

(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
2019年9月26日(26.09.2019)



(10) 国際公開番号

WO 2019/180804 A1

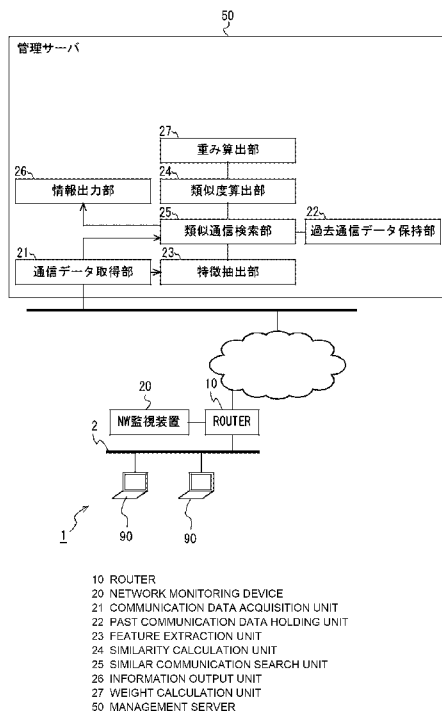
- (51) 国際特許分類:
H04L 12/70 (2013.01) *G06F 21/56* (2013.01)
- (21) 国際出願番号: PCT/JP2018/010961
- (22) 国際出願日: 2018年3月20日(20.03.2018)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 株式会社 P F U (PFU LIMITED) [JP/JP]; 〒9291192 石川県かほく市宇野気ヌ98番地の2 Ishikawa (JP).
- (72) 発明者: 高山 龍二 (TAKAYAMA, Ryuji); 〒9291192 石川県かほく市宇野気ヌ98番地の2 株式会社 P F U 内 Ishikawa (JP).
- (74) 代理人: 畑添 隆人, 外 (HATAZOE, Takahito et al.); 〒1020093 東京都千代田区平河町二丁目10番6号 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, 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, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, 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).

(54) Title: INFORMATION PROCESSING DEVICE, COMMUNICATION TYPE DISCRIMINATION METHOD, AND PROGRAM

(54) 発明の名称: 情報処理装置、通信種類判別方法及びプログラム



(57) Abstract: The present invention addresses the problem of classifying communication data which in the prior art had been difficult to discern as belonging to different types of communication. The information processing device is provided with: a communication data acquisition unit for acquiring communication data for communication by a terminal connected to a network; a past communication data holding unit for holding past communication data respectively related to a plurality of past instances of communication in association with the corresponding communication type information; a similarity calculation unit for calculating comprehensive similarity between communication of interest and the past instances of communication by comparing the communication data acquired by the communication data acquisition unit with the past communication data to calculate feature similarity for each discrimination point included in the communication, and weighting the calculated feature similarity for each discrimination point or each combination of discrimination points; a similar communication search unit for searching for one or more past instances of communication similar to the communication of interest, in accordance with the calculated comprehensive similarity; and an information output unit for outputting the communication type information corresponding to the retrieved past instance of communication.

添付公開書類：

一 国際調査報告（条約第21条(3)）

(57) 要約：従来の技術では異なる種類の通信であると判別することが難しかった通信データを分類することを課題とし、情報処理装置に、ネットワークに接続された端末による通信の通信データを取得する通信データ取得部と、複数の過去通信の夫々に係る過去通信データを、対応する通信種類情報と関連付けて保持する過去通信データ保持部と、通信データ取得部によって取得された通信データと、過去通信データとを比較して、通信に含まれる判別ポイント毎の特徴類似度を算出し、算出された特徴類似度を判別ポイント毎又は判別ポイントの組み合わせ毎に重み付けすることで、対象通信と過去通信との間の総合類似度を算出する類似度算出部と、算出された総合類似度に従って、対象通信に類似する1又は複数の過去通信を索出する類似通信検索部と、索出された過去通信に対応する通信種類情報を出力する情報出力部と、を備えた。

明 細 書

発明の名称： 情報処理装置、通信種類判別方法及びプログラム
技術分野

[0001] 本開示は、ネットワーク上の通信を分類するための技術に関する。

背景技術

[0002] 従来、検知されたアラートの通信情報と、過去のアラートの通信情報との類似性を示す技術が提案されている（特許文献1を参照）。

[0003] また、監視によって得られた通信データをテンプレートと比較し、類似度が閾値を超える場合に当該通信データとテンプレートとが関連すると判断する技術が提案されている（特許文献2を参照）。

先行技術文献

特許文献

[0004] 特許文献1：国際公開2016／092836号

特許文献2：米国特許公開2017／0230393号

発明の概要

発明が解決しようとする課題

[0005] 従来、シグネチャ照合によってシステムへの脅威を検知する技術が用いられているが、誤検知が少ないというメリットがある反面、未知のマルウェアを検知できないというデメリットがあった。一方、シグネチャ照合方式の弱点を克服するために、ネットワーク上の端末等の振る舞いの怪しさを判定することで、システムへの脅威を検知する技術も用いられているが、このような技術は、未知のマルウェアを検知可能であるというメリットの反面、誤検知が発生する可能性があるという問題がある。

[0006] これは、マルウェア等に係る通信の種類には、通信全体を見るとよく似ているが、特定の観点に絞って見ると異なる種類のマルウェアであると判別可能となるような、亜種や新種が多数存在するためである。このような亜種や新種を区別するために、教師あり学習技術を用いることも考えられるが、教

師あり学習において正確な識別が可能になる程度の大量のマルウェアの亜種や新種による通信データを、学習用データとして入手することは困難であり、教師あり学習技術を用いて多くの種類のマルウェアの亜種や新種を判別可能とすることは現実的ではない。

[0007] 本開示は、上記した問題に鑑み、従来の技術では異なる種類の通信であると判別することが難しかった通信データを分類することを課題とする。

課題を解決するための手段

[0008] 本開示の一例は、ネットワークに接続された端末による通信の通信データを取得する通信データ取得手段と、複数の過去通信の夫々に係る過去通信データを、対応する通信種類情報と関連付けて保持する過去通信データ保持手段と、前記通信データ取得手段によって取得された通信データと、前記過去通信データとを比較して、通信に含まれる判別ポイント毎の特徴類似度を算出し、算出された特徴類似度を該判別ポイント毎又は該判別ポイントの組み合わせ毎に重み付けすることで、該対象通信と該過去通信との間の総合類似度を算出する類似度算出手段と、算出された総合類似度に従って、対象通信に類似する1又は複数の過去通信を索出する類似通信検索手段と、索出された過去通信に対応する通信種類情報を出力する情報出力手段と、を備える情報処理装置である。

[0009] 本開示は、情報処理装置、システム、コンピューターによって実行される方法またはコンピューターに実行させるプログラムとして把握することが可能である。また、本開示は、そのようなプログラムをコンピューターその他の装置、機械等が読み取り可能な記録媒体に記録したものとしても把握できる。ここで、コンピューター等が読み取り可能な記録媒体とは、データやプログラム等の情報を電氣的、磁氣的、光学的、機械的または化学的作用によって蓄積し、コンピューター等から読み取ることができる記録媒体をいう。

発明の効果

[0010] 本開示によれば、従来の技術では異なる種類の通信であると判別することが難しかった通信データを分類することが可能となる。

図面の簡単な説明

[0011] [図1]実施形態に係るシステムの構成を示す概略図である。

[図2]実施形態に係るネットワーク監視装置および管理サーバーのハードウェア構成を示す図である。

[図3]第一の実施形態に係る管理サーバーの機能構成の概略を示す図である。

[図4]実施形態に係る通信種類判別処理の流れの概要を示すフローチャートである。

[図5]実施形態に係る特徴抽出処理の流れの概要を示すフローチャートである。

[図6]第一の実施形態に係る類似通信検索処理の流れの概要を示すフローチャートである。

[図7]実施形態に係る類似度算出処理の流れの概要を示すフローチャートである。

[図8]実施形態に係る重み算出処理の流れの概要を示すフローチャートである。

[図9]第二の実施形態に係る管理サーバーの機能構成の概略を示す図である。

[図10]第二の実施形態に係る類似通信検索処理の流れの概要を示すフローチャートである。

[図11]第三の実施形態に係る管理サーバーの機能構成の概略を示す図である。

[図12]第三の実施形態に係る類似通信検索処理の流れの概要を示すフローチャートである。

[図13]第四の実施形態に係る管理サーバーの機能構成の概略を示す図である。

[図14]第四の実施形態に係る類似通信検索処理の流れの概要を示すフローチャートである。

[図15]実施形態に係るシステムの構成のバリエーションを示す概略図である。

発明を実施するための形態

[0012] 以下、本開示に係る情報処理装置、通信種類判別方法及びプログラムの実施の形態を、図面に基づいて説明する。但し、以下に説明する実施の形態は、実施形態を例示するものであって、本開示に係る情報処理装置、通信種類判別方法及びプログラムを以下に説明する具体的構成に限定するものではない。実施にあたっては、実施の態様に応じた具体的構成が適宜採用され、また、種々の改良や変形が行われてよい。

[0013] 本実施形態では、本開示に係る情報処理装置、通信種類判別方法及びプログラムを、ネットワーク上で不正な活動を行っている端末を発見し、通信遮断やアラート通知等の対処を行うためのシステムにおいて実施した場合の実施の形態について説明する。但し、本開示に係る情報処理装置、通信種類判別方法及びプログラムは、ネットワーク上の通信を分類するために広く用いることが可能であり、本開示の適用対象は、本実施形態において示した例に限定されない。

[0014] [第一の実施形態]

<システムの構成>

図1は、本実施形態に係るシステム1の構成を示す概略図である。本実施形態に係るシステム1は、複数の情報処理端末90（以下、「ノード90」と称する）が接続されるネットワークセグメント2と、ノード90に係る通信を監視するためのネットワーク監視装置20（通信監視装置）と、ルータ10を介してネットワークセグメント2と通信可能に接続された管理サーバー50と、を備える。

[0015] 本実施形態において、ネットワーク監視装置20は、スイッチまたはルータ（図1に示した例では、ルータ）のモニタリングポート（ミラーポート）に接続されることで、ノード90によって送受信されるパケットやフレーム等の通信データを取得し、取得した通信データを管理サーバー50に送信する。この場合、ネットワーク監視装置20は、取得したパケットを転送しないパッシブモードで動作してもよい。

- [0016] 管理サーバー５０は、ネットワーク監視装置２０から情報を収集し、ネットワーク監視装置２０を管理する。なお、外部ネットワークには、更に検疫サーバーが設けられ、ネットワークセグメント２に接続されたノード９０に対して検疫サービスを提供してもよいし、業務サーバーが設けられ、ノード９０に対して業務のためのサービスを提供してもよい（図示は省略する）。
- [0017] 本実施形態に係るシステム１では、ノード９０から接続される各種サーバーは、インターネットや広域ネットワークを介して遠隔地において接続されたものであり、例えばＡＳＰ（Application Service Provider）によって提供されるが、これらのサーバーは、必ずしも遠隔地に接続されたものである必要はない。例えば、これらのサーバーは、ノード９０やネットワーク監視装置２０が存在するローカルネットワーク上に接続されていてもよい。
- [0018] 図２は、本実施形態に係るネットワーク監視装置２０および管理サーバー５０のハードウェア構成を示す図である。なお、図２においては、ネットワーク監視装置２０および管理サーバー５０以外の構成（ルータ１０、ノード９０等）については、図示を省略している。ネットワーク監視装置２０および管理サーバー５０は、それぞれ、ＣＰＵ（Central Processing Unit）１１ａ、１１ｂ、ＲＡＭ（Random Access Memory）１３ａ、１３ｂ、ＲＯＭ（Read Only Memory）１２ａ、１２ｂ、ＥＥＰＲＯＭ（Electrically Erasable and Programmable Read Only Memory）やＨＤＤ（Hard Disk Drive）等の記憶装置１４ａ、１４ｂ、ＮＩＣ（Network Interface Card）１５ａ、１５ｂ等の通信ユニット、等を備えるコンピューターである。
- [0019] 図３は、本実施形態に係る管理サーバー５０の機能構成の概略を示す図である。管理サーバー５０は、記憶装置１４ｂに記録されているプログラムが、ＲＡＭ１３ｂに読み出され、ＣＰＵ１１ｂによって実行されて、管理サーバー５０に備えられた各ハードウェアが制御されることで、通信データ取得

部 2 1、過去通信データ保持部 2 2、特徴抽出部 2 3、類似度算出部 2 4、類似通信検索部 2 5、情報出力部 2 6、及び重み算出部 2 7 を備える情報処理装置として機能する。なお、本実施形態及び後述する他の実施形態では、管理サーバー 5 0 の備える各機能は、汎用プロセッサである CPU 1 1 b によって実行されるが、これらの機能の一部または全部は、1 または複数の専用プロセッサによって実行されてもよい。また、本実施形態では、本開示に係る情報処理装置が管理サーバー 5 0 として実施される例について説明したが、本開示に係る情報処理装置が備える上記各機能の一部又は全部は、ネットワーク監視装置 2 0 に備えられていてもよい。

[0020] 通信データ取得部 2 1 は、監視対象ネットワークに設置されたネットワーク監視装置 2 0 によってキャプチャされ、管理サーバー 5 0 宛に送信された通信データを取得する。この通信データは、監視対象ネットワークに接続されたノード 9 0 による通信の通信データである。

[0021] 過去通信データ保持部 2 2 は、複数の過去通信の夫々に係る過去通信データを、対応する通信種類情報と関連付けて保持する。より具体的には、過去通信データ保持部 2 2 には、過去通信毎に、当該過去通信の ID、特徴データ、通信データ、対応する通信種類情報（マルウェア名／誤検知種別等）、当該マルウェア等への対処方法、等が関連づけられて保持されている。なお、ここで保持される過去通信データは、本開示に係るシステム 1 の内部に蓄積されたデータであってもよいし、外部のデータベースに蓄積されたデータが取得されたものであってもよい。

[0022] 特徴抽出部 2 3 は、通信データ取得部 2 1 によって取得された通信データから、通信に含まれる判別ポイント毎に、予め設定された特徴抽出方法を用いて特徴を抽出することで、対象通信の特徴データを生成する。

[0023] 類似度算出部 2 4 は、通信に含まれる判別ポイント毎に、対象通信の特徴データと過去通信の特徴データとを比較して特徴類似度を算出し、算出された特徴類似度を当該判別ポイント毎又は当該判別ポイントの組み合わせ毎に重み付けすることで、当該対象通信と当該過去通信との間の総合類似度を算

出する。

[0024] 類似通信検索部 25 は、算出された総合類似度に従って、対象通信に類似する 1 又は複数の過去通信を索出する。

[0025] 情報出力部 26 は、索出された過去通信に対応する通信種類情報、及び当該過去通信に関連付けられた対処方法を取得して出力する。

[0026] 重み算出部 27 は、判別ポイント毎の重みを算出する。より具体的には、重み算出部 27 は、類似通信検索部 25 によって索出される過去通信の対象範囲に含まれる過去通信同士のペアを複数作成し、通信種類情報が一致する過去通信ペアについて類似度算出部 24 が算出する総合類似度が 100% に近づき、通信種類情報が一致しない過去通信ペアについて類似度算出部 24 が算出する総合類似度が 0% に近づくように、重回帰分析を実行することによって、判別ポイント毎の重みを算出する。

[0027] <処理の流れ>

次に、本実施形態に係る管理サーバー 50 によって実行される処理の流れを説明する。なお、以下に説明する処理の具体的な内容および処理順序は、本開示を実施するための一例である。具体的な処理内容および処理順序は、本開示の実施の形態に応じて適宜選択されてよい。

[0028] 図 4 は、本実施形態に係る通信種類判別処理の流れの概要を示すフローチャートである。本フローチャートに示された処理は、ノード 90 による通信の通信データが、ネットワーク監視装置 20 によって取得され、管理サーバー 50 に入力されたことを契機として開始される。

[0029] ステップ S1 及びステップ S2 では、通信データが受信され、通信の特徴が抽出される。通信データ取得部 21 は、ネットワーク監視装置 20 によって取得されて、管理サーバー 50 宛に送信された、ノード 90 による通信データを受信（ステップ S1）する。以下、ここで取得された通信データに係る通信を、「対象通信」と称する。通信データが受信されると、特徴抽出部 23 は、受信された通信データから、通信に含まれる判別ポイント毎に特徴抽出処理を実行する（ステップ S2）。特徴抽出処理では、判別ポイント毎

の特徴が抽出され、特徴データが生成されるが、処理の詳細については後述する。その後、処理はステップS 3へ進む。

[0030] ステップS 3では、類似通信検索処理が実行される。類似通信検索部2 5は、過去通信データ保持部2 2によって保持されている過去通信の特徴データと、ステップS 2で生成された特徴データとを比較することで、ステップS 1で取得された通信データに類似する過去通信を索出する。なお、本実施形態では、過去通信の特徴データが過去通信データ保持部2 2によって保持されている例について説明したが、過去通信の特徴データは、本フローチャートに示された処理が実行される都度抽出されてもよい。類似通信検索処理の詳細については後述する。類似通信検索部2 5は、処理結果として、索出された過去通信のID、及び対象通信との類似度（総合類似度）を出力する。その後、処理はステップS 4へ進む。

[0031] ステップS 4では、検索結果が出力される。情報出力部2 6は、ステップS 3で索出された過去通信のIDに基づいて、当該過去通信の通信データ、対応する通信種類情報（マルウェア名／誤検知種別等）、及び当該マルウェア等への対処方法をデータベースから取得し、当該過去通信のID及び対象通信との類似度と併せて出力する。ユーザーは、出力内容から、対象通信の種類（マルウェア名等）、及び当該マルウェア等への対処方法を知ることができる。なお、対象通信の通信データ及び対応する通信種類情報等は、過去通信データとして過去通信データ保持部2 2に蓄積されてよい。その後、本フローチャートに示された処理は終了する。

[0032] 図5は、本実施形態に係る特徴抽出処理の流れの概要を示すフローチャートである。本フローチャートは、図4のステップS 2に示された特徴抽出処理の詳細を説明するものである。

[0033] ステップS 101では、特徴抽出設定が読み込まれる。特徴抽出部2 3は、判別ポイント毎に予め設定されている特徴抽出設定を、記憶装置1 4 bから読み込む。具体的には、特徴抽出設定には、判別ポイントの識別子（例えば、特徴名）及び当該特徴を抽出するための特徴抽出方法が含まれる。なお

、特徴抽出方法には、特徴抽出に用いられる分析手法（アルゴリズム）の他、特徴として設定されるパラメーターのルールも含まれてよい。その後、処理はステップS102へ進む。

[0034] ここで、判別ポイントは、通信内容のうち、当該通信の種類（例えば、マルウェア名）を特定する際に参照されるポイントであり、例えば、HTTPメソッド、URIパス、HTTPバージョン、リクエストパラメータ名、リクエストパラメータ値、リクエストヘッダー名、等である。本実施形態では、これらの判別ポイント毎に、適切な特徴抽出方法が予め設定されている。

[0035] 特徴抽出手法には、判別ポイントの種類に応じて適切な分析手法が割り当てられることが好ましい。採用可能な分析手法は限定されず、例えば、カテゴリカルデータ解析、文字Nグラム／単語Nグラム等を利用したテキスト分析、離散化を用いた分析、等が採用されてよい。例えば、URIパスの特徴抽出手法としてはNグラムが用いられ、HTTPバージョンの特徴抽出手法としてはカテゴリカルデータ解析が用いられてよい。

[0036] ステップS102及びステップS103では、特徴が抽出され、出力される。特徴抽出部23は、対象通信の通信データを、特徴抽出設定に従って分析することで、判別ポイント毎の特徴を抽出し、特徴データを生成する（ステップS102）。なお、特徴抽出の結果、対象通信の判別ポイント毎に設定される特徴データ（パラメータ）は、特徴抽出設定内のパラメータルールに従って決定されてよい。例えば、判別ポイントがカテゴリカルである場合、当該判別ポイントの特徴データにはカテゴリ毎に予め設定された値が与えられるし、その他、特徴データには、該当する判別ポイントの特徴抽出方法毎に適した特徴データ（例えば、ベクトル等）が与えられる。特徴抽出が完了すると、特徴抽出部23は、判別ポイント毎に、判別ポイントの識別子（例えば、特徴名）及び特徴データ（ベクトル等のパラメータ）を出力する（ステップS103）。その後、本フローチャートに示された処理は終了する。

[0037] 図6は、本実施形態に係る類似通信検索処理の流れの概要を示すフローチ

ャートである。本フローチャートは、図4のステップS3に示された類似通信検索処理の詳細を説明するものである。

[0038] ステップS201及びステップS202では、類似検索設定が読み込まれ、検索対象の過去通信が取得される。類似通信検索部25は、対象通信の類似通信検索のために予め設定された類似検索設定を、記憶装置14bから読み込む（ステップS201）。具体的には、類似検索設定には、検索範囲、判別ポイント毎に設定された重み（係数）、及び足切り条件が含まれる。次に、類似通信検索部25は、過去通信データ保持部22に蓄積された過去通信のうち、ステップS201で設定された検索範囲内の過去通信を、検索対象としてRAM13bに読み込む（ステップS202）。その後、処理はステップS203へ進む。

[0039] ここで、検索範囲には、過去通信データ保持部22に蓄積された過去通信を検索する際の検索範囲が設定される。本実施形態では、「過去通信データ保持部22が保持する過去通信データ全て」を検索対象とする例について説明するが、検索範囲は、期間や通信環境その他の条件を用いて絞り込まれてもよい。

[0040] 重みは、判別ポイント毎、または複数の判別ポイントの組み合わせ毎に設定される。本実施形態では、例えば、「HTTPメソッド：0.04」、「URIパス：0.08」、「HTTPバージョン：0.07」、「リクエストパラメータ名：0.07」、「リクエストパラメータ値：0.050」・・・等のように設定される。

[0041] 足切り条件は、検索結果をフィルタリングするために用いられる条件であり、検索結果の件数の範囲や類似度の範囲を設定することが出来る。例えば、「総合類似度上位30件」や「総合類似度N%以上」のような足切り条件が設定されてよい。

[0042] ステップS203では、類似度算出処理が実行される。類似度算出部24は、対象通信と、ステップS202で読み込まれた複数の過去通信の夫々との間の類似度（総合類似度）を算出する。類似度算出処理の詳細については

後述する。その後、処理はステップS 2 0 4へ進む。

[0043] ステップS 2 0 4では、対象通信に類似する過去通信が索出される。類似通信検索部2 5は、ステップS 2 0 3で対象通信との類似度が算出された各過去通信の総合類似度を参照し、設定された足切り条件に従って過去通信を絞り込むことで、対象通信に類似する過去通信を索出する。例えば、対象通信との総合類似度が算出された各過去通信のうち、総合類似度上位3 0件の過去通信が、類似通信検索処理による検索結果として特定される。その後、本フローチャートに示された処理は終了する。

[0044] 図7は、本実施形態に係る類似度算出処理の流れの概要を示すフローチャートである。本フローチャートは、図6のステップS 2 0 3に示された類似度算出処理の詳細を説明するものである。

[0045] ステップS 3 0 1では、判別ポイント毎の類似度（以下、「特徴類似度」と称する）が算出される。類似度算出部2 4は、判別ポイント毎に、対象通信について抽出された特徴データと、検索対象となった過去通信の特徴データとを比較し、特徴類似度を算出する。具体的には、判別ポイントのパラメータの種類毎に適切な類似度算出手法が採用されることが好ましく、例えば、ユークリッド距離等、一般的な距離・類似度法が用いられてよい。全ての検索対象過去通信について特徴類似度が算出されると、処理はステップS 3 0 2へ進む。

[0046] ステップS 3 0 2では、総合類似度が算出される。類似度算出部2 4は、ステップS 3 0 1で算出された特徴類似度、及び判別ポイント毎に与えられた重みに基づいて、過去通信毎の総合類似度を算出する。本実施形態では、総合類似度として、特徴類似度の加重平均を算出することとする（具体的には、以下の式を参照）。

$$\text{総合類似度} = \text{SUM} (\text{特徴類似度}_1 * \text{重み}_1 : \text{特徴類似度}_n * \text{重み}_n) / \text{SUM} (\text{重み}_1 : \text{重み}_n)$$

ここで、「特徴類似度_n」は、判別ポイントnの特徴類似度を示し、「重み_n」は、判別ポイントnの重みを示す。

[0047] なお、総合類似度は、特徴類似度に、判別ポイント毎又は判別ポイントの組み合わせ毎に与えられる重みを反映して算出されたものであればよく、本実施形態において説明した例に限定されない。また、総合類似度は、単一の指標で表現されるものに限定されず、複数の指標の組み合わせによって表現されるものであってもよい。総合類似度が算出されると、本フローチャートに示された処理は終了する。

[0048] なお、本実施形態において、判別ポイント毎に与えられる重みは、類似通信検索処理が実行される前に予め類似検索設定として読み込まれるが、判別ポイント毎の重みは、類似通信検索処理中に動的に算出されてもよい。類似通信検索処理中に重みを算出する例については、第三の実施形態以降で後述するが、重みの算出方法については、事前算出であるか動的算出であるかにかかわらず、同じ重み算出処理を用いて算出可能である。

[0049] 図8は、本実施形態に係る重み算出処理の流れの概要を示すフローチャートである。本フローチャートは、類似検索設定に予め設定される重みを算出するための重み算出処理、又は後述する図12のステップS603及び図14のステップS704に示された重み算出処理の詳細を説明するものである。

[0050] ステップS401及びステップS402では、重み算出対象範囲が設定され、重み算出対象の過去通信が取得される。重み算出部27は、重み算出に用いられる過去通信データの取得範囲である重み算出対象範囲を設定する（ステップS401）。ここで設定される重み算出対象範囲は、当該重み算出処理で算出された重みが用いられる類似通信検索処理の検索範囲を含むことが好ましい。このため、本実施形態において、重み算出部27は、当該重み算出処理で算出された重みが用いられる類似通信検索処理において設定される検索範囲と同一の範囲を、重み算出対象範囲として取得する。次に、重み算出部27は、過去通信データ保持部22に蓄積された過去通信のうち、ステップS401で設定された重み算出対象範囲内の過去通信を、RAM13bに読み込む（ステップS402）。その後、処理はステップS403へ進

む。

[0051] ステップS403では、過去通信のペアが作成される。重み算出部27は、ステップS402で読み込まれた過去通信のペアを複数作成する。この際、重み算出部27は、過去通信の通信種類情報（マルウェア名／誤検知種別等）を参照して、互いに同一の通信種類情報を有する過去通信のペア、及び互いに異なる通信種類情報を有する過去通信のペア、を含む複数のペアを作成する。例えば、重み算出部27は、マルウェアAの過去通信とマルウェアAの過去通信とからなるペア、マルウェアAの過去通信とマルウェアBの過去通信とからなるペア、となるように、過去通信のペアを複数作成する。その後、処理はステップS404へ進む。

[0052] ステップS404では、特徴類似度が算出される。重み算出部27は、ステップS403で作成されたペア毎に、ペア間で判別ポイント毎の特徴類似度を算出する。具体的には、マルウェアA－A間の特徴類似度として、特徴類似度 $_{AA1}$ から特徴類似度 $_{AAn}$ を、マルウェアA－B間の特徴類似度として、特徴類似度 $_{AB1}$ から特徴類似度 $_{ABn}$ を、算出する。特徴抽出及び特徴類似度算出の具体的な方法については、上記で説明したものと概略同様であるため、説明を省略する。その後、処理はステップS405へ進む。

[0053] ステップS405では、判別ポイント毎の重みが算出される。重み算出部27は、ペア毎に以下の式を立て、通信種類情報が一致するペアについての総合類似度（マルウェアA－A間の総合類似度）が100%に近づき、通信種類情報が一致しないペアについての総合類似度（マルウェアA－B間の総合類似度）が0%に近づくように、重回帰分析を実行することによって、判別ポイント毎の重みを算出する。

マルウェアA－A間の総合類似度＝SUM（特徴類似度 $_{AA1}$ ＊重み $_1$ ：特徴類似度 $_{AAn}$ ＊重み $_n$ ）／SUM（重み $_1$ ：重み $_n$ ）

マルウェアA－B間の総合類似度＝SUM（特徴類似度 $_{AB1}$ ＊重み $_1$ ：特徴類似度 $_{ABn}$ ＊重み $_n$ ）／SUM（重み $_1$ ：重み $_n$ ）

ここで、「特徴類似度 $_{AAn}$ 」は、マルウェアA－A間における判別ポイント

n の特徴類似度を示し、「特徴類似度 $_{ABn}$ 」は、マルウェア A－B 間における判別ポイント n の特徴類似度を示す。

[0054] このようにすることで、例えば、マルウェア A とマルウェア B とが、通信全体を見るとよく似ているような場合であっても、これらのマルウェアを区別するために重要な判別ポイントが重視されるような重みを、自動的に算出することができる。

[0055] [第二の実施形態]

次に、第二の実施形態を説明する。第二の実施形態では、上記第一の実施形態で説明した内容を重複する項目については、同一の符号を付して説明を省略する。

[0056] 上記第一の実施形態において、重み算出処理を説明したが、上記説明した重み算出処理を用いる場合、重み算出対象範囲が広すぎる場合には、適切な重みを算出することが困難となる。

[0057] そこで、本実施形態及び後述する第四の実施形態では、クラスタリング処理（グループ分け処理）を採用し、検索範囲及び重み算出対象範囲を、「通信全体を見るとよく似ているが、実際には異なる種類の通信が含まれる」グループ内に限定した上で、当該グループ内で正確に通信種類を特定するのに重視すべき（重みを大きくするべき）判別ポイントを探すこととした。

[0058] ここで、グループ分けに使用される手法には、例えば、クラスタリングや教師あり学習が用いられる。本開示に係る技術は、大量の学習データを用いることなく正確に通信種類（マルウェア名等）を判別することを目的としているが、通信種類判別の前段階であるグループ分けでは、類似する通信データを大雑把にグループ分けできれば十分であるため、比較的少量の学習データを用いて作られたクラスタリングモデル／学習モデルでのグループ分けであっても、効果を発揮できる。なお、このグループ分けでは、「通信全体をみると似ていないが、実際には同種」のマルウェア等が異なるグループにグループ分けされる可能性があるが、このグループは同一グループ内で通信種別を判別するための適切な重みを算出しやすくするためのグループであるた

め、問題とはならない。

[0059] <システムの構成>

図9は、本実施形態に係る管理サーバー50の機能構成の概略を示す図である。管理サーバー50は、記憶装置14bに記録されているプログラムが、RAM13bに読み出され、CPU11bによって実行されて、管理サーバー50に備えられた各ハードウェアが制御されることで、第一の実施形態で説明した通信データ取得部21、過去通信データ保持部22、特徴抽出部23、類似度算出部24、類似通信検索部25、情報出力部26、及び重み算出部27に加えて、モデル保持部28、グループ分け部29、及び設定部30を更に備える情報処理装置として機能する。

[0060] モデル保持部28は、複数の過去通信のグループ分けモデルを保持する。本開示に係る実施形態では、モデル保持部28は、複数の過去通信を学習データとして用いて、当該複数の過去通信を所定数のクラスタにクラスタリングするように生成されたクラスタリングモデルを、グループ分けモデルとして保持する。

[0061] グループ分け部29は、グループ分けモデルを用いることで、対象通信を、複数の過去通信がグループ分けされた何れかのグループにグループ分けする。本開示に係る実施形態では、グループ分け部29は、クラスタリングモデルを用いて、対象通信が属するクラスタのクラスタ番号を特定する。

[0062] 設定部30は、対象通信がグループ分けされたグループに応じて、類似通信検索部25によって索出される過去通信の対象範囲、及び類似度算出部24において用いられる判別ポイント毎の重みの少なくとも何れかを設定する。本開示に係る実施形態では、設定部30は、クラスタ番号のクラスタに属する過去通信を、対象範囲として設定する、

[0063] <処理の流れ>

次に、本実施形態に係る管理サーバー50によって実行される処理の流れを説明する。なお、以下に説明する処理の具体的な内容および処理順序は、本開示を実施するための一例である。具体的な処理内容および処理順序は、

本開示の実施の形態に応じて適宜選択されてよい。

[0064] 図10は、本実施形態に係る類似通信検索処理の流れの概要を示すフローチャートである。本フローチャートは、図4のステップS3に示された類似通信検索処理の詳細を説明するものである。

[0065] ステップS501では、対象通信が属するクラスタが特定される。グループ分け部29は、過去通信の通信データ及び／又は特徴データを用いて学習済みのクラスタリングモデルを用いて、クラスタリング処理（グループ分け処理）を行い、対象通信が属するクラスタを特定する。

[0066] 具体的には、グループ分け部29は、予め、過去通信の通信データ及び／又は特徴データを用いて、過去通信を所定数（例えば、50）のクラスタにクラスタリングすることが可能な学習済みクラスタリングモデルを用意しておく。クラスタリングのアルゴリズムは特に限定されないが、例えばk平均法等を用いることが出来る。そして、グループ分け部29は、対象通信の通信データ及び／又は特徴データをクラスタリングモデルに入力することで、当該対象通信が属するクラスタ番号を取得する。

[0067] このようにすることで、対象通信を、「通信全体を見るとよく似ているが、実際には異なる種類の通信が含まれる」グループに振り分けることが出来る。なお、本実施形態では、グループ分けのためにクラスタリングを採用しているが、対象通信を、「通信全体を見るとよく似ているが、実際には異なる種類の通信が含まれる」グループに振り分けることが可能であれば、具体的なグループ分け手段は所謂クラスタリングに限定されない。その後、処理はステップS502へ進む。

[0068] ステップS502では、クラスタ番号に応じた類似検索設定が読み込まれる。設定部30は、対象通信の類似通信検索のために、クラスタ番号毎に予め設定された類似検索設定を、記憶装置14bから読み込む。ここで、検索範囲には、上記説明した第一の実施形態と異なり、「対象通信が属するクラスタの過去通信（例えば、対象通信のクラスタ番号が1であれば、クラスタ番号1に属する過去通信全て）」が設定される。また、判別ポイント毎の重

みについては、クラスタ毎に、当該クラスタに属する過去通信を重み算出対象範囲として重み算出処理（図8を参照）を実行することで事前に算出された重みが設定される。その後、処理はステップS503へ進む。

[0069] ステップS503からステップS505の処理は、第一の実施形態で説明したステップS202からステップS204の処理と概略同様であるため、説明を省略する。ステップS501及びステップS502の処理によって、検索処理の母体が「通信全体を見るとよく似ているが、実際には異なる種類の通信が含まれる」グループに絞り込まれており、且つ、このようなグループを重み算出対象範囲として算出された重みが用いられるので、ステップS503からステップS505において、「通信全体を見るとよく似ているが、実際には異なる種類の通信」同士をより正確に判別可能であり、適切な過去通信がより索出されやすくなっている。フィルタリングによって類似する通信が索出されると、本フローチャートに示された処理は終了する。

[0070] [第三の実施形態]

次に、第三の実施形態を説明する。第三の実施形態では、上記他の実施形態で説明した内容を重複する項目については、同一の符号を付して説明を省略する。

[0071] 上記第一の実施形態及び第二の実施形態では、検索が1回で終了するが、1回の検索で適切な検索結果が得られないこともある。このため、適切な検索結果が得られるまで条件（検索範囲等）を変化させながら検索処理を複数回実行することが考えられる。しかし、上記第一の実施形態及び第二の実施形態では、重みは設定される検索範囲に合わせて事前に算出されるものであるため、処理中に変化する検索範囲に対して十分に適切な重み付けを提供できない。

[0072] そこで、本実施形態及び後述する第四の実施形態では、類似通信検索処理において、検索範囲を徐々に絞り込みながら検索処理を複数回実行し、絞り込まれた検索範囲に適した重みを動的に算出することとした。

[0073] <システムの構成>

図 1 1 は、本実施形態に係る管理サーバー 5 0 の機能構成の概略を示す図である。管理サーバー 5 0 は、記憶装置 1 4 b に記録されているプログラムが、RAM 1 3 b に読み出され、CPU 1 1 b によって実行されて、管理サーバー 5 0 に備えられた各ハードウェアが制御されることで、第一の実施形態で説明した通信データ取得部 2 1、過去通信データ保持部 2 2、特徴抽出部 2 3、類似度算出部 2 4、類似通信検索部 2 5、情報出力部 2 6、及び重み算出部 2 7 に加えて、終了判定部 3 1、及び設定変更部 3 2 を更に備える情報処理装置として機能する。

[0074] 終了判定部 3 1 は、類似通信検索部 2 5 によって、フィルタリング設定（足切り条件）に従って索出された、対象通信に類似する 1 又は複数の過去通信が、索出された過去通信の正確性を担保するために設定された終了条件を満たしているか否かを判定する。

[0075] 設定変更部 3 2 は、索出された過去通信が終了条件を満たしていないと判定された場合に、類似通信検索部 2 5 によって索出される過去通信の対象範囲、及びフィルタリング設定の少なくとも何れかを変更する。

[0076] <処理の流れ>

次に、本実施形態に係る管理サーバー 5 0 によって実行される処理の流れを説明する。なお、以下に説明する処理の具体的な内容および処理順序は、本開示を実施するための一例である。具体的な処理内容および処理順序は、本開示の実施の形態に応じて適宜選択されてよい。

[0077] 図 1 2 は、本実施形態に係る類似通信検索処理の流れの概要を示すフローチャートである。本フローチャートは、図 4 のステップ S 3 に示された類似通信検索処理の詳細を説明するものである。

[0078] ステップ S 6 0 1 では、類似検索設定が読み込まれる。類似通信検索部 2 5 は、対象通信の類似通信検索のために予め設定された類似検索設定を、記憶装置 1 4 b から読み込む。本実施形態において、類似検索設定には、検索範囲、足切り条件、検索終了条件、及び設定更新内容が含まれる。なお、第一の実施形態及び第二の実施形態において設定された、判別ポイント毎の重

みについては、本実施形態では類似検索処理中に動的に算出されるため、設定されない。その後、処理はステップS 6 0 2へ進む。

[0079] 検索範囲及び足切り条件の設定内容については、第一の実施形態で説明した通りであるため、説明を省略する。

[0080] 検索終了条件は、類似通信検索処理が出力する検索結果が所定の品質を満たしているか否かを判定するための条件であり、例えば、「総合類似度上位5件の通信種類情報（マルウェア名等）が全て同じ、又は処理の繰り返し回数が3回より大」のように設定することが出来る。

[0081] 設定更新内容は、検索結果が検索終了条件を満たさずに繰り返した場合に、類似検索設定を更新するための設定内容である。このため、設定更新内容には、2周目以降に設定される検索範囲、足切り条件、及び検索終了条件が含まれる。

[0082] ステップS 6 0 2では、検索対象の過去通信が取得される。類似通信検索部25は、過去通信データ保持部22に蓄積された過去通信のうち、類似検索設定に設定された検索範囲内の過去通信を、検索対象としてRAM13bに読み込む。その後、処理はステップS 6 0 3へ進む。

[0083] ステップS 6 0 3では、重み算出処理が実行される。重み算出部27は、類似検索設定に現在設定されている検索範囲（ステップS 6 0 7で更新されている場合には、更新後の検索範囲）を、重み算出対象範囲として、判別ポイント毎の重みを算出する。即ち、本フローチャートに示された類似通信検索処理では、対象通信と過去通信との類似度を算出するために用いられる重みが、処理が周回する毎に変化する検索範囲に応じて、動的に算出される。重み算出処理の詳細については図8を用いて説明済みであるため、説明を省略する。その後、処理はステップS 6 0 4へ進む。

[0084] ステップS 6 0 4では、類似度算出処理が実行される。類似度算出部24は、対象通信と、ステップS 6 0 2で読み込まれた複数の過去通信の夫々との間の類似度（総合類似度）を算出する。この際に用いられる判別ポイント毎の重みは、ステップS 6 0 3において、検索範囲に応じて動的に算出され

たものである。類似度算出処理の詳細については図 7 を用いて説明済みであるため、説明を省略する。その後、処理はステップ S 6 0 5 へ進む。

[0085] ステップ S 6 0 5 では、対象通信に類似する過去通信が索出される。類似通信検索部 2 5 は、ステップ S 6 0 4 で対象通信との総合類似度が算出された各過去通信の総合類似度を参照し、類似検索設定に設定された足切り条件に従って、過去通信を絞り込む。その後、処理はステップ S 6 0 6 へ進む。

[0086] ステップ S 6 0 6 及びステップ S 6 0 7 では、検索終了条件が満たされない場合に、類似検索設定が更新される。終了判定部 3 1 は、足切り条件に従って絞り込まれた過去通信や、繰り返し回数等が、検索終了条件を満たしているか否かを判定する（ステップ S 6 0 6）。検索終了条件が満たされた場合、本フローチャートに示された処理は終了する。一方、検索終了条件が満たされない場合、設定変更部 3 2 は、設定更新内容に従って、類似検索設定を更新する（ステップ S 6 0 7）。この際、類似検索設定は、次の繰り返し処理での検索結果が検索終了条件をより満たしやすくなるような内容に構成されることが好ましい。また、設定変更部 3 2 は、処理の繰り返し回数を 1 回加算する。その後、処理はステップ S 6 0 2 へ進む。

[0087] 即ち、本実施形態では、検索終了条件を満たすまで、類似検索設定が更新されながら、ステップ S 6 0 2 からステップ S 6 0 7 の検索処理が繰り返し実行される。例えば、1 周目の検索処理では、類似検索設定として、以下の内容が設定される。

検索範囲：過去通信データ保持部 2 2 が保持する過去通信データ全て

足切り条件：総合類似度上位 1 0 0 件

検索終了条件：総合類似度上位 5 件の通信種類情報が全て同じ、又は処理の繰り返し回数が 3 回より大

[0088] そして、1 周目の検索処理で検索終了条件が満たされなかった場合、設定更新内容が参照され、2 周目の類似検索設定は例えば以下のように更新される。

検索範囲：1 周目の検索結果（1 0 0 件）

足切り条件：総合類似度上位８０％

検索終了条件：総合類似度上位５件の通信種類情報が全て同じ、又は処理の繰り返し回数が３回より大

[0089] 即ち、本実施形態によれば、十分な品質の検索結果（例えば、「総合類似度上位５件の通信種類情報が全て同じ」）が得られるか、又は繰り返し回数が上限（例えば、３回）に達するまで、類似検索設定の更新、重み算出処理、類似度算出処理、及び検索処理が繰り返し実行される。

[0090] [第四の実施形態]

次に、第四の実施形態を説明する。第四の実施形態では、上記他の実施形態で説明した内容を重複する項目については、同一の符号を付して説明を省略する。

[0091] 本実施形態では、第二の実施形態における適切な検索範囲の設定、及び第三の実施形態における動的な重み算出の両方の効果を得るために、類似通信検索処理において、クラスタリングによって検索範囲を絞り込み、その後、検索範囲を徐々に絞り込みながら検索処理を複数回実行し、絞り込まれた検索範囲に適した重みを動的に算出することとしている。

[0092] <システムの構成>

図１３は、本実施形態に係る管理サーバー５０の機能構成の概略を示す図である。管理サーバー５０は、記憶装置１４ｂに記録されているプログラムが、ＲＡＭ１３ｂに読み出され、ＣＰＵ１１ｂによって実行されて、管理サーバー５０に備えられた各ハードウェアが制御されることで、通信データ取得部２１、過去通信データ保持部２２、特徴抽出部２３、類似度算出部２４、類似通信検索部２５、情報出力部２６、重み算出部２７、モデル保持部２８、グループ分け部２９、設定部３０、終了判定部３１、及び設定変更部３２を備える情報処理装置として機能する。

[0093] <処理の流れ>

次に、本実施形態に係る管理サーバー５０によって実行される処理の流れを説明する。なお、以下に説明する処理の具体的な内容および処理順序は、

本開示を実施するための一例である。具体的な処理内容および処理順序は、本開示の実施の形態に応じて適宜選択されてよい。

[0094] 図14は、本実施形態に係る類似通信検索処理の流れの概要を示すフローチャートである。本フローチャートは、図4のステップS3に示された類似通信検索処理の詳細を説明するものである。

[0095] ステップS701では、対象通信が属するクラスタが特定される。グループ分け部29は、過去通信の通信データ及び／又は特徴データを用いて学習済みのクラスタリングモデルを用いて、クラスタリング処理（グループ分け処理）を行い、対象通信が属するクラスタを特定する。クラスタリングモデルの作成、及びクラスタ特定の詳細については、図10のステップS501で説明したものと概略同様であるため、説明を省略する。その後、処理はステップS702へ進む。

[0096] ステップS702では、クラスタ番号に応じた類似検索設定が読み込まれる。設定部30は、対象通信の類似通信検索のために、クラスタ番号毎に予め設定された類似検索設定を、記憶装置14bから読み込む。本実施形態において、類似検索設定には、検索範囲、足切り条件、検索終了条件、及び設定更新内容が含まれる。なお、判別ポイント毎の重みについては、第三の実施形態と同様、類似検索処理中に動的に算出されるため、設定されない。その後、処理はステップS703へ進む。

[0097] ここで、検索範囲には、例えば「対象通信が属するクラスタの過去通信（例えば、対象通信のクラスタ番号が1であれば、クラスタ番号1に属する過去通信全て）」が設定される。なお、検索終了条件及び設定更新内容の設定内容については、第三の実施形態で説明した通りであるため、説明を省略する。

[0098] ステップS703からステップS708の処理は、第三の実施形態で説明したステップS602からステップS607の処理と概略同様であるため、説明を省略する。検索終了条件が満たされると、本フローチャートに示された処理は終了する。

[0099] <バリエーション>

上記説明した第一から第四の実施形態では、ネットワーク監視装置 20 が、スイッチまたはルータのモニタリングポート（ミラーポート）に接続されることでノード 90 によって送受信されるパケットやフレーム等を取得し、取得したパケットを転送しないパッシブモードで動作する例について説明した（図 1 を参照）。但し、上記実施形態に示したネットワーク構成は、本開示を実施するための一例であり、実施にあたってはその他のネットワーク構成が採用されてもよい。

[0100] 例えば、ネットワーク監視装置 20 は、モニタリングポート（ミラーポート）に接続されず、単にネットワークセグメント 2 に接続されている場合であっても、ネットワークセグメント 2 を流れるフレームを、自身の MAC アドレス宛でないものも含めて全て取得することで、ノード 90 によって送受信されるパケットやフレーム等を取得することが出来る。この場合も、ネットワーク監視装置 20 は、パッシブモードで動作する。また、例えば、ネットワーク監視装置 20 は、ネットワークセグメント 2 のスイッチまたはルータと、その上位にある他のスイッチまたはルータと、の間に接続されることで、通過するパケットやフレーム等を取得してもよい（図 15 を参照）。この場合、ネットワーク監視装置 20 は、取得したパケットのうち、遮断しなくてもよいパケットについては転送するインラインモードで動作する。また、ネットワーク監視装置 20 は、ルータまたはスイッチに内包されてもよい。

[0101] なお、本実施形態では、ネットワークを流れるパケットを取得して、上記した各種の検知エンジンによりリアルタイムで検知を行う実施形態について説明したが、本開示の適用範囲は、リアルタイム検知に限定されない。例えば、ネットワークを流れる通信に係るデータを蓄積しておいて、蓄積されたデータに対して上記した各種の検知エンジンによる処理を行うこととしてもよい。

[0102] <効果>

上記説明した第一から第四の実施形態によれば、従来の技術では異なる種類の通信であると判別することが難しかった通信データを分類することが可能となる。

符号の説明

- [0103] 20 ネットワーク監視装置
 50 管理サーバー

請求の範囲

- [請求項1] ネットワークに接続された端末による通信の通信データを取得する通信データ取得手段と、
- 複数の過去通信の夫々に係る過去通信データを、対応する通信種類情報と関連付けて保持する過去通信データ保持手段と、
- 前記通信データ取得手段によって取得された通信データと、前記過去通信データとを比較して、通信に含まれる判別ポイント毎の特徴類似度を算出し、算出された特徴類似度を該判別ポイント毎又は該判別ポイントの組み合わせ毎に重み付けすることで、該対象通信と該過去通信との間の総合類似度を算出する類似度算出手段と、
- 算出された総合類似度に従って、対象通信に類似する1又は複数の過去通信を索出する類似通信検索手段と、
- 索出された過去通信に対応する通信種類情報を出力する情報出力手段と、
- を備える情報処理装置。
- [請求項2] 前記類似通信検索手段によって索出される過去通信の対象範囲に含まれる過去通信同士のペアを複数作成し、通信種類情報が一致する過去通信ペアについて前記類似度算出手段が算出する総合類似度が100%に近づき、通信種類情報が一致しない過去通信ペアについて前記類似度算出手段が算出する総合類似度が0%に近づくように、重回帰分析を実行することによって、判別ポイント毎の重みを算出する重み算出手段を更に備える、
- 請求項1に記載の情報処理装置。
- [請求項3] 前記複数の過去通信のグループ分けモデルを保持するモデル保持手段と、
- 前記グループ分けモデルを用いることで、前記対象通信を、前記複数の過去通信がグループ分けされた何れかのグループにグループ分けするグループ分け手段と、

前記対象通信がグループ分けされたグループに応じて、前記類似通信検索手段によって索出される過去通信の対象範囲、及び前記類似度算出手段において用いられる判別ポイント毎の重みの少なくとも何れかを設定する設定手段と、

を更に備える、請求項 1 又は 2 に記載の情報処理装置。

[請求項4]

前記モデル保持手段は、前記複数の過去通信を学習データとして用いて、該複数の過去通信を所定数のクラスタにクラスタリングするように生成されたクラスタリングモデルを、前記グループ分けモデルとして保持し、

前記グループ分け手段は、前記クラスタリングモデルを用いて、前記対象通信が属するクラスタのクラスタ番号を特定し、

前記設定手段は、前記クラスタ番号のクラスタに属する過去通信を、前記対象範囲として設定する、

請求項 3 に記載の情報処理装置。

[請求項5]

前記類似通信検索手段は、フィルタリング設定に更に従って、対象通信に類似する 1 又は複数の過去通信を索出し、

前記類似通信検索手段によって索出された過去通信が、索出された過去通信の正確性を担保するために設定された終了条件を満たしているか否かを判定する終了判定手段と、

索出された過去通信が前記終了条件を満たしていないと判定された場合に、前記類似通信検索手段によって索出される過去通信の対象範囲、及び前記フィルタリング設定の少なくとも何れかを変更する設定変更手段と、

を更に備える、請求項 1 から 4 の何れか一項に記載の情報処理装置。

[請求項6]

前記通信データ取得手段によって取得された通信データから、通信に含まれる判別ポイント毎に特徴を抽出することで、対象通信の特徴データを生成する特徴抽出手段を更に備え、

前記類似度算出手段は、前記判別ポイント毎に、前記対象通信の特徴データと前記過去通信の特徴データとを比較して特徴類似度を算出し、算出された特徴類似度を該判別ポイント毎又は該判別ポイントの組み合わせ毎に重み付けすることで、該対象通信と該過去通信との間の総合類似度を算出する、

請求項 1 から 5 の何れか一項に記載の情報処理装置。

[請求項7] 前記特徴抽出手段は、判別ポイント毎に予め設定された特徴抽出方法を用いて、判別ポイント毎の特徴データを抽出する、

請求項 6 に記載の情報処理装置。

[請求項8] 前記通信データ取得手段は、監視対象ネットワークに設置された通信監視装置によってキャプチャされ、該情報処理装置宛に送信された通信データを取得する、

請求項 1 から 7 の何れか一項に記載の情報処理装置。

[請求項9] 前記情報出力手段は、索出された過去通信に関連付けられた対処方法を取得して出力する、

請求項 1 から 8 の何れか一項に記載の情報処理装置。

[請求項10] コンピューターが、

ネットワークに接続された端末による通信の通信データを取得する通信データ取得ステップと、

複数の過去通信の夫々に係る過去通信データを、対応する通信種類情報と関連付けて保持する過去通信データ保持ステップと、

前記通信データ取得ステップで取得された通信データと、前記過去通信データとを比較して、通信に含まれる判別ポイント毎の特徴類似度を算出し、算出された特徴類似度を該判別ポイント毎又は該判別ポイントの組み合わせ毎に重み付けすることで、該対象通信と該過去通信との間の総合類似度を算出する類似度算出ステップと、

算出された総合類似度に従って、対象通信に類似する 1 又は複数の過去通信を索出する類似通信検索ステップと、

索出された過去通信に対応する通信種類情報を出力する情報出力ステップと、

を実行する通信種類判別方法。

[請求項11]

コンピュータが、

前記類似通信検索ステップで索出される過去通信の対象範囲に含まれる過去通信同士のペアを複数作成し、通信種類情報が一致する過去通信ペアについて前記類似度算出ステップで算出される総合類似度が100%に近づき、通信種類情報が一致しない過去通信ペアについて前記類似度算出ステップで算出される総合類似度が0%に近づくように、重回帰分析を実行することによって、判別ポイント毎の重みを算出する重み算出ステップを更に実行する、

請求項10に記載の通信種類判別方法。

[請求項12]

コンピュータが、

前記複数の過去通信のグループ分けモデルを保持するモデル保持ステップと、

前記グループ分けモデルを用いることで、前記対象通信を、前記複数の過去通信がグループ分けされた何れかのグループにグループ分けするグループ分けステップと、

前記対象通信がグループ分けされたグループに応じて、前記類似通信検索ステップで索出される過去通信の対象範囲、及び前記類似度算出ステップで用いられる判別ポイント毎の重みの少なくとも何れかを設定する設定ステップと、

を更に実行する、請求項10又は11に記載の通信種類判別方法。

[請求項13]

前記モデル保持ステップでは、前記複数の過去通信を学習データとして用いて、該複数の過去通信を所定数のクラスタにクラスタリングするように生成されたクラスタリングモデルを、前記グループ分けモデルとして保持し、

前記グループ分けステップでは、前記クラスタリングモデルを用い

て、前記対象通信が属するクラスタのクラスタ番号を特定し、

前記設定ステップでは、前記クラスタ番号のクラスタに属する過去通信を、前記対象範囲として設定する、

請求項 1 2 に記載の通信種類判別方法。

[請求項14]

前記類似通信検索ステップでは、フィルタリング設定に更に従って、対象通信に類似する 1 又は複数の過去通信を索出し、

コンピュータが、

前記類似通信検索ステップで索出された過去通信が、索出された過去通信の正確性を担保するために設定された終了条件を満たしているか否かを判定する終了判定ステップと、

索出された過去通信が前記終了条件を満たしていないと判定された場合に、前記類似通信検索ステップで索出される過去通信の対象範囲、及び前記フィルタリング設定の少なくとも何れかを変更する設定変更ステップと、

を更に実行する、請求項 1 0 から 1 3 の何れか一項に記載の通信種類判別方法。

[請求項15]

コンピュータを、

ネットワークに接続された端末による通信の通信データを取得する通信データ取得手段と、

複数の過去通信の夫々に係る過去通信データを、対応する通信種類情報と関連付けて保持する過去通信データ保持手段と、

前記通信データ取得手段によって取得された通信データと、前記過去通信データとを比較して、通信に含まれる判別ポイント毎の特徴類似度を算出し、算出された特徴類似度を該判別ポイント毎又は該判別ポイントの組み合わせ毎に重み付けすることで、該対象通信と該過去通信との間の総合類似度を算出する類似度算出手段と、

算出された総合類似度に従って、対象通信に類似する 1 又は複数の過去通信を索出する類似通信検索手段と、

索出された過去通信に対応する通信種類情報を入力する情報出力手段

として機能させるためのプログラム。

[請求項16]

コンピュータを、

前記類似通信検索手段によって索出される過去通信の対象範囲に含まれる過去通信同士のペアを複数作成し、通信種類情報が一致する過去通信ペアについて前記類似度算出手段が算出する総合類似度が100%に近づき、通信種類情報が一致しない過去通信ペアについて前記類似度算出手段が算出する総合類似度が0%に近づくように、重回帰分析を実行することによって、判別ポイント毎の重みを算出する重み算出手段として更に機能させる、

請求項15に記載のプログラム。

[請求項17]

コンピュータを、

前記複数の過去通信のグループ分けモデルを保持するモデル保持手段と、

前記グループ分けモデルを用いることで、前記対象通信を、前記複数の過去通信がグループ分けされた何れかのグループにグループ分けするグループ分け手段と、

前記対象通信がグループ分けされたグループに応じて、前記類似通信検索手段によって索出される過去通信の対象範囲、及び前記類似度算出手段において用いられる判別ポイント毎の重みの少なくとも何れかを設定する設定手段

として更に機能させる、請求項15又は16に記載のプログラム。

[請求項18]

前記モデル保持手段は、前記複数の過去通信を学習データとして用いて、該複数の過去通信を所定数のクラスタにクラスタリングするように生成されたクラスタリングモデルを、前記グループ分けモデルとして保持し、

前記グループ分け手段は、前記クラスタリングモデルを用いて、前

記対象通信が属するクラスタのクラスタ番号を特定し、

前記設定手段は、前記クラスタ番号のクラスタに属する過去通信を、前記対象範囲として設定する、

請求項 17 に記載のプログラム。

[請求項19]

前記類似通信検索手段は、フィルタリング設定に更に従って、対象通信に類似する 1 又は複数の過去通信を索出し、

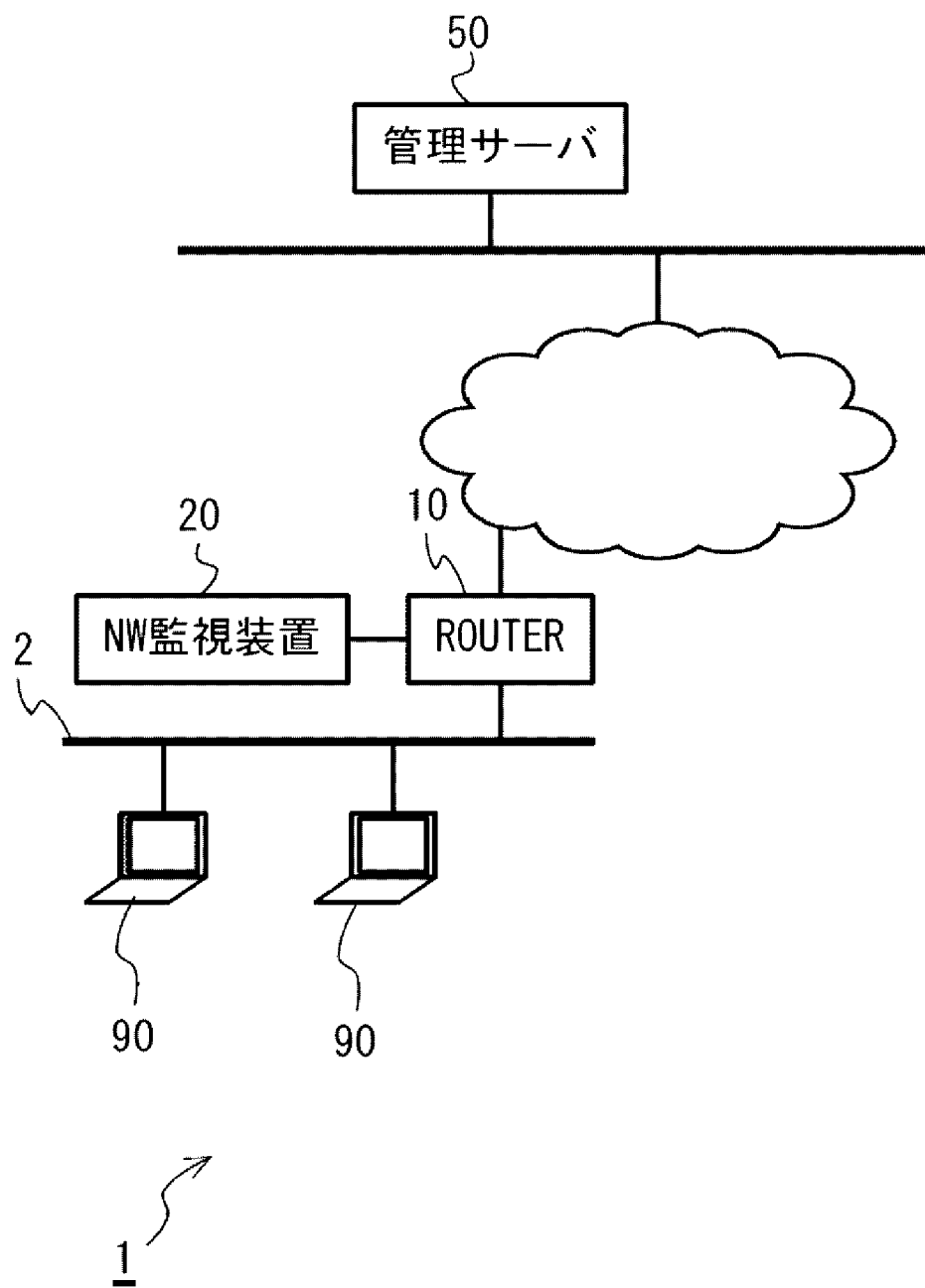
コンピューターを、

前記類似通信検索手段によって索出された過去通信が、索出された過去通信の正確性を担保するために設定された終了条件を満たしているか否かを判定する終了判定手段と、

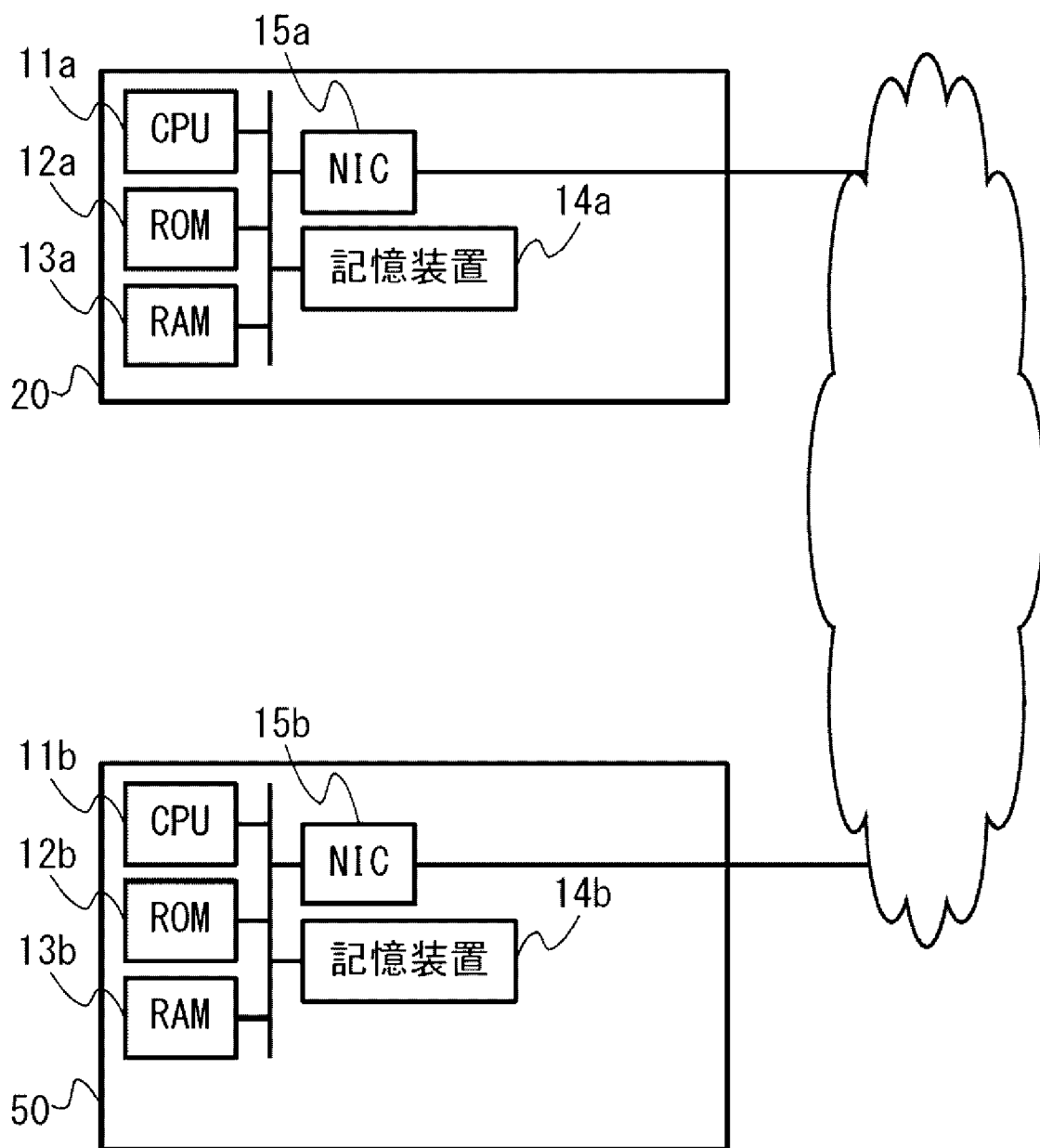
索出された過去通信が前記終了条件を満たしていないと判定された場合に、前記類似通信検索手段によって索出される過去通信の対象範囲、及び前記フィルタリング設定の少なくとも何れかを変更する設定変更手段

として更に機能させる、請求項 15 から 18 の何れか一項に記載のプログラム。

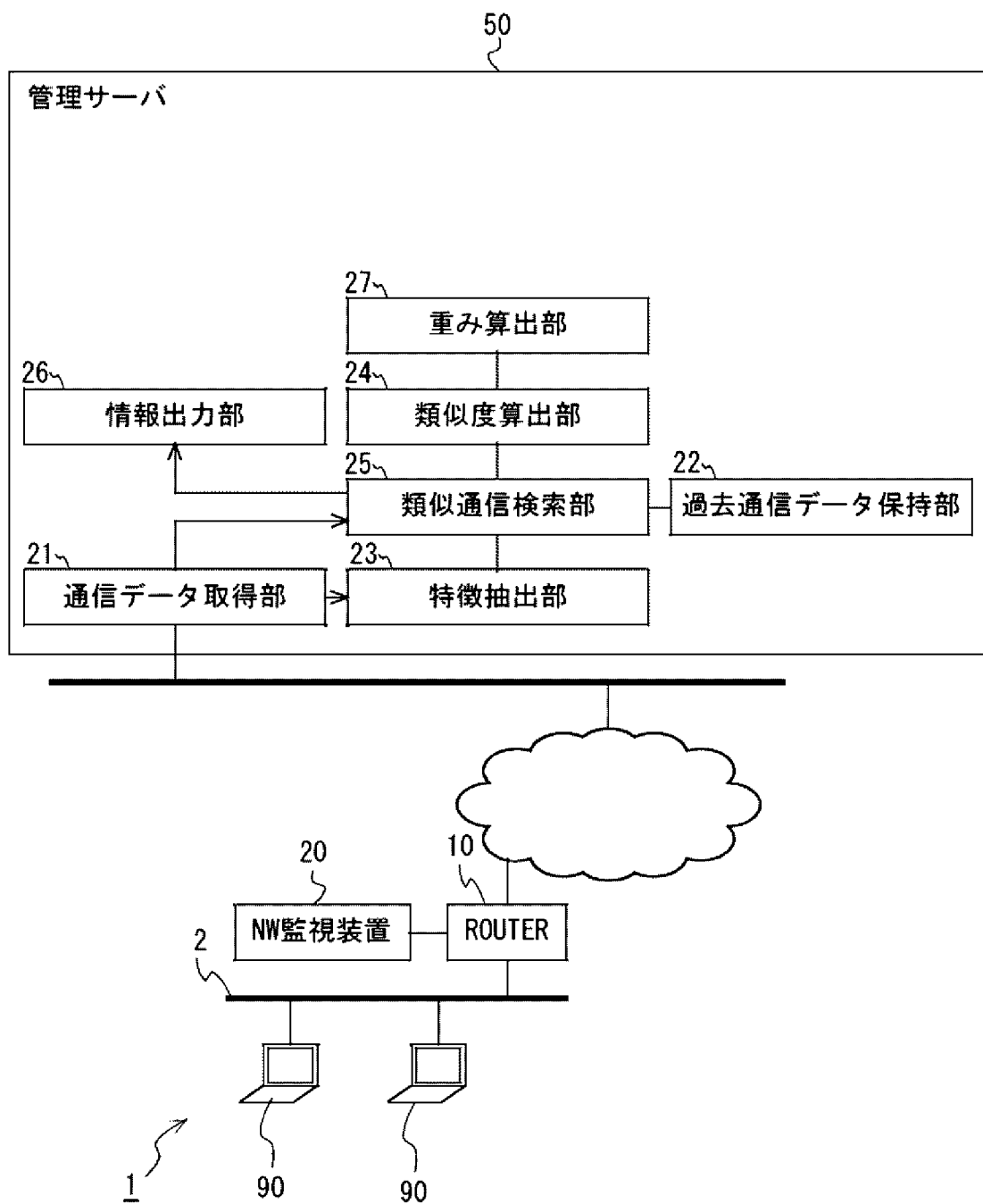
[図1]



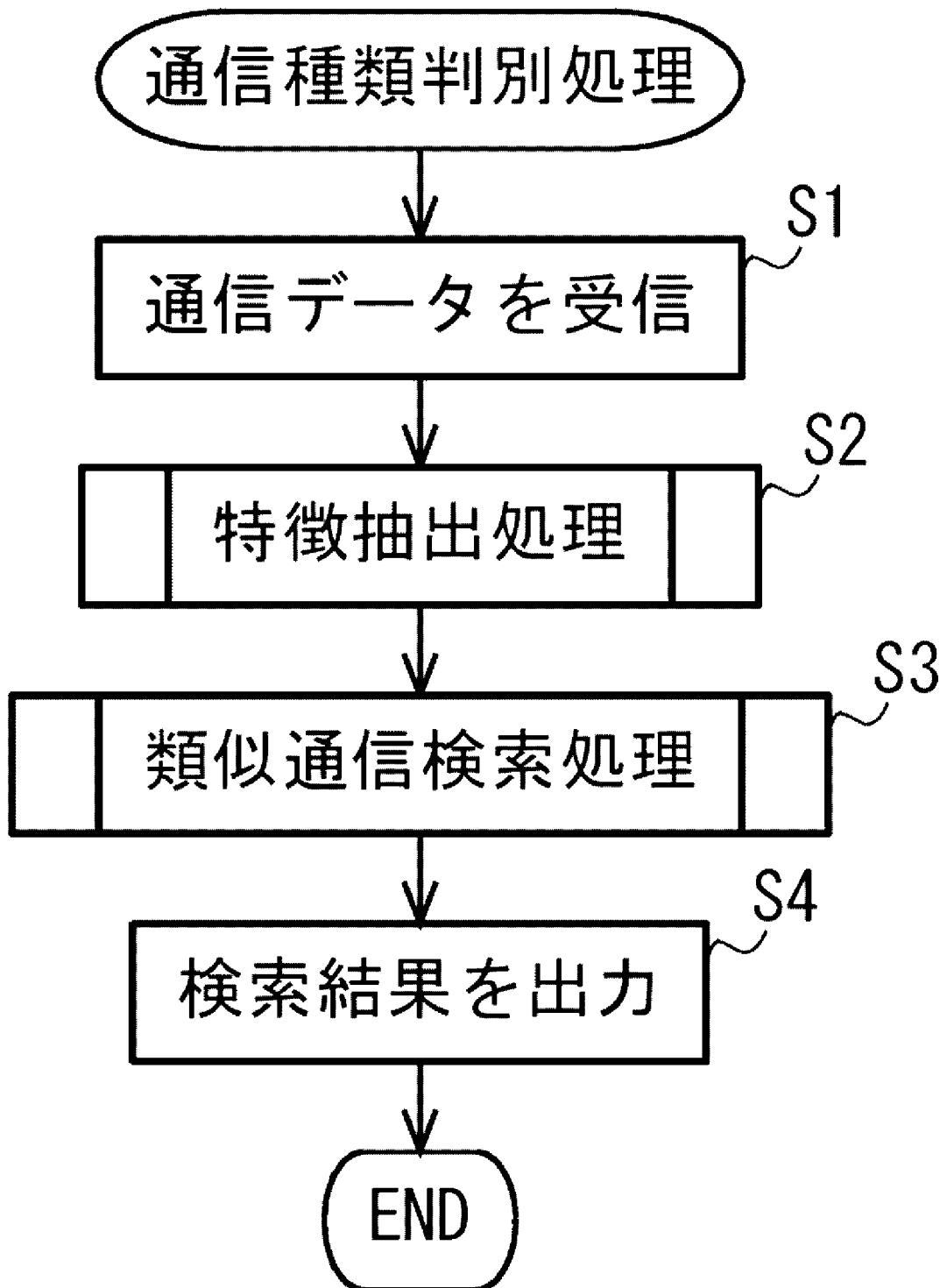
[図2]



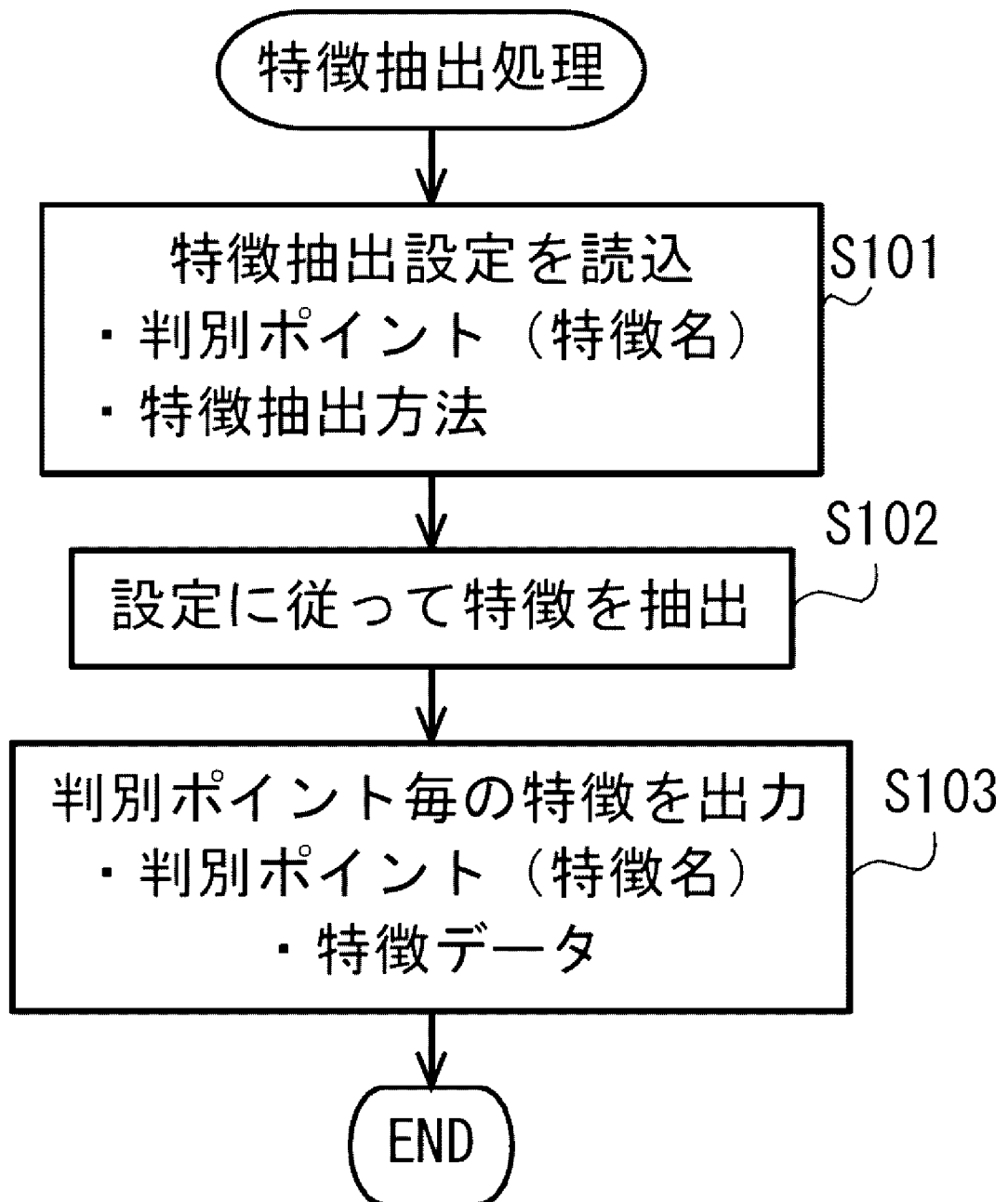
[図3]



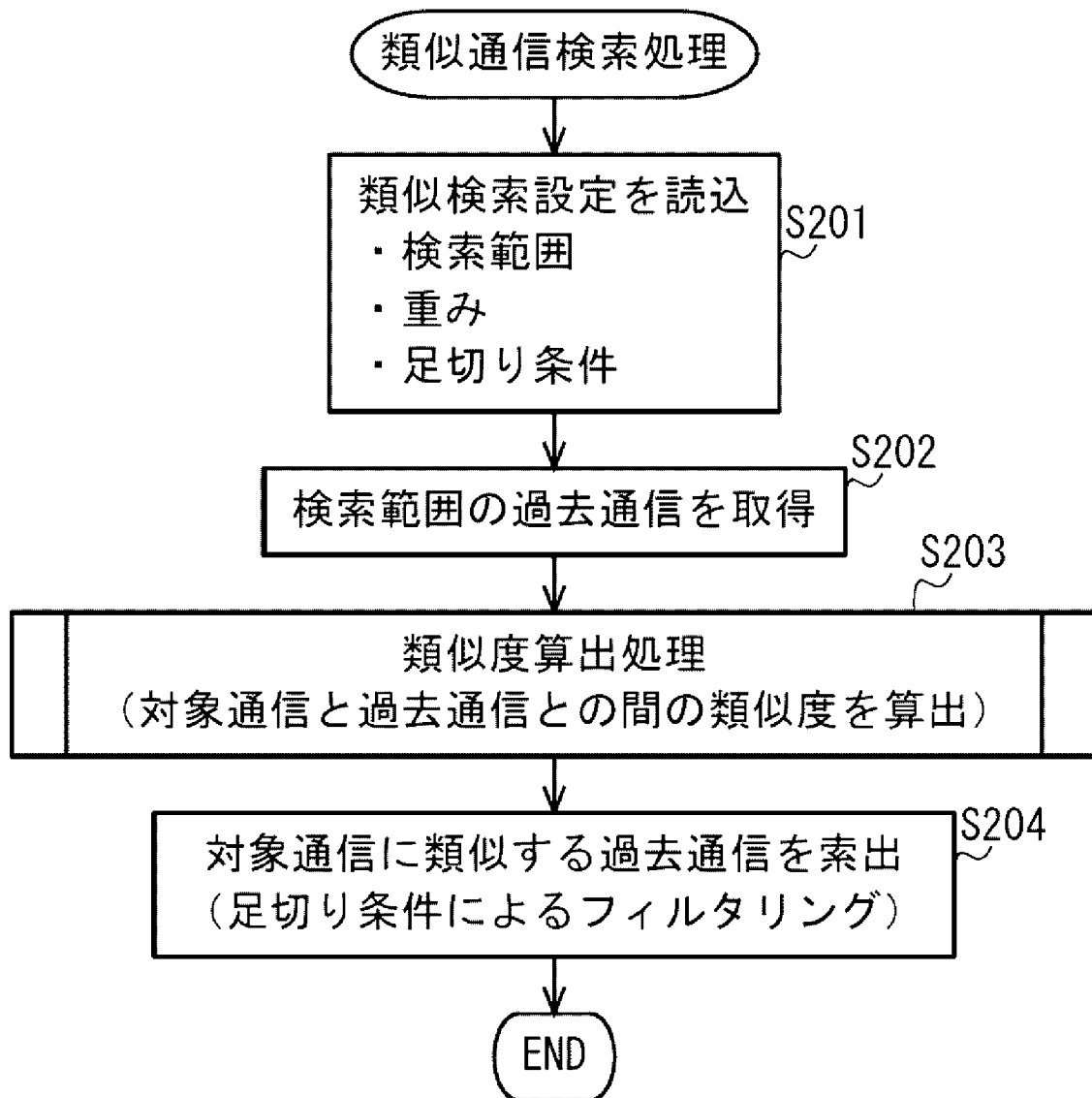
[図4]



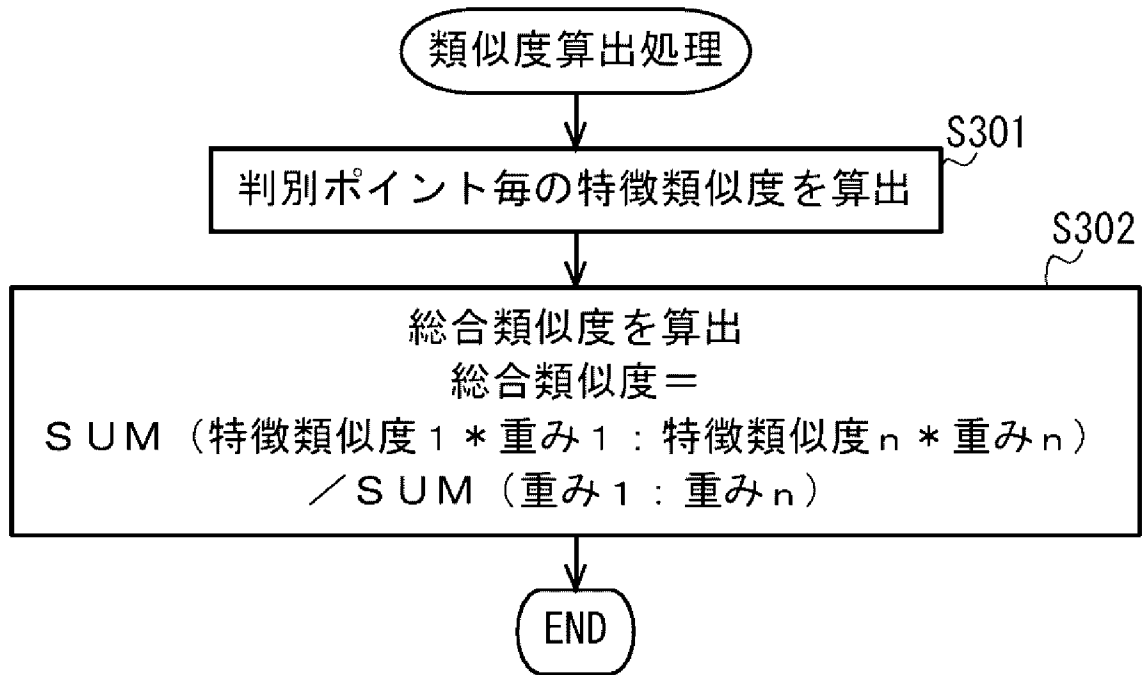
[図5]



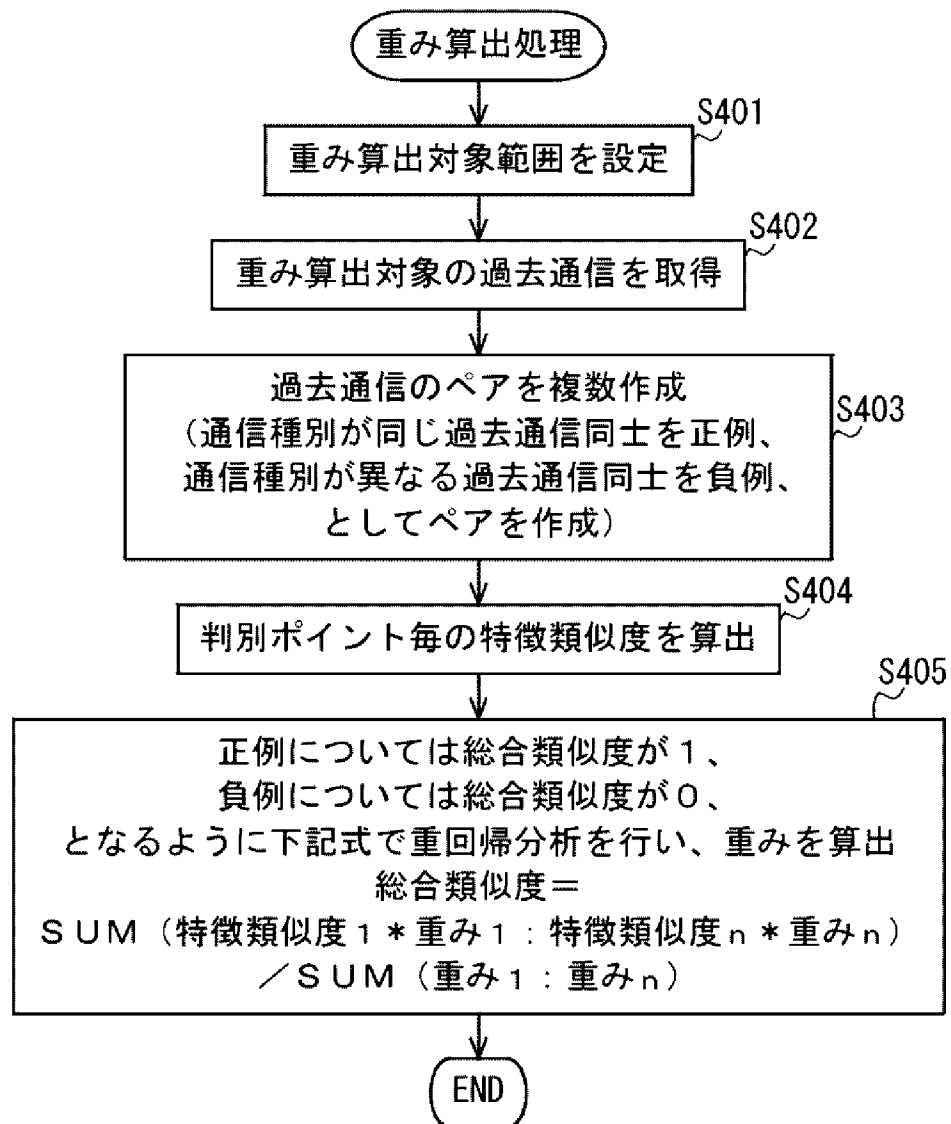
[図6]



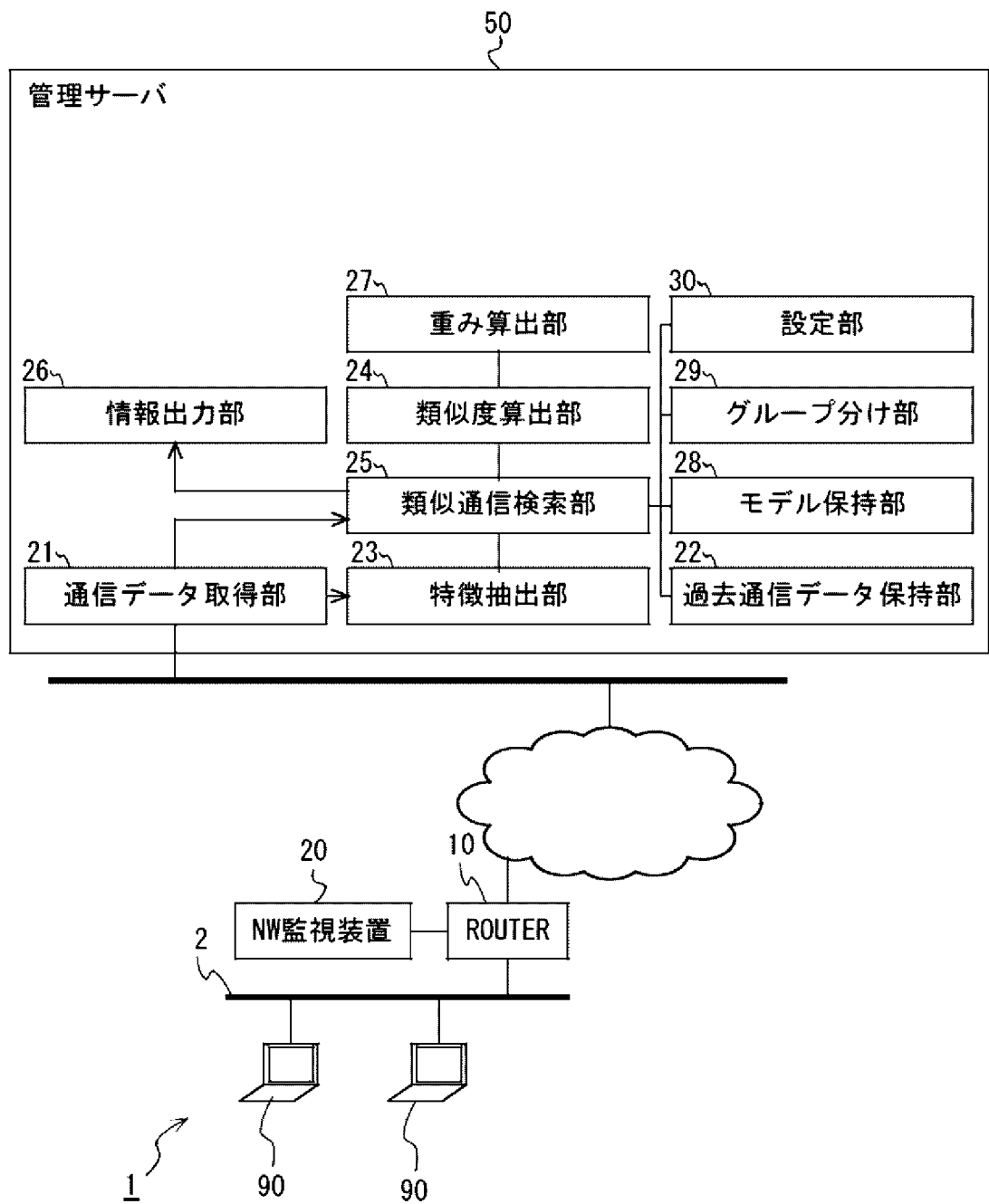
[図7]



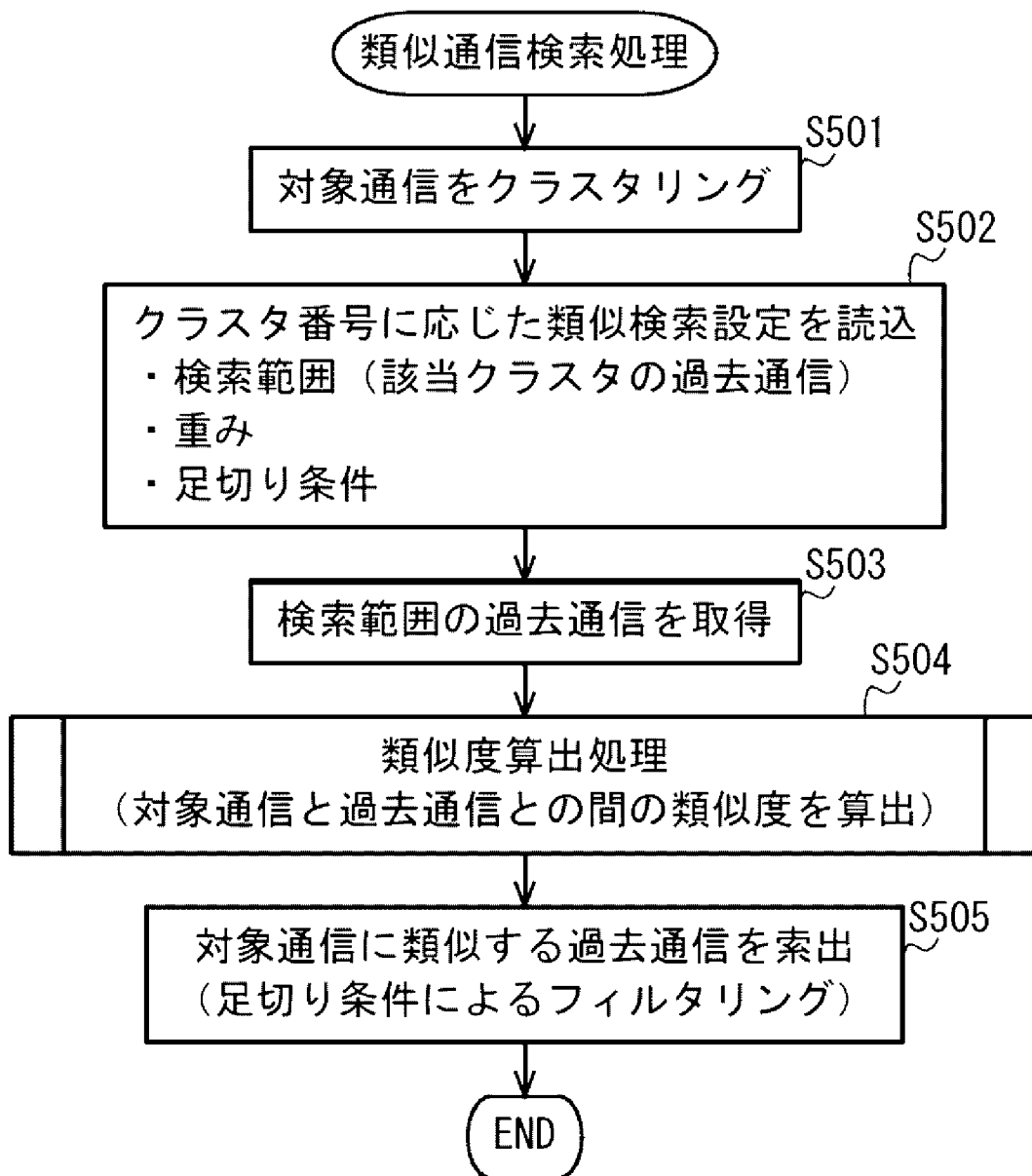
[図8]



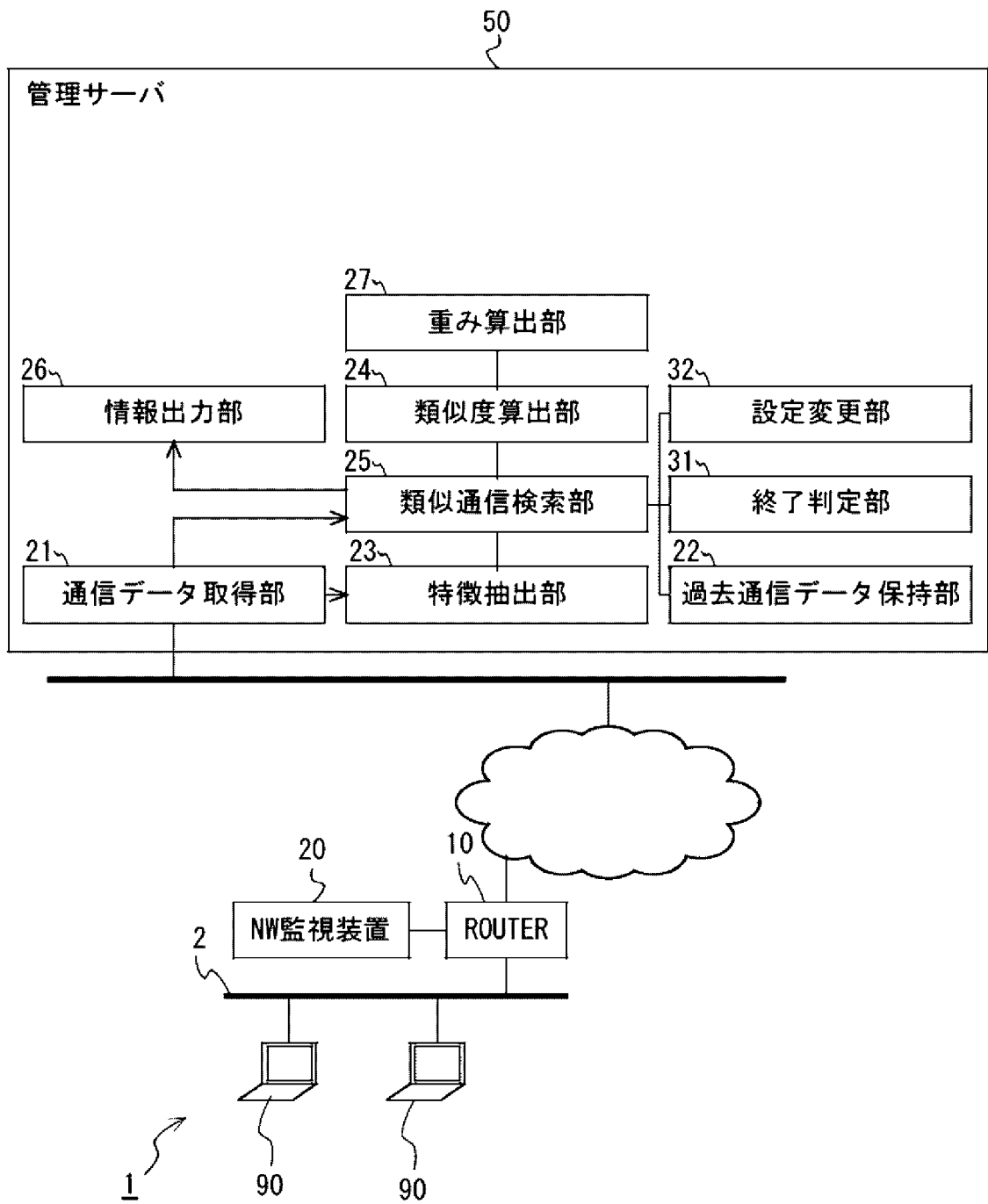
[図9]



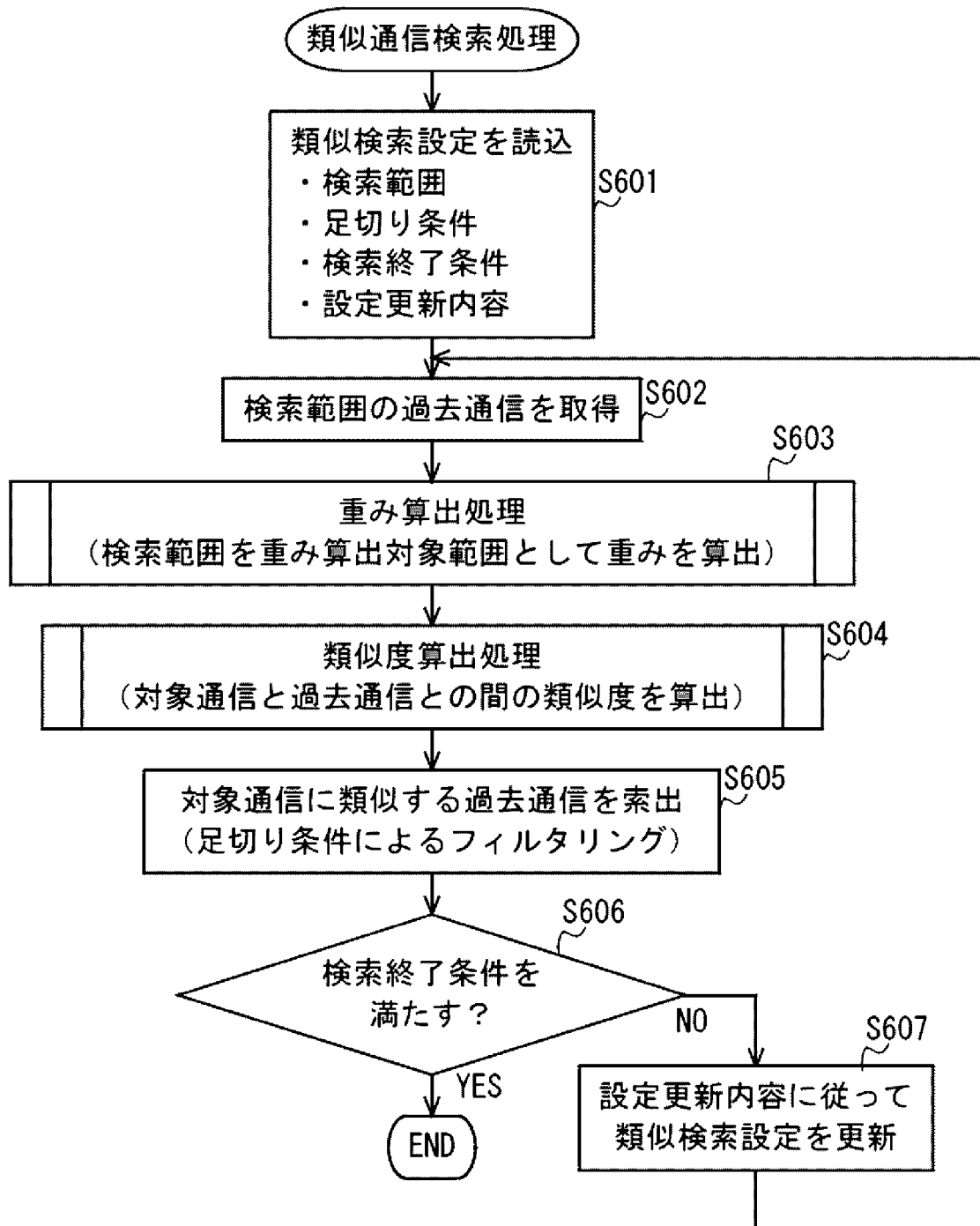
[図10]



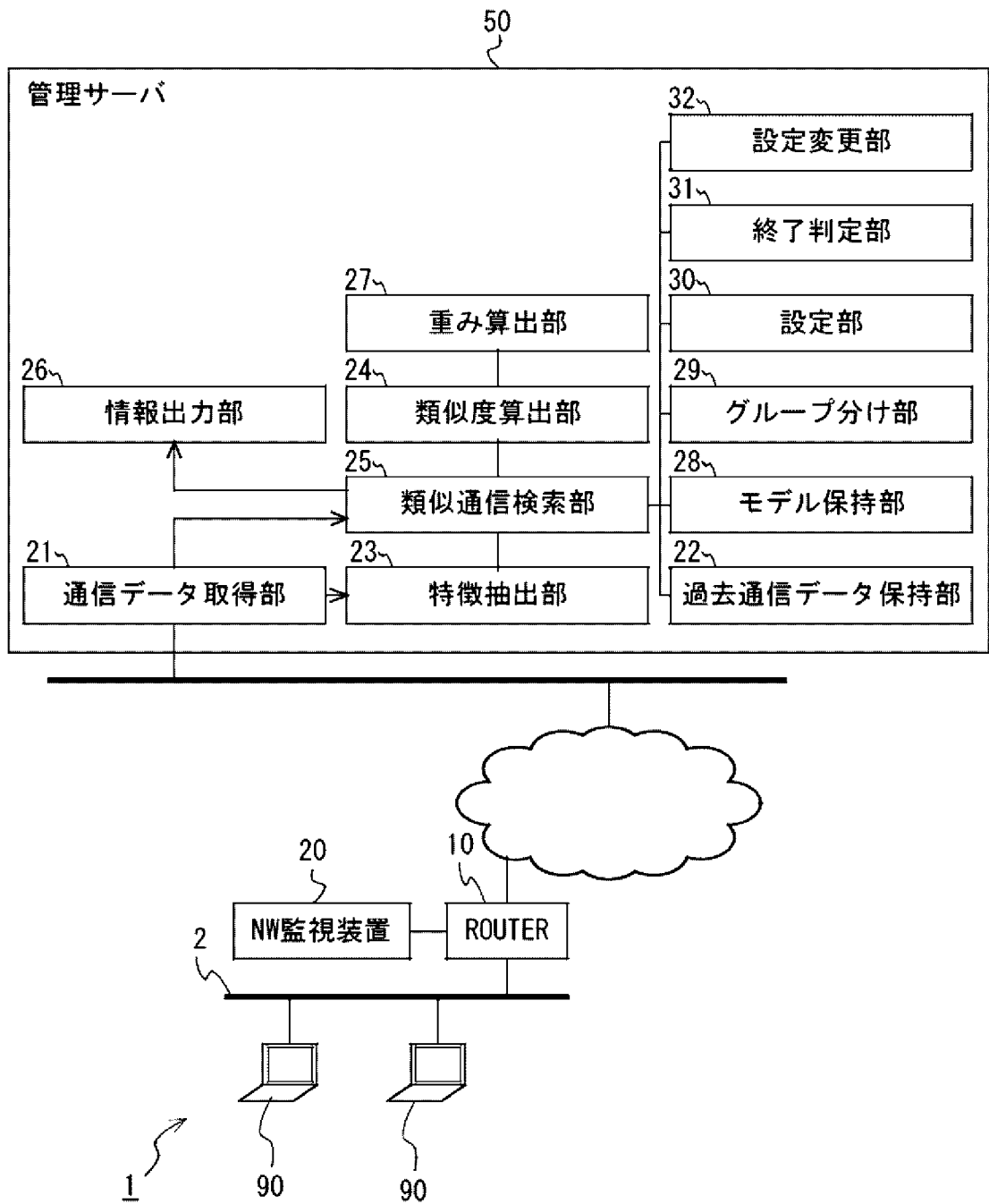
[図11]



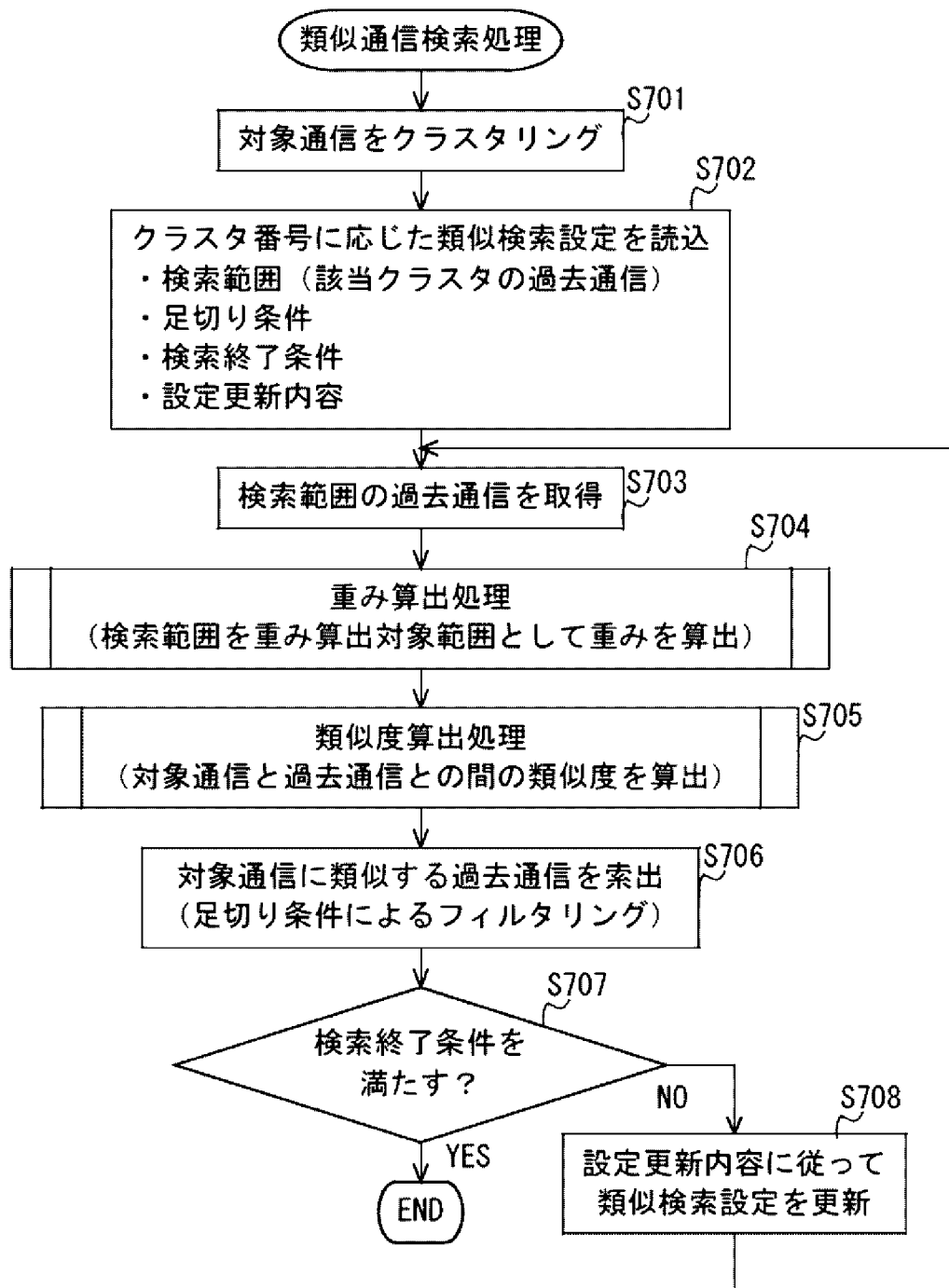
[図12]



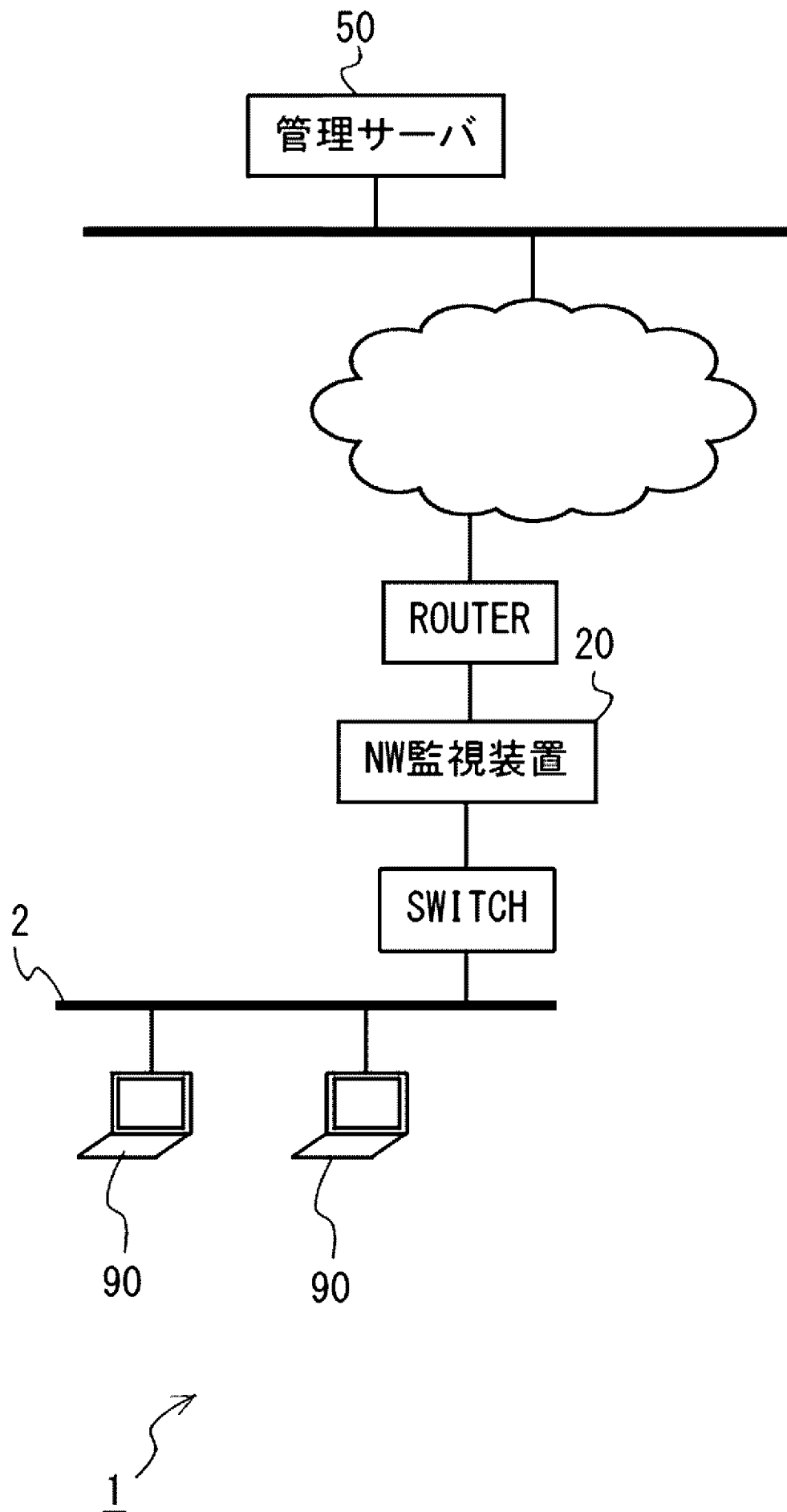
[図13]



[図14]



[図15]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2018/010961

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. H04L12/70 (2013.01) i, G06F21/56 (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. H04L12/70, G06F21/56

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 1971-2018

Registered utility model specifications of Japan 1996-2018

Published registered utility model applications of Japan 1994-2018

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 A	JP 2008-176753 A (NATIONAL INSTITUTE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY) 31 July 2008, claim 1, fig. 1, 2 (Family: none)	1, 6-10, 15 2-5, 11-14, 16-19
Y A	JP 2017-37555 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 16 February 2017, claims 1, 4, paragraph [0038] & JP 6039768 B1	1, 6-10, 15 2-5, 11-14, 16-19
A	JP 2017-142744 A (NEC CORP.) 17 August 2017, entire text, all drawings (Family: none)	1-19
A	US 9781144 B1 (FIREEYE, INC.) 03 October 2017, entire text, all drawings (Family: none)	1-19

☐ 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
12.06.2018

Date of mailing of the international search report
19.06.2018

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. H04L12/70(2013.01)i, G06F21/56(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, G06F21/56

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	1971-2018年
日本国実用新案登録公報	1996-2018年
日本国登録実用新案公報	1994-2018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 A	JP 2008-176753 A（独立行政法人情報通信研究機構）2008.07.31, 請求項1、図1, 2（ファミリーなし）	1, 6-10, 15 2-5, 11-14, 16 -19
Y A	JP 2017-37555 A（日本電信電話株式会社）2017.02.16, 請求項1, 4, 0038 & JP 6039768 B1	1, 6-10, 15 2-5, 11-14, 16 -19
A	JP 2017-142744 A（日本電気株式会社）2017.08.17, 全文、全図（ファミリーなし）	1-19

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

の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

12.06.2018

国際調査報告の発送日

19.06.2018

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

8 5 2 3

電話番号 03-3581-1101 内線 3596

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	US 9781144 B1 (FIREEYE, INC.,) 2017.10.03, 全文、全図 (ファミリーなし)	1-19