





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.

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).

- (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

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

推定装置は、観測される DNS 応答に基づいて、ドメイン名ごとの DNS 要求の数を計測し、観測されるフローについて IP アドレスごとのフロー数及びデータ量の合計を計測し、前記ドメイン名又は IP アドレスをノードとし、ドメイン名及び IP アドレスの間の対応関係を枝とするグラフを生成し、ノードを構成する各 IP アドレスのデータ量と、中間ノードを除くノードを構成するドメイン名の DNS 要求の数との関係を示す変換行列を推定し、ドメイン名ごとの要求数に各ドメイン名のフローごとのデータ量を乗じた値が、変換行列に IP アドレスごとのデータ量を乗じた値に等しいとして、各ドメイン名のフローごとのデータ量を求め、当該データ量に、ドメイン名ごとの DNS 要求の数に乗じて、ドメイン名ごとのデータ量を推定する。

## 明 細 書

発明の名称：推定装置、推定方法、及び記録媒体

### 技術分野

[0001] 本発明は、推定装置、推定方法、及び記録媒体に関する。

### 背景技術

[0002] 近年、動画配信サービスをはじめとする、多数のユーザを抱えるインターネット上のサービスにおいて、コンテンツの配信元サーバを地理的、又はネットワーク的に分散させることで、トラヒックの分散、サーバ負荷の分散、あるいは遅延時間の低減を図る仕組みが導入されている。特に、CDN（Contents Delivery Network）は、配信元サーバを分散させることに特化したサービスであり、現在、多くのインターネット・サービスが、CDNを利用してコンテンツの配信を行っている。

[0003] 一方で、CDNでは配信元サーバの各IPアドレスが複数のサービスで共用される場合がある。この運用形態により、CDN経由のトラヒックの統計情報を知りたい場合において、フロー情報に含まれるIPアドレスの情報では、そのIPアドレスに紐付く配信元サービスを一意に特定することが困難である。

[0004] 上記の課題を解決する方法の一つとして、DPI（Deep Packet Inspection）が存在する。DPIはパケットのペイロード領域（OSI（Open Systems Interconnection）参照モデルではレイヤ5から7に該当する領域）をも含めたパケット分析方式全般を指す。DPIを用いて、パケットのペイロード領域に含まれるサービス名や識別子を抽出することにより、そのパケットのサービス名の特定が可能である。例えば、HTTP（HyperText Transfer Protocol）パケットの場合は、HTTPリクエスト・パケットのHTTPヘッダに含まれる「HOSTフィールド」を参照することで、そのパケットのサービス名を特定できる。

[0005] しかしながら、パケットのペイロード領域にサービスを特定する情報が含

まれていない場合には、DPIを適用してサービス名を特定するのは困難である。また、パケットが暗号化されておりペイロード領域を参照できない場合においても、DPIの適用は困難である。更に、DPIでは、全てのパケットについて、ペイロード領域も含めてキャプチャを行い、なおかつ平行して分析を行うため、測定装置のコストと負荷が高く、大容量トラフィックへの適用は現実的には困難である。

[0006] 一方、DPIを用いずに、かつ、暗号化通信にも適用できる方式として、DNS (Domain Name System) ログとフロー情報とを突合してフローのサービスを識別する方式が提案されている（例えば、非引用文献4参照）。非引用文献4では、DNSログとフロー情報とが同一地点で収集される。その上で、あるユーザが送受信したフローと、該フローの直近に該ユーザが送受信したDNSパケットの問い合わせドメイン名とを紐付けることで、該フローのドメイン名（すなわち、サービスの識別名）が推定される。この方式はフロー情報とDNSログとを同一地点で収集するという条件において、HTTP通信では75%から97%の推定精度、暗号化通信（TLS (Transport Layer Security)）では74%から96%の推定精度を示している。

## 先行技術文献

### 非特許文献

[0007] 非特許文献1: Domain Names: Implementation and Specification, <https://www.ietf.org/rfc/rfc1035.txt>

非特許文献2: Cisco Systems NetFlow Services Export Version9, <http://www.ietf.org/rfc/rfc3954.txt>

非特許文献3: InMon Corporation's sFlow: A Method for Monitoring Traffic in Switched and Routed Networks, <https://www.ietf.org/rfc/rfc3176.txt>

非特許文献4: Bermudez, Ignacio N., et al, "DNS to the rescue: discerning content and services in a tangled web", In: Proceedings of the 2012 ACM conference on Internet measurement conference, ACM, 2012, p. 413-426, 20

12.

非特許文献5 : P.-A. Absil, et al., (2007), Optimization Algorithms on Matrix Manifolds, pp.10-14, ISBN 978-0-691-13298-3.

非特許文献6 : Russell, Stuart J., Norvig Peter, (2003), Artificial Intelligence: A Modern Approach (3rd ed.), Upper Saddle River, New Jersey: Prentice Hall, pp.122-125, ISBN 978-0-13-604259-4.

## 発明の概要

### 発明が解決しようとする課題

[0008] しかしながら、フロー情報とDNSログとが異なる場所で収集される環境においては、フローと直近のDNSパケットの突合は困難である。したがって、引用文献4に記載された方式では、斯かる環境において収集されたフロー情報とDNSログとに基づいて、サービス名単位（例えば、ドメイン名単位）のトラフィック量を推定するのは困難である。

[0009] 本発明は、上記の点に鑑みてなされたものであって、ドメイン名ごとのトラフィック量の推定のための情報収集に関する制約を緩和することを目的とする。

### 課題を解決するための手段

[0010] そこで上記課題を解決するため、推定装置は、ネットワークにおいて観測されるDNS応答に基づいて、ドメイン名ごとのDNS要求の数を計測する第一の計測部と、ネットワークにおいて観測されるフローについて、IPアドレスが共通する単位ごとに、フロー数及びデータ量の合計を計測する第二の計測部と、前記DNS応答に含まれるドメイン名又はIPアドレスをノードとし、前記ドメイン名及び前記IPアドレスの間の対応関係を枝とするグラフを生成する生成部と、前記グラフのノードを構成する各IPアドレスに関して前記第二の計測部によって計測されたデータ量と、前記グラフの中間ノードを除くノードを構成するドメイン名に関して前記第一の計測部によって計測されたDNS要求の数との関係を示す変換行列を推定する行列推定部と、前記ドメイン名ごとの要求数に前記各ドメイン名のフローごとのデータ

量を乗じた値が、前記変換行列に前記IPアドレスごとのデータ量を乗じた値に等しいとする関係に基づいて、前記各ドメイン名のフローごとのデータ量を求め、当該データ量に、ドメイン名ごとのDNS要求の数を乗じて、ドメイン名ごとのデータ量の推定値を算出する算出部と、を有する。

### 発明の効果

[0011] ドメイン名ごとのトラフィック量の推定のための情報収集に関する制約を緩和することができる。

### 図面の簡単な説明

[0012] [図1]第一の実施の形態における推定装置のハードウェア構成例を示す図である。

[図2]第一の実施の形態における推定装置の機能構成例を示す図である。

[図3]グラフ生成部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4]DNSログ記憶部の構成例を示す図である。

[図5]フロー情報記憶部の構成例を示す図である。

[図6]DNS応答テーブルに基づく有向グラフの一例を示す図である。

[図7]連結グラフの抽出を説明するための図である。

[図8]連結グラフの各ノードへの属性情報の付与を説明するための図である。

[図9]トラフィック量推定部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10]連結グラフの各ノードへの番号の割り振りを説明するための図である。

[図11]連結グラフに基づく隣接行列の一例を示す図である。

[図12]フロー変換行列の初期値の抽出例を示す図である。

[図13]ドメイン名ごとのトラフィック量の推定結果の出力例を示す図である。

[図14]第三の実施の形態における推定装置の機能構成例を示す図である。

### 発明を実施するための形態

[0013] 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、第一の実

施の形態における推定装置のハードウェア構成例を示す図である。図１の推定装置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、ＣＰＵ１０４、及びインタフェース装置１０５等を有する。

[0014] 推定装置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。

[0015] 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。ＣＰＵ１０４は、メモリ装置１０３に格納されたプログラムに従って推定装置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。

[0016] 図２は、第一の実施の形態における推定装置の機能構成例を示す図である。図２において、推定装置１０は、ＤＮＳ応答収集部１１、ＤＮＳ応答統計処理部１２、フロー情報収集部１３、フロー情報統計処理部１４、グラフ生成部１５、トラヒック量推定部１６、及び推定値変換部１７等を有する。これら各部は、推定装置１０にインストールされる１以上のプログラムが、ＣＰＵ１０４に実行させる処理により実現される。推定装置１０は、また、ＤＮＳログ記憶部１２１、フロー情報記憶部１２２、及びグラフ情報記憶部１２３を利用する。ＤＮＳログ記憶部１２１、フロー情報記憶部１２２、及びグラフ情報記憶部１２３は、例えば、図１の補助記憶装置１０２、又は推定装置１０にネットワークを介して接続可能な記憶装置等を用いて実現可能である。

- [0017] DNS 応答収集部 11 は、ネットワーク上において観測される DNS (Domain Name System) 応答パケットを収集する。DNS 応答パケットの収集方法は、特定のものに限定されない。例えば、DNS キャッシュサーバ又はネットワーク中継装置（例えば、ルータ、スイッチ）より、パケットキャプチャによって DNS 応答パケットが収集されてもよい。
- [0018] DNS 応答統計処理部 12 は、一定期間において収集された DNS 応答パケットについて統計処理を実行する。統計処理の結果は、DNS ログ記憶部 121 に記憶される。
- [0019] フロー情報収集部 13 は、ネットワーク上において観測される任意のフロー（例えば、クライアントとサーバとの間のフロー）に関する情報（以下、「フロー情報」という。）をネットワークを介してフロー情報を収集する。サーバとは、例えば、Web サーバやコンテンツを配信するサーバ等である。フローとは、一つの意味の有るメッセージ（例えば、要求又は応答等）を構成するパケットの集合をいう。したがって、一つのフローを構成するパケットの宛先、送信元の IP アドレス、及びポート番号は共通する。フロー情報は、例えば、ルータに搭載されたフロー情報収集機能を用いて収集されてもよい。又は、フロー情報収集部 13 がパケットをキャプチャして、キャプチャ情報からフロー情報を生成してもよい。フロー情報の通信プロトコルとして NetFlow（非特許文献 2 参照）や sFlow（非特許文献 3 参照）が挙げられるが、「送信元 IP アドレス」、「宛先 IP アドレス」、「フロー数」、「バイト数」の情報が取得可能であれば、フロー情報の収集方法やプロトコルは特定のものに限定されない。なお、本実施の形態では、フローが通信の最小単位として扱われるが、パケットが通信の最小単位として扱われてもよい。すなわち、以下の説明におけるフローは、パケットに置き換えられてもよい。
- [0020] フロー情報統計処理部 14 は、一定期間において収集されたフロー情報について統計処理を実行する。統計処理の結果は、DNS ログ記憶部 121 に記憶される。

- [0021] 以上から明らかなように、本実施の形態における推定装置 10 は、DNS 応答パケットとフロー情報との両方を収集可能なネットワークであれば、どのようなネットワークに接続されていてもよい。斯かるネットワークの一例として、ISP (Internet Service Provider) のバックボーン・ネットワーク、企業内ネットワーク、大学ネットワーク、データセンタ・ネットワーク等が挙げられる。
- [0022] グラフ生成部 15 は、グラフ変換部 151、連結グラフ抽出部 152、及び属性情報付与部 153 等を含む。グラフ変換部 151 は、DNS ログ記憶部 121 に記憶されている情報を、有向グラフに変換する。より詳しくは、グラフ変換部 151 は、DNS 応答パケットに含まれているドメイン名（別名をも含む。）又は IP アドレスをノードとし、当該ドメイン名と IP アドレス、又は当該ドメイン名同士の対応関係を枝とする有向グラフを生成する。
- [0023] 連結グラフ抽出部 152 は、グラフ変換部 151 によって生成された有向グラフから、連結グラフを抽出する。すなわち、連結グラフ抽出部 152 は、グラフ変換部 151 によって生成された有向グラフを、1 以上の連結グラフの単位に分解する。連結グラフとは、有向グラフの枝を辿ることで到達可能な各ノードと、当該各ノード間を接続する枝との集合である。有向グラフの分割アルゴリズムはグラフ理論の一般的な方式が適用可能である。
- [0024] 属性情報付与部 153 は、各連結グラフの各ノードに属性情報を付与する。各ノードは、ドメイン名又は IP アドレスである。したがって、各ノードには、当該ノードのドメイン名又は IP アドレスに関して、DNS ログ記憶部 121 又はフロー情報記憶部 122 に記憶されている情報が属性情報として付与される。
- [0025] なお、グラフ生成部 15 によって生成された有向グラフ（各連結グラフ）を示す情報は、グラフ情報記憶部 123 に記憶される。
- [0026] トラヒック量推定部 16 は、グラフ情報記憶部 123 に記憶されている情報が示す有向グラフに基づいて、サービスごとのトラヒック量を推定する。

一般的に、ドメイン名によって各サービスを識別することができる。すなわち、サービスごとにドメイン名は異なる。但し、別名によってサービスを識別するのは困難である。一つのサービスに対して複数の別名が設定される可能性が有るからである。したがって、本実施の形態では、別名を除くドメイン名ごとに、トラフィック量が推定される。図2において、トラフィック量推定部16は、フロー数分配推定部161及びトラフィック量算出部162を含む。これら各部の機能の詳細については後述される。

[0027] 推定値変換部17は、トラフィック量推定部16で推定されたドメイン名ごとのトラフィック量の内部表現を、人間が確認可能なテキスト等の表現形式に変換する。例えば、推定値変換部17は、変換後の表現形式によって、ドメイン名ごとのトラフィック量を可視化する。

[0028] 以下、推定装置10が実行する処理手順について説明する。図3は、グラフ生成部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3の開始時において、ログ情報記憶部には、期間t11において収集されたDNS応答パケットに関してDNS応答統計処理部12によって計測された統計情報が記憶されており、フロー情報記憶部122には、期間t12において収集されたフロー情報に関してフロー情報統計処理部14によって計測された統計情報が記憶されている。ここで、期間t11と期間t12とは、同じであってもよいし、異なってもよい。期間t11と期間t12とが同じであるとは、それぞれの期間の開始時期及び終了時期が一致することをいう。期間t11と期間t12とが異なるとは、それぞれの期間の開始時期及び終了時期の少なくともいずれか一方が異なることをいう。

[0029] ステップS101において、グラフ変換部151は、DNSログ記憶部121に記憶されている情報（以下、「DNSログ」という。）を読み出す。

[0030] 図4は、DNSログ記憶部の構成例を示す図である。図4に示されるように、DNSログ記憶部121には、DNS要求テーブルT1と、DNS応答テーブルT2とが記憶されている。

[0031] DNS要求テーブルT1には、DNS応答のQuestionセクション

に記述された情報（すなわち、DNS要求に関する情報）が記憶されている。具体的には、DNS要求テーブルT1には、観測されたドメイン名及びクエリタイプの組み合わせの単位ごとに、要求数及びユーザ数が記憶されている。

[0032] ドメイン名（厳密にはFQDN（Fully Qualified Domain Name）であり、以下のドメイン名についても同じ。）は、DNS要求において問い合わせの対象とされたドメイン名（すなわち、名前解決の対象のドメイン名）である。クエリタイプは、当該DNS要求において問い合わせの対象とされたレコードである。「A」は、Aレコードを示す。図4には示されていないが、AAAAレコードが問い合わせの対象とされた場合のクエリタイプは、「AAAA」となる。要求数は、「ドメイン名」及び「クエリタイプ」に対応するQuestionセクションを含むDNS応答の数である。なお、1つのDNS応答は、複数のDNS応答パケットによって構成される場合も有るため、ここでは、DNS応答の数としている。ユーザ数は、「ドメイン名」及び「クエリタイプ」に対応するQuestionセクションを含むDNS応答に係るDNS応答パケットの宛先IPアドレスの種類数である。すなわち、ユーザ数は、DNS要求元の種類数である。「ドメイン名」及び「クエリタイプ」ごとの要求数及びユーザ数は、DNS応答統計処理部12によって計測される。なお、図4において、要求数及びユーザ数の値は、便宜上、記号によって示されているが、実際には数値である。

[0033] 一方、DNS応答テーブルT2には、DNS応答のAnswerセクションに記述されたAレコード、AAAAレコード、又はCNAMEレコード等に関する情報が記憶されている。具体的には、DNS応答テーブルT2には、観測されたドメイン名、レコードタイプ、及びレコードデータの組み合わせの単位ごとに、TTL（Time To Live）が記憶されている。

[0034] ドメイン名は、名前解決の対象とされたドメイン名である。レコードタイプは、DNS応答に含まれているレコードのタイプである。レコードデータは、DNS応答に含まれているレコードの値（すなわち、ドメイン名に対し

て対応付けられている値)である。当該レコードがAレコード又はA A A Aレコードである場合(すなわち、レコードタイプが「A」又は「A A A A」である場合)、レコードデータの値はI Pアドレス(I P v 4のI Pアドレス又はI P v 6のI Pアドレス)である。当該レコードがC N A M Eレコードである場合(すなわち、レコードタイプが「C N A M E」である場合)、レコードデータの値は、別名である。T T Lは、ドメイン名のキャッシュの有効期限の最大値の推定値である。有効期限の最大値とは、「D N S権威サーバ」のゾーンファイルの設定ファイルにおいて定義されている値であり、ネットワーク管理者等によって設定される。観測されるD N S応答パケット内の各レコードのT T Lは、当該D N S応答パケットの観測時点での残り秒数であるため、必ずしも最大値ではない。そこで、D N S応答統計処理部12は、期間t 11において観測されたD N S応答パケットに含まれているAレコード又はC N A M Eレコードのうち、ドメイン名、レコードタイプ、及びレコードデータが重複するレコードごとに、観測されたT T Lの中で最大の値を計測し、計測結果を、当該レコードのT T Lの最大値として推定する。

[0035] 続いて、グラフ変換部151は、フロー情報記憶部122に記憶されているフロー情報の統計情報(以下、「フロー情報ログ」という。)を読み出す(ステップS102)。

[0036] 図5は、フロー情報記憶部の構成例を示す図である。図5に示されるように、フロー情報記憶部122には、観測されたフローのI Pアドレスごとに(すなわち、観測されたフローについてI Pアドレスが共通する単位ごとに)、フロー数、ユーザ数、バイト数が記憶されている。

[0037] I Pアドレスは、サーバ側のI Pアドレスである。サーバ側のI Pアドレスは、観測されたフローの送信元I Pアドレス又は宛先I Pアドレスである。推定装置10がI S P等によって運用される場合、各クライアントのI Pアドレスは、I S Pによって割り当てられる。換言すれば、推定装置10は、クライアントのI Pアドレスの一覧を保持することができる。フロー情報

統計処理部 14 は、斯かる一覧に基づいて、送信元 IP アドレス又及び宛先 IP アドレスのいずれがサーバの IP アドレスであるのかを特定してもよい。

[0038] フロー数は、「IP アドレス」に係るフローの数である。ユーザ数は、「IP アドレス」に係るフローのクライアント側の IP アドレスの数である。バイト数は、「IP アドレス」に係るフローのバイト数の総和（データ量）である。

[0039] フロー数、ユーザ数、及びバイト数は、フロー情報統計処理部 14 によって計測される。なお、図 5 において、フロー数、ユーザ数、及びバイト数の値は、便宜上、記号によって示されているが、実際には数値である。

[0040] 続いて、グラフ変換部 151 は、DNS 応答テーブル T2 の内容を、有向グラフに変換する（ステップ S103）。

[0041] 図 6 は、DNS 応答テーブルに基づく有向グラフの一例を示す図である。図 6 のグラフ g1 は、図 4 に示した DNS 応答テーブル T2 に対応する。すなわち、グラフ g1 の各ノードは、DNS 応答テーブル T2 のドメイン名又はレコードデータ（IP アドレス若しくは別名）である。具体的には、ノード n1、ノード n2、及びノード n3 は、ドメイン名のノードである。ノード n4 及びノード n5 は、IP アドレスのノードである。

[0042] また、グラフ g1 は、DNS 応答テーブル T2 におけるドメイン名とレコードデータとの対応関係を示すと共に、ドメイン名からレコードデータへの向きを有する有向枝を含む。

[0043] 続いて、連結グラフ抽出部 152 は、グラフ変換部 151 によって生成された有向グラフを、連結グラフの単位に分解し、各連結グラフを抽出する（ステップ S104）。

[0044] 図 7 は、連結グラフの抽出を説明するための図である。ステップ S103 において生成される有向グラフが、仮に、図 7 の左側の破線の矩形で囲まれたものである場合、図 7 の右側に示されるように、4 つの連結グラフが抽出される。なお、本実施の形態では、グラフ g1 が、そのまま 1 つの連結グラ

フとして抽出される。

[0045] 続いて、属性情報付与部 153 は、各連結グラフの各ノードに対して属性情報を付与する（ステップ S105）。

[0046] 図 8 は、連結グラフの各ノードへの属性情報の付与を説明するための図である。図 8 に示されるように、ドメイン名の各ノードに対しては、DNS 要求テーブル T1 において当該ノードに対応するレコードの要求数及びユーザ数と、DNS 応答テーブル T2 において当該ノードに対応する TTL とが付与される。図 8 では、「＜要求数＞／＜ユーザ数＞／＜TTL＞」の形式で、付与された属性情報が示されている。

[0047] 一方、IP アドレスの各ノードに対しては、フロー情報記憶部 122 において当該ノードに対応するレコードのフロー数、ユーザ数、及びバイト数が付与される。図 8 では、「＜フロー数＞／＜ユーザ数＞／＜バイト数＞」の形式で、付与された属性情報が示されている。

[0048] なお、期間 t11 と期間 t12 との時間が相互に異なる場合、又は各期間の時間幅が相互に異なる場合、各 IP アドレスのノード（ノード n4 及びノード n5）に付与されたフロー数の合計と、中間ノードを除くドメイン名のノード（ノード n1 及びノード n2）に付与された要求数の合計とが一致しなくなる可能性が有る。この場合、後述されるフロー数分配推定部 161 によって推定されるフロー変換行列 H の推定精度が劣化する可能性が有る。そこで、属性情報付与部 153 は、連結グラフごとに、中間ノードを除く各ドメイン名のノードの要求数の合計が 1 となるように、当該各ノードの要求数を正規化すると共に、IP アドレスの各ノードのフロー数の合計が 1 となるように、IP アドレスの各ノードのフロー数を正規化するようにしてもよい。

[0049] なお、グラフ生成部 15 によって生成された各連結グラフを示す情報（以下、「グラフ情報」という。）は、グラフ情報記憶部 123 に記憶される。グラフ情報の生成周期は、DNS 応答統計処理部 12 による DNS 応答パケットの統計処理の周期以上であり、かつ、フロー情報統計処理部 14 による

フロー情報の統計処理の周期以上であれば、どのような周期であってもよい。グラフ情報の生成時期（以下、「グラフ生成時期」という。）が訪れた時点において、DNSログ記憶部121に最後に記憶されたDNSログ（DNS要求テーブルT1及びDNS応答テーブルT2）と、フロー情報記憶部122に最後に記憶されたフロー情報ログとが用いられて、グラフ情報が生成されてもよい。

[0050] 各グラフ生成時期に生成されたグラフ情報は、当該グラフ生成時期に対応付けられてグラフ情報記憶部123に記憶される。例えば、期間t11と期間t12とに対応するグラフ生成時期をグラフ生成時期t1とすると、期間t11におけるDNSログと期間t12におけるフロー情報ログとに基づいて生成されたグラフ情報は、グラフ生成時期t1に対応付けられてグラフ情報記憶部123に記憶される。また、期間t11よりの後の一定期間である期間t21においてDNSログ記憶部121に記憶されたDNSログと、期間t12よりの後の一定期間である期間t22においてフロー情報記憶部122に記憶されたフロー情報ログとに基づいて生成される各連結グラフを示すグラフ情報は、期間t21及び期間t22とに対応するグラフ生成時期t2に対応付けられてグラフ情報記憶部123に記憶される。

[0051] 続いて、トラヒック量推定部16が実行する処理手順について説明する。図9は、トラヒック量推定部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。なお、図9は、各連結グラフについて実行されるが、本実施の形態では、図6のグラフg1が処理対象とされる。また、図9の処理の実行のタイミングは、図3の処理の実行タイミングと非同期であってもよい。

[0052] ステップS201において、フロー数分配推定部161は、変数tに1を代入する。変数tは、処理対象とするグラフ生成時期と、ステップS202以降の実行回数とを識別するための変数である。

[0053] 続いて、フロー数分配推定部161は、グラフ情報記憶部123から、t番目のグラフ生成時期に対応するグラフ情報を取得する（ステップS202

）。ここでは、グラフ生成時期  $t_1$  における、図 8 のグラフ  $g_1$  を示すグラフ情報が取得される。続くステップ S 203～ステップ S 207 において、フロー数分配推定部 161 は、グラフ  $g_1$  の形状的特徴と、グラフ  $g_1$  の各ノードに付与された属性情報（すなわち、DNS ログ及びフロー情報ログ）とに基づき、ドメイン名ごとの要求数と IP アドレスごとのフロー数との関係を示すフロー変換行列  $H$  を推定する。本実施の形態では、グラフ  $g_1$  において、中間ノードを除くドメイン名のノード（すなわち、別名ではないノード）であるノード  $n_1$  とノード  $n_2$  とのそれぞれのトラフィック量を求めるためにそれぞれのフロー数を知りたいところ、フロー変換行列  $H$  は、ノード  $n_4$  とノード  $n_5$  とにおいて既知であるフロー数が、ノード  $n_1$  とノード  $n_2$  とに対してどのように分配されるべきであるのかを推定するための行列である。なお、中間ノードを除くドメイン名のノードが 1 つである連結グラフ（例えば、図 7 の右側の上から 2 番目の連結グラフ及び上から 4 番目の連結グラフ）の場合、中間ノードを除くドメイン名のノードのフロー数は、IP アドレスのノードのフロー数を単純に合算することで求めることが可能であるため、以降の処理は実行されなくてもよい。

[0054] ステップ S 203 において、フロー数分配推定部 161 は、グラフ  $g_1$  の各ノードに番号を割り振る。各ノードの番号が 1 つの連結グラフ内で重複しなければ、番号の割り振り方には制限が無い。

[0055] 図 10 は、連結グラフの各ノードへの番号の割り振りを説明するための図である。図 10 においては、ノード  $n_1$ 、 $n_2$ 、 $n_3$ 、 $n_4$ 、 $n_5$  の順に、(1)、(2)、(3)、(4)、(5) の番号が割り振られた例が示されている。

[0056] 続いて、フロー数分配推定部 161 は、グラフ  $g_1$  を隣接行列  $A$  に変換する (S 204)。隣接行列  $A$  の行数と列数は等しく、かつ、その行数及び列数は、ステップ S 203 において割り振られた番号の個数（すなわち、グラフ  $g_1$  のノードの数）に等しい。

[0057] 図 11 は、連結グラフに基づく隣接行列の一例を示す図である。図 11 に

示される隣接行列  $A$  において、行方向がソース側（有向枝が出る側）であり、列方向がシンク側（有向枝が入る側）である。隣接行列  $A$  への変換方法は、基本的には、一般に知られている汎用的な方法が利用されてよい。但し、本実施の形態において、隣接行列  $A$  は、一般的な隣接行列に対して以下の（a）及び（b）に示される拡張がなされる。

（a）ノード  $i$  からノード  $j$  に有向枝が存在し、ノード  $j$  の入次数が  $n$  の場合、 $A_{ij} = 1/n$  とする。

（b）ノード  $i$  の入次数が 0 の場合、ノード  $i$  の対角成分である  $A_{ii} = 1$  とする。

[0058] なお、図 11 において、0 である成分の値は、便宜上、空欄とされている。

[0059] 例えば、 $A_{13}$ 、 $A_{23}$ 、 $A_{34}$ 、及び  $A_{35}$  の値は、上記（a）に基づいて割り当てられた値である。すなわち、 $A_{13}$ 、 $A_{23}$  は、それぞれ、ノード  $n1$  からノード  $n3$  への有向枝、ノード  $n2$  からノード  $n3$  への有向枝に対応するが、ノード  $n3$  の入次数（ノード  $n3$  に入ってくる有向枝の数）は、2 である。したがって、 $A_{13}$  及び  $A_{23}$  の値は、 $1/2 = 0.5$  となる。また、 $A_{34}$ 、 $A_{35}$  は、それぞれ、ノード  $n3$  からノード  $n4$  への有向枝、ノード  $n3$  からノード  $n4$  への有向枝に対応するが、ノード  $n4$  及びノード  $n5$  のいずれについても入次数は 1 である。したがって、 $A_{34}$  及び  $A_{35}$  の値は、 $1/1 = 1$  となる。

[0060] 一方、 $A_{11}$  及び  $A_{22}$  の値は、上記（b）に基づいて割り当てられた値である。すなわち、ノード  $n1$  及びノード  $n2$  のいずれについても、入次数は 0 である。したがって、 $A_{11}$  及び  $A_{22}$  の値は 1 とされている。

[0061] 続いて、フロー数分配推定部 161 は、グラフ  $g1$  の最大ホップ長の数だけ隣接行列  $A$  を掛け合わせて、行列  $A'$  を求める（ステップ S205）。すなわち、フロー数分配推定部 161 は、 $A' = A^n$  を演算する（ $n$  はグラフ  $g1$  の最大ホップ長）。なお、隣接行列  $A$  を  $n$  乗する意義は、一般的なグラフ理論に詳しい。

[0062] 続いて、フロー数分配推定部 161 は、行列  $A'$  から、グラフ  $g_1$  における中間ノードを除くドメイン名のノードに対応する行と、IP アドレスのノードに対応する列との組み合わせに対応する成分を抽出する（ステップ S206）。抽出された成分が構成する行列は、フロー変換行列  $H$  の初期値  $H_{init}$  とされる。

[0063] 図 12 は、フロー変換行列の初期値の抽出例を示す図である。図 12 では、行列  $A'$  における、(1) 行及び (2) 行と、(4) 列及び (5) 列との重複部分の成分が、初期値  $H_{init}$  として抽出される例が示されている。

[0064] 続いて、フロー数分配推定部 161 は、初期値  $H_{init}$  と、グラフ  $g_1$ （図 8）において初期値  $H_{init}$  の各行に対応するドメイン名のノードに付与されている要求数と、グラフ  $g_1$  において初期値  $H_{init}$  の各列に対応する IP アドレスに対応するノードに付与されているフロー数とに基づいて、フロー変換行列  $H$  を推定する（ステップ S207）。フロー変換行列  $H$  は、当該要求数及び当該フロー数に含まれるノイズ等により一意に定まるものではないため、本実施の形態では、最適化問題を解くことにより、フロー変換行列  $H$  の局所解が推定される。例えば、以下の式 (1) は、最適化問題を解く場合の定式化の一例である。

[0065] [数1]

$$H = \operatorname{argmin}_H (\|Y - HX\|_F + \alpha \|H - H_{init}\|_F) \cdots (1)$$

$$X = \begin{pmatrix} F4 \\ F5 \end{pmatrix} \quad Y = \begin{pmatrix} Q1 \\ Q2 \end{pmatrix}$$

ここで、 $\|\cdots\|_F$  は、フロベニウス・ノルムを表す。

また、 $\alpha$  は、初期状態（初期値  $H_{init}$ ）の重視度を定めるパラメータである。すなわち、フロー変換行列  $H$  について、初期値  $H_{init}$  にできるだけ近い解

を得たい場合に、 $\alpha$ の値は大きくされる。

[0066] 更に、ベクトル $X$ は、グラフ $g_1$ において初期値 $H_{init}$ の各列に対応するIPアドノード（ノード $n_4$ 、ノード $n_5$ ）に付与されているフロー数（ $F_4$ 、 $F_5$ ）である。ベクトル $Y$ は、グラフ $g_1$ において初期値 $H_{init}$ の各行に対応するドメイン名のノード（ノード $n_1$ 、ノード $n_2$ ）に付与されている要求数（ $Q_1$ 、 $Q_2$ ）である。

[0067] すなわち、フロー変換行列 $H$ は、IPアドレスごとのフロー数を、ドメイン名ごとのフロー数に変換するための行列である。

[0068] なお、最適化問題を解く場合、フロー変換行列 $H$ の初期値は、任意の値であってもよい。例えば、全ての成分の値が0である行列が初期値とされてもよい。この場合、ステップ $S_{203} \sim S_{206}$ は実行されなくてもよい。但し、本実施の形態のように、初期値 $H_{init}$ を求めることで、フロー変換行列 $H$ について、グラフ $g_1$ の状態（すなわち、DNS応答パケットやフロー情報の観測状態）に即した解が得られる可能性を高めることができる。その結果、ドメイン名ごとのトラフィック量の推定精度の向上を期待することができる。

[0069] なお、式（1）は、最適化問題として解く場合の定式化の一例であり、式（1）からフロー変換行列 $H$ に関する更新式を導出することで局所解を求めることができる。他の例として、ヒルクライム法等の最適化問題を解く方式が適用されてもよい（例えば、非特許文献6参照）。

[0070] 続いて、フロー数分配推定部161は、変数 $t$ の値が、グラフ $g_1$ の中間ノードを除くドメイン名のノードの数に達したか否かを判定する（ステップ $S_{208}$ ）。変数 $t$ の値が該当ノードの数に達していない場合（ステップ $S_{208}$ でNo）、フロー数分配推定部161は、変数 $t$ に1を加算して（ステップ $S_{209}$ ）、ステップ $S_{202}$ 以降を繰り返す。すなわち、次のグラフ生成時期に関して、ステップ $S_{202}$ 以降が実行される。本実施の形態において、該当ノードの数は、ノード $n_1$ とノード $n_2$ との2つである。したがって、例えば、生成時期 $t_2$ に関してステップ $S_{202}$ 以降が実行され

、生成時期  $t_2$  に対するフロー変換行列  $H$  が生成される。

[0071] このように、フロー変換行列  $H$  は、グラフ  $g_1$  の中間ノードを除くドメイン名のノードの数だけ、グラフ生成時期が相互に異なるグラフ情報に基づいて生成される。例えば、本実施の形態では、グラフ生成時期  $t_1$  に対するフロー変換行列  $H_1$  と、グラフ生成時期  $t_2$  に対するフロー変換行列  $H_2$  とが生成される。この理由については後述される。

[0072] ステップ  $S_{210}$  以降では、フロー変換行列  $H$  等が用いられて、トラヒック量算出部  $162$  によって、ドメイン名ごとのトラヒック量の推定値が算出される。

[0073] ステップ  $S_{210}$  において、トラヒック量算出部  $162$  は、値が未知であるデータ（すなわち、値が推定されるべきデータ）、及び値が既知のデータ（すなわち、グラフ  $g_1$  の各ノードに属性情報として付与されているデータ）のそれぞれについて変数を定義する。本実施の形態では、以下のように変数が定義される。

[既知のデータ]

$R_a$  : 各 IP アドレスのフロー毎バイト数

$S_a$  : 各 IP アドレスのフロー数

$B_a$  : 各 IP アドレスのバイト数

$S_d$  : 各ドメイン名のフロー数

なお、各ドメイン名のフロー数  $S_d$  は、DNS ログからは得られないものの、要求数で近似可能であるため、その値が代入される。

[未知のデータ]

$R_d$  : 各ドメイン名のフロー毎バイト数

$B_d$  : 各ドメイン名のバイト数

本実施の形態において、既知の各データの値は、以下の通りとなる。

$R_a = (B_4 / F_4, B_5 / F_5)$

$S_a = (F_4, F_5)$

$B_a = (B_4, Q_5)$

$$S_d = (Q_1, Q_2)$$

なお、グラフ  $g_1$  に関して、中間ノードを除くドメイン名のノードの数は 2 であるため、 $R_d$  及び  $B_d$  の次元数は 2 である。

[0074] 続いて、トラフィック量算出部 162 は、各ドメインのフロー毎バイト数  $R_d$  を未知変数として、フロー変換行列  $H$  を用いてトラフィック量の推定式 (2) を生成する (ステップ S211)。

[0075] [数2]

$$\begin{aligned} S_d \otimes R_d &= H \cdot S_a \otimes R_a \\ &= H \cdot B_a \end{aligned} \quad \dots (2)$$

\*  $\otimes$  はベクトルの要素毎の積であり、  
ベクトル  $A, B$  に対して以下のように定義

$$A \otimes B = A \cdot B^T$$

$R_d$  に含まれる未知変数の数は、各隣接行列の中間ノードを除くドメイン名の数と一致する。推定式 (2) を展開することにより連立一次方程式が得られる。

[0076] なお、連立一次方程式は、 $R_d$  の次元数分生成される。具体的には、中間ノードを除くドメイン名のノードの数だけ生成されているフロー変換行列  $H$  のそれぞれが用いられて、連立一次方程式が生成される。したがって、本実施の形態では、グラフ生成時期  $t_1$  とグラフ生成時期  $t_2$  とに対する以下の 2 つの連立一次方程式 (3) が生成される。

[0077]

[数3]

$$\begin{aligned} t1: Sd1 \otimes Rd &= H1 \cdot Ba1 \\ t2: Sd2 \otimes Rd &= H2 \cdot Ba2 \end{aligned} \quad \dots (3)$$

連立一次方程式（３）において、 $Sd1$ 、 $H1$ 、 $Ba1$ は、グラフ生成時期  $t1$  に対応する  $Sd$ 、フロー変換行列  $H$ 、 $Ba$  である。 $Sd2$ 、 $H2$ 、 $Ba2$ は、グラフ生成時期  $t2$  に対応する  $Sd$ 、フロー変換行列  $H$ 、 $Ba$  である。このように、フロー変換行列、 $Sd$ 、及び  $Ba$  は、グラフ  $g1$  の中間ノードを除くドメイン名のノードの数分だけ必要となるため、ステップ  $S202 \sim S207$  は、その数分繰り返されるのである。

[0078] なお、ここでは、 $Rd$  が時刻に対して不変であること又は変化が小さいことが仮定されている。斯かる仮定は、一般的に妥当であるものと考えられる。フロー毎バイト数は、サービスに依存するものであり、時刻に対する依存度は小さいと考えられるからである。例えば、動画サイトと単なるテキスト情報のサイトを比較すれば、動画サイトの方がフロー毎バイト数が大きいと考えられ、その関係は、時刻に対して不変である可能性が高いと考えられる。したがって、本実施の形態は、フロー毎バイト数について、時刻に対する依存度が低い環境に対して好適である。

[0079] 続いて、トラフィック量算出部 162 は、連立一次方程式（３）を解くことで、 $Rd$  を求める（ステップ  $S212$ ）。但し、観測される  $DNS$  応答パケットやフロー情報のノイズの影響等により  $Rd$  は一意に定まらない場合がある。その場合には、非特許文献 6 の局所解を求める方法が利用されてもよい。

[0080] 続いて、トラフィック量算出部 162 は、ステップ  $S212$  において得られた  $Rd$  に対して、 $Sd$  を乗算することにより、 $Bd$  を求める（ステップ  $S2$

13)。すなわち、以下の演算が実行される。

[0081] [数4]

$$Bd = Sd \otimes Rd$$

ここで、Bdは、目的とする、ドメイン名ごとのトラフィック量である。

[0082] なお、ドメイン名ごとのトラフィック量の推定方法として、上記以外の方法が採用されてもよい。例えば、非特許文献5のような最適化アルゴリズム等の要素技術が用いられてもよい。

[0083] 推定値変換部17は、上記のように推定された、ドメイン名ごとのトラフィック量を、例えば、図13に示されるような形式で出力する。

[0084] 図13は、ドメイン名ごとのトラフィック量の推定結果の出力例を示す図である。図13では、グラフg1の他に、グラフg2が示されている。グラフg2は、グラフg1と共に抽出された連結グラフであるとする。この場合、非中間ノードの4つのドメイン名について、トラフィック量が得られる。推定値変換部17は、これらのトラフィック量を、例えば、テーブルT3に示される形式で可視化する。テーブルT3には、ドメイン名ごとに推定トラフィック量が示されている。なお、テーブルT3における行の配列順序は、推定トラフィック量によってソートされてもよい。すなわち、ドメイン名ごとの推定トラフィック量のランキングが出力されてもよい。出力形態は、所定のものに限定されない。例えば、表示装置に表示されてもよいし、プリンタに印刷されてもよい。又は、ネットワークを介して他のコンピュータに送信されてもよい。

[0085] 上述したように、第一の実施の形態によれば、統計化された入力データ（DNSログ及びフロー情報ログ）が、グラフ生成部15によって有向グラフに変換され、その有向グラフの特性を活かしつつ、トラフィック量推定部16

によって、統計情報に基づいてドメイン名ごとのトラフィック量が推定される。これら一連の処理の中では、個々のユーザやパケット単位の紐付けは必要とされない。すなわち、本実施の形態では、各DNS応答パケットと各フロー情報とを一つずつ紐付けるのではなく、DNS応答パケットの統計情報と、フロー情報の統計情報とを関連付けることにより、ドメイン名ごとのトラフィック量が推定される。したがって、DNS応答パケット及びフロー情報のそれぞれの収集場所や収集時刻の違いによりDNSログとフロー情報のそれぞれの送信ユーザ群が異なるケースにおいても対応可能である。よって、ドメイン名ごとのトラフィック量の推定のための情報収集に関する制約を緩和することができる。

[0086] その結果、例えば、DNS応答パケットとフロー情報とを異なる場所及び時間で収集せざるを得ない制約が有る大規模ネットワークにおいて、運用者は、サービス単位のトラフィック量を容易に把握することができ、ネットワーク監視・運用コストの削減、回線輻輳時の原因の切り分けの迅速化、あるいはネットワークの設備投資の最適化等の応用的な効果を期待することができる。

[0087] 次に、第二の実施の形態について説明する。第二の実施の形態では第一の実施の形態と異なる点について説明する。第二の実施の形態において特に言及されない点については、第一の実施の形態と同様でもよい。

[0088] 第二の実施の形態でのトラフィック量算出部162は、は、図9のステップS211において、以下の推定式(4)を生成する。

[0089] [数5]

$$\begin{aligned} \underline{Sd} \otimes \underline{Rd} &= \underline{Sa} \otimes \underline{Ra} \\ &= \underline{Ba} \quad \dots (4) \end{aligned}$$

すなわち、第二の実施の形態では、フロー変換行列Hは利用されない。し

たがって、図9のステップS201～S208は実行されなくてもよい。また、図3の処理も実行されなくてもよい。この場合、グラフ生成時期ごとの連立一次方程式は、以下の通りとなる。

[0090] [数6]

$$\begin{aligned} t1: Sd1 \otimes Rd &= Ba1 \\ t2: Sd3 \otimes Rd &= Ba2 \\ t3: Sd3 \otimes Rd &= Ba3 \end{aligned}$$

その他については、第一の実施の形態と同様でよい。

[0091] 第二の実施の形態は、十分な観測データ（DNS応答パケット及びフロー情報）が蓄積されている場合に適用されることが好ましい。観測データが多ければ多いほど、フロー変換行列Hが利用されなくても、ドメイン名ごとのトラフィック量の推定精度の劣化が大きくなるのを回避できると可能性が高くなるからである。

[0092] 次に、第三の実施の形態について説明する。第三の実施の形態では第一又は第二の実施の形態と異なる点について説明する。第三の実施の形態において特に言及されない点については、第一又は第二の実施の形態と同様でもよい。

[0093] 図14は、第三の実施の形態における推定装置の機能構成例を示す図である。図14中、図2と同一部分には同一符号を付し、その説明は省略する。図14において、推定装置10は、更に、要求数補正部18及びフロー情報補正部19を有する。

[0094] 要求数補正部18は、DNSログに含まれる要求数とTTLとを用いて、クライアントによるサーバへのアクセス数又はフロー数を推定し、推定結果によって、DNSログ内の要求数を置き換える。すなわち、DNSログに含まれる要求数は、クライアントによるサーバへのアクセス数又はフロー数と

は異なる。通常、ユーザがWebページを閲覧するなどして、あるURL (Uniform Resource Locator) にアクセスする際には、当該URLに含まれるドメイン名に対して名前解決が試みられる。但し、一般にユーザの端末にはDNS応答をキャッシュする仕組み (スタブリゾルバ) が存在するため、クライアントによるWebページへのアクセス数に対して、観測されるDNS応答パケットは大幅に少なくなる。

[0095] そこで、要求数補正部18は、TTLと要求数とを用いて、実際のサーバへのアクセス数を推定する。例えば、TTLによって特定される期間において行われた名前解決の回数を推定し、推定結果を当初の要求数に加算することで、当該要求数が補正されてもよい。

[0096] したがって、第三の実施の形態において、図4に示されるDNS要求テーブルT1の要求数は、補正後の値となる。

[0097] 一方、フロー情報補正部19は、フロー情報ログのフロー数、ユーザ数、及びバイト数を補正する。すなわち、ルータ等の一般的なネットワーク機器では、性能上の理由により定常的にサンプリングを行いながらフロー情報が収集される。例えば、1000パケットごとに1パケットに関するフロー情報が収集される。このようにサンプリングされたフロー情報に基づくフロー数、ユーザ数、及びバイト数は、実際の値よりも小さいものとなる。

[0098] そこで、フロー情報補正部19は、サンプリング前のフロー数、ユーザ数、及びバイト数を推定し、推定結果によって、フロー情報ログの内容を補正する。例えば、サンプリングの割合に基づいて補正が行われてもよい。サンプリングの割合が1/1000であれば、フロー数、ユーザ数、及びバイト数の値を1000倍することにより、補正が行われてもよい。又は他の方法によって補正が行われてもよい。

[0099] したがって、第三の実施の形態において、図4に示されるDNS応答テーブルT2の要求数は、補正後の値となる。

[0100] 上述したように、第三の実施の形態によれば、DNSログの要求数やフロー情報ログの値を実際の値に近づけることができるため、ドメイン名ごとの

トラヒック量の推定精度の向上を期待することができる。

[0101] なお、上記各実施の形態において、DNS 応答統計処理部 12 は、第一の計測部の一例である。フロー情報統計処理部 14 は、第二の計測部の一例である。グラフ生成部 15 は、生成部の一例である。フロー数分配推定部 161 は、行列推定部の一例である。トラヒック量算出部 162 は、算出部の一例である。

[0102] 以上、本発明の実施例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、特許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。

[0103] 本出願は、2015年2月17日に出願された日本国特許出願第2015-028744号に基づきその優先権を主張するものであり、同日本国特許出願の全内容を参照することにより本願に援用する。

## 符号の説明

|        |     |             |
|--------|-----|-------------|
| [0104] | 10  | 推定装置        |
|        | 11  | DNS 応答収集部   |
|        | 12  | DNS 応答統計処理部 |
|        | 13  | フロー情報収集部    |
|        | 14  | フロー情報統計処理部  |
|        | 15  | グラフ生成部      |
|        | 16  | トラヒック量推定部   |
|        | 17  | 推定値変換部      |
|        | 18  | 要求数補正部      |
|        | 19  | フロー情報補正部    |
|        | 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 | CPU         |

|       |             |
|-------|-------------|
| 1 0 5 | インタフェース装置   |
| 1 2 1 | D N S ログ記憶部 |
| 1 2 2 | フロー情報記憶部    |
| 1 2 3 | グラフ情報記憶部    |
| 1 5 1 | グラフ変換部      |
| 1 5 2 | 連結グラフ抽出部    |
| 1 5 3 | 属性情報付与部     |
| 1 6 1 | フロー数分配推定部   |
| 1 6 2 | トラヒック量算出部   |
| B     | バス          |

## 請求の範囲

- [請求項1]        ネットワークにおいて観測されるDNS応答に基づいて、ドメイン名ごとのDNS要求の数を計測する第一の計測部と、
- ネットワークにおいて観測されるフローについて、IPアドレスが共通する単位ごとに、フロー数及びデータ量の合計を計測する第二の計測部と、
- 前記DNS応答に含まれるドメイン名又はIPアドレスをノードとし、前記ドメイン名及び前記IPアドレスの間の対応関係を枝とするグラフを生成する生成部と、
- 前記グラフのノードを構成する各IPアドレスに関して前記第二の計測部によって計測されたデータ量と、前記グラフの中間ノードを除くノードを構成するドメイン名に関して前記第一の計測部によって計測されたDNS要求の数との関係を示す変換行列を推定する行列推定部と、
- 前記ドメイン名ごとの要求数に前記各ドメイン名のフローごとのデータ量を乗じた値が、前記変換行列に前記IPアドレスごとのデータ量を乗じた値に等しいとする関係に基づいて、前記各ドメイン名のフローごとのデータ量を求め、当該データ量に、ドメイン名ごとのDNS要求の数を乗じて、ドメイン名ごとのデータ量の推定値を算出する算出部と、
- を有することを特徴とする推定装置。
- [請求項2]        前記行列推定部は、前記グラフを隣接行列に変換し、前記隣接行列を前記グラフの最大ホップ長の数だけ掛け合わせて得られる行列のうち、前記グラフの中間ノードを除くドメイン名のノードとIPアドレスのノードとの組み合わせに対応する成分を前記変換行列の初期値として抽出し、抽出された初期値と、前記グラフのノードを構成する各IPアドレスに関して前記第二の計測部によって計測されたデータ量と、前記グラフの中間ノードを除くノードを構成するドメイン名に関

して前記第一の計測部によって計測されたDNS要求の数とに基づいて、前記変換行列を推定し、

前記隣接行列は、入次数が0であるノードの対角成分の値を1とし、入次数が1以上であるノードに入る枝に対応する成分の値を、当該枝の数で1を除すことにより得られる値とする、  
ことを特徴とする請求項1記載の推定装置。

[請求項3] 前記生成部は、前記DNS応答に含まれるドメイン名又はIPアドレスをノードとし、前記ドメイン名及び前記IPアドレスの間の対応関係を枝とする1以上の連結グラフを生成し、

前記行列推定部は、前記連結グラフごとに、前記変換行列を推定し、

前記算出部は、前記連結グラフごとに、当該連結グラフの中間ノードを除くノードを構成するドメイン名ごとの推定値を算出する、  
ことを特徴とする請求項1又は2記載の推定装置。

[請求項4] ネットワークにおいて観測されるDNS応答に基づいて、ドメイン名ごとのDNS要求の数を計測する第一の計測部と、

ネットワークにおいて観測されるフローについて、IPアドレスが共通する単位ごとに、フロー数及びデータ量の合計を計測する第二の計測部と、

前記ドメイン名ごとの要求数に前記各ドメイン名のフローごとのデータ量を乗じた値が、前記IPアドレスごとのデータ量に等しいとする関係に基づいて、前記各ドメイン名のフローごとのデータ量を求め、当該データ量に、ドメイン名ごとのDNS要求の数を乗じて、ドメイン名ごとのデータ量の推定値を算出する算出部と、  
を有することを特徴とする推定装置。

[請求項5] コンピュータが、

ネットワークにおいて観測されるDNS応答に基づいて、ドメイン名ごとのDNS要求の数を計測する第一の計測手順と、

ネットワークにおいて観測されるフローについて、ＩＰアドレスが共通する単位ごとに、フロー数及びデータ量の合計を計測する第二の計測手順と、

前記ＤＮＳ応答に含まれるドメイン名又はＩＰアドレスをノードとし、前記ドメイン名及び前記ＩＰアドレスの間の対応関係を枝とするグラフを生成する生成手順と、

前記グラフのノードを構成する各ＩＰアドレスに関して前記第二の計測手順によって計測されたデータ量と、前記グラフの中間ノードを除くノードを構成するドメイン名に関して前記第一の計測手順によって計測されたＤＮＳ要求の数との関係を示す変換行列を推定する行列推定手順と、

前記ドメイン名ごとの要求数に前記各ドメイン名のフローごとのデータ量を乗じた値が、前記変換行列に前記ＩＰアドレスごとのデータ量を乗じた値に等しいとする関係に基づいて、前記各ドメイン名のフローごとのデータ量を求め、当該データ量に、ドメイン名ごとのＤＮＳ要求の数を乗じて、ドメイン名ごとのデータ量の推定値を算出する算出手順と、

を実行することを特徴とする推定方法。

[請求項6]

コンピュータが、

ネットワークにおいて観測されるＤＮＳ応答に基づいて、ドメイン名ごとのＤＮＳ要求の数を計測する第一の計測手順と、

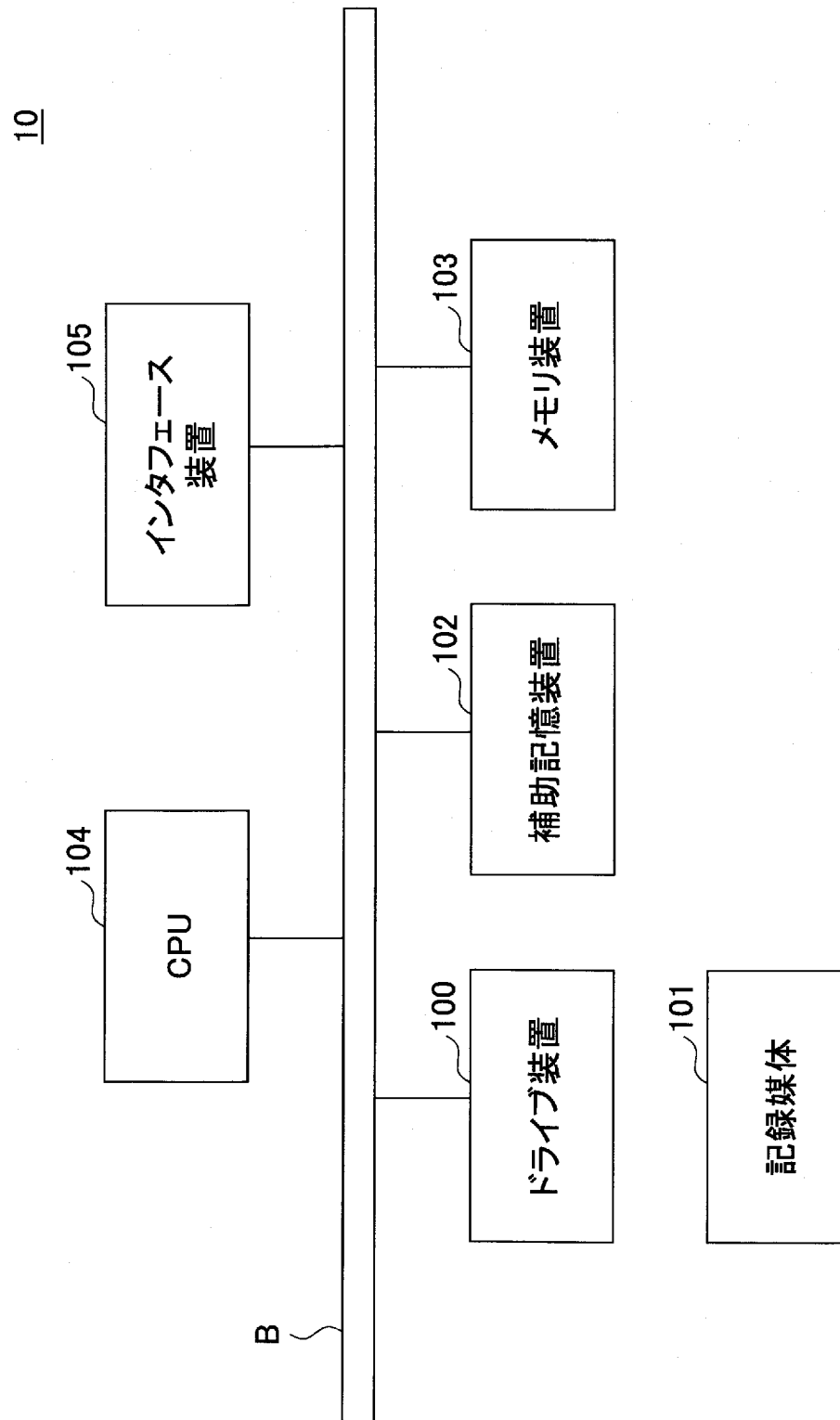
ネットワークにおいて観測されるフローについて、ＩＰアドレスが共通する単位ごとに、フロー数及びデータ量の合計を計測する第二の計測手順と、

前記ドメイン名ごとの要求数に前記各ドメイン名のフローごとのデータ量を乗じた値が、前記ＩＰアドレスごとのデータ量に等しいとする関係に基づいて、前記各ドメイン名のフローごとのデータ量を求め、当該データ量に、ドメイン名ごとのＤＮＳ要求の数を乗じて、ドメ

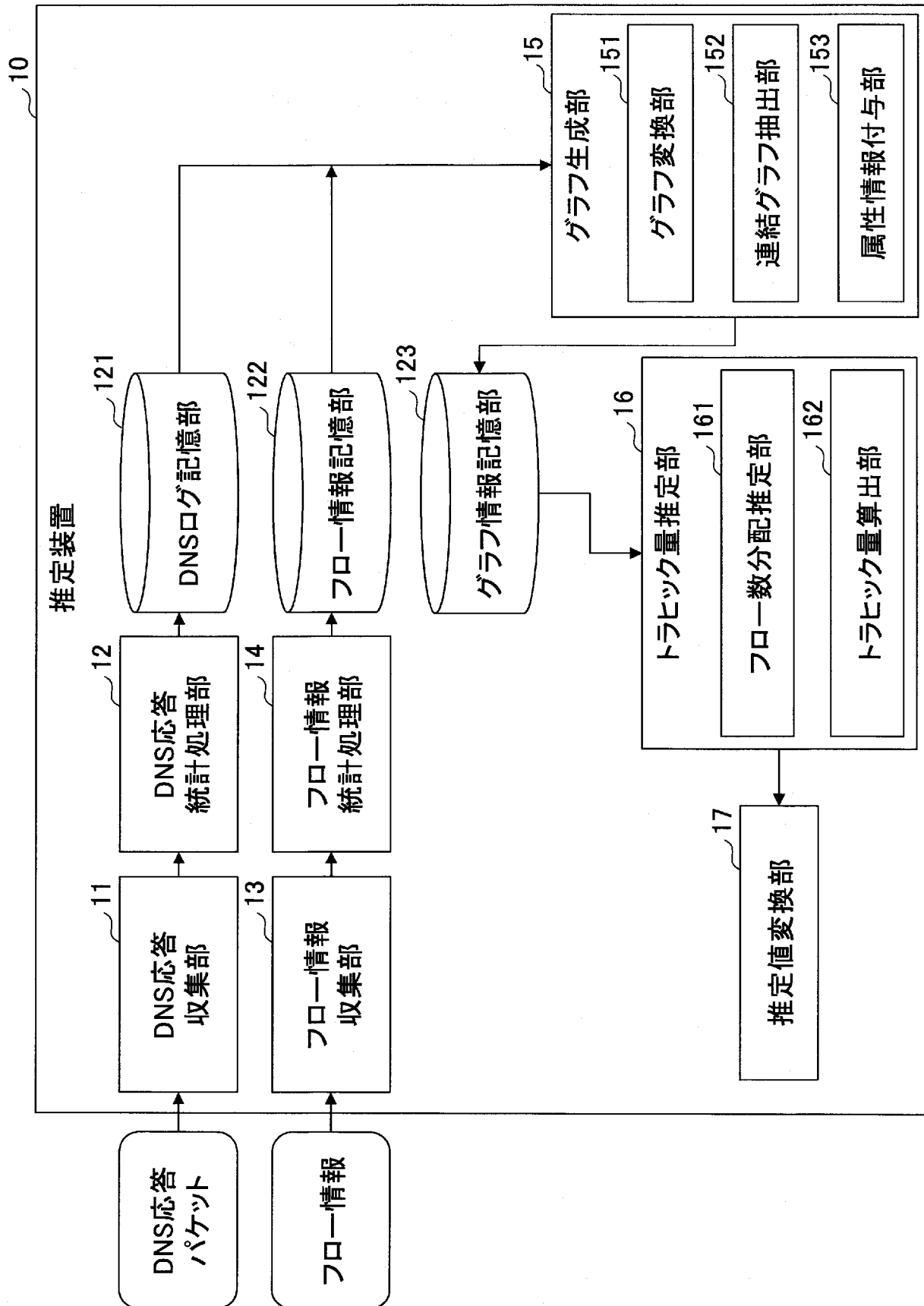
イン名ごとのデータ量の推定値を算出する算出手順と、  
を実行することを特徴とする推定方法。

[請求項7]       請求項1乃至4いずれか一項記載の各部としてコンピュータを機能  
させるためのプログラムを記録したコンピュータ読み取り可能な記録  
媒体。

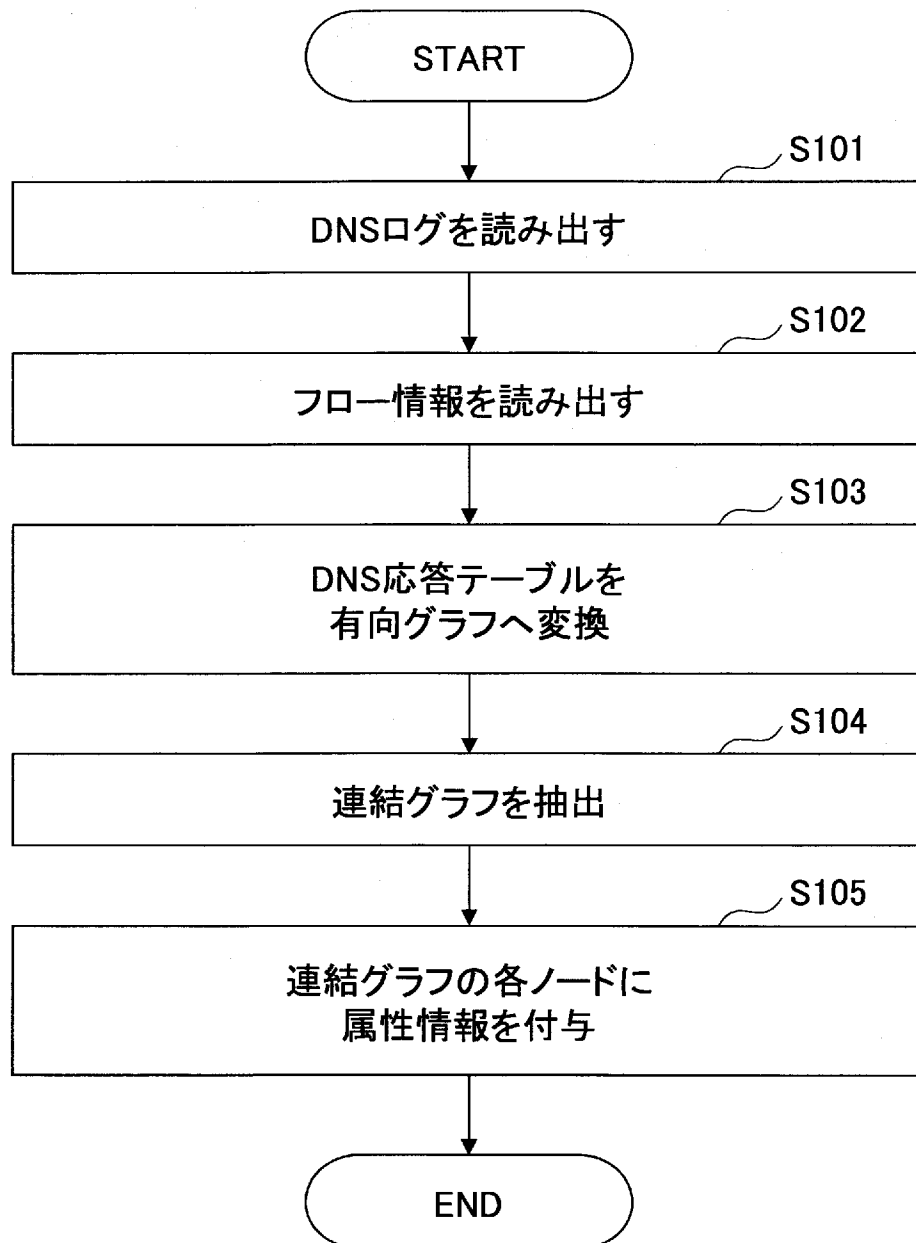
[図1]



[図2]



[図3]



[図4]

121

T1

| ドメイン名       | クエリタイプ | 要求数 | ユーザ数 |
|-------------|--------|-----|------|
| www.A.com   | A      | Q1  | U1   |
| www.B.com   | A      | Q2  | U2   |
| www.cdn.net | A      | Q3  | U3   |

T2

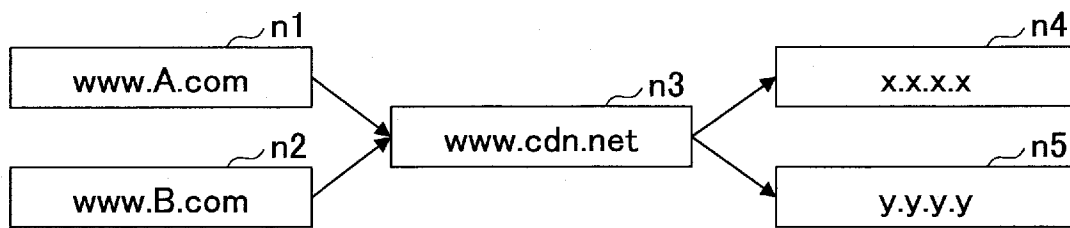
| ドメイン名       | TTL  | レコードタイプ | レコードデータ     |
|-------------|------|---------|-------------|
| www.A.com   | 3600 | CNAME   | www.cdn.net |
| www.B.com   | 600  | CNAME   | www.cdn.net |
| www.cdn.net | 30   | A       | x.x.x.x     |
| www.cdn.net | 30   | A       | y.y.y.y     |

[図5]

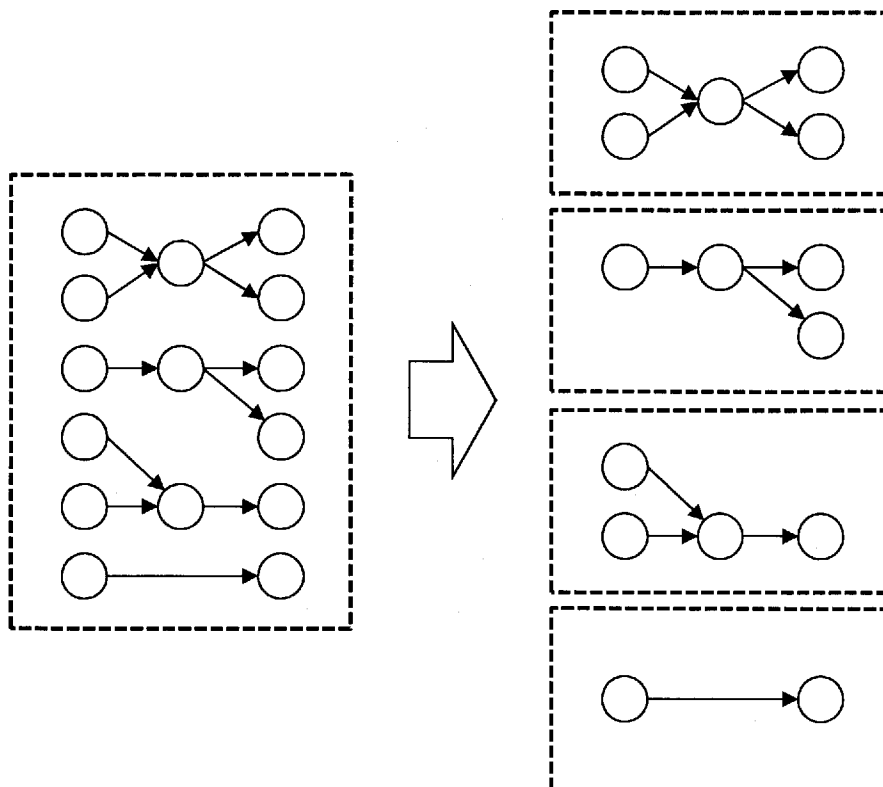
122

| IPアドレス  | フロー数 | ユーザ数 | バイト数 |
|---------|------|------|------|
| x.x.x.x | F4   | U4   | B4   |
| y.y.y.y | F5   | U5   | B5   |

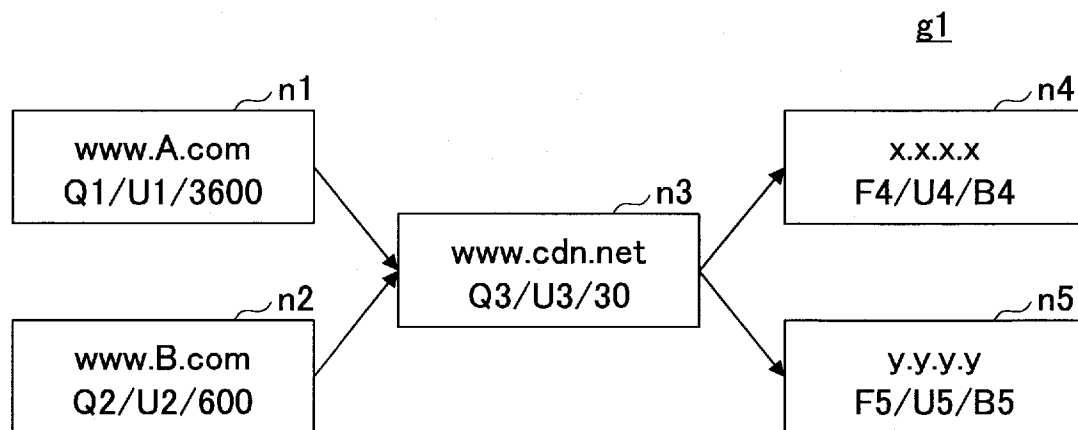
[図6]



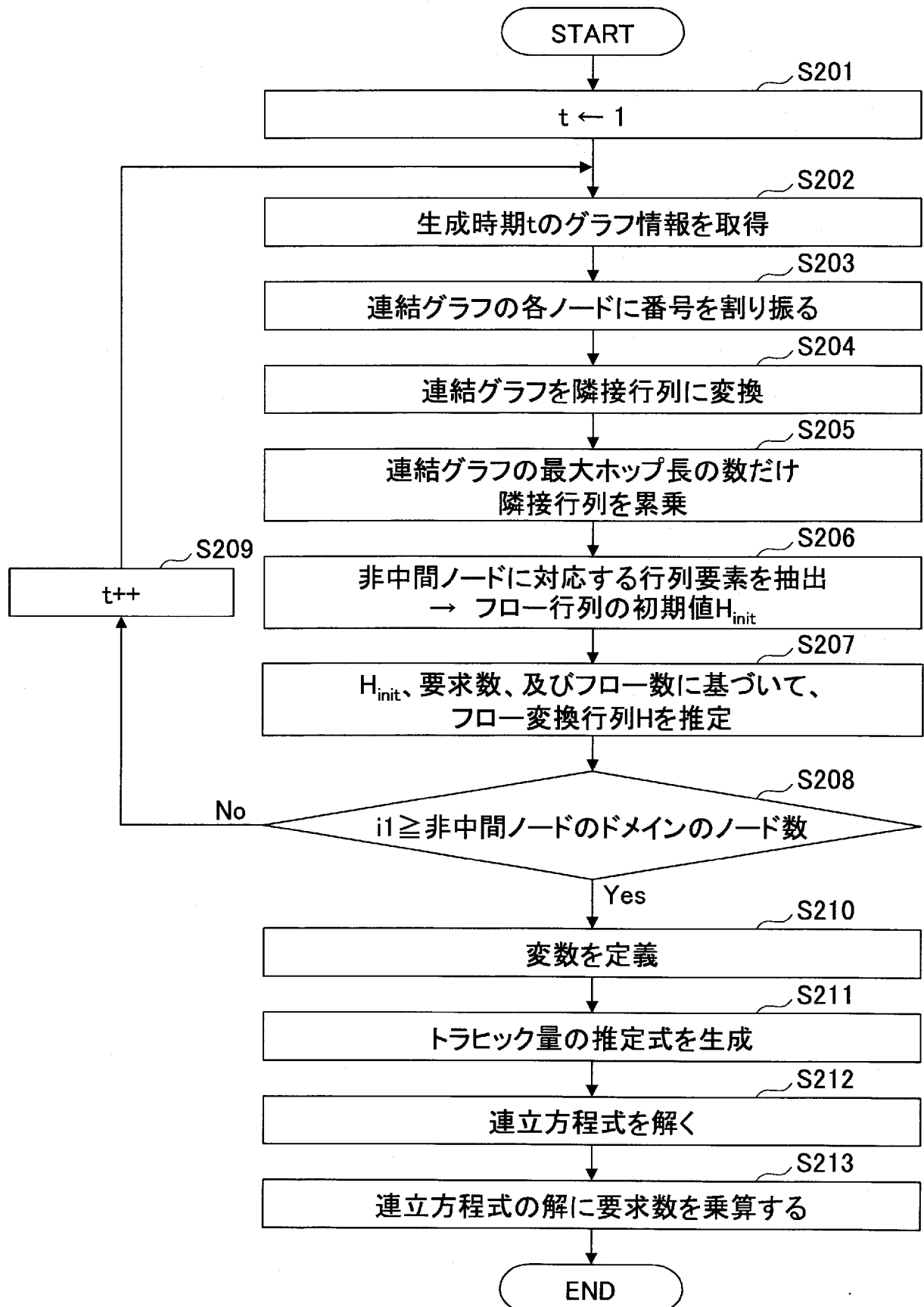
[図7]



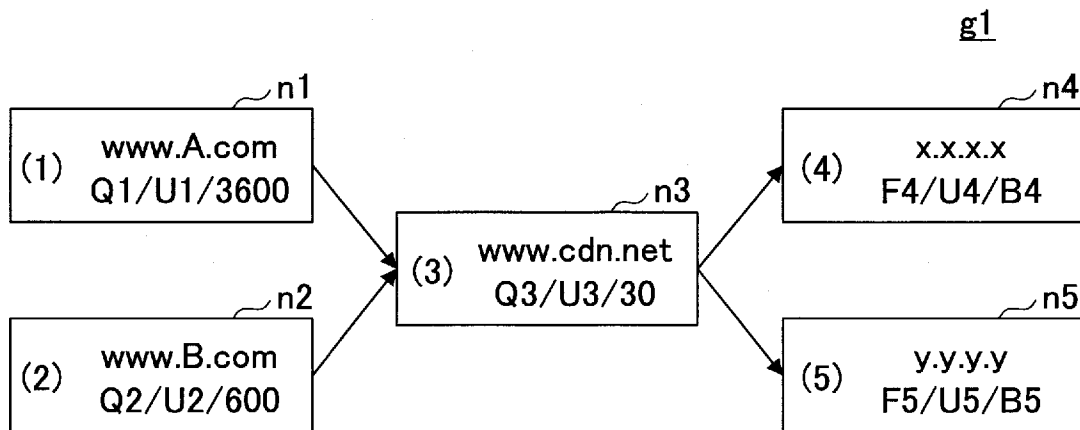
[図8]



[図9]



[図10]



[図11]

シンク側 (j)

|          | (1) | (2) | (3) | (4) | (5) |
|----------|-----|-----|-----|-----|-----|
| ソース側 (i) |     |     |     |     |     |
| (1)      | 1   |     | 0.5 |     |     |
| (2)      |     | 1   | 0.5 |     |     |
| (3)      |     |     |     | 1   | 1   |
| (4)      |     |     |     |     |     |
| (5)      |     |     |     |     |     |

$A =$

[図12]

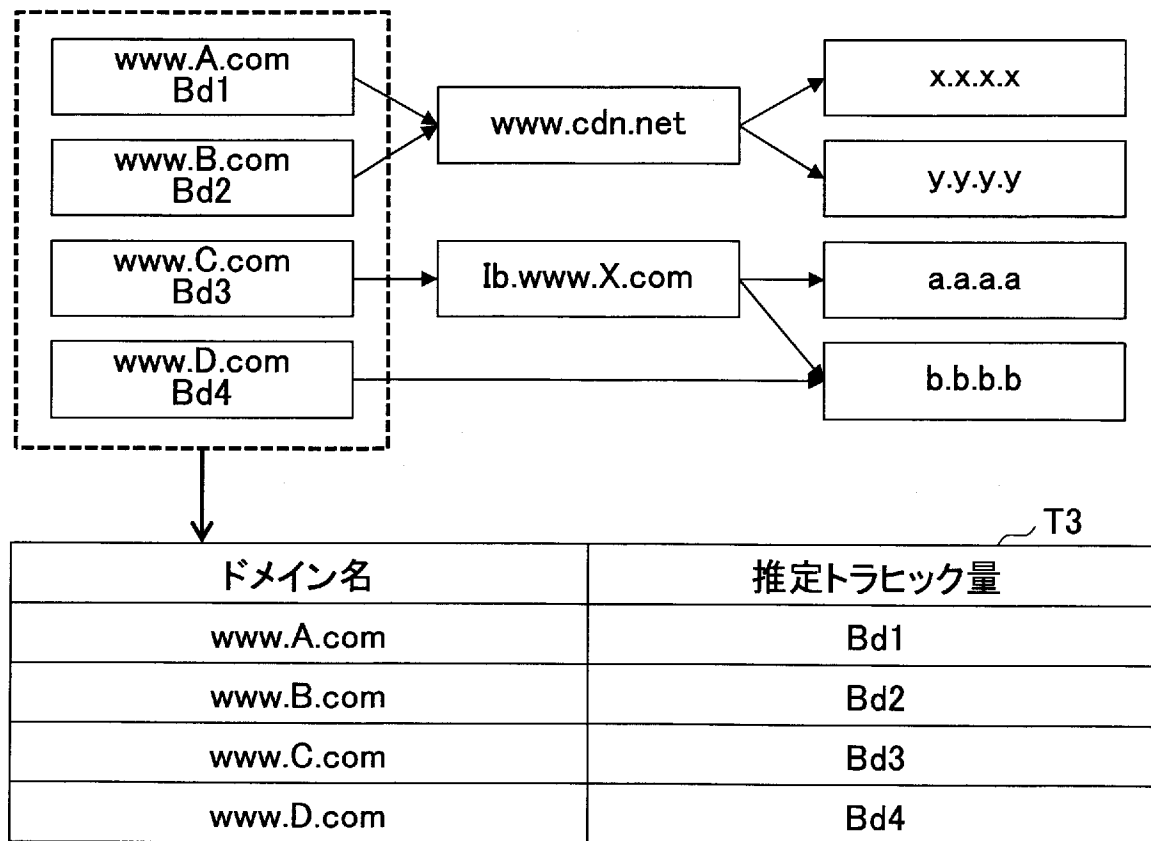
|     | (1) | (2) | (3) | (4) | (5) |
|-----|-----|-----|-----|-----|-----|
| (1) | 1   |     | 0.5 | 0.5 | 0.5 |
| (2) |     | 1   | 0.5 | 0.5 | 0.5 |
| (3) |     |     |     |     |     |
| (4) |     |     |     |     |     |
| (5) |     |     |     |     |     |

$A' =$

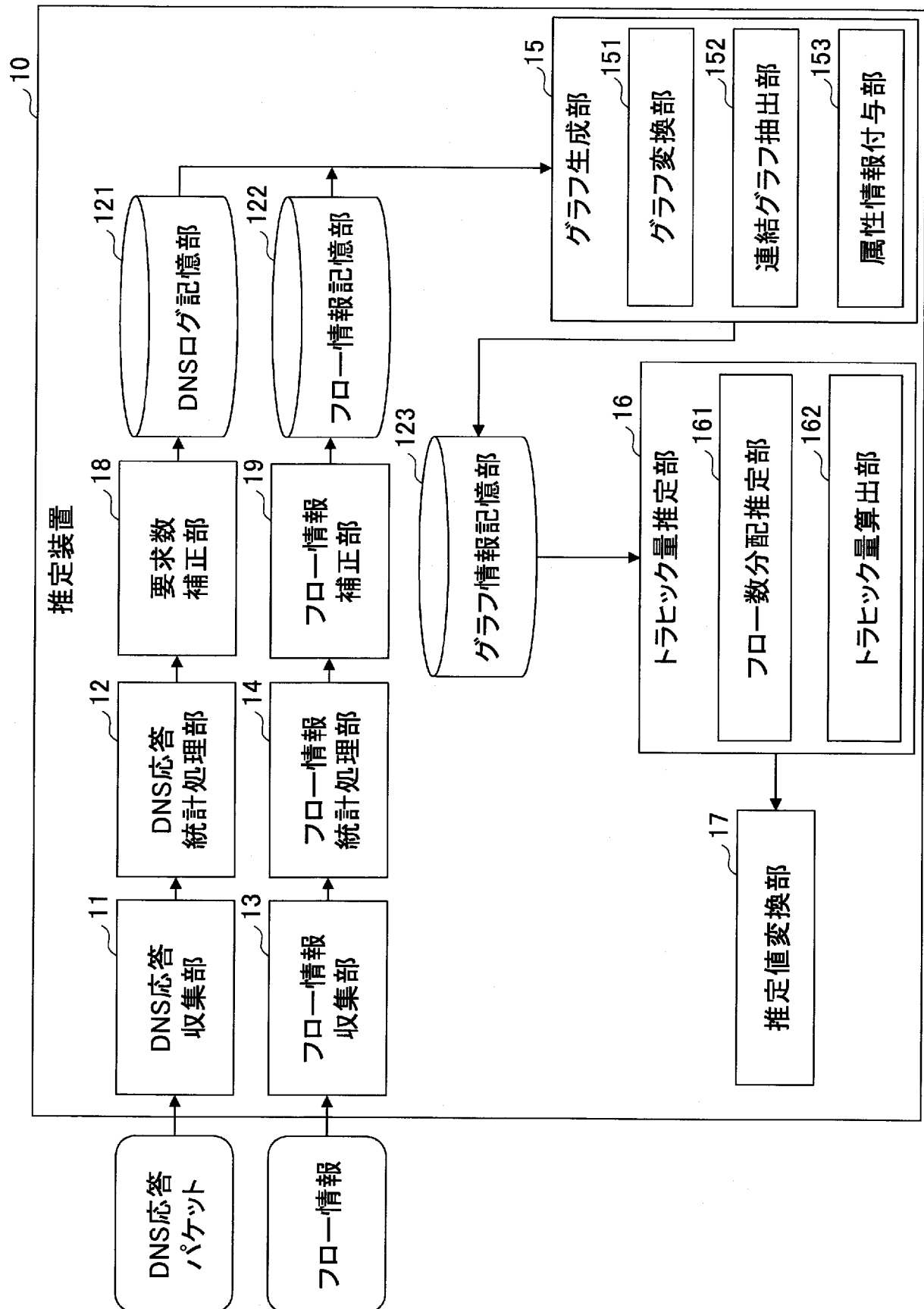
|     | (4) | (5) |
|-----|-----|-----|
| (1) | 0.5 | 0.5 |
| (2) | 0.5 | 0.5 |

$= H_{init}$

[図13]



[図14]



# INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2016/054373

## A. CLASSIFICATION OF SUBJECT MATTER

H04L12/70(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)

H04L12/70

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 | 1996-2016 |
| Kokai Jitsuyo Shinan Koho | 1971-2016 | Toroku Jitsuyo Shinan Koho | 1994-2016 |

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 2013-157931 A (Nippon Telegraph and Telephone Corp.),<br>15 August 2013 (15.08.2013),<br>entire text; all drawings<br>(Family: none)                                | 1-7                   |
| A         | JP 2006-129533 A (Niksun, Inc.),<br>18 May 2006 (18.05.2006),<br>entire text; all drawings<br>& US 2002/0105911 A1<br>entire text; all drawings<br>& WO 2000/031963 A1 | 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  
15 April 2016 (15.04.16)

Date of mailing of the international search report  
26 April 2016 (26.04.16)

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

International application No.

PCT/JP2016/054373

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 2006-13876 A (Nippon Telegraph and Telephone Corp.),<br>12 January 2006 (12.01.2006),<br>entire text; all drawings<br>(Family: none)                                                                                     | 1-7                   |
| A         | JP 2005-348416 A (Lucent Technologies Inc.),<br>15 December 2005 (15.12.2005),<br>entire text; all drawings<br>& US 2005/0270984 A1<br>entire text; all drawings<br>& US 2005/0270985 A1 & EP 1603273 A1<br>& EP 1603274 A1 | 1-7                   |

## A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int.Cl. H04L12/70(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. H04L12/70

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 | 1971-2016年 |
| 日本国実用新案登録公報 | 1996-2016年 |
| 日本国登録実用新案公報 | 1994-2016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<br>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<br>請求項の番号 |
|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| A               | JP 2013-157931 A (日本電信電話株式会社) 2013.08.15, 全文、全図 (ファミリーなし)                                                                                     | 1-7            |
| A               | JP 2006-129533 A (ニクスン インコーポレイテッド) 2006.05.18, 全文、全図 & US 2002/0105911 A1 全文、全図 & WO 2000/031963 A1                                           | 1-7            |
| A               | JP 2006-13876 A (日本電信電話株式会社) 2006.01.12, 全文、全図 (ファミリーなし)                                                                                      | 1-7            |
| A               | JP 2005-348416 A (ルーセント テクノロジーズ インコーポレイテッド) 2005.12.15, 全文、全図 & US 2005/0270984 A1 全文、全図 & US 2005/0270985 A1 & EP 1603273 A1 & EP 1603274 A1 | 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

|                                                                |                                                                     |
|----------------------------------------------------------------|---------------------------------------------------------------------|
| 「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

15.04.2016

国際調査報告の発送日

26.04.2016

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)

宮島 郁美

5 X

8523

電話番号 03-3581-1101 内線 3596