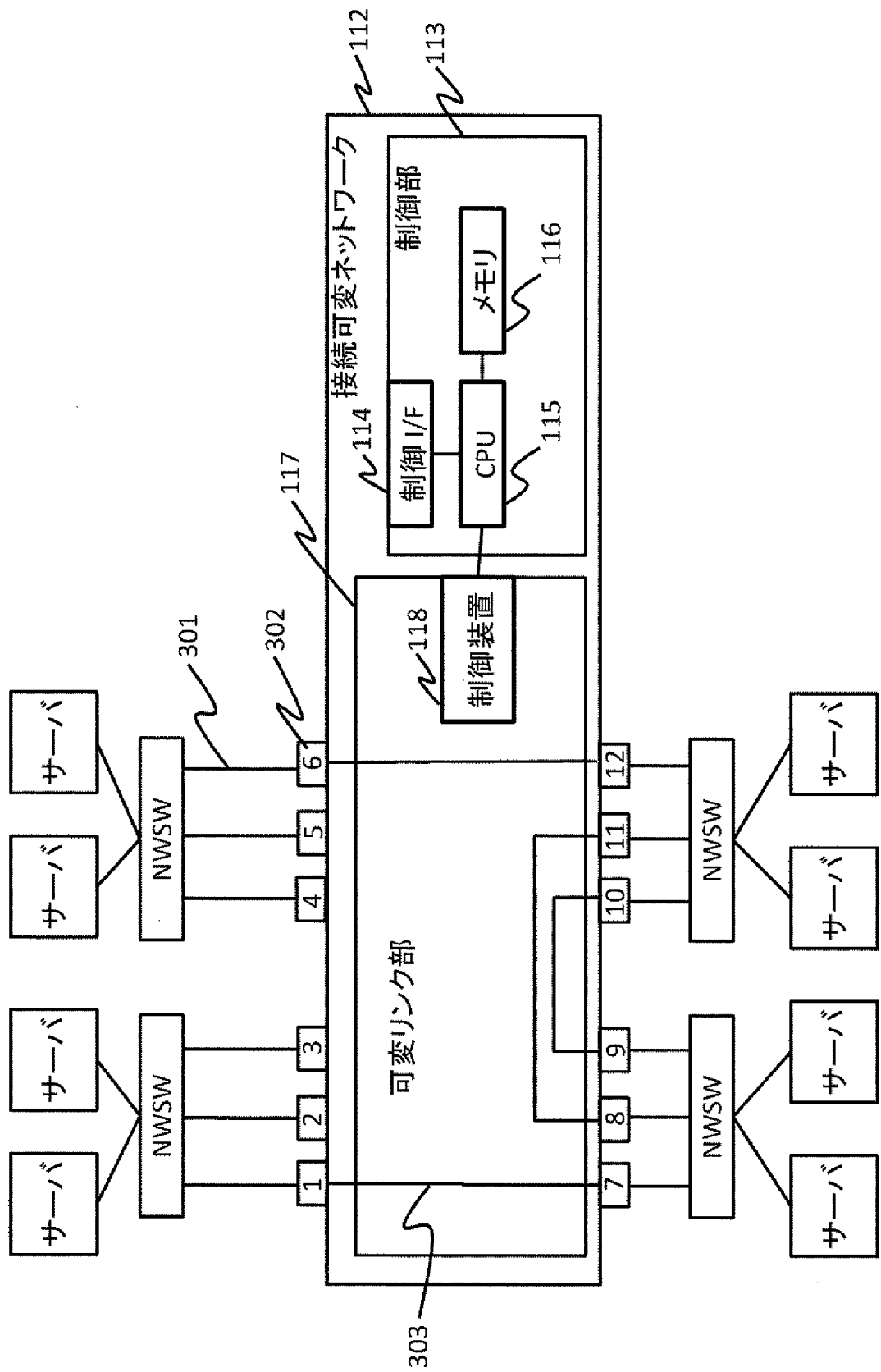
		SW番号 →			
SW番号 ↓		1	2	3	4
	1	-	5	0	2
	2	5	-	10	0
	3	0	10	-	5
	4	2	0	5	-

[図11]

図11

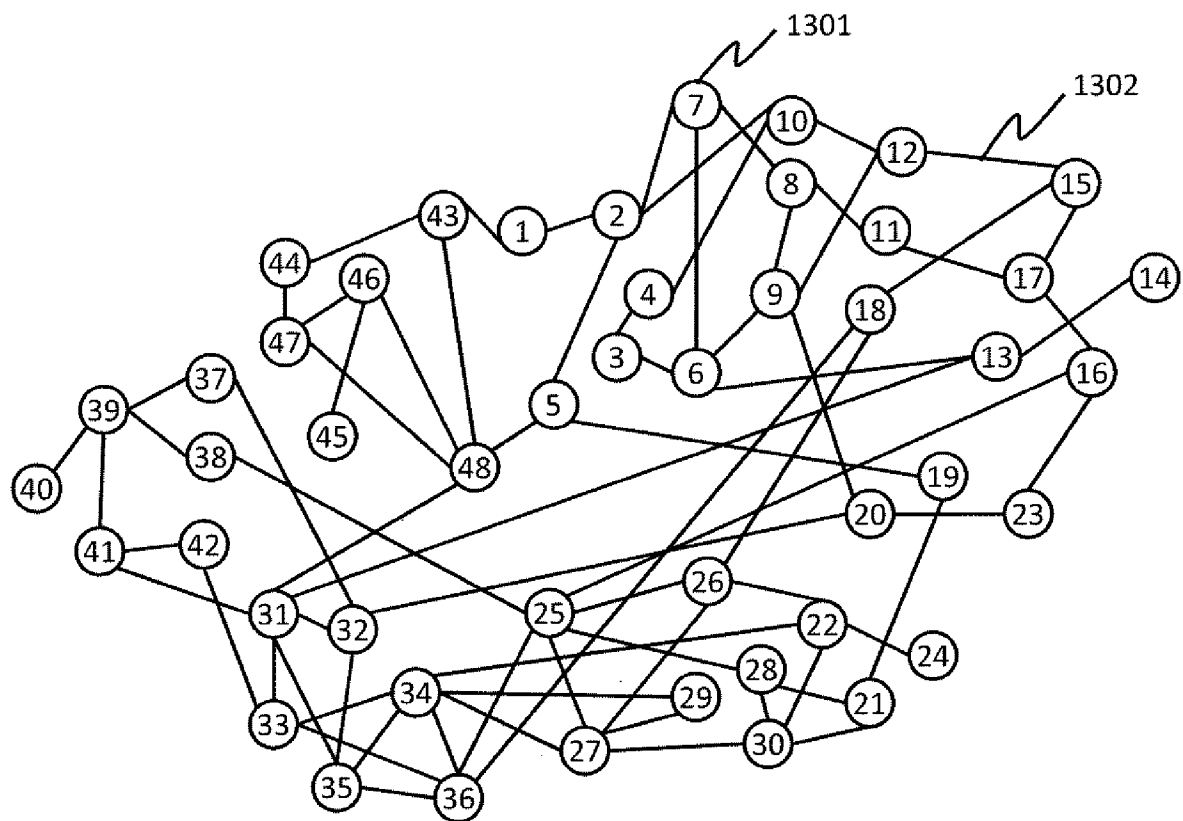
ネットワークスイッチのペア		通信量	割合	リンク本数
1	2	5	0.23	1
1	3	0	0	0
1	4	2	0.09	0
2	3	10	0.45	2
2	4	0	0	0
3	4	5	0.22	1

[図12]



[図13]

図13



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2012/084019

A. CLASSIFICATION OF SUBJECT MATTER

H04L12/701(2013.01) i, G06F15/173(2006.01) i

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)

H04L12/701, G06F15/173

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

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2013
Kokai Jitsuyo Shinan Koho	1971-2013	Toroku Jitsuyo Shinan Koho	1994-2013

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 2012-164259 A (Fujitsu Ltd.), 30 August 2012 (30.08.2012), paragraphs [0014] to [0023]; fig. 1 (Family: none)	1-7
A	JP 2004-171234 A (Toshiba Corp.), 17 June 2004 (17.06.2004), paragraphs [0042] to [0044] & US 2004/0098718 A1 & CN 1503150 A	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

“T” 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
22 February, 2013 (22.02.13)

Date of mailing of the international search report
12 March, 2013 (12.03.13)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

A. 発明の属する分野の分類（国際特許分類（I P C））
Int.Cl. H04L12/701 (2013.01) i, G06F15/173 (2006.01) i

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（I P C））

Int.Cl. H04L12/701, G06F15/173

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

日本国実用新案公報	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 3 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 3 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 3 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 2012-164259 A（富士通株式会社）2012.08.30, [0014]-[0023]段落、図1（ファミリーなし）	1 - 7
A	JP 2004-171234 A（株式会社東芝）2004.06.17, [0042]-[0044]段落 & US 2004/0098718 A1 & CN 1503150 A	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。

* 引用文献のカテゴリー	の日の後に公表された文献
「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	「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矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「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	「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
「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）	「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
「O」口頭による開示、使用、展示等に言及する文献	「&」同一パテントファミリー文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願	

国際調査を完了した日

2 2 . 0 2 . 2 0 1 3

国際調査報告の発送日

1 2 . 0 3 . 2 0 1 3

国際調査機関の名称及びあて先
日本国特許庁（I S A / J P）
郵便番号 1 0 0 - 8 9 1 5
東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

永井 啓司

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 9 6

5 X

3 6 5 6