

(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  
2013年7月18日(18.07.2013)



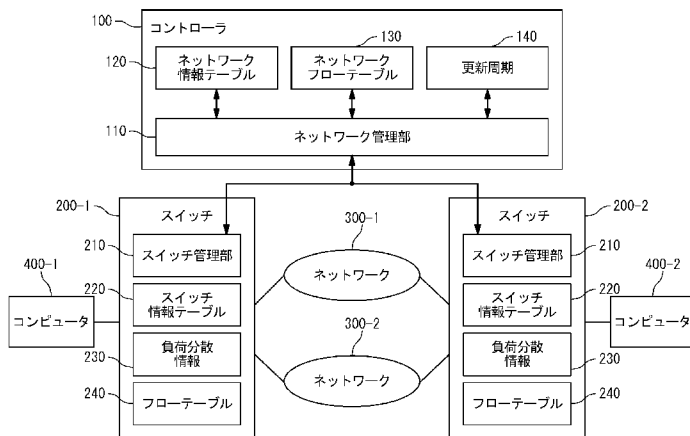
(10) 国際公開番号  
WO 2013/105551 A1

- (51) 国際特許分類:  
H04L 12/721 (2013.01) H04L 12/801 (2013.01)  
H04L 12/70 (2013.01) H04L 12/911 (2013.01)
- (21) 国際出願番号: PCT/JP2013/050118
- (22) 国際出願日: 2013年1月8日(08.01.2013)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:  
特願 2012-003551 2012年1月11日(11.01.2012) JP
- (71) 出願人(米国を除く全ての指定国について): 日本電気株式会社(NEC CORPORATION) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 Tokyo (JP).
- (72) 発明者; および  
(71) 出願人(米国についてのみ): 蒲倉 正憲 (KABAKURA Masanori) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP).
- (74) 代理人: 工藤 実(KUDOH Minoru); 〒1400013 東京都品川区南大井六丁目24番10号カドヤビル6階 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: COMPUTER SYSTEM, CONTROLLER, SWITCH, COMMUNICATION METHOD, AND RECORDING MEDIUM IN WHICH NETWORK MANAGEMENT PROGRAM HAS BEEN STORED

(54) 発明の名称: コンピュータシステム、コントローラ、スイッチ、通信方法、及びネットワーク管理プログラムが格納された記録媒体



- 100 Controller
- 110 Network management unit
- 120 Network information table
- 130 Network flow table
- 140 Update period
- 200-1, 200-2 Switch
- 210 Switch management unit
- 220 Switch information table
- 230 Load distribution information
- 240 Flow table
- 300-1, 300-2 Network
- 400-1, 400-2 Computer

(57) Abstract: This computer system is provided with: a switch which transfers testing-use data to another switch via each of a plurality of networks in order to acquire transfer performance for each of the plurality of networks; and a controller which sets, in the flow table for the switch, a flow entry in which, among the plurality of networks, a network for which the transfer performance is greater than or equal to a predetermined threshold has been defined as a transfer path to the other switch. As a result, load distribution of a computer system is achieved using OpenFlow technology.

(57) 要約: 本発明によるコンピュータシステムは、複数のネットワークのそれぞれを介して他のスイッチに検査用データを転送して、複数のネットワークのそれぞれの転送性能を取得するスイッチと、複数のネットワークのうち、転送性能が所定の閾値以上のネットワークを他のスイッチへの転送経路として規定したフローエントリを、スイッチのフローテーブルに設定するコントローラとを具備する。これにより、オープンフロー技術を利用してコンピュータシステムの負荷分散が実現される。

WO 2013/105551 A1

## 明 細 書

### 発明の名称：

コンピュータシステム、コントローラ、スイッチ、通信方法、及びネットワーク管理プログラムが格納された記録媒体

### 技術分野

[0001] 本発明は、コンピュータシステム、及び通信方法に関し、特に、オープンフロー（オープンフロー）技術を利用したコンピュータシステム、及びその通信方法に関する。

### 背景技術

[0002] 複数のネットワークによって冗長化されたコンピュータシステムは、障害が発生したときに代替ネットワークによって運用することができるため、高い可用性を有する。又、平常運用時についても複数のネットワークから通信に使用するものを選択することによって、負荷分散を実現し通信の高速化が可能になる。

[0003] システム内で効率よく通信をするためには、できるだけ性能の高いネットワークを優先的に使用する必要がある。例えば、通信回線の性能情報に基づいて使用経路を決定することで通信効率を高めることができる。

[0004] 性能情報に基づいて使用経路を決定するシステムが、例えば特開2007-189615に記載されている（特許文献1参照）。特許文献1に記載のシステムは、ネットワーク負荷に基づいて経路を算出することにより、効率の良い通信路を設定する。

[0005] 特許文献1に記載のネットワーク監視支援装置は、ネットワーク機器を監視して、障害を検出するとともに障害の影響範囲を解析し、当該影響範囲を回避する通信経路を算出する。この際、ネットワーク監視支援装置は、トラフィックの利用状況及びCPU利用率等のネットワークの負荷に基づいて経路を算出する。

[0006] 特許文献1に記載のシステムでは、障害による影響範囲を回避しつつ、ト

ラフィックの利用状況に基づいて使用経路を設定している。このため、障害を回避しながらトラフィック利用率の低い経路を選択することができる。しかしながら、運用されていない経路は、トラフィックが発生していないため、使用経路として選択され得るが、選択された経路が必ずしも転送性能の良い経路とは限らない。このため、特許文献1に記載のシステムでは、ネットワーク全体に障害が発生したときの即時対応については考慮されていない。

[0007] 又、計算機システムの一例が、特開2000-348005（特許文献2参照）や特開2007-287064（特許文献3参照）に記載されている。特許文献2や特許文献3には、複数経路を持つストレージシステムについて、負荷状況や障害発生状況に応じて、動的に使用パスを切り替え、可用性の向上や、性能の向上を実現するシステムが記載されている。

### 先行技術文献

#### 特許文献

[0008] 特許文献1：特開2007-189615

特許文献2：特開2000-348005

特許文献3：特開2007-287064

#### 非特許文献

[0009] 非特許文献1：OpenFlow Switch Specification Version 1.0.0 (Wire Protocol 0x01) December 31, 2009

### 発明の概要

[0010] しかし、特許文献2、3の記載のシステムでは、使用する経路が予め設定されているため、平常状態で優先的に使用する経路を自動的に決定することができない。又、動的な構成変更が発生した場合の動作を想定していないため、パス障害が発生したときや、動的な構成変更が発生したときに、使用経路を決定することができない。更に、特許文献2、3の計算機システムはストレージシステムを対象としているため、FCoE (Fibre Chan

nel over Ethernet（登録商標）など、同一ネットワーク上にI/Oが混在するような場合について考慮されていない。

[0011] 以上のことから、本発明の目的は、複数のネットワーク間において、より高性能なネットワークへの通信が可能なネットワークシステムを提供することにある。

[0012] 本発明によるコンピュータシステムは、コントローラと、複数のネットワークを介して他のスイッチに接続され、前記コントローラによって設定されたフローエントリに従って、受信パケットの中継処理を行うスイッチとを具備する。スイッチは、複数のネットワークのそれぞれを介して他のスイッチに検査用データを転送して、複数のネットワークのそれぞれの転送性能を取得する。又、スイッチは、転送性能を複数のネットワークのそれぞれに対応付けてコントローラに通知する。コントローラは、複数のネットワークのうち、転送性能が所定の閾値以上のネットワークを他のスイッチへの転送経路として規定したフローエントリを、スイッチのフローテーブルに設定する。

[0013] 本発明による通信方法は、複数のネットワークを介して他のスイッチに接続され、コントローラによって設定されたフローエントリに従って、受信パケットの中継処理を行うスイッチを具備するコンピュータシステムにおいて実行される。本発明による通信方法は、スイッチが、複数のネットワークのそれぞれを介して他のスイッチに検査用データを転送して、複数のネットワークのそれぞれの転送性能を取得するステップと、転送性能を複数のネットワークのそれぞれに対応付けてコントローラに通知するステップと、コントローラが、複数のネットワークのうち、転送性能が所定の閾値以上のネットワークを他のスイッチへの転送経路として規定したフローエントリを、スイッチのフローテーブルに設定するステップとを具備する。

[0014] 本発明によれば、複数のネットワーク間において、より高性能なネットワークへの通信することが可能となる。

### 図面の簡単な説明

[0015] 上記発明の目的、効果、特徴は、添付される図面と連携して実施の形態の

記述から、より明らかになる。

[図1]図1は、本発明によるコンピュータシステムの実施の形態における構成の一例を示す図である。

[図2]図2は、本発明に係るネットワーク情報テーブルの構造を示す図である。

[図3]図3は、本発明に係るネットワークフローテーブルの構造を示す図である。

[図4]図4は、本発明に係るスイッチ情報テーブルの構造を示す図である。

[図5]図5は、本発明に係る負荷分散情報の構成を示す図である。

[図6]図6は、本発明に係るフローテーブルの構造を示す図である。

[図7]図7は、本発明によるコンピュータシステムの実施の形態における構成の他の一例を示す図である。

[図8]図8は、本発明に係るスイッチにおける接続変更時の初期設定動作を示すフローチャートである。

[図9]図9は、本発明に係るコントローラにおける接続変更時の設定動作を示すフローチャートである。

[図10]図10は、本発明に係るネットワーク情報テーブルの一例を示す図である。

[図11]図11は、本発明に係るスイッチ情報テーブルの一例（設定前）を示す図である。

[図12]図12は、本発明に係るスイッチ情報テーブル設定（更新）動作を示すフローチャートである。

[図13]図13は、本発明に係るスイッチ情報テーブルの他の一例（設定後）を示す図である。

[図14]図14は、本発明に係るネットワーク情報テーブルの設定（更新）動作を示すフローチャートである。

[図15]図15は、本発明に係るネットワーク情報テーブルの一例を示す図である。

[図16]図16は、本発明に係るネットワークフローテーブル及びフローテーブルの設定（更新）動作を示すフローチャートである。

[図17]図17は、本発明に係る負荷分散情報の一例を示す図である。

[図18]図18は、本発明に係る送信元スイッチビュー又はスイッチ間経路ビューの一例を示す図である。

[図19]図19は、本発明に係るネットワーク情報テーブルの更に他の一例を示す図である。

[図20]図20は、本発明に係るネットワークフローテーブルの一例を示す図である。

[図21]図21は、本発明に係るフローテーブルの一例を示す図である。

[図22]図22は、本発明によるコンピュータシステムの実施の形態における構成の更に他の一例を示す図である。

### 発明を実施するための形態

#### [0016] (概要)

本発明によるコントローラは、スイッチにおいて測定されたネットワークの転送性能（例えば転送速度）を取得し、取得した転送性能と負荷分散ポリシーに基づいて、通信に利用するネットワークを自動選択する。通信に利用するネットワークを指定したフローエントリは、オープンフロープロトコルに従ってスイッチに設定される。これにより複数の通信ネットワークから転送性能の良好なネットワークを、通信経路として自動選択することが可能となる。又、コントローラは、定期的にネットワークの転送性能の取得及びスイッチへの最適経路（最適フロー）の設定を行う。これにより、構成が動的に変更されるネットワークにおいても、最適な経路を利用した通信に自動変更され得る。

[0017] 以下、添付図面を参照しながら本発明の実施の形態を説明する。図面において同一、又は類似の参照符号は、同一、類似、又は等価な構成要素を示す。

#### [0018] (コンピュータシステムの構成)

図1を参照して、本発明によるコンピュータシステムの構成を説明する。図1は、本発明によるコンピュータシステムの実施の形態における構成を示す図である。本発明によるコンピュータシステムは、オープンフローを利用して通信経路の構築及びパケットデータの転送制御を行う。図1を参照して、本発明によるコンピュータシステムは、オープンフローコントローラ100（以下、コントローラ100と称す）、複数のオープンフロースイッチ200-1、200-2（以下、スイッチ200-1、200-2と称す）、ネットワーク300-1、300-2、複数のクライアント端末400-1、400-2（以下、コンピュータ400-1、400-2と称す）を具備する。又、スイッチ200-1、200-2を区別せずに総称する場合はスイッチ200と称し、ネットワーク300-1、300-2を区別せずに総称する場合はネットワーク300と称し、コンピュータ400-1、400-2を区別せずに総称する場合はコンピュータ400と称して説明する。尚、図1に示すスイッチ200、ネットワーク300、コンピュータ400は、それぞれ2つ設けられているが、これに限らず、任意の数が設けられても良い。

[0019] コンピュータ400は、図示しないCPU、ネットワークインタフェース（I/F）、及びメモリ（記録媒体）を備えるコンピュータ装置であり、メモリ（記録媒体）内のプログラムを実行することで他のコンピュータ400との間で通信を行う。

[0020] 本発明によるコンピュータシステムでは、複数のネットワーク300を経由して接続されたスイッチ200を介してコンピュータ400間の通信が行われる。複数のネットワーク300を経由して接続されたスイッチ200間の通信は、オープンフロー（プログラマブルフローとも称す）プロトコルに従って制御される。コントローラ100とスイッチ200とは制御用のネットワークで接続される。

[0021] 詳細には、コントローラ100は、オープンフロー技術により、システム内におけるパケット転送に係る通信経路パケット転送処理を制御する。オー

プンフロー技術とは、コントローラ（ここではコントローラ100）が、ルーティングポリシー（フローエントリ：フロー+アクション）に従い、マルチレイヤ及びフロー単位の経路情報をスイッチ200に設定し、経路制御やノード制御を行う技術を示す（詳細は、非特許文献1を参照）。これにより、経路制御機能がルータやスイッチから分離され、コントローラによる集中制御によって最適なルーティング、トラフィック管理が可能となる。オープンフロー技術が適用されるスイッチ200は、従来のルータやスイッチのようにパケットやフレームの単位ではなく、END2ENDのフローとして通信を取り扱う。

[0022] 本発明によるコントローラ100は、図示しないCPU、ネットワークインタフェース（I/F）、及びメモリ（記録媒体）を備えるコンピュータ装置である。コントローラ100の図示しないメモリ（記録媒体）には、ネットワーク情報テーブル120、ネットワークフローテーブル130、更新周期140、ネットワーク管理プログラムが記録される。コントローラ100では、図示しないCPUによってネットワーク管理プログラムが実行されることで、ネットワーク管理部110の機能が実現される。

[0023] ネットワーク管理部110は、スイッチ200から取得するネットワーク300の接続状況や負荷状況に基づいて、通信経路となるネットワーク300を決定する。次に、ネットワーク管理部110は、スイッチ200が保持するフローテーブル240にフローエントリ（ルール241+アクション242）を設定又は更新することで、当該スイッチ200の動作（例えばパケットデータの中継動作や破棄）を制御する。

[0024] 詳細には、ネットワーク管理部110は、スイッチ200からのファーストパケットの受信通知や負荷情報に基づいて制御対象を決めるルール、動作を規定するアクション、及び通信経路を設定、削除又は更新する。ここでファーストパケットとは、スイッチ200のフローテーブル240に設定されたルール241に適合しないパケットデータを示す。

[0025] ルール241には、例えば、TCP/IPのパケットデータにおけるヘッ



ダ情報に含まれる、OSI (Open Systems Interconnection) 参照モデルのレイヤ1からレイヤ4のアドレスや識別子の組み合わせが規定される。例えば、図9に示すレイヤ1の物理ポート、レイヤ2のMACアドレス、レイヤ3のIPアドレス、レイヤ4のポート番号、VLANタグ (VLAN id) のそれぞれの組み合わせがルール241として設定される。尚、VLANタグには、優先順位 (VLAN Priority) が付与されていても良い。

[0026] ここで、ルール241に設定されるポート番号等の識別子やアドレス等は、所定の範囲で設定されても構わない。又、宛先や送信元のアドレス等を区別してルール241として設定されることが好ましい。例えば、MAC宛先アドレスの範囲や、接続先のアプリケーションを特定する宛先ポート番号の範囲、接続元のアプリケーションを特定する送信元ポート番号の範囲がルール241として設定される。更に、データ転送プロトコルを特定する識別子をルール241として設定してもよい。

[0027] アクション242には、例えばTCP/IPのパケットデータを処理する方法が規定される。例えば、受信パケットデータを中継するか否かを示す情報や、中継する場合はその送信先となるスイッチ200やネットワーク300が設定される。又、アクション242には、パケットデータの複製や、破棄することを指示する情報が設定されてもよい。

[0028] 図2は、本発明に係るネットワーク情報テーブル120の構造を示す図である。図2を参照して、ネットワーク情報テーブル120には、送信元スイッチ識別子121、送信先スイッチ識別子122、ネットワーク識別子123、及び転送速度124が対応付けられて1つのエントリとして格納される。本発明に係るスイッチ200には一意のスイッチ識別子が付与され、ネットワーク300には一意のネットワーク識別子が付与される。送信元スイッチ識別子121は、対応する転送速度124を測定するためのパケットの転送元となるスイッチ200の識別子であり、送信先スイッチ識別子122は、当該パケットの送信先のスイッチ200の識別子である。ネットワーク識

別子123は、対応する送信元スイッチ識別子121及び送信先スイッチ識別子122を持つスイッチ200間に設けられたネットワーク300の識別子である。転送速度124は、対応するネットワーク識別子123をもつネットワーク300を介したスイッチ200間の転送速度である。例えば、所定の間隔で処理されたスイッチ200間通信の平均処理速度が転送速度124としてネットワーク情報テーブル120に記録される。

[0029] 図3は、本発明に係るネットワークフローテーブル130の構造を示す図である。ネットワークフローテーブル130には、スイッチ200がどのネットワークを通して通信するかを示す情報が記録される。詳細には、ネットワークフローテーブル130には、送信元スイッチ識別子131をキーとしたエントリとして、送信先スイッチ識別子132と、ネットワーク識別子リスト133が対応付けられて格納される。送信元スイッチ識別子131は、パケットの送信元のスイッチ200の識別子であり、送信先スイッチ識別子132は、当該パケットの送信先のスイッチ200の識別子である。ネットワーク識別子リスト133には、対応する送信元スイッチ識別子131及び送信先スイッチ識別子132のスイッチ200間に設けられたネットワーク300のネットワーク識別子が記録される。ネットワーク識別子リスト133には、任意の個数のネットワーク識別子が格納される。

[0030] 更新周期140は、ネットワーク管理部110がスイッチ200から転送性能を取得する周期を含む。更新周期140は、スイッチ200毎に異なる周期であっても良いし、全てのスイッチ200の共通の周期であっても良い。

[0031] スイッチ200は、ネットワーク300を介した通信を制御する。詳細には、スイッチ200は、コンピュータ400やネットワーク300を介して送信されたデータの転送先を決定する。この際、スイッチ200は、オープンフロープロトコルに従い、コントローラ100によって設定（更新）されたフローテーブル240に従って受信パケットの処理方法（アクション）を決定する。

[0032] 本発明によるスイッチ200は、図示しないCPU、ネットワークインタフェース（I/F）、及びメモリ（記録媒体）を備える。スイッチ200の図示しないメモリ（記録媒体）には、スイッチ管理プログラム、スイッチ情報テーブル220、負荷分散情報230、フローテーブル240が格納される。スイッチ200では、図示しないCPUによってスイッチ管理プログラムが実行されることで、スイッチ管理部210の機能が実現される。

[0033] スイッチ管理部210は、自身のフローテーブル240に基づいてデータの転送先となるネットワーク300を決定し、当該ネットワーク300を経由して通信を実施する。本発明によるシステムでは、リンクアグリゲーション等の既存技術を用いて、複数のネットワーク300を経由して通信を実施することができるものとする。又、スイッチ管理部210は、自身が接続される（使用する）ネットワーク300の転送性能を取得するために定期的に検査用データを送信する。更に、スイッチ管理部210は、コントローラ100からのフローテーブル設定（更新）命令や、転送性能の取得命令に応答してフローテーブル240の設定（更新）や転送性能の取得及びコントローラへの通知を行う。

[0034] 図4は、本発明に係るスイッチ情報テーブル220の構造を示す図である。図4を参照して、スイッチ情報テーブル220には、送信先スイッチ識別子221、ネットワーク識別子222、及び転送性能（転送容量223、転送時間224）が対応付けられてエントリとして格納される。送信先スイッチ識別子221は、パケットの転送先となるスイッチ200の識別子である。ネットワーク識別子222は、当該転送に利用したネットワーク300の識別子である。転送容量223は、当該転送において、所定の時間内に転送されたデータサイズの累計を示す。転送時間224は、当該転送に要した時間の合計を示す。例えば、転送容量223として転送時間224内に転送されたデータサイズが当該転送時間224に対応付けられてスイッチ情報テーブル220に記録される。

[0035] 図5は、本発明に係る負荷分散情報230の構成を示す図である。図5を

参照して、負荷分散情報 230 は、検査周期 231、最大ネットワーク数 232、及び転送閾値 233 を含む。検査周期 231 は、スイッチ 200 が転送性能（例えば転送容量 223 や転送時間 224）を取得するための通信を実施する周期を示す情報である。最大ネットワーク数 232 は、スイッチ 200 が接続（利用）可能なネットワークの最大数を示す情報である。転送閾値 233 は、使用可能なネットワークのうち、最大性能のネットワーク 300 における利用可能なパスの割合を示す情報である。負荷分散情報 230 は、予め設定されることが好ましく、自ら又はコントローラ 100 によって設定又は更新されても良い。

[0036] 図 6 は、本発明に係るフローテーブル 240 の構造を示す図である。フローテーブル 240 には、ルール 241 及びアクション 242 がフローエントリとして設定される。アクション 242 にはデータの転送先となるスイッチの識別子（送信先スイッチ識別子 243）や、転送に使用するネットワーク 300 の識別子のリスト（ネットワーク識別子リスト 244）が含まれる。ネットワーク識別子リスト 244 には、任意の個数のネットワーク識別子が登録される。

[0037] 受信パケットのヘッダ情報が、フローテーブル 240 に記録されたルール 241 に適合（一致）する場合、当該パケットデータは、スイッチ管理部 210 によって、アクション 242 に規定されたネットワーク 300 を介してアクション 242 に規定された他のスイッチ 200 に転送される。

[0038] 図 1 を参照して、ネットワーク 300-1、300-2 はそれぞれ独立したネットワークであり、スイッチ 200 間で利用（接続）されるネットワークの数はこれに限らない。例えば、図 7 に示すように、スイッチ 200-1 とスイッチ 200-2 の間に 4 つのネットワーク 300-1~300-4 が設けられても良い。

[0039] （動作）

次に、図 7 に示されるコンピュータシステムを参照して、本発明によるコンピュータシステムの動作を説明する。以下では、一例としてスイッチ 20

0-1、200-2のスイッチ識別子を“スイッチx”、“スイッチy”とし、ネットワーク300-1~300-4のネットワーク識別子を“ネットワーク1”~“ネットワーク4”として説明する。

[0040] 先ず、図8を参照して、スイッチ200における初期設定動作を説明する。図8は、本発明に係るスイッチ200における接続変更時の初期設定動作を示すフローチャートである。ここでは当初、スイッチ200におけるスイッチ情報テーブル220及びフローテーブル240には何も設定されていない状態であり、負荷分散情報230には、運用方針に応じた値が予め設定されているものとする。

[0041] 新たにシステムに追加されたスイッチ200のスイッチ管理部210は、自身のスイッチ識別子、及び自身に接続されたネットワークのネットワーク識別子をコントローラ100のネットワーク管理部110に通知する（ステップS101）。ここで通知されるネットワーク識別子は、例えば、ネットワーク300に接続されたポート番号と、当該ネットワーク300を介して接続される他のスイッチ200のポート番号との組合せを特定する情報である。

[0042] スイッチ200は、通知したスイッチ識別子及びネットワーク識別子に応じてコントローラ100から送信されるネットワーク情報を取得する（ステップS102）。ネットワーク情報は、通知したスイッチ識別子及びネットワーク識別子に基づいて更新登録されたネットワーク情報テーブル120内の情報のうち、転送速度124を除く情報を示す。ここで、スイッチ200は、ネットワーク情報テーブル120のうち、新たに登録されたネットワーク情報のみを取得しても良いし、ネットワーク情報テーブル120に登録された全てのネットワーク情報を取得してもよい。

[0043] スイッチ200は、取得したネットワーク情報に基づいてスイッチ情報テーブル220を更新する（ステップS103）。詳細には、スイッチ200は、取得したネットワーク情報のうち、自身のスイッチ識別子に一致する送信元スイッチ識別子121を含むエントリを、スイッチ情報テーブル220

に登録する。このとき、スイッチ200は、スイッチ情報テーブル220に未登録の送信先スイッチ識別子及びネットワーク識別子を含むエントリを新規に登録する。ここで、コントローラ100から送信されたネットワーク情報に含まれる送信先スイッチ識別子122及びネットワーク識別子222は、送信先スイッチ識別子221及びネットワーク識別子222としてスイッチ情報テーブル220に新規登録される。又、スイッチ情報テーブル220に新規登録されるエントリの転送容量223及び転送時間224は“0”に設定される。

[0044] 次に、図9を参照して、コントローラ100における初期設定動作を説明する。図9は、本発明に係るコントローラ100における接続変更時の初期設定動作を示すフローチャートである。ここでは当初、コントローラ100におけるネットワーク情報テーブル120、及び、ネットワークフローテーブル130には何も設定されていない状態であり、更新周期140には、運用方針に応じた値が予め設定されているものとする。

[0045] コントローラ100のネットワーク管理部110は、システムに新規接続されたスイッチ200から通知されるスイッチ識別子とネットワーク識別子をネットワーク情報テーブル120に新規登録する（ステップS201、S202）。詳細には、ネットワーク管理部110は、ネットワーク情報テーブル120に既に登録済みのスイッチ識別子を送信元スイッチ識別子121とし、スイッチ200から通知された未登録のスイッチ識別子を送信先スイッチ識別子122として対応付けてネットワーク情報テーブル120に新規登録する。この際、ネットワーク管理部110は、既に登録済みのネットワーク識別子や、スイッチ200から通知された未登録のネットワーク識別子を参照して、送信元スイッチ識別子121と送信先スイッチ識別子122に対応するスイッチ200の間に接続されたネットワークを特定する。そして、ネットワーク管理部110は、特定したネットワークの識別子をネットワーク識別子123として当該送信元スイッチ識別子121及び送信先スイッチ識別子122に対応付けてネットワーク情報テーブル120に新規登録す

る。尚、新規登録されるネットワーク情報テーブルのエントリ内の転送速度 124 は“0”に設定される。

[0046] ネットワーク管理部 110 は、システムに新規追加されたスイッチ 200 の識別子を送信元スイッチ識別子 121 とし、登録済みのスイッチ識別子を送信先スイッチ識別子 122 とし、当該スイッチ間に接続されるネットワークの識別子をネットワーク識別子 123、転送速度 124 を“0”とするエントリを新規登録する（ステップ S203）。すなわち、ネットワーク管理部 110 は、ステップ A202 で新規登録されたエントリの送信元スイッチ識別子 121 と送信先スイッチ識別子 122 を入れ替えたエントリをネットワーク情報テーブル 120 に新規登録する。

[0047] 続いてネットワーク管理部 110 は、ステップ S202、S203 において更新登録が完了したネットワーク情報テーブル 120 内のネットワーク情報をスイッチ 200 に送信する（ステップ S204）。ネットワーク管理部 110 は、ネットワーク情報テーブル 120 において新規登録したエントリのうち、スイッチ 200 及びネットワーク 300 の接続関係に係るネットワーク情報（ここでは、送信元スイッチ識別子 121、送信先スイッチ識別子 122、ネットワーク識別子 123）を、送信元スイッチ識別子 121 に対応するスイッチ 200 に送信する。尚、ネットワーク管理部 110 は、ネットワーク情報テーブル 120 に既に登録されている全てのエントリにおけるネットワーク情報を、対応するスイッチ 200 又は全てのスイッチ 200 に送信しても良い。

[0048] 一例として、図 7 に示すシステムにおいて、スイッチ 200-1 が新規追加された場合のネットワーク情報テーブル 120 の更新登録、及びスイッチ情報テーブル 220 の更新登録について説明する。ここでは、スイッチ 200-2 が既にシステムに接続されており、そのスイッチ識別子“スイッチ y”はネットワーク情報テーブル 120 に登録されているものとする。

[0049] スイッチ 200-1 は、自身のスイッチ識別子“スイッチ x”、自身に接続するネットワーク 300 のネットワーク識別子“ネットワーク 1”～“ネ

ットワーク4”をコントローラ100に通知する。コントローラ100は、登録済みのスイッチ識別子“スイッチy”を送信元スイッチ識別子121、通知されたスイッチ識別子“スイッチx”を送信先スイッチ識別子122とし、ネットワーク識別子“ネットワーク1”～“ネットワーク4”をネットワーク識別子123とし、転送速度を“0”とする4つのエントリを登録する。又、コントローラ100は、通知されたスイッチ識別子“スイッチx”を送信元スイッチ識別子121、登録済みのスイッチ識別子“スイッチy”を送信先スイッチ識別子122とし、ネットワーク識別子“ネットワーク1”～“ネットワーク4”をネットワーク識別子123とし、転送速度を“0”とする4つのエントリを登録する。これにより、図10のようにネットワーク情報テーブル120が更新登録される。

[0050] コントローラ100は、更新登録されたネットワーク情報テーブル120からネットワーク情報を抽出し、スイッチ200-1、200-2に送信する。例えば、コントローラ100は、ネットワーク情報テーブル120のうち、送信元スイッチ識別子121“スイッチx”に対応するエントリ内のネットワーク情報を、当該スイッチ識別子に対応するスイッチ200-1に送信する。又、コントローラ100は、送信元スイッチ識別子121“スイッチy”に対応するエントリ内のネットワーク情報を、当該スイッチ識別子に対応するスイッチ200-2に送信する。

[0051] 図11を参照して、スイッチ200-1は、コントローラ100から取得したネットワーク情報に基づき、送信先スイッチ識別子221を“スイッチy”、ネットワーク識別子222を“ネットワーク1”～“ネットワーク4”とし、転送容量223及び転送時間224を“0”とした4つのエントリを、スイッチ情報テーブル220に新規登録する。スイッチ200-2も同様に、送信先スイッチ識別子221を“スイッチx”としてエントリを自身のスイッチ情報テーブル220に新規登録する。

[0052] 上記一例では、ネットワーク情報テーブル120に新規登録されたネットワーク情報が、対応するスイッチ200に送信される場合について説明した



が、これに限らず、ネットワーク情報テーブル120に登録済みの全てのネットワーク情報が、登録済みの全てのスイッチ200に送信されても良い。この場合、スイッチ200では、送信元スイッチ識別子が自身のスイッチ識別子と一致するネットワーク情報のうち、送信先スイッチ識別子がスイッチ情報テーブル220に未登録のネットワーク情報のみがスイッチ情報テーブル220に新規登録される。

[0053] コントローラ100からスイッチ200への通信ができなくなったときの動作を説明する。コントローラ100のネットワーク管理部110は、定期的又は任意の時期にシステム内の全てのスイッチ200又は任意のスイッチ200に対し、接続確認のための信号を送信することが好ましい。この場合、ネットワーク管理部110は、当該信号に対するスイッチ200からの応答信号によって、スイッチ200がシステムに接続されているか否かを確認できる。あるいは、ネットワーク管理部110は、スイッチ200から転送性能が、所定の時間通知されない場合、当該スイッチ200との通信が不通である（又はシステムから離脱した）と判定してもよい。

[0054] ネットワーク管理部110は、スイッチ200への通信ができなくなった場合（又はシステムから離脱したと判定した場合）、ネットワーク情報テーブル120の全エントリを検索し、送信元スイッチ識別子121、または、送信先スイッチ識別子122が当該スイッチ200のスイッチ識別子と一致するエントリをネットワーク情報テーブル120から削除する。

[0055] 本発明によるコンピュータシステムでは、スイッチ200においてネットワーク300を介したスイッチ200間の転送性能（転送容量や転送時間）が測定され、コントローラ100に送信される。コントローラ100は、ネットワーク300の転送性能を参照して、通信経路として利用するネットワーク300を決定し、オープンフロープロトコルに従い、スイッチ200のフローテーブル240を設定（更新）する。

[0056] 先ず、図12を参照して、スイッチ200における転送性能の測定動作を説明する。図12は、本発明に係るスイッチ情報テーブル220における転

送容量 223 及び転送時間 224 の設定（更新）動作を示すフローチャートである。

- [0057] スイッチ管理部 210 は、他のスイッチ 200 への通信要求があったとき、フローテーブル 240 で規定されたネットワーク 300 を用いて通信を開始し、通信開始時刻を記録する（ステップ S301）。例えば、コンピュータ 400 から転送されたパケットデータが、フローテーブル 240 のルール 241 に適合する場合、当該ルールに対応するアクション 242 に従ったネットワーク 300 を利用してパケット転送が行われる。このときスイッチ管理部 210 は、パケット転送開始時刻を図示しないメモリ（記録媒体）に記録する。
- [0058] スイッチ 200 において通信が正常に完了すると、スイッチ管理部 210 は、通信完了時刻と通信開始時刻との差から、転送時間 224 を算出する（ステップ S302 Yes、S303）。一方、通信が正常に完了しなかったときは、転送時間 224 の計算は行わない（ステップ S302 No）。
- [0059] スイッチ管理部 210 は、ネットワーク情報テーブル 120 において、通信に使用したネットワーク 300 と通信先のスイッチ 200 に該当するエントリの転送容量 223 及び転送時間 224 を更新する（ステップ S304）。詳細には、スイッチ管理部 210 は、パケットに適合するアクション 242 から特定される送信先スイッチ識別子 243 及びネットワーク識別子リスト 244 に対応するエントリをネットワーク情報テーブル 120 から抽出し、当該エントリ内の転送容量 223 及び転送時間 224 を更新する。転送容量 223 は、通信開始から通信完了までに転送したパケットデータ量が記録され、転送時間 224 は、ステップ 303 において算出された転送時間が記録される。
- [0060] 例えば、図 7 に示すシステムにおいて、スイッチ 200-1 からネットワーク 300-1～300-4 を介してスイッチ 200-2 宛に通信が行われた場合、図 13 のように、スイッチ情報テーブル 220 が更新される。転送容量 223 及び転送時間 224 には、最新の転送容量及び転送速度が上書き

更新されても良いし、所定期間内の転送容量及び転送速度が累積されて記録されても良い。あるいは、転送容量 2 2 3 及び転送時間 2 2 4 には、転送容量及び転送速度の所定通信回数当たりの平均値や、所定のアルゴリズムによって計算された統計値が記録されても良い。

[0061] 図 1 2 に示す動作で転送性能（ここでは転送容量と転送時間）を更新するためには、何らかの通信が必要である。すなわち、通信に使用しないネットワーク 3 0 0 では情報が更新されない。このため、スイッチ管理部 2 1 0 は、定期的又は任意の時期に検査用の通信を実施し、自身に接続された全てのネットワーク 3 0 0 の転送性能を更新することが好ましい。例えば、スイッチ管理部 2 1 0 は、所定のデータ量の検査用データを、定期的又は任意の時期に、自身に接続されたネットワーク 3 0 0 を介して他のスイッチ 2 0 0 に転送し、図 1 2 と同様な動作により、当該ネットワーク 3 0 0 及び当該スイッチ 2 0 0 に対応する転送容量 2 2 3 及び転送時間 2 2 4 を更新する。定期的に検査用の通信を行う場合、スイッチ管理部 2 1 0 は、負荷分散情報 2 3 0 に設定された検査周期 2 3 1 の経過に応じて検査用パケットデータを送信する。これにより、スイッチ 2 0 0 は、他の装置からパケット転送等の通信要求が発生しない場合でも、ネットワーク 3 0 0 の転送性能を取得することが可能となる。

[0062] 次に、図 1 4 及び図 1 5 を参照して、本発明に係るネットワーク情報テーブル 1 2 0 の設定（更新）動作を説明する。図 1 4 は、本発明に係るネットワーク情報テーブル 1 2 0 の設定（更新）動作を示すフローチャートである。

[0063] ネットワーク管理部 1 1 0 は、更新周期 1 4 0 の経過に応じて、スイッチ 2 0 0 に対しスイッチ情報の取得指示を行う（ステップ S 4 0 1 Y e s、S 4 0 2）。スイッチ管理部 2 1 0 は、スイッチ情報の取得指示を受けたら、ネットワーク管理部 1 1 0 に対し、自身のスイッチ識別子、スイッチ情報テーブル 2 2 0 をスイッチ情報として送信する。

[0064] ネットワーク管理部 1 1 0 は、スイッチ情報の取得指示に対する応答が規

定時間内にない場合、応答のないスイッチ200への通信ができなくなったものとして、当該スイッチ200に対するエントリをネットワーク情報テーブル120から削除する（ステップS403No、S404）。一方、スイッチ情報の取得指示に対する応答として、スイッチ200からスイッチ情報が送信されると、ネットワーク管理部110は、送信されたスイッチ情報によりネットワーク情報テーブル120を更新する（ステップS403Yes、S405）。

[0065] 詳細には、ネットワーク管理部110は、受信したスイッチ識別子に一致する送信元スイッチ識別子121と、受信したスイッチ情報テーブル220の送信先スイッチ識別子221に一致する送信先スイッチ識別子122を含むエントリを、ネットワーク情報テーブル120から特定する。ネットワーク管理部110は、受信したスイッチ情報テーブル220で示されるネットワーク毎の転送容量223及び転送時間224から、ネットワーク毎の転送速度（転送容量／転送時間）を算出し、特定したエントリに記録する。

[0066] 例えば、図7に示すスイッチ200-1から図13に示すスイッチ情報テーブル220を受け取った場合、ネットワーク管理部110は、図15に示すネットワーク情報テーブル120のように、送信元スイッチ識別子121が、スイッチ200-1のスイッチ識別子“スイッチx”に一致し、送信先スイッチ識別子122が送信先スイッチ識別子221“スイッチy”に一致するエントリを更新する。ここでは、ネットワーク300-1～300-4（ネットワーク識別子“ネットワーク1”～“ネットワーク4”）に対応する転送速度124として、図13に示す転送容量223を転送時間224で除した値“1000MB/sec”、“800MB/sec”、“300MB/sec”、“200MB/sec”が設定される。ここで、転送容量223が“0”、すなわち、まったく通信が行われていない場合は、転送速度124は“0”に設定される。

[0067] ステップS403、S404、S405の処理をネットワーク情報テーブル120の全てのエントリについて実施することで、例えば、図7に示すよ

うに全てのエントリの転送速度124が更新される。

[0068] 次に、図16から図21を参照して、ネットワークフローテーブル130及びフローテーブル240の更新動作について説明する。図16は、本発明に係るネットワークフローテーブル130及びフローテーブル240の設定(更新)動作を示すフローチャートである。

[0069] ネットワーク管理部110は、更新周期140の経過に応じて、スイッチ200に対し負荷分散情報230の取得指示を行う(ステップS501Yes、S502)。スイッチ管理部210は、負荷分散情報230の取得指示を受けたら、ネットワーク管理部110に対し、自身のスイッチ識別子及び負荷分散情報230を送信する。尚、ステップS502の負荷分散情報230の取得指示のタイミングと、ステップS402のスイッチ情報の取得指示のタイミング、すなわち、ステップS502とステップS402の更新周期140は同じでも異なってもどちらでも良い。

[0070] ネットワーク管理部110は、負荷分散情報230の取得指示に対する応答が規定時間内にない場合、応答のないスイッチ200への通信ができなくなったものとして、当該スイッチ200に対するエントリをネットワーク情報テーブル120から削除する(ステップS503No、S504)。一方、負荷分散情報230の取得指示に対する応答として、スイッチ200からスイッチ識別子及び負荷分散情報230が送信されると、ネットワーク管理部110は、送信された負荷分散情報230に基づいてスイッチ間の通信に利用するネットワークを決定し、ネットワークフローテーブル130及び当該スイッチ200のフローテーブル240を設定(更新)する(ステップS503Yes、S505~S508)。

[0071] 先ず、ネットワーク管理部110は、ネットワーク情報テーブル120において、負荷分散情報230の通知元のスイッチ200のスイッチ識別子と一致する送信元スイッチ識別子121のエントリ(ネットワーク情報)のうち、送信先スイッチ識別子122毎に最大の転送速度124を最大転送速度として抽出する(ステップS505)。

- [0072] ステップS505の具体例を説明する。ネットワーク管理部110は、受信したスイッチ識別子に一致する送信元スイッチ識別子121に対応するエントリを、送信元スイッチビューとしてネットワーク情報テーブル120から抽出する。ネットワーク管理部110は、送信元スイッチビューに現れる、すべての送信先スイッチ識別子122を送信先スイッチリストとして作成する。
- [0073] ネットワーク管理部110は、所定の送信先スイッチ識別子122を含むエントリを送信元スイッチビューから抽出し、スイッチ間経路ビューとする。スイッチ間経路ビューは、送信元スイッチ識別子121に対応するスイッチ200から所定の送信先スイッチ識別子122のスイッチ200へ到るネットワーク300のネットワーク識別子123と、そのネットワーク300の転送速度124が登録されている。ネットワーク管理部110は、抽出したスイッチ間経路ビューの全エントリの中で、転送速度124の値が最大のものを探し、当該スイッチ間におけるネットワーク300の最大転送速度とする。ネットワーク管理部110は、このようなスイッチ間ネットワークの最大転送速度を、送信先スイッチリストに登録された全ての送信先スイッチに対して行う。これにより、負荷分散情報230の通知元のスイッチ200から他のスイッチ200への転送速度が最大となるネットワーク300が、送信先スイッチ毎に特定される。
- [0074] 次に、ネットワーク管理部110は、負荷分散情報230とステップS505で算出した最大転送速度から、許容転送速度を算出する。例えば、ネットワーク管理部110は、負荷分散情報230の転送閾値233と最大転送速度を乗じた値を許容転送速度として算出する。具体例として転送閾値233が60%であり、最大転送速度が1000MB/secであるとき、許容転送速度は600MB/secとなる。このようなスイッチ間ネットワークの許容転送速度を、送信先スイッチリストに登録された全ての送信先スイッチに対して行う。これにより、負荷分散情報230の通知元のスイッチ200から他のスイッチ200への許容転送速度が、送信先スイッチ毎に特定さ

れる。

[0075] 次に、ネットワーク管理部 110 は、負荷分散情報 230 とステップ S506 で算出した許容転送速度に基づいて通信に使用するネットワークを絞り込み、ネットワークフローテーブル 130 に登録（更新）する（ステップ S507）。ここでは、ネットワーク管理部 110 は、スイッチ間経路ビューのすべてのエントリについて、ステップ S506 で求めた許容転送速度と転送速度 124 を比較し、転送速度 124 が許容転送速度以上のエントリに絞り込む。このとき、転送速度 124 が 0 のエントリは絞り込み結果から除外する。又、転送速度 124 が許容転送速度以上のエントリ数が、負荷分散情報 230 の最大ネットワーク数 232 より多い場合はさらに絞り込みが行われる。例えば、許容転送速度によって絞り込まれた結果から、転送速度 124 が高い順に、最大ネットワーク数 232 の個数のエントリまで絞り込まれる。このようなスイッチ間通信で使用するネットワークの絞り込み（スイッチ間経路ビューの絞り込み）を、送信先スイッチリストに登録された全ての送信先スイッチに対して行う。これにより、負荷分散情報 230 の通知元のスイッチ 200 から他のスイッチ 200 への通信に利用するネットワーク 300 が、送信先スイッチ毎に特定される。

[0076] ネットワーク管理部 110 は、上記処理によって絞り込んだエントリで規定されたネットワーク 300 を、当該エントリで規定された送信元スイッチ及び送信先スイッチの間で使用するネットワークとして設定する。詳細には、当該エントリの送信元スイッチ識別子及び送信先スイッチ識別子は、ネットワークフローテーブル 130 の送信元スイッチ識別子 131 及び送信先スイッチ識別子 132 として登録され、当該エントリのネットワーク識別子は、これらに対応するネットワーク識別子リスト 133 に登録される。このようなネットワークフローテーブル 130 への登録を、送信先スイッチリストに登録された全ての送信先スイッチに対して行う。これにより、負荷分散情報 230 の通知元のスイッチ 200 から他のスイッチ 200 への通信に利用するネットワーク 300 が、送信先スイッチ毎にネットワークフローテーブ

ル 130 に登録される。

- [0077] ステップ S503、S504、S505、S506、S507 の処理をネットワーク情報テーブル 120 の全てのエントリについて実施することで、コントローラ 100 と通信可能な全てのスイッチ 200 間の各々で利用するネットワーク 300 をネットワークフローテーブル 130 に登録することができる。
- [0078] ネットワーク管理部 110 は、更新したネットワークフローテーブル 130 に基づいてスイッチ 200 のフローテーブル 240 を更新する（ステップ S508）。ネットワーク管理部 110 は、スイッチ管理部 210 へ、更新対象のフローテーブルを送信する。ここでは、送信元スイッチ識別子 131 に一致するスイッチ識別子のスイッチ 200 に、当該送信元スイッチ識別子 131 に対応付けられた送信先スイッチ識別子 132 及びネットワーク識別子リスト 133 が送信される。スイッチ管理部 210 は、送信された送信先スイッチ識別子 132 及びネットワーク識別子リスト 133 を、送信先スイッチ識別子 243 及びネットワーク識別子リスト 244 として、自身のフローテーブル 240 に更新設定する。
- [0079] スwitch管理部 210 は、他のスイッチ 200 への通信があったとき、フローテーブル 240 を参照し、送信先スイッチ識別子 243 が当該他のスイッチ 200 のスイッチ識別子と一致するエントリのネットワーク識別子リスト 244 に記載されているネットワークを使用して通信を行う。ネットワーク識別子リスト 244 に複数の通信経路が含まれている場合はリンクアグリゲーションなどの従来技術を用いて通信を行う。具体的な通信方式については本発明の範囲外のため、省略する。
- [0080] 次に、ステップ S505～S508 のネットワークの決定動作について、具体例を用いて説明する。ここでは、図 7 に示すシステムにおいて、スイッチ 200-1 から送信された負荷分散情報 230 に基づいてスイッチ 200-1 に設定するフローを決定する動作について説明する。
- [0081] ネットワーク管理部 110 は、図 17 に示す負荷分散情報 230 及びスイ



タッチ識別子“スイッチx”をスイッチ200-1から受け取ると、送信元スイッチ識別子121が“スイッチx”と一致するエントリを、スイッチxに関する送信元スイッチビューとして、ネットワーク情報テーブル120から抽出する。ここで抽出された送信元スイッチビューの一例を図18に示す。

[0082] ネットワーク管理部110は、送信元スイッチビューに現れる、すべての送信先スイッチ識別子が格納された送信先スイッチリストを作成する。図14を参照すると、送信元スイッチビューには送信先スイッチ識別子122として“スイッチy”のみが登場するため、送信先スイッチリストには送信先スイッチ識別子“スイッチy”が1個だけ登録される。

[0083] ネットワーク管理部110は、作成した送信元スイッチビューの送信先スイッチ識別子122がスイッチ識別子“スイッチy”と一致するエントリを抽出する。ここでは、送信元スイッチビューのすべてのエントリが該当するため、図14の送信元スイッチビューがそのまま、スイッチ200-1（“スイッチx”）を送信元とし、スイッチ200-2（“スイッチy”）を送信先としたスイッチ間経路ビューとなる。

[0084] 続いて、ネットワーク管理部110は最大転送速度の抽出及び許容転送速度の計算を行う。ここでは、図18に示すスイッチ間経路ビューから、送信先スイッチ200-2に対する最大転送速度として1000MB/secが抽出される。又、図17に示す負荷分散情報230の転送閾値233が60%であることがことから、許容転送速度は600MB/secとなる。

[0085] ネットワーク管理部110は、使用するネットワークの絞り込みを行う。ここでは、図18に示すスイッチ間経路ビューから、転送速度124が許容転送速度“600MB/sec”に満たないエントリが除外される。これにより“ネットワーク3”と“ネットワーク4”を含むエントリは、利用ネットワークの候補から除外される。

[0086] 続いて、ネットワーク管理部110は、残されたスイッチ間経路ビューのエントリ数をチェックする。エントリ数は2つであり、図13に示す負荷分散情報230から、最大ネットワーク数232は“2”であるので、十分な

個数まで絞り込めており、絞り込み処理は終了する。本処理により絞り込まれた後のスイッチ間経路ビューを図19に示す。

[0087] ネットワーク管理部110は、図19のように絞り込まれたスイッチ間経路ビューに登場するすべてのネットワーク識別子123“ネットワーク1”、“ネットワーク2”を、使用ネットワークリストとして生成する。

[0088] ネットワーク管理部110は、使用ネットワークリストに基づいてネットワークフローテーブル130を更新する。更新されたネットワークフローテーブル130を図20に示す。スイッチ200-1の送信先スイッチリストには1つの要素“スイッチy”のみであるため、他の送信先スイッチに対する上記処理は行われず、図20に示すネットワークフローテーブル130が更新対象フローテーブルとしてスイッチ200-1に送信される。スイッチ200-1のスイッチ管理部210は、自身のフローテーブル240を、送信された更新対象フローテーブルに置き換える。これにより、スイッチ200-1のフローテーブル240（のアクション242）は図21に示すように書き換えられる。以降、スイッチ200-1は、スイッチ200-2を送信先としてパケット転送する場合、転送速度が許容速度600MB/secよりも速いネットワーク300-1、300-2のみを利用することとなる。

[0089] 本発明によるコンピュータシステムでは、スイッチ200において収集したネットワークの負荷状況をコントローラ100で集約し、その結果に基づいたスイッチ200のフロー制御（使用ネットワークの決定）をコントローラ100によって行っている。このため、本発明によれば、通信ネットワークが存在するシステム（例示：ストレージシステム）において、通信に使用するネットワークを負荷分散ポリシーに応じて自動的に判断できる。

[0090] 又、本発明に係るコントローラ100は、スイッチ200において測定された通信転送速度によって負荷状況を把握しているため、複数のスイッチ200がネットワークの帯域を共有している場合でも、負荷状況を判断することが可能である。

[0091] 更に、本発明に係るコントローラ100は、規定間隔でネットワークフローテーブルの更新を行っている。このため、ネットワークの障害や、通信速度の低下が発生に関わらず、自動的に適切な経路（通信状態の良好な経路）に切り替えることが可能である。本発明によるコンピュータシステムでは、スイッチ200の追加や移動、削除に例示されるように、動的に構成が変更されても、構成状況をコントローラ100において把握できる。コントローラ100は、構成変更後の通信経路における転送性能を定期的を取得し、最適経路によるフロー制御を実現できる。このため、ネットワーク内に障害が発生した場合でも、最適な通信経路を利用したスイッチ間通信が可能となる。

[0092] 又、ネットワーク内の全通信について転送速度等の統計情報を取得しているため、FCoEなど、同一ネットワーク上にI/Oが混在するような場合についても、最適な通信経路を利用することが可能となる。

[0093] 以上、本発明の実施の形態を詳述してきたが、具体的な構成は上記実施の形態に限られるものではなく、本発明の要旨を逸脱しない範囲の変更があっても本発明に含まれる。図1や図7に示したシステムの一部では、スイッチ200を2つとしていたが、これに限らず3つ以上の任意の個数のスイッチについても拡張可能である。例えば、4つのスイッチ200-1~200-4を備えるシステムの一部を図22に示す。図22に示す一例では、4つのスイッチ200-1~200-4のそれぞれの上に2つのネットワーク330-1、330-2が接続されている。又、スイッチ200-1~200-4には、それぞれコンピュータ400-1~400-4が接続されている。このようなシステムでも、上述の実施例と同様の動作により、最適ネットワークを利用したコンピュータ間通信が実現できる。

[0094] 尚、本出願は、日本出願番号2012-003551に基づき、日本出願番号2012-003551における開示内容は引用により本出願に組み込まれる。

## 請求の範囲

[請求項1]

コントローラと、

複数のネットワークを介して他のスイッチに接続され、前記コントローラによって設定されたフローエントリに従って、受信パケットの中継処理を行うスイッチと、

を具備し、

前記スイッチは、前記複数のネットワークのそれぞれを介して前記他のスイッチに検査用データを転送して、前記複数のネットワークのそれぞれの転送性能を取得し、前記転送性能を前記複数のネットワークのそれぞれに対応付けて前記コントローラに通知し、

前記コントローラは、前記複数のネットワークのうち、前記転送性能が所定の閾値以上のネットワークを前記他のスイッチへの転送経路として規定したフローエントリを、前記スイッチのフローテーブルに設定する

コンピュータシステム。

[請求項2]

請求項1に記載のコンピュータシステムにおいて、

前記スイッチは、前記コントローラによって定期的に行われる前記転送性能の取得指示に応じて、前記転送性能の取得及び通知を行う

コンピュータシステム。

[請求項3]

請求項1又は2に記載のコンピュータシステムにおいて、

前記コントローラは、前記転送性能に基づいた前記スイッチに対するフローエントリの設定を定期的に行う

コンピュータシステム。

[請求項4]

請求項1から3のいずれか1項に記載のコンピュータシステムにおいて、

前記コントローラは、前記スイッチから送信された転送性能に基づいて許容転送速度を算出し、転送性能が前記許容転送速度以上のネットワークを前記スイッチが利用するネットワークとして、前記スイッ

チのフローテーブルに設定する

コンピュータシステム。

[請求項5]

請求項4に記載のコンピュータシステムにおいて、

前記コントローラは、転送性能が前記許容転送速度以上のネットワークのうち、転送性能が上位のネットワークから、前記スイッチが利用可能な最大ネットワーク数以下のネットワークを、前記スイッチが利用するネットワークとして決定する

コンピュータシステム。

[請求項6]

請求項1から5のいずれか1項に記載のコンピュータシステムで利用されるコントローラ。

[請求項7]

請求項1から5のいずれか1項に記載のコンピュータシステムで利用されるスイッチ。

[請求項8]

コントローラと、

複数のネットワークを介して他のスイッチに接続され、前記コントローラによって設定されたフローエントリに従って、受信パケットの中継処理を行うスイッチと、

を具備するコンピュータシステムにおいて実行される通信方法であって、

前記スイッチが、複数のネットワークのそれぞれを介して他のスイッチに検査用データを転送して、前記複数のネットワークのそれぞれの転送性能を取得するステップと、

前記転送性能を前記複数のネットワークのそれぞれに対応付けて前記コントローラに通知するステップと、

前記コントローラが、前記複数のネットワークのうち、前記転送性能が所定の閾値以上のネットワークを前記他のスイッチへの転送経路として規定したフローエントリを、前記スイッチのフローテーブルに設定するステップと

を具備する

通信方法。

[請求項9]

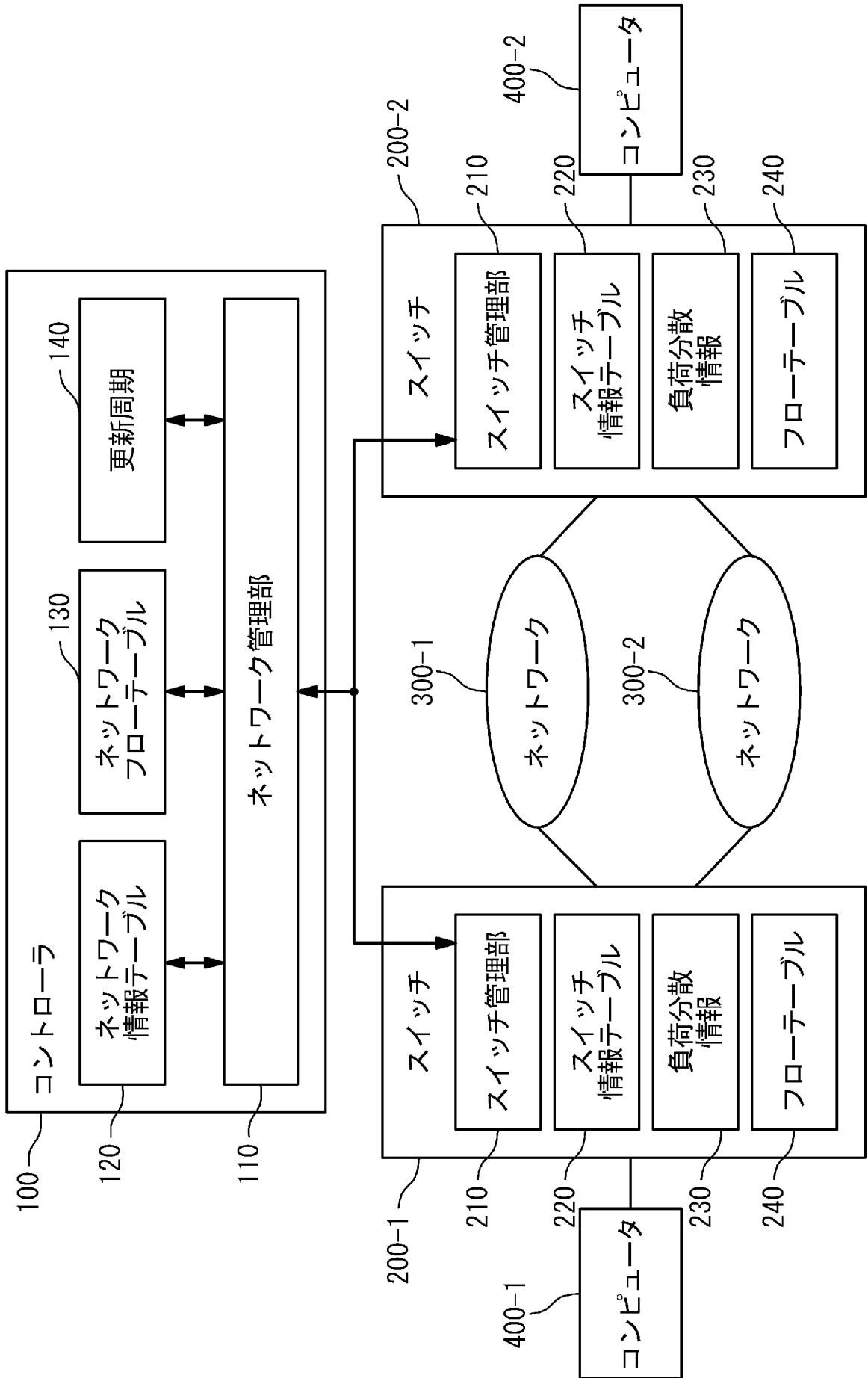
スイッチに接続された複数のネットワークのそれぞれの転送性能を前記スイッチから取得するステップと、

前記スイッチは、受信パケットをフローテーブルで規定されたネットワークに転送し、

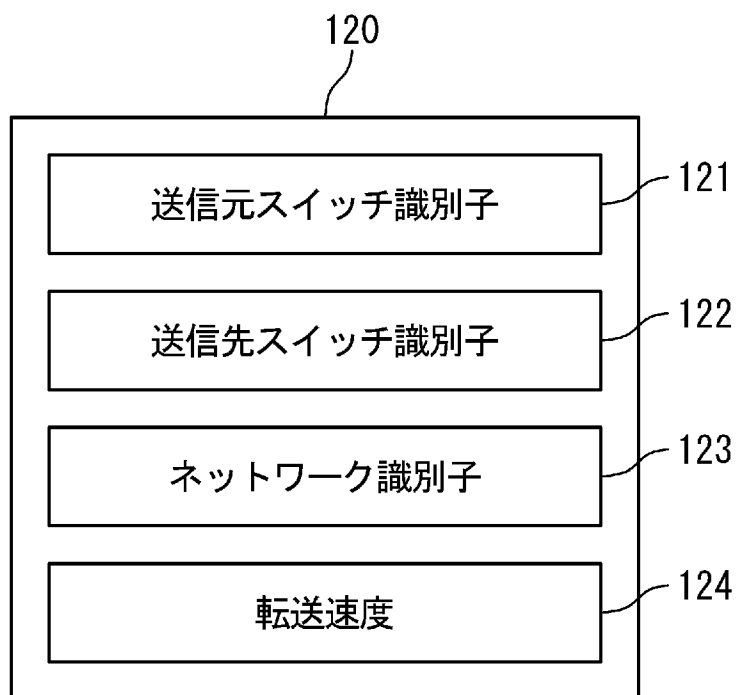
転送性能が閾値以上のネットワークを転送経路として規定したフローエントリを、前記スイッチの前記フローテーブルに設定するステップと

をコンピュータに実行させるネットワーク管理プログラムが格納された記録媒体。

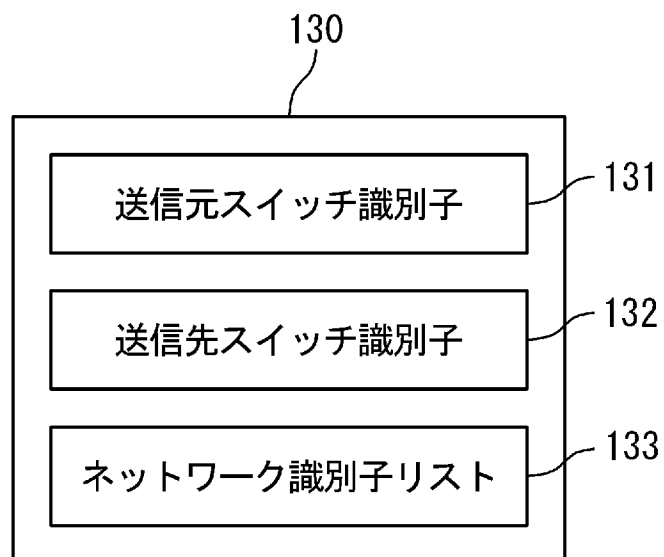
[図1]



[図2]

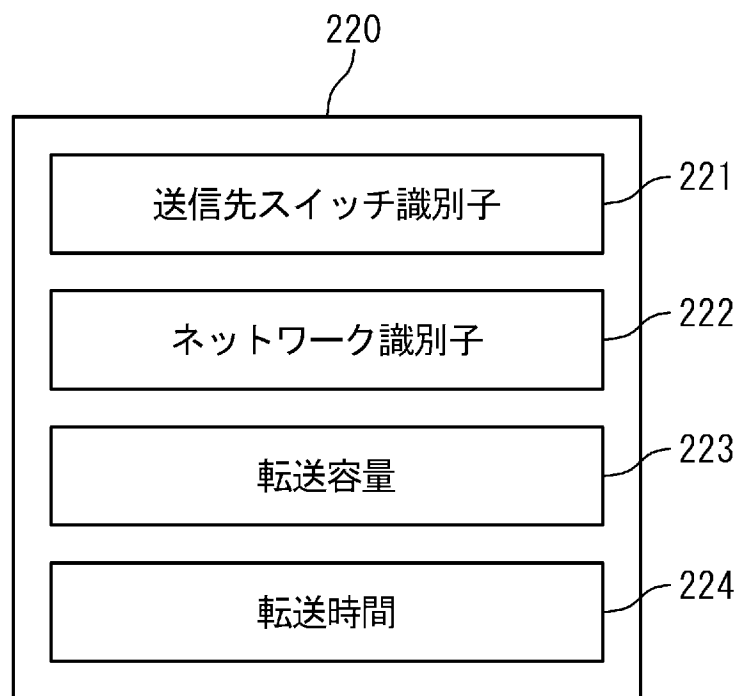


[図3]

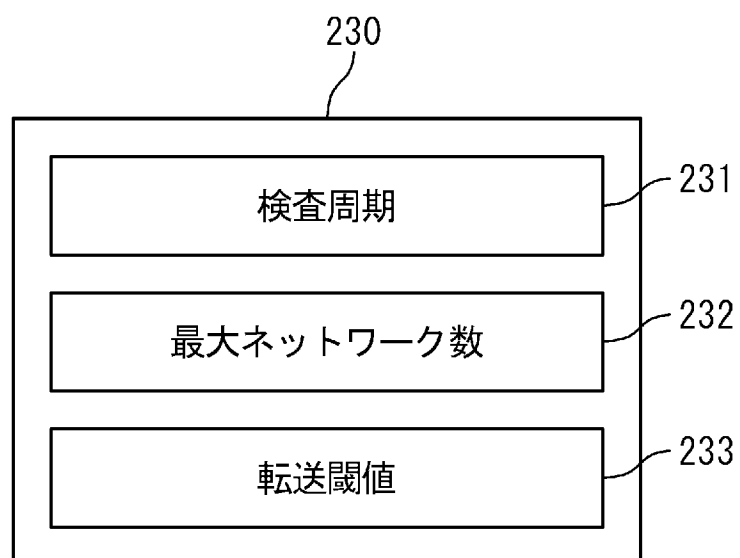




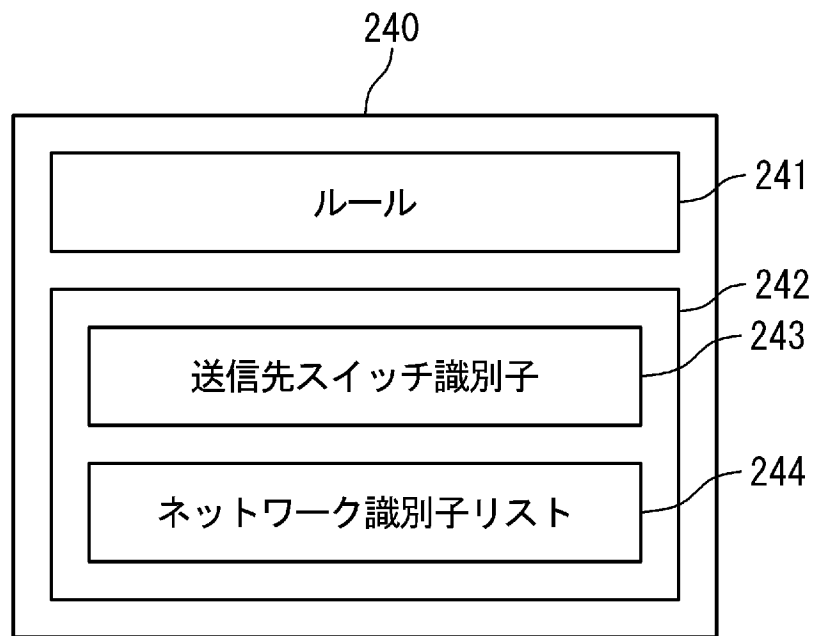
[図4]



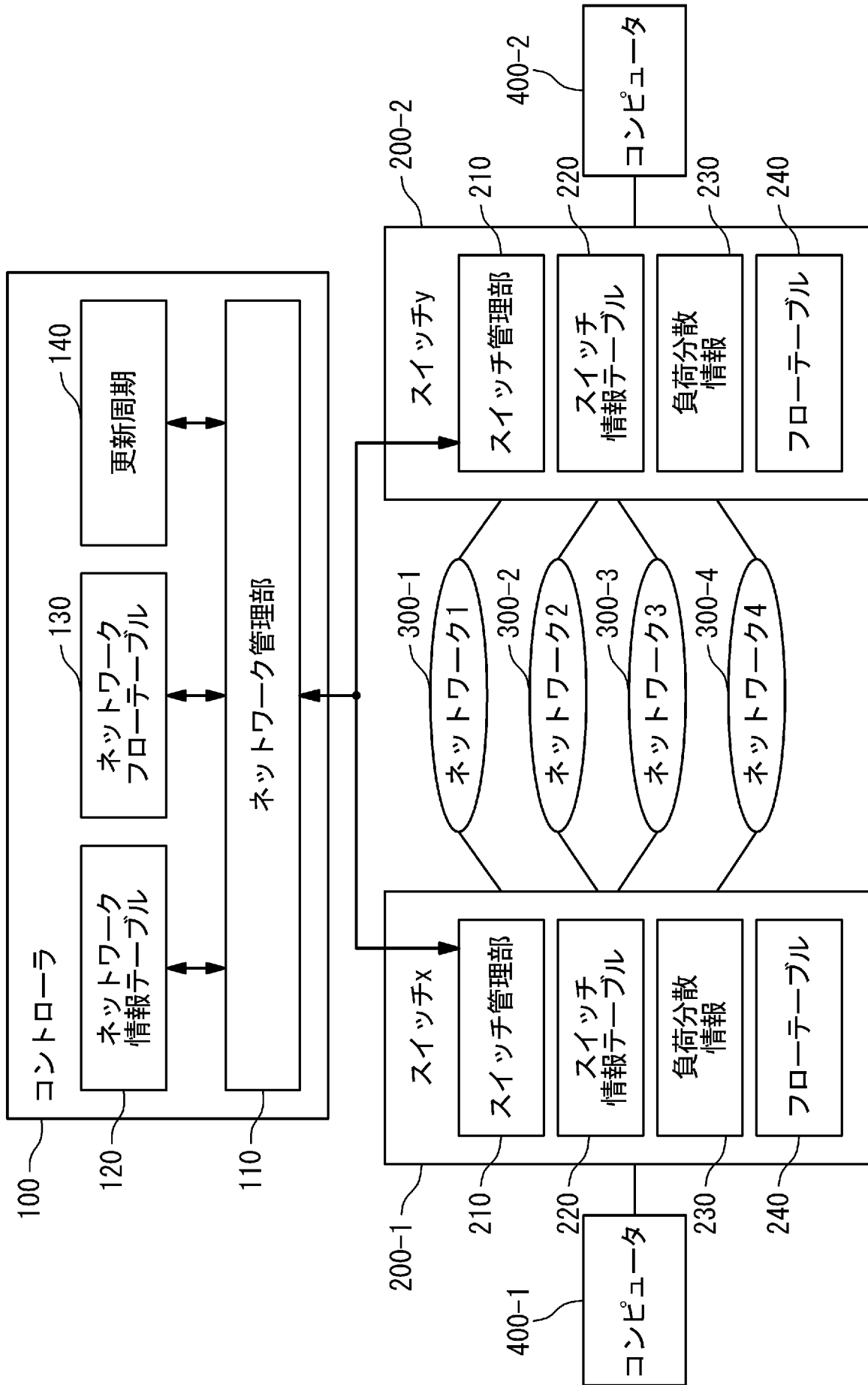
[図5]



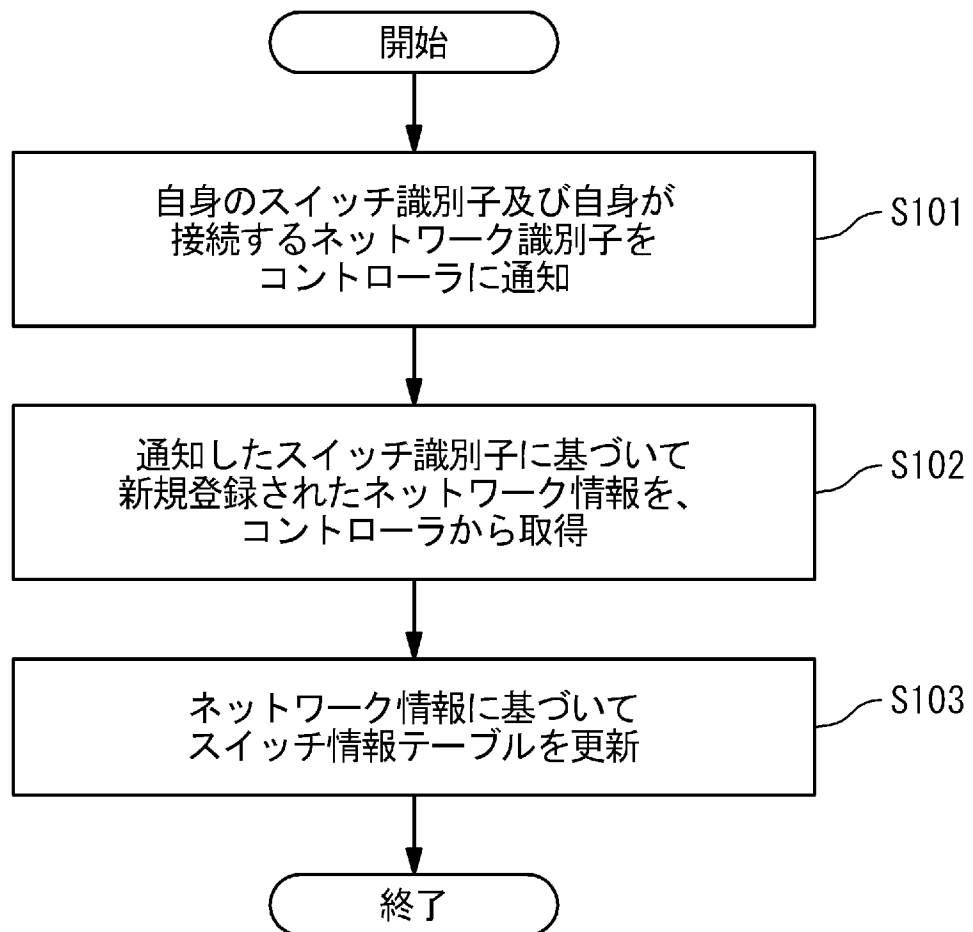
[図6]



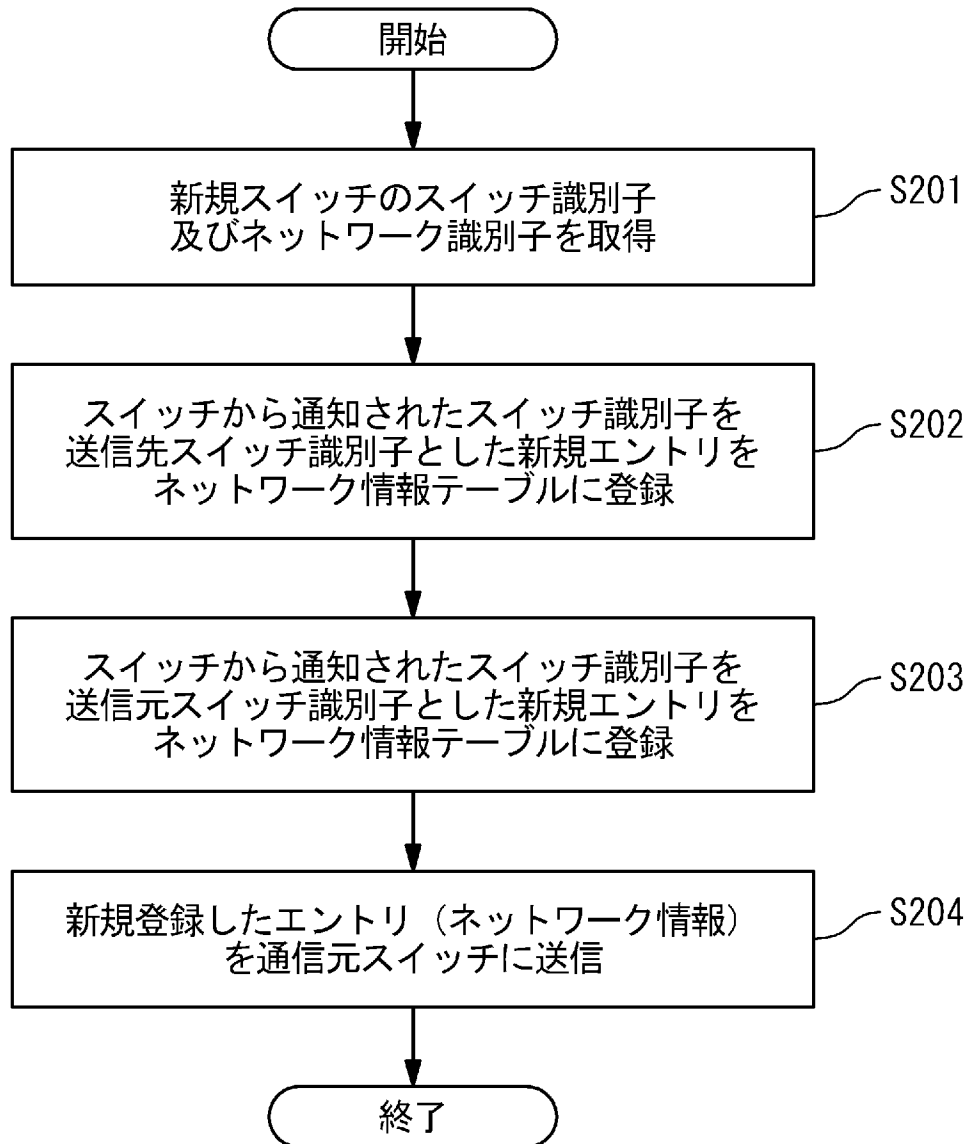
[図7]



[図8]



[図9]



[図10]

120

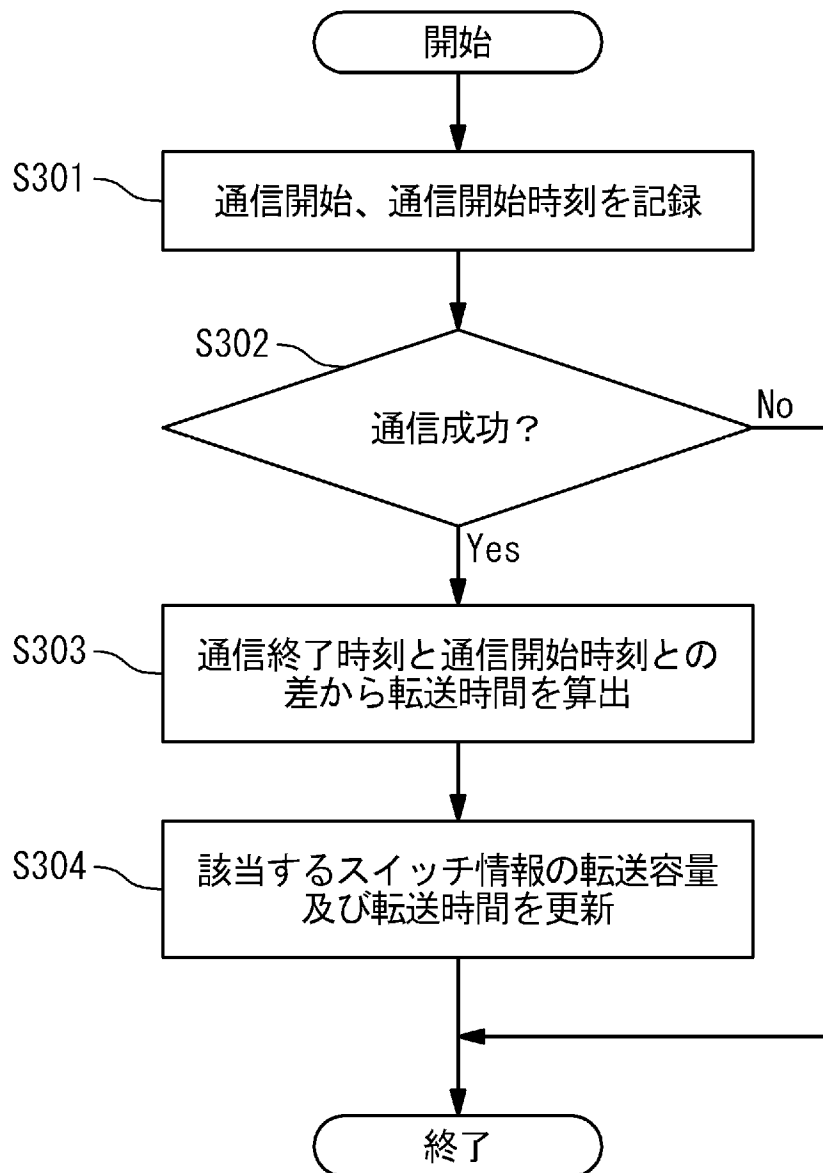
送信元 スイッチ 識別子	送信先 スイッチ 識別子	ネットワーク 識別子	転送速度
スイッチx	スイッチy	ネットワーク1	0
スイッチx	スイッチy	ネットワーク2	0
スイッチx	スイッチy	ネットワーク3	0
スイッチx	スイッチy	ネットワーク4	0
スイッチy	スイッチx	ネットワーク1	0
スイッチy	スイッチx	ネットワーク2	0
スイッチy	スイッチx	ネットワーク3	0
スイッチy	スイッチx	ネットワーク4	0

[図11]

220

送信先 スイッチ 識別子	ネットワーク 識別子	転送容量	転送時間
スイッチy	ネットワーク1	0	0
スイッチy	ネットワーク2	0	0
スイッチy	ネットワーク3	0	0
スイッチy	ネットワーク4	0	0

[図12]



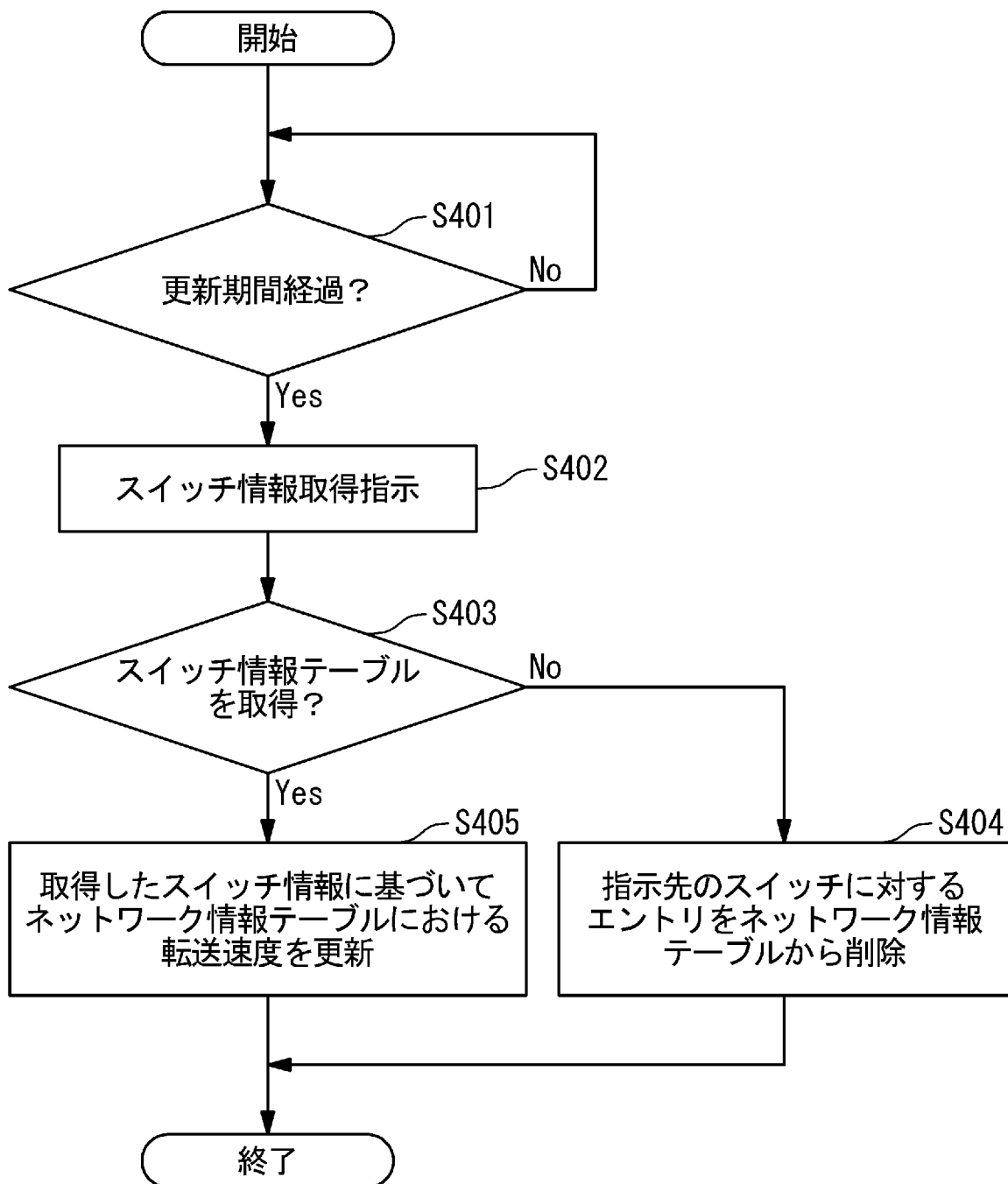


[図13]

220

送信先 スイッチ 識別子	ネットワーク 識別子	転送容量	転送時間
スイッチy	ネットワーク1	2000MB	2000ms
スイッチy	ネットワーク2	1600MB	2000ms
スイッチy	ネットワーク3	300MB	1000ms
スイッチy	ネットワーク4	200MB	1000ms

[図14]

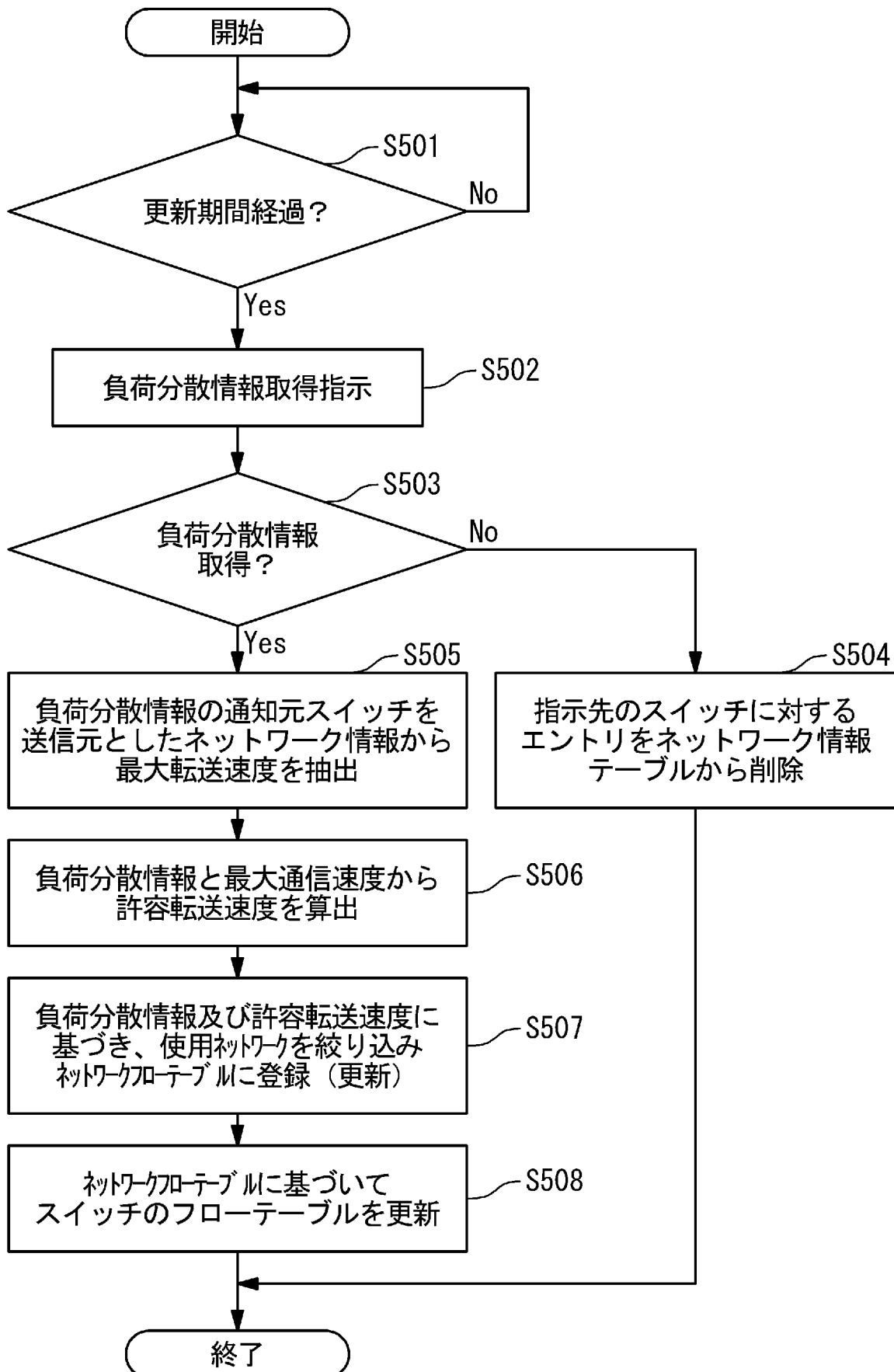


[図15]

120

送信元 スイッチ 識別子	送信先 スイッチ 識別子	ネットワーク 識別子	転送速度
スイッチx	スイッチy	ネットワーク1	1000MB/sec
スイッチx	スイッチy	ネットワーク2	800MB/sec
スイッチx	スイッチy	ネットワーク3	300MB/sec
スイッチx	スイッチy	ネットワーク4	200MB/sec
スイッチy	スイッチx	ネットワーク1	500MB/sec
スイッチy	スイッチx	ネットワーク2	450MB/sec
スイッチy	スイッチx	ネットワーク3	250MB/sec
スイッチy	スイッチx	ネットワーク4	150MB/sec

[図16]



[図17]

230

231	232	233
検査周期	最大 ネットワーク数	転送閾値
30sec	2	60%

[図18]

121	122	123	124
送信元 スイッチ 識別子	送信先 スイッチ 識別子	ネットワーク 識別子	転送速度
スイッチx	スイッチy	ネットワーク1	1000MB/sec
スイッチx	スイッチy	ネットワーク2	800MB/sec
スイッチx	スイッチy	ネットワーク3	300MB/sec
スイッチx	スイッチy	ネットワーク4	200MB/sec

[図19]

121 送信元 スイッチ 識別子	122 送信先 スイッチ 識別子	123 ネットワーク 識別子	124 転送速度
スイッチx	スイッチy	ネットワーク1	1000MB/sec
スイッチx	スイッチy	ネットワーク2	800MB/sec

[図20]

130

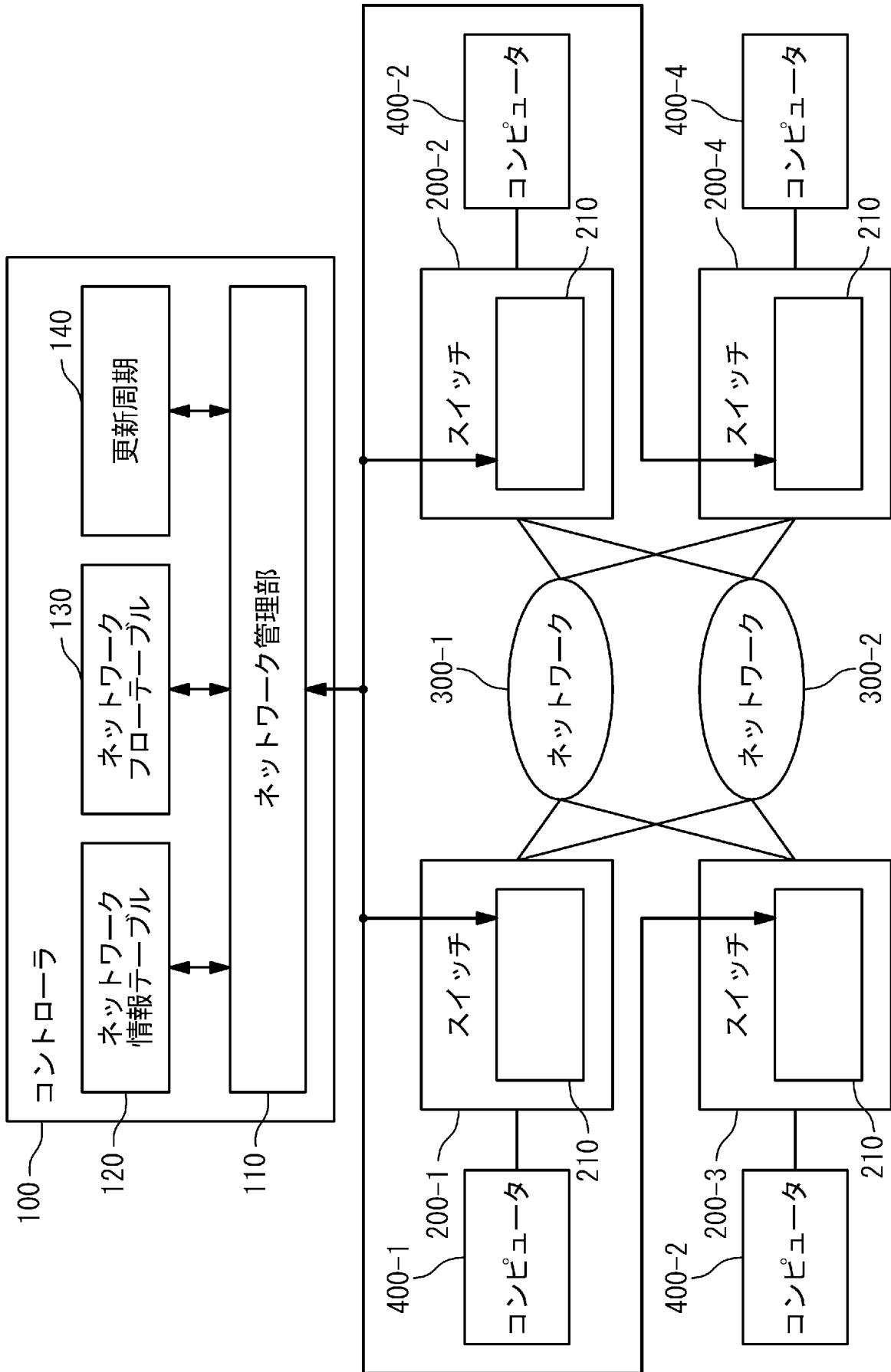
131	送信元スイッチ識別子	スイッチx
132	送信先スイッチ識別子	スイッチy
133	ネットワーク識別子リスト	(ネットワーク1、ネットワーク2)

[図21]

242

243	送信先スイッチ識別子	スイッチY
244	ネットワーク識別子リスト	ネットワーク1、ネットワーク2

[図22]



**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/JP2013/050118

**A. CLASSIFICATION OF SUBJECT MATTER**

H04L12/721(2013.01)i, H04L12/70(2013.01)i, H04L12/801(2013.01)i,  
H04L12/911(2013.01)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)

H04L12/721, H04L12/70, H04L12/801, H04L12/911

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.
Y	Seok Hong Min et al., Implementation of a Programmable Service Composition Network using NetFPGA-based OpenFlow Switches, 1st Asia NetFPGA Developers Workshop, 2010.06, [retrieved on 2013-02-21], Retrieved from the Internet: <URL: http://fif.kr/AsiaNetFPGAs/paper/1-1.pdf>	1-3, 6-9
Y	JP 2006-60579 A (Fujitsu Ltd.), 02 March 2006 (02.03.2006), claims 2, 3; paragraphs [0125] to [0144] & US 2006/0039335 A1	1-3, 6-9

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  
21 February, 2013 (21.02.13)

Date of mailing of the international search report  
05 March, 2013 (05.03.13)

Name and mailing address of the ISA/  
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2013/050118

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2004-236198 A (Fujitsu Ltd.), 19 August 2004 (19.08.2004), entire text & US 2004/0184483 A1 & EP 1443722 A2 & DE 602004024732 D & CN 1525701 A	1-9
A	JP 2002-64536 A (Nippon Telegraph and Telephone Corp.), 28 February 2002 (28.02.2002), abstract (Family: none)	1-9

A. 発明の属する分野の分類 (国際特許分類 (IPC))  
 Int.Cl. H04L12/721 (2013.01)i, H04L12/70 (2013.01)i, H04L12/801 (2013.01)i, H04L12/911 (2013.01)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 (IPC))  
 Int.Cl. H04L12/721, H04L12/70, H04L12/801, H04L12/911

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 1922-1996年  
 日本国公開実用新案公報 1971-2013年  
 日本国実用新案登録公報 1996-2013年  
 日本国登録実用新案公報 1994-2013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
Y	Seok Hong Min et al., Implementation of a Programmable Service Composition Network using NetFPGA-based OpenFlow Switches, 1st Asia NetFPGA Developers Workshop, 2010.06, [retrieved on 2013-02-21], Retrieved from the Internet: <URL: http://fif.kr/AsiaNetFPGAs/paper/1-1.pdf>	1-3, 6-9
Y	JP 2006-60579 A (富士通株式会社) 2006.03.02, 【請求項2】, 【請求項3】, 【0125】 - 【0144】 & US 2006/0039335 A1	1-3, 6-9

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  
 「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  
 「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  
 「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)  
 「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日の後に公表された文献  
 「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  
 「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  
 「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  
 「&」同一パテントファミリー文献

国際調査を完了した日 21.02.2013 国際調査報告の発送日 05.03.2013

国際調査機関の名称及びあて先 日本国特許庁 (ISA/J P) 郵便番号100-8915 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) 速水 雄太 電話番号 03-3581-1101 内線 3596	5 X	3365
---	--	-----	------

C (続き) . 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 2004-236198 A (富士通株式会社) 2004.08.19, 全文 & US 2004/0184483 A1 & EP 1443722 A2 & DE 602004024732 D & CN 1525701 A	1-9
A	JP 2002-64536 A (日本電信電話株式会社) 2002.02.28, 要約 (ファミリーなし)	1-9