

(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
2020年2月13日(13.02.2020)



(10) 国際公開番号

WO 2020/031996 A1

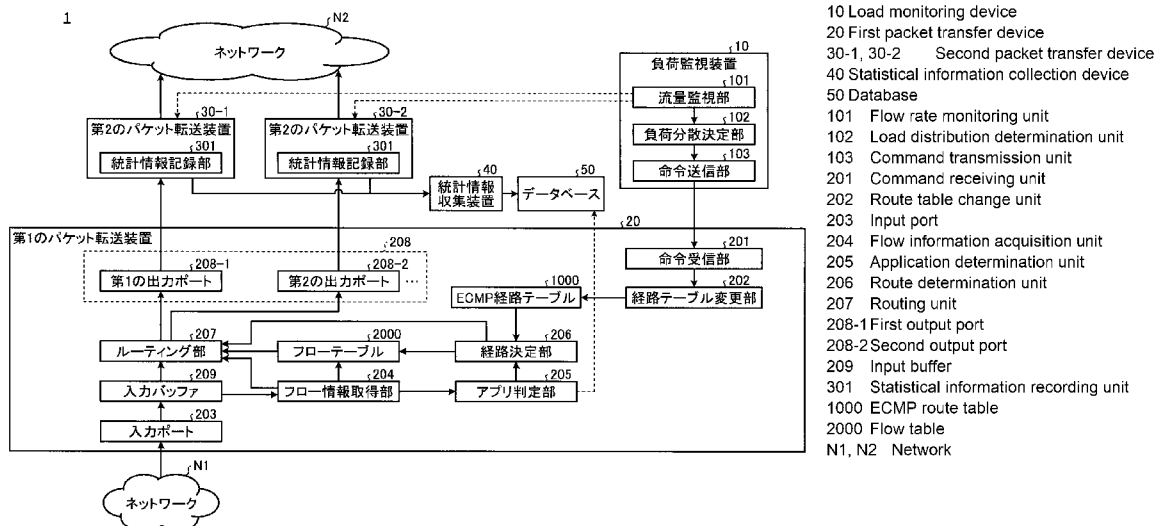
- (51) 国際特許分類:
H04L 12/803 (2013.01) H04L 12/851 (2013.01)
- (21) 国際出願番号: PCT/JP2019/030813
- (22) 国際出願日: 2019年8月6日(06.08.2019)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2018-148940 2018年8月7日(07.08.2018) JP
- (71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).
- (72) 発明者: 木村 明寛 (KIMURA, Akihiro); 〒1808585 東京都武蔵野市緑町3丁目9-1 1

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9-1 1 N T T 知的財産センタ内 Tokyo (JP). 土屋 英雄(TSUCHIYA, Hideo); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 N T T 知的財産センタ内 Tokyo (JP). 岡田 昭宏 (OKADA, Akihiro); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 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9-1 1 N T T 知的財産センタ内 Tokyo (JP).

(74) 代理人: 特許業務法人酒井国際特許事務所 (SAKAI INTERNATIONAL PATENT OFFICE); 〒1000013 東京都千代田区霞が関3丁目8番1号 虎の門三井ビルディング Tokyo (JP).

(54) Title: LOAD DISTRIBUTION SYSTEM AND LOAD DISTRIBUTION METHOD

(54) 発明の名称: 負荷分散システム及び負荷分散方法



(57) Abstract: The present invention is characterized by including: a storage means for storing, in association with packet destination information, information pertaining to a flow characteristic indicating whether a flow constituted by packets tends to have many short packets or tends to have many long packets; a determination means for, upon receipt of a packet by an input port, determining, on the basis of the storage means, the flow characteristic of the flow constituted by the packet; and an output means for, in accordance with the flow characteristic determined by the determination means, outputting the packet received by the input port to, from among a plurality of packet transfer devices, a packet transfer device with a load state indicating that the load due to the flow with said flow characteristic would be small.

(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, JO, 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, 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, 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))

(57) 要約: パケットにより構成されるフローがショートパケットが多い傾向のフロー又はロングパケットが多い傾向のフローのいずれであることを示すフロー特性に関する情報を該パケットの宛先情報に対応付けて記憶する記憶手段と、入力ポートがパケットを受信すると、該パケットにより構成されるフローのフロー特性を前記記憶手段に基づいて判定する判定手段と、前記判定手段により判定されたフロー特性に応じて、複数のパケット転送装置のうち、前記フロー特性のフローによる負荷が少ないことを示す負荷状態のパケット転送装置に対して前記入力ポートが受信したパケットを出力する出力手段と、を有することを特徴とする。

明 細 書

発明の名称： 負荷分散システム及び負荷分散方法

技術分野

[0001] 本発明は、負荷分散システム及び負荷分散方法に関する。

背景技術

[0002] ルータ等のパケット転送装置の負荷を分散させる技術が従来から知られている。例えば、パケットの転送先として複数のパケット転送装置と複数の転送経路とを用意し、それぞれの転送経路にトラヒックを均等に印加することで、パケット転送装置の負荷を分散する技術が知られている。また、このような負荷分散技術として、E C M P (Equal Cost Multi Path) 又はE C M Pルーティングと呼ばれる技術も知られている。E C M Pは、上述した複数の転送経路のうち、等コストとなる転送経路にトラヒックを均等に印加する負荷分散技術である。

[0003] E C M Pを用いた負荷分散システムにおいて、パケット転送装置の出力物理ポートの出力バッファを監視し、この出力バッファの使用量又はパケット流量が閾値を超えた場合、転送経路を切り替える技術が知られている（例えば、特許文献1）。

[0004] パケット転送装置における転送容量の指標として、b p s (bits per second) で表されるデータ量の転送容量と、p p s (packets per second) で表されるパケット数の転送容量とがある。また、これらの転送容量には、パケット転送装置に固有の上限値がある。例えば、パケット転送装置をネットワーク仮想化技術により実現した場合、b p sは物理N I C (Network Interface Card) のラインレートに律速されることがあり、p p sはO S (Operating System) やアプリケーション等のソフトウェアのパケット処理速度に律速されることがある。

[0005] E C M Pでは、フロー特性、すなわちロングパケット中心のフローかショートパケット中心のフローかを考慮していないため、パケット転送装置が転

送するパケット長に偏りがある場合に、パケット転送装置の転送容量を最大限に活用することができないことがある。ここで、フローとは、フロー情報が同一のパケット（トラヒック）又はパケット群のことを言う。フロー情報は、パケットに含まれるヘッダ情報により特定され、例えば、5-tuple（送信元IP（Internet Protocol）アドレス、宛先IPアドレス、送信元ポート番号、宛先ポート番号、プロトコルタイプ）のことを言う。

[0006] 例えば、動画等のトラヒックはロングパケットが多いという性質がある。このため、動画等のトラヒックが集中した場合、データ量の転送容量は枯渇する一方で、パケット数の転送容量には余裕がある状態となる。他方で、例えば、VoIP（Voice over Internet Protocol）やテレメトリング等のトラヒックはショートパケットが多いという性質がある。このため、VoIPやテレメトリング等のトラヒックが集中した場合、データ量の転送容量には余裕がある一方で、パケット数の転送容量が枯渇する状態となる。

[0007] そこで、フロー特性を考慮して負荷を分散する技術がある。例えば、宅内装置（Customer Premises Equipment）の一部機能を仮想化するvCPE（Virtualized CPE）をホストするサーバの負荷分散を、サーバのbps及びppsの観点での負荷状態とフローのパケット長に基づいて行う負荷分散方式がある（例えば、非特許文献1）。

先行技術文献

特許文献

[0008] 特許文献1：特開2004-350078号公報

非特許文献

[0009] 非特許文献1：木村 明寛、宮本 克真、河野 伸也、土屋 英雄、岡田 昭宏、「フロー特性に基づくvCPEサーバの負荷分散方式」、電子情報通信学会2018年総合大会講演論文集，Vol. 2018年__通信（2），p. 906-909

発明の概要

発明が解決しようとする課題

[0010] フロー特性に基づいて負荷分散を行う場合、いかにフロー特性を効率良く正確に判定するかが課題となる。

[0011] 本発明の実施の形態は、上記の点に鑑みてなされたもので、フロー特性を効率良く正確に判定することを目的とする。

課題を解決するための手段

[0012] 上記目的を達成するため、本発明の実施の形態は、パケットにより構成されるフローがショートパケットが多い傾向のフロー又はロングパケットが多い傾向のフローのいずれであることを示すフロー特性に関する情報を該パケットの宛先情報に対応付けて記憶する記憶手段と、入力ポートがパケットを受信すると、該パケットにより構成されるフローのフロー特性を前記記憶手段に基づいて判定する判定手段と、前記判定手段により判定されたフロー特性に応じて、複数のパケット転送装置のうち、前記フロー特性のフローによる負荷が少ないことを示す負荷状態のパケット転送装置に対して前記入力ポートが受信したパケットを出力する出力手段と、を有することを特徴とする。

発明の効果

[0013] フロー特性を効率良く正確に判定することを目的とする。

図面の簡単な説明

[0014] [図1]図1は、本発明の実施の形態における負荷分散システムの全体構成の一例を示す図である。

[図2]図2は、実施例1における負荷分散システムの機能構成の一例を示す図である。

[図3]図3は、実施例1におけるデータベースの一例を示す図である。

[図4]図4は、実施例1におけるECMP経路テーブルの一例を示す図である。

[図5]図5は、実施例1におけるECMP経路テーブルの変更処理の流れの一例を説明するための図である。

[図6]図6は、実施例1におけるパケット転送処理の流れの一例を説明するための図である。

[図7]図7は、実施例1におけるフロー特性の登録処理の流れの一例を説明するための図である。

[図8]図8は、新規登録処理の流れの一例を説明するための図である。

[図9]図9は、実施例2におけるデータベースの一例を示す図である。

[図10]図10は、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。

発明を実施するための形態

[0015] 以下、本発明の実施の形態について説明する。本発明の実施の形態では、フロー特性に応じて、負荷分散を行う負荷分散システム1について説明する。また、以降では、フロー特性には、ロングパケットが多いフローであることを示す「高bpsフロー」と、ショートパケットが多いフローであることを示す「高ppsフロー」とがあるものとする。

[0016] <全体構成>

まず、本発明の実施の形態における負荷分散システム1の全体構成について、図1を参照しながら説明する。図1は、本発明の実施の形態における負荷分散システム1の全体構成の一例を示す図である。

[0017] 図1に示すように、本発明の実施の形態における負荷分散システム1には、負荷監視装置10と、第1のパケット転送装置20と、第2のパケット転送装置30と、統計情報収集装置40と、データベース50とが含まれる。

[0018] 負荷監視装置10は、例えばコンピュータ又はコンピュータシステム等であり、負荷分散対象である第2のパケット転送装置30のトラフィック流量を監視する。そして、負荷監視装置10は、第2のパケット転送装置30のトラフィック流量が所定の閾値を超えた場合（又は、所定の閾値を下回った場合）に、負荷分散の実施（又は負荷分散の解除）を示す命令を第1のパケット転送装置20に送信する。

[0019] 第1のパケット転送装置20は、例えばルータ等であり、ネットワークN1側からのパケットを第2のパケット転送装置30に転送する。このとき、

第1の packets 転送装置 20 は、負荷分散に用いられる所定のテーブルを参照して、複数の第2の packets 転送装置 30 のうち、フロー特性に応じた第2の packets 転送装置 30 に当該 packets を転送する。

[0020] また、第1の packets 転送装置 20 は、負荷監視装置 10 からの命令に応じて、負荷分散に用いられる所定のテーブルを変更する。

[0021] 第2の packets 転送装置 30 は、例えばルータ等であり、第1の packets 転送装置 20 から転送された packets をネットワーク N2 側に転送する。以降では、複数の第2の packets 転送装置 30 の各々を区別する場合は、「第2の packets 転送装置 30-1」、「第2の packets 転送装置 30-2」等と表す。

[0022] 統計情報収集装置 40 は、第2の packets 転送装置 30 からフロー特性を通知するフロー特性通知を受信してデータベース 50 を更新する。データベース 50 は、フロー特性に関する統計情報を記憶する。

[0023] なお、図1に示す負荷分散システム1の構成は一例であって、他の構成であっても良い。例えば、第1の packets 転送装置 20 及び第2の packets 転送装置 30 は、ネットワーク仮想化技術によってコンピュータ上等に構築された仮想的なルータ等であっても良い。

[0024] (実施例1)

以降では、本発明の実施の形態における負荷分散システム1の実施例1について説明する。

[0025] <機能構成>

まず、実施例1における負荷分散システム1の機能構成について、図2を参照しながら説明する。図2は、実施例1における負荷分散システム1の機能構成の一例を示す図である。

[0026] 図2に示すように、実施例1における負荷監視装置10は、流量監視部101と、負荷分散決定部102と、命令送信部103とを有する。

[0027] 流量監視部101は、第2の packets 転送装置 30 の統計情報から、これらの第2の packets 転送装置 30 のトラフィック流量をデータ量 (bps) 及

びパケット数（pps）それぞれの観点で監視する。すなわち、流量監視部 101 は、各第 2 のパケット転送装置 30 について、監視結果として、データ量（bps）及びパケット数（pps）を得る。以降では、監視結果とした得られたデータ量（bps）を「bps 結果値」、監視結果として得られたパケット数（pps）を「pps 結果値」と表す。

[0028] 負荷分散決定部 102 は、負荷分散を実施又は解除するかを決定する。すなわち、負荷分散決定部 102 は、bps 結果値が、予め設定された bps 閾値を超えたか又は下回ったかを判定する。同様に、負荷分散決定部 102 は、pps 結果値が、予め設定された pps 閾値を超えたか又は下回ったかを判定する。

[0029] そして、負荷分散決定部 102 は、bps 結果値が bps 閾値を超えたと判定した場合、負荷分散を実施すると決定する。一方で、負荷分散決定部 102 は、bps 結果値が bps 閾値を下回ったと判定した場合、負荷分散を解除すると決定する。

[0030] 同様に、負荷分散決定部 102 は、pps 結果値が pps 閾値を超えたと判定した場合、負荷分散を実施すると決定する。一方で、負荷分散決定部 102 は、pps 結果値が pps 閾値を下回ったと判定した場合、負荷分散を解除すると決定する。

[0031] 命令送信部 103 は、負荷分散決定部 102 によって負荷分散を実施すると決定された場合又は負荷分散を解除すると判定された場合、所定の命令を第 1 のパケット転送装置 20 に送信する。

[0032] ここで、命令には、「閾値種別」と、「命令種別」と、「対象装置の識別情報」とが含まれる。「閾値種別」は、監視結果が超えた閾値又は監視結果が下回った閾値の種別であり、bps 閾値を示す情報又は pps 閾値を示す情報のいずれかが設定される。「命令種別」は、負荷分散の実施又は解除のいずれであるかを示す情報が設定される。「対象装置の識別情報」は、監視結果が閾値を超えた又は下回った第 2 のパケット転送装置 30 の識別情報であり、例えば、当該第 2 のパケット転送装置 30 の IP アドレスが設定され

る。ただし、「対象装置の識別情報」には、IPアドレス以外にも、例えば、当該第2の packets 転送装置 30 の ID、MAC (Media Access Control) アドレス、ホスト名等が設定されても良い。

[0033] また、図2に示すように、実施例1における第2の packets 転送装置 30 は、統計情報記録部 301 を有する。統計情報記録部 301 は、第2の packets 転送装置 30 がフローの転送を完了すると、当該フローのフロー情報 (5-tuple) とフロー特性を統計情報収集装置 40 にフロー特性通知を用いて通知する。

[0034] 統計情報収集装置 40 は、フロー特性通知に基づいてフローを特定し、特定したフローの特性に関する統計情報がデータベース 50 に登録されていれば当該統計情報を更新し、登録されていない場合は、フロー特性通知に基づいてレコードを作成し、データベース 50 に追加する。

[0035] 図3は、実施例1におけるデータベース 50 の一例を示す図である。図3に示すように、データベース 50 には、レコードとしてフロー特性統計情報が格納されている。また、フロー特性統計情報には、「宛先 IP アドレス」と、「宛先ポート」と、「フロー特性のカウンタ」と、「スコア」とが含まれる。

[0036] 「宛先 IP アドレス」には、パケットの宛先の IP アドレスが設定される。「宛先ポート」には、パケットの出力ポートが設定される。「フロー特性のカウンタ」には、「ショート (A)」と「ロング (B)」の2つのカウンタが含まれる。ショートカウンタには、ショートパケット中心のフロー数が設定され、ロングカウンタには、ロングパケット中心のフロー数が設定される。

[0037] 「スコア」には、ショートカウンタの値 (A) からロングカウンタの値 (B) を引いた差の値 (A - B) を、両カウンタの合計値 (A + B) で割って正規化した値が設定される。「スコア」が正の値であれば「宛先 IP アドレス」と「宛先ポート」で識別されるフローの特性はショートパケット中心であり、「スコア」が負の値であれば「宛先 IP アドレス」と「宛先ポート」

で識別されるフローの特性はロングパケット中心である。

[0038] 図3に示す例では、「宛先IPアドレス」が「10.0.0.10」であり「宛先ポート」が「5004」であるフローについて、ショートパケット中心のフロー数は「1000」であり、ロングパケット中心のフロー数は「0」であり、「スコア」は「1」である。

[0039] 統計情報収集装置40は、フロー特性通知に含まれるフロー特性が「高ppsフロー」である場合には、ショートカウンタに「1」を加え、フロー特性通知に含まれるフロー特性が「高bpsフロー」である場合には、ロングカウンタに「1」を加える。また、統計情報収集装置40は、レコードを追加する場合には、フロー特性通知に含まれるフロー特性に基づいてショートカウンタ又はロングカウンタを「1」に設定し、他のカウンタを「0」に設定する。

[0040] データベース50の検索速度を上げるためには、レコード数を減らすことが重要となる。5tuple全ての組み合わせでレコードを作成するとレコード数が膨大となるため、統計情報収集装置40は、「宛先IPアドレス」と「宛先ポート」の宛先情報のみを用いてレコードを作成する。

[0041] ただし、宛先情報のみでレコードを作成するとレコード数を抑制できるが、5tupleでレコードを作成する場合に比べてフロー特性の判定精度が低下する。例えば、宛先情報は同一でも送信元によってフロー特性が異なる場合にフロー特性の判定精度が低下する。そこで、統計情報収集装置40は、レコード作成のキーを宛先情報に限定することでレコード数を抑制しつつ、フロー特性の判定精度が高いレコードを優先的にデータベース50に残すようにレコードの置換を行う。

[0042] すなわち、統計情報収集装置40は、データベース50が一杯の状態でレコードの置換が必要な場合に、削除するレコードを「スコア」の絶対値 a に基づいて選択する。統計情報収集装置40は、まず $a=0$ であるレコードを選択する。そして、 $a=0$ であるレコードがない場合には、統計情報収集装置40は、レコードが最後に参照されてからの経過時間に w/a を加算した

後、直近で最も使用されなかったレコードから削除していくLRU (Least Recently Used) に基づいてレコードを選択する。ここで、 w は任意の正数である。

[0043] このように、統計情報収集装置40は、 a が小さいレコード、すなわちフロー特性に偏りが少なくフロー特性の判定精度が低いレコードがLRUで不利に評価されるように経過時間を a に基づいて変更してLRUを適用する。このため、統計情報収集装置40は、フロー特性の判定精度が高いレコードがデータベース50に残るようにすることができる。

[0044] なお、統計情報収集装置40及びデータベース50は、例えば、第1の packets 転送装置20に含まれても良い。また、データベース50は、例えば、RAM (Random Access Memory) や補助記憶装置、専用のメモリ装置等により実現される。

[0045] また、図2に示すように、実施例1における第1の packets 転送装置20は、命令受信部201と、経路テーブル変更部202と、入力ポート203と、フロー情報取得部204と、アプリ判定部205と、経路決定部206と、ルーティング部207と、出力ポート208と、入力バッファ209とを有する。これら各機能部は、例えば、第1の packets 転送装置20にインストールされた1以上のプログラムがCPU等を実行させる処理により実現される。

[0046] 命令受信部201は、負荷監視装置10からの命令を受信する。経路テーブル変更部202は、命令受信部201が受信した命令に応じて、ECMP 経路テーブル1000を変更する。ECMP 経路テーブル1000は、エントリとして経路情報を格納しているテーブルである。なお、ECMP 経路テーブル1000は、例えば、補助記憶装置や専用のメモリ装置等に格納されている。実施例1におけるECMP 経路テーブル1000の詳細については後述する。

[0047] 入力ポート203は、ネットワークN1側からの packets を受信する。そして、入力ポート203は、受信した packets を入力バッファ209に格納

する。なお、入力バッファ209は、受信したパケットを一時的に格納しておくための記憶領域であり、例えばRAMや補助記憶装置、専用のメモリ装置等により実現される。

[0048] フロー情報取得部204は、入力バッファ209に格納されているパケットのフロー情報（例えば、5-tuple）からハッシュ値を生成する。また、フロー情報取得部204は、フローテーブル2000を参照して、生成したハッシュ値と一致するハッシュ値が含まれるエントリが存在するか否かを判定する。フローテーブル2000は、エントリとして、ハッシュ値と、出力ポート208を示す情報（出力ポート情報）とが対応付けられたフロー転送情報を格納しているテーブルである。出力ポート情報は、複数の出力ポート208のうちのいずれか1つの出力ポート208を特定するための情報である。なお、フローテーブル2000は、例えば、RAMや補助記憶装置、専用のメモリ装置等に格納されている。

[0049] そして、フロー情報取得部204は、該当のエントリが存在すると判定した場合、当該エントリをルーティング部207に通知する。一方で、フロー情報取得部204は、該当のエントリが存在しないと判定した場合、当該パケットを新規フローのパケットと判断して、該当のエントリが存在しない旨をアプリ判定部205に通知する。このとき、フロー情報取得部204は、例えば、5-tuple又は当該パケットのL7情報（例えば、L7のヘッダ情報等）とハッシュ値とをアプリ判定部205に通知する。

[0050] なお、フローテーブル2000に格納されているエントリ（フロー転送情報）は、例えば、経路決定部206等によって所定の時間毎に削除（クリア）されても良い。

[0051] アプリ判定部205は、フロー情報取得部204からの通知に応じて、宛先情報をキーとしてデータベース50を検索し、新規フローのフロー特性を特定する。アプリ判定部205は、「スコア」が正の値であれば新規フローをショートパケット中心のフローと特定し、「スコア」が負の値であれば新規フローをロングパケット中心のフローと特定する。そして、アプリ判定部

205は、特定したフロー特性を経路決定部206に通知する。このとき、アプリ判定部205は、例えば、5-tupleと、フロー特性と、ハッシュ値とを経路決定部206に通知する。

[0052] 経路決定部206は、ECMP経路テーブル1000を参照して、アプリ判定部205が特定したフロー特性から出力ポート208を決定する。また、経路決定部206は、決定した出力ポート208を示す情報と、ハッシュ値とを対応付けたフロー転送情報をフローテーブル2000に格納する。そして、経路決定部206は、フローテーブル2000に格納したエントリ（フロー転送情報）をルーティング部207に通知する。

[0053] ルーティング部207は、フロー情報取得部204又は経路決定部206から通知されたエントリ（フロー転送情報）に含まれる出力ポート情報を取得する。そして、ルーティング部207は、取得した出力ポート情報が示す出力ポート208に当該パケットを送信する。

[0054] 出力ポート208は、ルーティング部207から受け取ったパケットを出力する。ここで、出力ポート208は、例えば、第2のパケット転送装置30毎に存在する。例えば、出力ポート208には、第2のパケット転送装置30-1に対応する出力ポート208-1、第2のパケット転送装置30-2に対応する出力ポート208-2等がある。なお、出力ポート208は、物理ポートであっても良いし、論理ポートであっても良い。

[0055] ここで、実施例1におけるECMP経路テーブル1000について、図4を参照しながら説明する。図4は、実施例1におけるECMP経路テーブル1000の一例を示す図である。

[0056] 図4に示すように、実施例1におけるECMP経路テーブル1000には、エントリとして経路情報が格納されている。また、経路情報には、「宛先ネットワーク」と、「サブネットマスク」と、「ゲートウェイのリスト」と、「出力ポート情報」と、「データ量（bps）の負荷状態」と、「パケット数（pps）の負荷状態」とが含まれる。

[0057] 「宛先ネットワーク」には、パケットの転送先のネットワークのIPアド

レス又はdefaultが設定される。「サブネットマスク」には、宛先ネットワークに対応するサブネットマスクが設定される。

[0058] 「ゲートウェイのリスト」には、宛先ネットワークに対応するゲートウェイ（第2の packets 転送装置 30）のIPアドレスが設定される。なお、「ゲートウェイのリスト」には、1つの宛先ネットワークに対して、1つ以上のゲートウェイのIPアドレスが設定される。

[0059] 「出力ポート情報」には、「ゲートウェイのリスト」に設定されている各ゲートウェイ（のIPアドレス）に対する出力ポート 208 を示す情報がそれぞれ設定される。

[0060] 図4に示す例では、ゲートウェイ「192.168.1.0」に対して、出力ポート 208-1 を示す「eth0」が対応付けられている。これは、ゲートウェイ「192.168.1.0」から packets を送信する場合、出力ポート 208-1 から送信されることを示している。

[0061] 同様に、例えば、ゲートウェイ「192.168.1.1」に対して、出力ポート 208-2 を示す「eth1」が対応付けられている。これは、ゲートウェイ「192.168.1.1」から packets を送信する場合、出力ポート 208-2 から送信されることを示している。他のゲートウェイについても同様である。

[0062] 「データ量 (bps) の負荷状態」には、「ゲートウェイのリスト」に設定されている各ゲートウェイ（のIPアドレス）に対して、当該ゲートウェイにおけるデータ量 (bps) の負荷状態が設定される。ここで、図4に示す例では、「データ量 (bps) の負荷状態」に「0」が設定されている場合、該当のゲートウェイにおけるデータ量 (bps) の負荷状態が「低負荷」であることを示している。一方で、「データ量 (bps) の負荷状態」に「1」が設定されている場合、該当のゲートウェイにおけるデータ量 (bps) の負荷状態が「高負荷」であることを示している。

[0063] 「パケット数 (pps) の負荷状態」には、「ゲートウェイのリスト」に設定されている各ゲートウェイ（のIPアドレス）に対して、当該ゲートウ

エイにおけるパケット数（pps）の負荷状態が設定される。ここで、図4に示す例では、「パケット数（pps）の負荷状態」に「0」が設定されている場合、該当のゲートウェイにおけるパケット数（pps）の負荷状態が「低負荷」であることを示している。一方で、「パケット数（pps）の負荷状態」に「1」が設定されている場合、該当のゲートウェイにおけるパケット数（pps）の負荷状態が「高負荷」であることを示している。

[0064] このように、実施例1におけるECMP経路テーブル1000には、宛先ネットワーク毎に、当該宛先ネットワークに対する1以上のゲートウェイと、これらの各ゲートウェイの出力ポート情報及び負荷状態とが含まれる経路情報が格納されている。

[0065] 後述するように、実施例1では、各ゲートウェイ（第2の packets 転送装置30）の負荷状態が閾値を超えた場合（又は下回った場合）に、当該ゲートウェイに対応する負荷状態を変更すると共に、パケットを転送する際に、フロー特性に応じて低負荷状態のゲートウェイからパケットを送信することで、負荷分散が行われる。

[0066] <ECMP経路テーブルの変更処理>

以降では、実施例1におけるECMP経路テーブル1000の変更処理について、図5を参照しながら説明する。図5は、実施例1におけるECMP経路テーブル1000の変更処理の流れの一例を説明するための図である。

[0067] まず、負荷監視装置10の流量監視部101は、第2の packets 転送装置30の統計情報から、これらの第2の packets 転送装置30のbps結果値及びpps結果値を取得する（ステップS101）。ここで、流量監視部101は、例えば、所定の時間毎に、各第2の packets 転送装置30のbps結果値及びpps結果値を取得する。なお、統計情報は、例えば、netstat等により取得することができる。

[0068] 次に、負荷監視装置10の負荷分散決定部102は、負荷分散を実施又は解除するかを決定する（ステップS102）。すなわち、負荷分散決定部102は、bps結果値がbps閾値を超えたか又は下回ったかを判定する。

同様に、負荷分散決定部 102 は、pps 結果値が pps 閾値を超えたか又は下回ったかを判定する。なお、bps 閾値及び pps 閾値は、全ての第 2 のパケット転送装置 30 で同一の値が決められても良いし、第 2 のパケット転送装置 30 毎に値が決められても良い。又は、例えば、1 以上の第 2 のパケット転送装置 30 を所定の基準でグループ化した上で、グループ毎に値が決められても良い。

[0069] そして、負荷分散決定部 102 は、bps 結果値が bps 閾値を超えたと判定した場合、負荷分散を実施すると決定する。一方で、負荷分散決定部 102 は、bps 結果値が bps 閾値を下回ったと判定した場合、負荷分散を解除すると決定する。

[0070] 同様に、負荷分散決定部 102 は、pps 結果値が pps 閾値を超えたと判定した場合、負荷分散を実施すると決定する。一方で、負荷分散決定部 102 は、pps 結果値が pps 閾値を下回ったと判定した場合、負荷分散を解除すると決定する。

[0071] ステップ S102 において、負荷分散の実施及び負荷分散の解除のいずれも決定されなかった場合、負荷監視装置 10 は、上記のステップ S101 に戻る。

[0072] 一方で、ステップ S102 において、負荷分散の実施又は負荷分散の解除の少なくとも一方が決定された場合、負荷監視装置 10 の命令送信部 103 は、所定の命令を第 1 のパケット転送装置 20 に送信する（ステップ S103）。

[0073] 第 1 のパケット転送装置 20 の命令受信部 201 は、負荷監視装置 10 からの命令を受信する（ステップ S104）。

[0074] 次に、第 1 のパケット転送装置 20 の経路テーブル変更部 202 は、命令受信部 201 が受信した命令に応じて、ECMP 経路テーブル 1000 を変更する（ステップ S105）。

[0075] 具体的には、例えば、「閾値種別」に「bps 閾値」、「命令種別」に「負荷分散の実施」、「対象装置の識別情報」に「192.168.1.0」

がそれぞれ設定されている通知を受信した場合、経路テーブル変更部202は、ECMP経路テーブル1000の各エントリについて、IPアドレスが「192.168.1.0」であるゲートウェイに対する「データ量(bps)の負荷状態」を「1」に変更する。

[0076] 同様に、例えば、「閾値種別」に「bps閾値」、「命令種別」に「負荷分散の解除」、「対象装置の識別情報」に「192.168.1.2」がそれぞれ設定されている通知を受信した場合、経路テーブル変更部202は、ECMP経路テーブル1000の各エントリについて、IPアドレスが「192.168.1.2」であるゲートウェイに対する「データ量(bps)の負荷状態」を「0」に変更する。

[0077] 同様に、例えば、「閾値種別」に「pps閾値」、「命令種別」に「負荷分散の実施」、「対象装置の識別情報」に「192.168.1.1」がそれぞれ設定されている通知を受信した場合、経路テーブル変更部202は、ECMP経路テーブル1000の各エントリについて、IPアドレスが「192.168.1.1」であるゲートウェイに対する「パケット数(pps)の負荷状態」を「1」に変更する。

[0078] 同様に、例えば、「閾値種別」に「pps閾値」、「命令種別」に「負荷分散の解除」、「対象装置の識別情報」に「192.168.1.0」がそれぞれ設定されている通知を受信した場合、経路テーブル変更部202は、ECMP経路テーブル1000の各エントリについて、IPアドレスが「192.168.1.0」であるゲートウェイに対する「パケット数(pps)の負荷状態」を「0」に変更する。

[0079] 以上のように、実施例1における負荷分散システム1は、負荷分散対象である第2のパケット転送装置30のbps結果値又はpps結果値が閾値を超えた場合又は下回った場合、ECMP経路テーブル1000に格納されている各エントリにおいて、これらの第2のパケット転送装置30の負荷状態を変更する。これにより、ECMP経路テーブル1000に格納されている各エントリによって、各第2のパケット転送装置30の負荷状態(データ量

(bps)の負荷状態及びパケット数(pps)の負荷状態)が管理される。

[0080] <パケット転送処理>

以降では、実施例1におけるパケット転送処理について、図6を参照しながら説明する。図6は、実施例1におけるパケット転送処理の流れの一例を説明するための図である。

[0081] まず、第1のパケット転送装置20の入力ポート203は、ネットワークN1側からのパケットを受信する(ステップS201)。

[0082] 次に、第1のパケット転送装置20の入力ポート203は、上記のステップS201で受信したパケットを入力バッファ209に格納する(ステップS202)。

[0083] 次に、第1のパケット転送装置20のフロー情報取得部204は、入力バッファ209に格納されているパケットのフロー情報(例えば、5-tuple)からハッシュ値を生成する(ステップS203)。なお、ハッシュ値の生成に用いられるハッシュ関数は任意のものを用いることができる。

[0084] 次に、第1のパケット転送装置20のフロー情報取得部204は、上記のステップS203で生成したハッシュ値と一致するハッシュ値が含まれるエントリ(フロー転送情報)がフローテーブル2000に格納されているか否かを判定する(ステップS204)。

[0085] ステップS204において、当該エントリがフローテーブル2000に格納されていると判定された場合、第1のパケット転送装置20は、後述するステップS208の処理に進む。

[0086] 一方で、ステップS204において、当該エントリがフローテーブル2000に格納されていないと判定された場合、第1のパケット転送装置20のアプリ判定部205は、データベース50を用いて、当該パケットのフロー特性を特定する(ステップS205)。すなわち、アプリ判定部205は、当該パケットによるフローのフロー特性が「高bpsフロー」又は「高ppsフロー」のいずれであるかを特定する。アプリ判定部205は、当該パケ

ットによるフローに対応する「スコア」が正であるか負であるかに基づいてフロー特性を特定することができる。

[0087] 次に、第1の packets 転送装置 20 の経路決定部 206 は、ECMP 経路テーブル 1000 を参照して、上記のステップ S 205 で特定したフロー特性から出力ポート 208 を決定する（ステップ S 206）。経路決定部 206 は、例えば、次の S 1 ～ S 3 の手順により出力ポート 208 を決定する。

[0088] S 1) まず、経路決定部 206 は、アプリ判定部 205 から通知された 5-tuple に含まれる宛先 IP アドレスのネットワークアドレスに対応するエントリを ECMP 経路テーブル 1000 から検索する。このとき、経路決定部 206 は、例えば、ロングストマッチでエントリを検索する。当該ネットワークアドレスに対応するエントリが検索された場合、経路決定部 206 は、このエントリを参照する。一方で、当該ネットワークアドレスに対応するエントリが検索されなかった場合、経路決定部 206 は、宛先ネットワークに「default」が設定されているエントリを参照する。

[0089] S 2) 次に、経路決定部 206 は、フロー特性に応じて、参照しているエントリに含まれる「ゲートウェイのリスト」の中から低負荷状態のゲートウェイを検索する。

[0090] 例えば、上記のステップ S 205 で特定したフロー特性が「高 bps フロー」である場合、経路決定部 206 は、当該エントリに含まれる「ゲートウェイのリスト」のうち、「データ量 (bps) の負荷状態」に「0」が設定されているゲートウェイを上から順番に検索する。

[0091] 一方で、例えば、上記のステップ S 205 で特定したフロー特性が「高pps フロー」である場合、経路決定部 206 は、当該エントリに含まれる「ゲートウェイのリスト」のうち、「パケット数 (pps) の負荷状態」に「0」が設定されているゲートウェイを上から順番に検索する。

[0092] S 3) そして、経路決定部 206 は、上記の S 2 で検索されたゲートウェイに対応する出力ポート情報が示す出力ポート 208 を、当該パケットの出力ポート 208 に決定する。

[0093] なお、上記のS 2でゲートウェイが検索されなかった場合、経路決定部206は、高負荷状態のゲートウェイに対応する出力ポート情報が示す出力ポート208を、当該パケットの出力ポート208に決定する。例えば、上記のステップS 205で特定したフロー特性が「高bpsフロー」で、「データ量(bps)の負荷状態」に「0」が設定されているゲートウェイが存在しない場合、経路決定部206は、「データ量(bps)の負荷状態」に「1」が設定されているゲートウェイに対応する出力ポート情報が示す出力ポート208を、当該パケットの出力ポート208に決定する。同様に、例えば、上記のステップS 205で特定したフロー特性が「高ppsフロー」で、「パケット数(pps)の負荷状態」に「0」が設定されているゲートウェイが存在しない場合、経路決定部206は、「パケット数(pps)の負荷状態」に「1」が設定されているゲートウェイに対応する出力ポート情報が示す出力ポート208を、当該パケットの出力ポート208に決定する。

[0094] ここで、同一の宛先ネットワークに対する出力ポート208をラウンドロビンで変更したい場合は、経路決定部206は、決定した出力ポート208に対応するゲートウェイを、「ゲートウェイのリスト」の最後尾に移動させても良い。例えば、図4に示すECMP経路テーブル1000において、宛先ネットワーク「default」のゲートウェイ「192.168.1.0」の出力ポート情報「eth0」が出力ポート208に決定された場合、このゲートウェイ「192.168.1.0」をリストの最後尾（すなわち、ゲートウェイ「192.168.1.2」の後ろ）に移動させても良い。これにより、同一の宛先ネットワークに対する出力ポート208をラウンドロビンで変更することで、更なる負荷分散が期待できる。

[0095] 次に、第1のパケット転送装置20の経路決定部206は、アプリ判定部205から通知されたハッシュ値と、上記のステップS 206で決定した出力ポート208を示す情報（出力ポート情報）とを対応付けたフロー転送情報をフローテーブル2000に格納する（ステップS 207）。

[0096] 第1のパケット転送装置20のルーティング部207は、該当のエントリ

(上記のステップS 2 0 4 でハッシュ値が一致すると判定されたエントリ又は上記のステップS 2 0 7 で格納されたエントリ) から出力ポート情報を取得する(ステップS 2 0 8)。そして、ルーティング部2 0 7は、取得した出力ポート情報が示す出力ポート2 0 8に当該パケットを送信する。

[0097] 第1のパケット転送装置2 0の出力ポート2 0 8は、ルーティング部2 0 7から受け取ったパケットを送信する(ステップS 2 0 9)。

[0098] <フロー特性の登録処理>

以降では、実施例1におけるフロー特性の登録処理について、図7を参照しながら説明する。図7は、実施例1におけるフロー特性の登録処理の流れの一例を説明するための図である。

[0099] まず、統計情報収集装置4 0は、第2のパケット転送装置3 0からフロー特性通知を受信する(ステップS 3 0 1)。

[0100] 次に、統計情報収集装置4 0は、フロー特性通知に含まれる宛先情報で特定されるフローがデータベース5 0に登録されているか否かを判定する(ステップS 3 0 2)。

[0101] ステップS 3 0 2において、当該フローがデータベース5 0に登録されていると判定した場合、統計情報収集装置4 0は、フロー特性通知に含まれるフロー特性に基づいて、当該フローのレコードの「フロー特性のカウンタ」に「1」を加算する(ステップS 3 0 3)。統計情報収集装置4 0は、フロー特性通知に含まれるフロー特性が「高p p sフロー」である場合にはショートカウンタに「1」を加算し、フロー特性通知に含まれるフロー特性が「高b p sフロー」である場合にはロングカウンタに「1」を加算する。

[0102] 一方、ステップS 3 0 2において、当該フローがデータベース5 0に登録されていないと判定した場合、統計情報収集装置4 0は、当該フローのレコードを作成してデータベース5 0に登録する新規登録処理を行う(ステップS 3 0 4)。

[0103] 次に、統計情報収集装置4 0は、当該フローのレコードの「スコア」を更新する(ステップS 3 0 5)。

- [0104] 以上のように、統計情報収集装置40は、フロー特性通知に含まれる宛先情報で特定されるフローがデータベース50に登録されている場合には「フロー特性のカウンタ」に「1」を加算し、登録されていない場合にはレコードを作成してデータベース50に登録する。これにより、第1の packets 転送装置20は、受信したパケットのフロー特性をデータベース50を参照して特定することができる。
- [0105] 図8は、新規登録処理の流れの一例を説明するための図である。統計情報収集装置40は、フロー特性通知に含まれる宛先情報とフロー特性を用いてレコードを作成する（ステップS401）。統計情報収集装置40は、フロー特性通知に含まれるフロー特性が「高ppsフロー」である場合にはショートカウンタに「1」を設定してロングカウンタに「0」を設定し、フロー特性通知に含まれるフロー特性が「高bpsフロー」である場合にはロングカウンタに「1」を設定してショートカウンタに「0」を設定する。
- [0106] 次に、統計情報収集装置40は、レコードの追加が可能であるか否かを判定する（ステップS402）。
- [0107] ステップS402において、レコードの追加が可能である場合、統計情報収集装置40は、データベース50に作成したレコードを追加する（ステップS403）。
- [0108] 一方、ステップS402において、レコードの追加が不可能でない場合、統計情報収集装置40は、「スコア」が「0」のレコードがあるか否かを判定する（ステップS404）。
- [0109] ステップS404において、「スコア」が「0」のレコードがある場合、統計情報収集装置40は、「スコア」が「0」のレコードを、作成したレコードで置換する（ステップS405）。
- [0110] 一方、ステップS404において、「スコア」が「0」のレコードがない場合、統計情報収集装置40は、全てのレコードについて、「スコア」に基づく値 $b = w / a$ を経過時間に加算する（ステップS406）。そして、統計情報収集装置40は、bが加算された経過時間が最大のレコードを特定し

(ステップS407)、特定したレコードを、作成したレコードで置換する(ステップS408)。

[0111] 以上のように、実施例1における負荷分散システム1は、第1の packets 転送装置20から第2の packets 転送装置30に packets を転送する際に、当該 packets によるフローのフロー特性に応じて、低負荷状態の第2の packets 転送装置30に当該 packets を転送する。これにより、実施例1における負荷分散システム1では、第2の packets 転送装置30の負荷分散を実現することができる。このため、例えば、或る第2の packets 転送装置30に高bpsフローの packets が集中する事態や高ppsフローの packets が集中する事態を防止することができ、第2の packets 転送装置30の転送容量を最大限に活かすことができるようになる。

[0112] また、実施例1における負荷分散システム1は、フロー特性統計情報をデータベース50に記憶する。そして、第1の packets 転送装置20は、 packets を受信すると、受信した packets が構成するフローの特性をデータベース50を参照して判定する。このため、第1の packets 転送装置20は、フロー特性を効率良く正確に判定することができる。

[0113] また、統計情報収集装置40は第2の packets 転送装置30からフロー特性通知を受信する毎にデータベース50を更新するため、データベース50は、フロー特性統計情報を記憶することができる。

[0114] また、データベース50は5-tupleではなく宛先情報に対応付けて「フロー特性のカウンタ」及び「スコア」をフロー毎に記憶するため、第1の packets 転送装置20は、フロー特性を効率良く判定することができる。

[0115] また、統計情報収集装置40は、データベース50のレコードを置換するときに「スコア」が「0」であるレコードを削除対象とし、「スコア」が「0」であるレコードがない場合には、経過時間に w/a を加算した値を用いてLRUにより削除対象を選択する。このため、第1の packets 転送装置20は、レコード数を抑えつつフロー特性を正確に判定することができる。

[0116] (実施例2)

時間経過によってパケットを送受信するサーバの用途が変化すると、宛先情報が同一でもフロー特性が変化する。このような状況に対処するため、実施例2では、データベース50は、「フロー特性のカウンタ」を複数の期間に分けて記憶する。

[0117] 図9は、実施例2におけるデータベース50の一例を示す図である。図9に示すように、「フロー特性のカウンタ」は、0～1ヶ月前の情報をカウントするカウンタ、1～2ヶ月前の情報をカウントするカウンタ、及び2～3ヶ月前の情報をカウントするカウンタに分けられる。図9では1ヶ月であるが、期間は、2週間、2ヶ月など他の単位でも良い。また、「フロー特性のカウンタ」は、期間毎に「スコア」の計算に用いられる重み係数を有する。

[0118] 0～1ヶ月前のショートカウンタの値(A)とロングカウンタの値(B)と重み係数(C)、1～2ヶ月前のショートカウンタの値(D)とロングカウンタの値(E)と重み係数(F)、及び2～3ヶ月前のショートカウンタの値(G)とロングカウンタの値(H)と重み係数(I)を用いて「スコア」は以下のように計算される。

[0119] 「スコア」= { (A-B) × C + (D-E) × F + (G-H) × I } ÷ { (A+B) × C + (D+E) × F + (G+H) × I }

[0120] なお、図9では、C=10、F=5、I=1として「スコア」が計算されている。例えば、「宛先IPアドレス」が「20.0.0.15」であり「宛先ポート」が「443」であるフローの「スコア」は以下のように計算される。

[0121] 「スコア」= { (1000-0) × 10 + (0-1000) × 5 + (0-1000) × 1 } ÷ { (1000+0) × 10 + (0+1000) × 5 + (0+1000) × 1 } = { 10000-5000-1000 } ÷ { 10000+5000+1000 } = 0.25

[0122] 以上のように、実施例2における負荷分散システム1では、データベース50はショートカウンタとロングカウンタの組を期間毎に複数有し、統計情報収集装置40は「スコア」を各組のショートカウンタとロングカウンタの

差及び重み係数に基づいて計算する。このため、第１の packets 転送装置 20 は、フロー特性が変化する場合にもフロー特性を正確に判定することができる。

[0123] <ハードウェア構成>

最後に、本発明の実施の形態における負荷監視装置 10、第１の packets 転送装置 20、第２の packets 転送装置 30 及び統計情報収集装置 40 のハードウェア構成について説明する。本発明の実施の形態における負荷監視装置 10、第１の packets 転送装置 20、第２の packets 転送装置 30 及び統計情報収集装置 40 は、例えば図 10 に示すコンピュータ 500 を 1 台以上用いることにより実現される。図 10 は、コンピュータ 500 のハードウェア構成の一例を示す図である。

[0124] 図 10 に示すコンピュータ 500 は、CPU (Central Processing Unit) 501 と、RAM 502 と、ROM (Read Only Memory) 503 と、外部 I/F 504 と、通信 I/F 505 と、補助記憶装置 506 とを有する。これら各ハードウェアは、それぞれがバス B を介して通信可能に接続されている。

[0125] CPU 501 は、ROM 503 や補助記憶装置 506 等からプログラムやデータを RAM 502 上に読み出して処理を実行する演算装置である。

[0126] RAM 502 は、プログラムやデータを一時保持する揮発性の半導体メモリである。ROM 503 は、電源を切ってもプログラムやデータを保持することができる不揮発性の半導体メモリである。ROM 503 には、例えば、OS 設定やネットワーク設定等が格納されている。

[0127] 外部 I/F 504 は、外部装置とのインタフェースである。外部装置には、記録媒体 504a 等がある。コンピュータ 500 は、外部 I/F 504 を介して、記録媒体 504a 等の読み取りや書き込み等を行うことができる。

[0128] 記録媒体 504a には、例えば、フレキシブルディスク、CD (Compact Disc)、DVD (Digital Versatile Disk)、SD メモリカード (Secure Digital memory card)、USB (Universal Serial Bus) メモリカー

ド等がある。

[0129] 通信 I / F 5 0 5 は、他の装置と通信を行うためのインタフェースである。なお、第 1 のパケット転送装置 2 0 及び第 2 のパケット転送装置 3 0 は、複数の通信 I / F 5 0 5 を有する。

[0130] 補助記憶装置 5 0 6 は、例えば H D D (Hard Disk Drive) や S S D (Solid State Drive) 等であり、プログラムやデータを格納している不揮発性の記憶装置である。補助記憶装置 5 0 6 に格納されているプログラムやデータには、例えば、O S、当該 O S 上において各種機能を実現するアプリケーションプログラム等がある。

[0131] 本発明の実施の形態における負荷監視装置 1 0、第 1 のパケット転送装置 2 0、第 2 のパケット転送装置 3 0 及び統計情報収集装置 4 0 は、図 1 0 に示すコンピュータ 5 0 0 を 1 台以上用いることにより、上述した各種処理を実行することができる。例えば、負荷監視装置 1 0 が有する各機能部は、補助記憶装置 5 0 6 に格納された 1 以上のプログラムが C P U 5 0 1 等に行わせる処理により実現される。また、例えば、第 1 のパケット転送装置 2 0 が有する各機能部は、補助記憶装置 5 0 6 に格納された 1 以上のプログラムが C P U 5 0 1 等に行わせる処理や通信 I / F 5 0 5 等により実現される。

[0132] なお、負荷監視装置 1 0 は、例えばディスプレイ等の表示装置と、例えばキーボードやマウス等の入力装置とのうちの少なくとも一方を有していても良い。

[0133] 本発明は、具体的に開示された上記の実施の形態に限定されるものではなく、特許請求の範囲から逸脱することなく、種々の変形や変更が可能である。例えば、上記の実施の形態では、負荷分散を行うパケット転送装置を第 1 のパケット転送装置 2 0、負荷分散の対象となるパケット転送装置を第 2 のパケット転送装置 3 0 としたが、これは、或るパケット転送装置が常に第 1 のパケット転送装置 2 0 又は第 2 のパケット転送装置 3 0 のいずれかとして機能することを意味するものではない。例えば、同一のパケット転送装置が

第1の packets 転送装置 20 として機能し、かつ、第2の packets 転送装置 30 としても機能しても良い。

符号の説明

[0134]	1	負荷分散システム
	10	負荷監視装置
	20	第1の packets 転送装置
	30	第2の packets 転送装置
	40	統計情報収集装置
	50	データベース
	101	流量監視部
	102	負荷分散決定部
	103	命令送信部
	201	命令受信部
	202	経路テーブル変更部
	203	入力ポート
	204	フロー情報取得部
	205	アプリ判定部
	206	経路決定部
	207	ルーティング部
	208	出力ポート
	209	入力バッファ
	301	統計情報記録部
	N1、N2	ネットワーク

請求の範囲

- [請求項1] パケットにより構成されるフローがショートパケットが多い傾向のフロー又はロングパケットが多い傾向のフローのいずれであることを示すフロー特性に関する情報を該パケットの宛先情報に対応付けて記憶する記憶手段と、
- 入力ポートがパケットを受信すると、該パケットにより構成されるフローのフロー特性を前記記憶手段に基づいて判定する判定手段と、
- 前記判定手段により判定されたフロー特性に応じて、複数のパケット転送装置のうち、前記フロー特性のフローによる負荷が少ないことを示す負荷状態のパケット転送装置に対して前記入力ポートが受信したパケットを出力する出力手段と、
- を有することを特徴とする負荷分散システム。
- [請求項2] 前記パケット転送装置がフローの転送を完了すると、該フローの宛先情報とフロー特性を前記パケット転送装置から受信して前記記憶手段を更新する更新手段
- を有することを特徴とする請求項1に記載の負荷分散システム。
- [請求項3] 前記フロー特性に関する情報は、ショートパケットが多い傾向のフローの数を示すショートカウンタと、ロングパケットが多い傾向のフローの数を示すロングカウンタと、前記ショートカウンタとロングカウンタの差に基づくスコアを含み、
- 前記判定手段は、前記スコアに基づいて前記フロー特性を判定することを特徴とする請求項2に記載の負荷分散システム。
- [請求項4] 前記更新手段は、前記記憶手段のエントリを置換するとき、前記スコアが0であるエントリがある場合は、前記スコアが0であるエントリを置換されるエントリとして選択し、前記スコアが0であるエントリがない場合は、各エントリが前記記憶手段に登録されてからの経過時間に前記スコアの逆数に基づく第1の値を加えた第2の値を用いてLRUに従って、置換されるエントリを選択することを特徴とする

請求項3に記載の負荷分散システム。

[請求項5] 前記フロー特性に関する情報は、前記ショートカウンタとロングカウンタの組を期間毎に複数含み、前記スコアは、各組のショートカウンタとロングカウンタの差に基づく値であることを特徴とする請求項3又は4に記載の負荷分散システム。

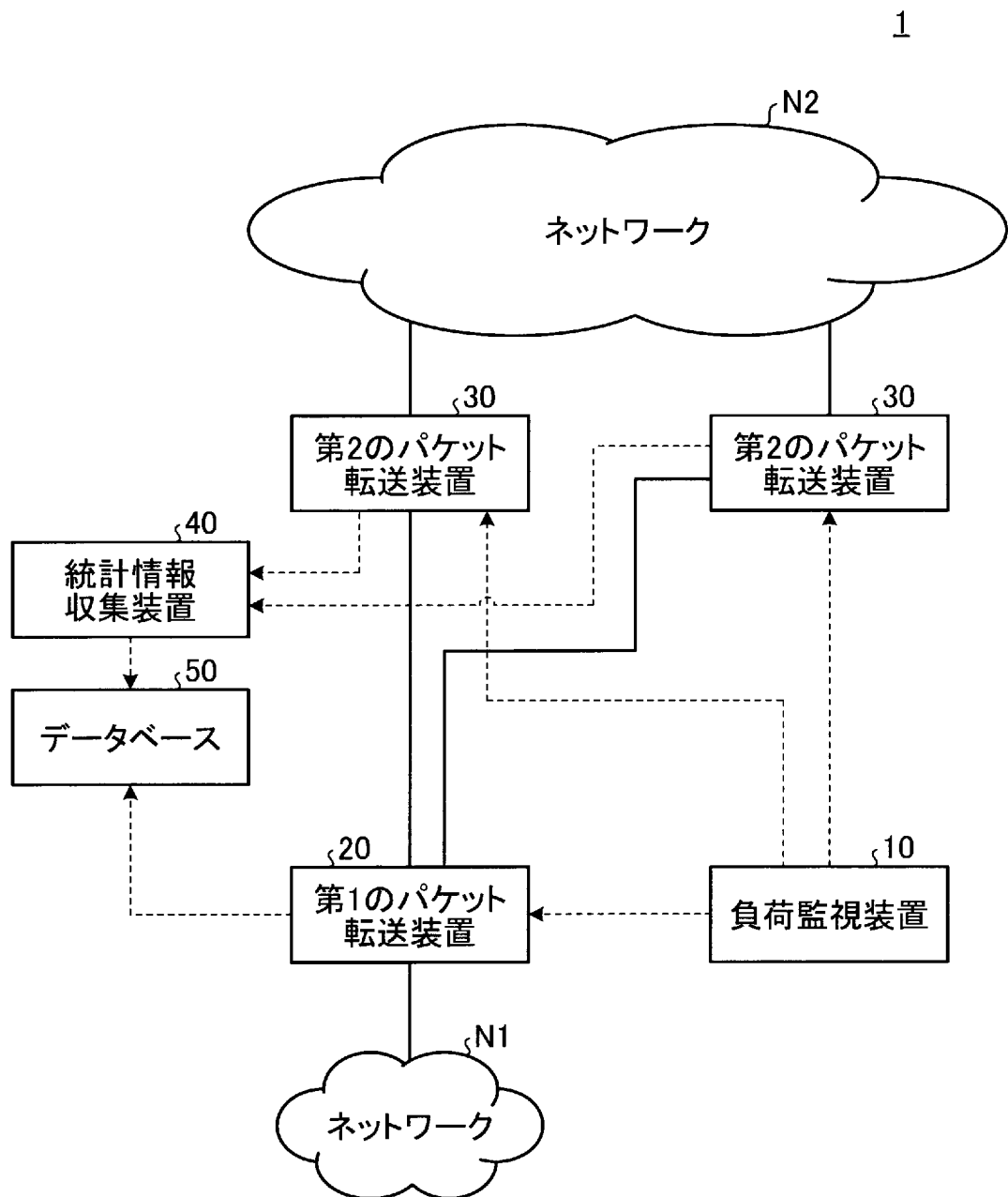
[請求項6] パケットにより構成されるフローがショートパケットが多い傾向のフロー又はロングパケットが多い傾向のフローのいずれであることを示すフロー特性に関する情報を該パケットの宛先情報に対応付けて記憶手段に記憶する手順と、

入力ポートがパケットを受信すると、該パケットにより構成されるフローのフロー特性を前記記憶手段に基づいて判定する判定手順と、

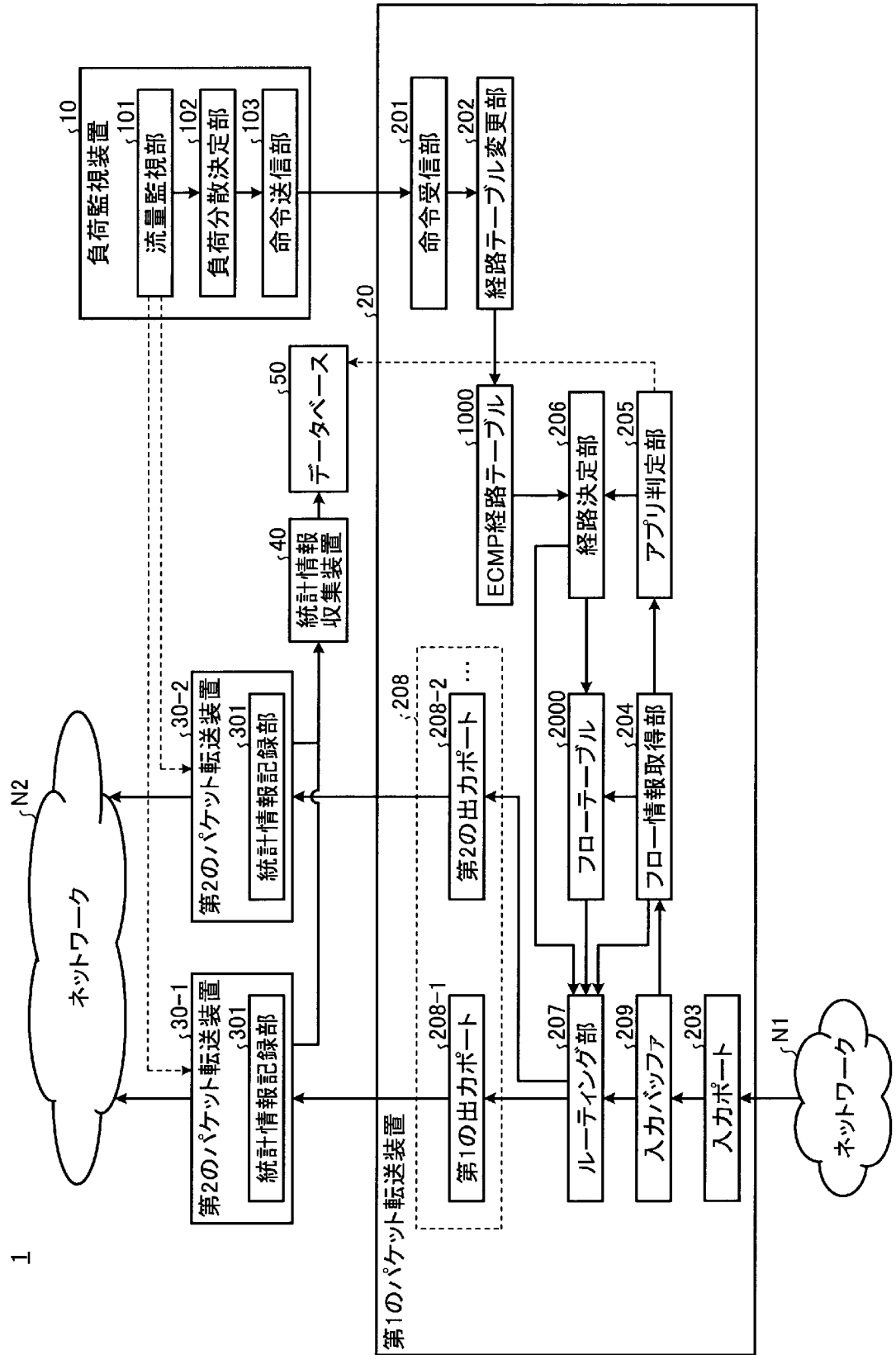
前記判定手順により判定されたフロー特性に応じて、複数のパケット転送装置のうち、前記フロー特性のフローによる負荷が少ないことを示す負荷状態のパケット転送装置に対して前記入力ポートが受信したパケットを出力する出力手順と、

を負荷分散システムが実行することを特徴とする負荷分散方法。

[図1]



[図2]



[図3]

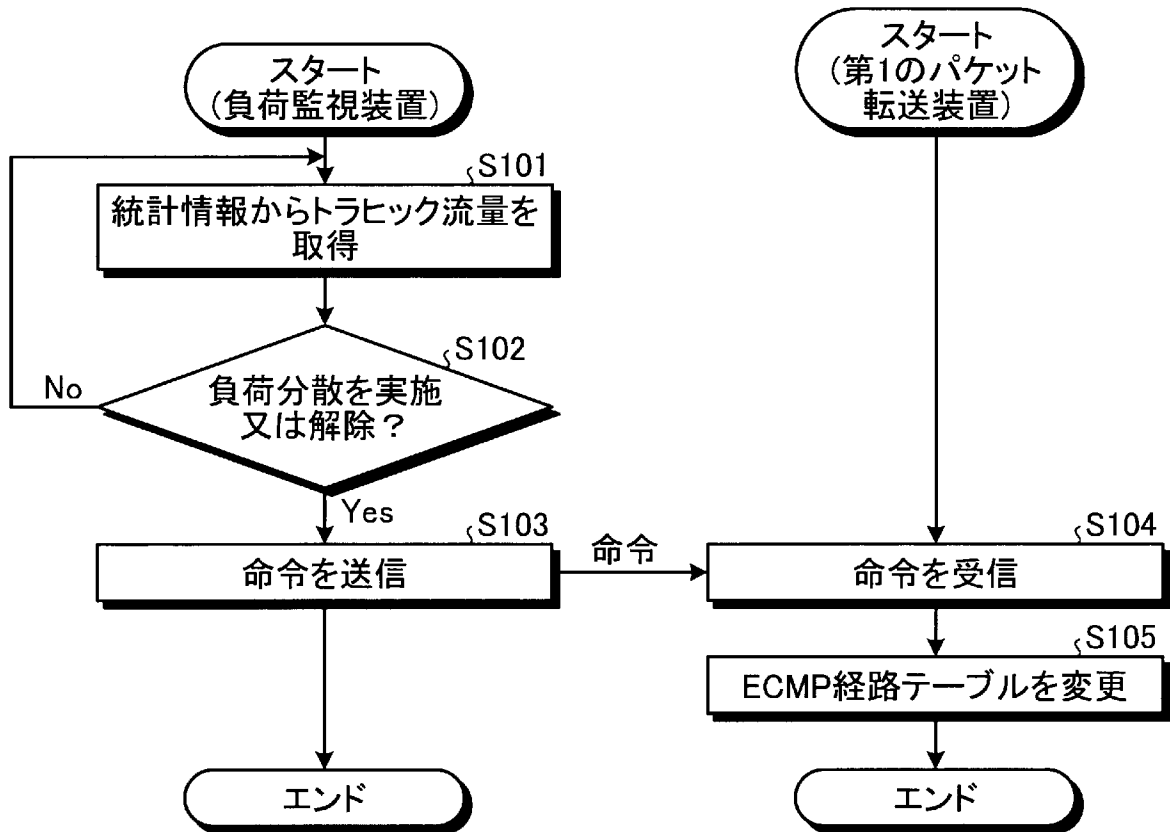
宛先IPアドレス	宛先ポート	フロー特性のカウンタ		スコア (A-B)÷(A+B)
		ショート(A)	ロング(B)	
10.0.0.10	5004	1000	0	1
20.0.0.15	443	490	510	-0.02
30.0.0.20	443	500	500	0
⋮	⋮	⋮	⋮	⋮

宛先 ネットワーク	サブネット マスク	ゲートウェイの リスト	出力ポート 情報	データ量(bps)の 負荷状態	パケット数(pps)の 負荷状態
default	-	192.168.1.0	eth0	0	1
		192.168.1.1	eth1	0	0
		192.168.1.2	eth2	1	0
10.0.0.0	255.255.255.0	192.168.1.3	eth3	0	1
∴	∴	∴	∴	∴	∴

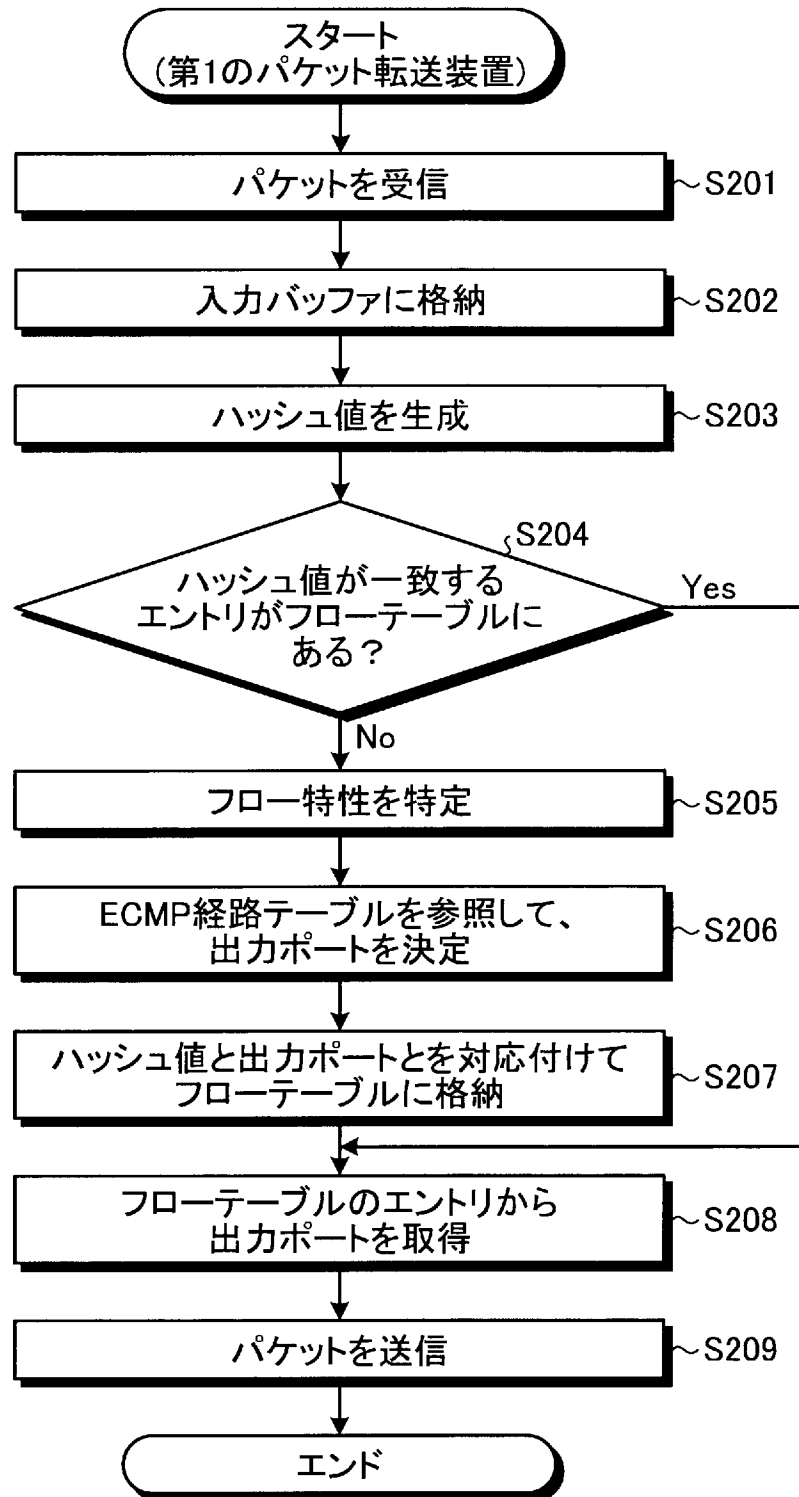
経路情報

{ 1000

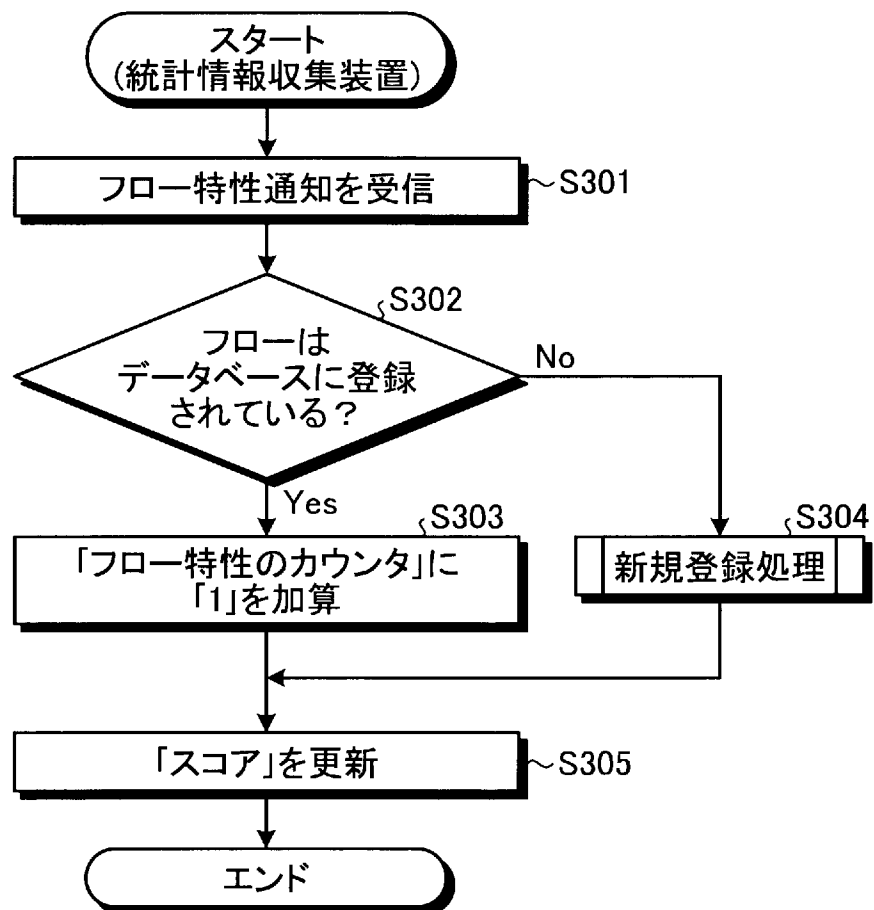
[図5]



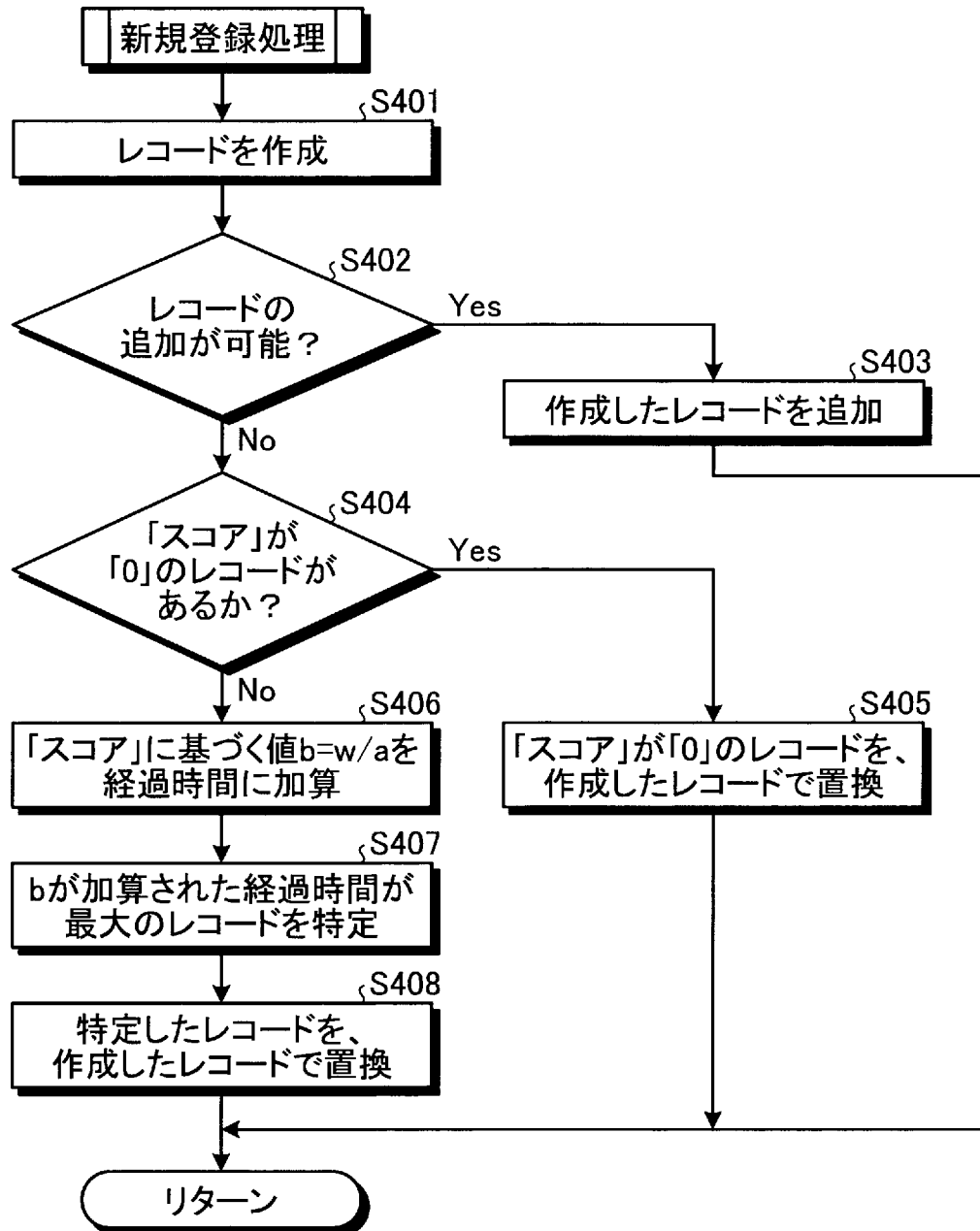
[図6]



[図7]



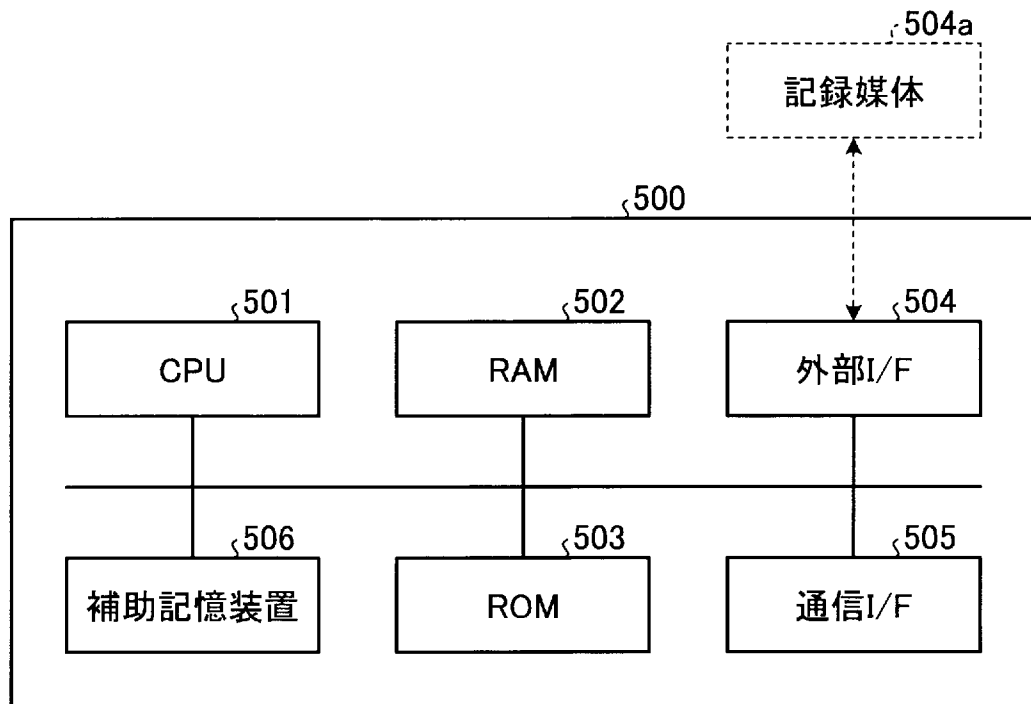
[図8]



[図9]

[illegible]

[図10]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2019/030813

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. H04L12/803 (2013.01) i, H04L12/851 (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)

Int.Cl. H04L12/803, H04L12/851

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	1971-2019
Registered utility model specifications of Japan	1996-2019
Published registered utility model applications of Japan	1994-2019

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 2016-525846 A (IBM) 25 August 2016, entire text, all drawings & US 2016/0191392 A1, entire text, all drawings & CN 104348740 A	1-6
A	JP 2015-233221 A (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) 24 December 2015, entire text, all drawings (Family: none)	1-6

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 30.08.2019	Date of mailing of the international search report 10.09.2019
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.

A. 発明の属する分野の分類（国際特許分類（IPC））

Int.Cl. H04L12/803(2013.01)i, H04L12/851(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. H04L12/803, H04L12/851

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	1971-2019年
日本国実用新案登録公報	1996-2019年
日本国登録実用新案公報	1994-2019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 2016-525846 A（インターナショナル・ビジネス・マシーンズ・コーポレーション）2016.08.25, 全文, 全図 & US 2016/0191392 A1 全文, 全図 & CN 104348740 A	1-6
A	JP 2015-233221 A（日本電信電話株式会社）2015.12.24, 全文, 全図（ファミリーなし）	1-6

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	「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

30.08.2019

国際調査報告の発送日

10.09.2019

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）

野元 久道

5 X

9184

電話番号 03-3581-1101 内線 3596