



(51) International Patent Classification:
H04M 3/436 (2006.01)

(21) International Application Number:
PCT/EP2022/057916

(22) International Filing Date:
25 March 2022 (25.03.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2104475.5 30 March 2021 (30.03.2021) GB

(71) Applicant: **BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY** [GB/GB]; 1 Braham Street, London E1 8EE (GB).

(72) Inventor: **WALKER, Joanne**; 9th Floor, One Braham, Braham Street, London E1 8EE (GB).

(74) Agent: **BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY, INTELLECTUAL PROPERTY DEPARTMENT**; 9th Floor, One Braham, Braham Street, London E1 8EE (GB).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): 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, IT, JM, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (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).

Published:

- with international search report (Art. 21(3))
- in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE

(54) Title: COMMUNICATIONS NETWORK

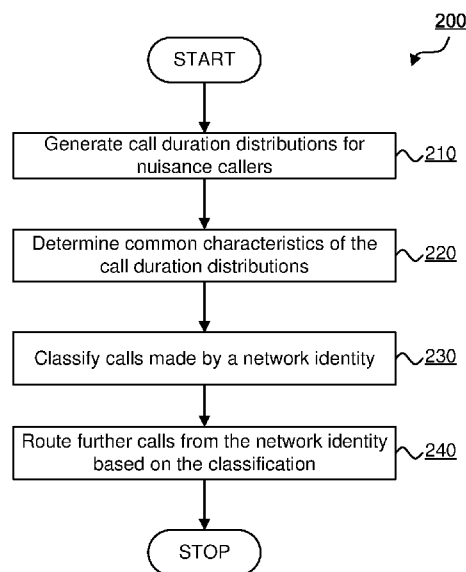


FIGURE 2

(57) Abstract: A method, computer system and computer program for routing calls from a network identity in a communications network are provided. The method uses a dimensionality reduction technique to determine one or more common characteristics of a plurality of nuisance call duration distributions. The nuisance call duration distributions are each generated from a plurality of calls placed by a respective nuisance caller. The method classifies a plurality of calls made by the network identity based on a similarity between a call duration distribution of the plurality of calls and the common characteristics of the nuisance call duration distributions. The method routes further calls from the network identity in accordance with the classification.

Communications Network

Field of the Invention

The present invention relates to a method of routing calls from a network identity in a communications network. In particular, the present invention classifies a plurality of calls made by the network identity and routes further calls from the network identity in accordance with the classification.

Background to the Invention

Nuisance (or unwanted) telephone calls have become problematic for many users and it is known to use telephone answering machines or systems to block calls which are from unidentified numbers, numbers held in a list of blocked numbers (or alternatively numbers which are not present on a list of allowed numbers), or numbers of a particular category (such as, for example, numbers associated with international calls). However, it would be preferable if such calls could be blocked within the operator's networks, such that the customer does not need to take significant action or acquire specific equipment in order to avoid being disturbed by nuisance calls.

Commonly, nuisance telephone calls are automated telephone calls. That is to say, calls which are automatically initiated by a computer dialler (which may also be referred to as a robot or automatic dialler) as opposed to human-generated calls which are dialled (or otherwise initiated) by a human operator. In some cases, when an automated call is answered by a called party, the computer dialler will play a pre-recorded message to the called party. In other cases, the computer dialler will connect the called party to a human operator (or agent) once the automated telephone call has been answered. These approaches are also combined in some cases, with the computer dialler playing a pre-recorded message and then connecting the call to a human operator. Where a human operator is involved, they may typically structure their conversation according to a pre-planned script. Through the use of computer diallers to initiate nuisance calls, a significant number of nuisance calls can be generated within a limited period of time.

Summary of the Invention

It is therefore desirable to provide a mechanism that allows unwanted nuisance calls to be detected and routed in an appropriate manner.

According to a first aspect of the invention, there is provided a method of routing calls from a network identity in a communications network, the method comprising: using a

dimensionality reduction technique to determine one or more common characteristics of a plurality of nuisance call duration distributions, each of the nuisance call duration distributions being generated from a plurality of calls placed by a respective nuisance caller; classifying a plurality of calls made by the network identity based on a similarity between a call duration distribution of the plurality of calls and the common characteristics of the nuisance call duration distributions; and routing further calls from the network identity in accordance with the classification.

The plurality of nuisance call duration distributions may comprise a plurality of sets of nuisance call duration distributions, each set being associated with a respective period of time, wherein the nuisance call duration distributions in each set are generated from a plurality of calls placed by nuisance callers during the respective period of time. Using the dimensionality reduction technique to determine one or more common characteristics of the plurality of nuisance call duration distributions may comprise using the dimensionality reduction technique to determine one or more common characteristics of each set of nuisance call duration distributions. The classification of the plurality of calls made by the network identity may be based on a similarity between the call duration distribution of the plurality of calls and the common characteristics of each set of nuisance call duration distributions. The plurality of calls may be classified as nuisance calls when the similarity between the call duration distribution and any of the nuisance call duration distributions is greater than a predetermined threshold.

The method may further comprise combining the common characteristics of each set of nuisance call duration distributions. The classification of the plurality of calls may be based on a similarity between the call duration distribution of the plurality of calls and the combined common characteristics of the plurality of sets of nuisance call duration distributions.

The respective periods of time for the plurality of sets of nuisance call duration distributions may be consecutive adjoining preceding time periods. The periods of time may each have a duration of a day.

The respective periods of time for the plurality of sets of nuisance call duration distributions are consecutive occurrences of a periodically occurring time period. The periodically recurring time period may be daily and the periods of time may have a duration which covers a particular portion of the day. The particular portion of the day may be one of: morning; early-afternoon; late-afternoon; and evening.

The common characteristics may be represented by principal components obtained from carrying out a principal component analysis of the nuisance call duration distributions. The similarity between the call duration distribution of the plurality of calls and the common characteristics may be determined by: projecting the call duration distribution onto a domain defined by the principal components that represent the common characteristics to create a projected call duration distribution; projecting the projected call duration distribution back into its original domain to create a reconstructed call duration distribution; and determining a measure of loss between the reconstructed call duration distribution and the call duration distribution of the plurality of calls, the measure of loss indicating the similarity between the call duration distribution and the common characteristics.

The network identity making the plurality of calls to be classified may be selected for classification in response to a total number of calls made by the network identity in a particular period of time exceeding a predetermined threshold.

According to a second aspect of the invention, there is provided a computer system comprising a processor and a memory storing computer program code for performing the steps of a method according to the first aspect.

According to a third aspect of the invention, there is provided a computer program which, when executed by one or more processors, is arranged to carry out a method according to the first aspect.

Brief Description of the Figures

In order that the present invention may be better understood, embodiments thereof will now be described, by way of example only, with reference to the accompanying drawings, in which:

Figure 1 is a schematic depiction of an exemplary telephone network within which embodiments of the invention may operate; and

Figure 2 is a flowchart describing how a method according to the present invention can be implemented.

Detailed Description of Embodiments

Figure 1 is a schematic depiction of an exemplary telephone network 100 within which embodiments of the invention may operate. The exemplary telephone network 100 is a conventional telephone network comprising a plurality of core exchanges 110, a plurality of

local exchanges 120, a plurality of customer telephony terminals 130, one or more domestic gateways 140, one or more international gateways 150, one or more voicemail servers 160, one or more call data stores 170 and one or more analyst terminals 180.

The core exchanges 110 are interconnected by a plurality of communications links 190.

- 5 Each of the plurality of core exchanges 110 are further connected to one or more local exchanges by further communications links 190 (although, for the sake of clarity, not all of the core exchanges 110 illustrated in figure 1 are shown as being connected to local exchanges 120).

- 10 The local exchanges 120 are each connected to a respective core exchange 110 via a respective communication link 190. Each of the local exchanges 120 is also connected to a respective subset of the customer telephony terminals 130 via yet further communications links 190 (although again, for the sake of clarity, this is not shown for each of the local exchanges 120 in figure 1).

- 15 The customer telephony terminals 130 are each connected to a respective local exchange 120 via a respective communication link 190. The customer telephony terminals 130 can include devices such as telephones, private branch exchanges (PBX), conference phones, computer diallers, fax machines, modems, answering machines and so on.

- 20 The domestic gateways 140 are each connected to one or more other telephony networks (not shown) in the same country. The domestic gateways 140 enable calls to be routed between the telephone network 100 and the other telephony networks. That is to say, the domestic gateways 140 enable the customer telephony terminals 130 within the telephone network 100 to place calls to and/or receive calls from telephony terminals within the other telephony networks.

- 25 The international gateways 150 are each connected to one or more other international telephony networks (not shown). The international gateways 150 enable calls to be routed between the telephone network 100 and the other international telephony networks. That is to say, the international gateways 150 enable the customer telephony terminals 130 within the telephone network 100 to place calls to and/or receive calls from telephony terminals within the other international telephony networks.

- 30 The voicemail servers 160 are connected to the telephone network 100 via respective communications links 190. They may be connected at any point in the telephone network 100, such as at core exchange 110 as shown in figure 1. Although not illustrated in figure 1, voicemail servers 160 may also or alternatively be connected to a local exchange 120. Each

of the voicemail servers 160 provides a voicemail facility to a plurality of customers of the telephone network 100. For example, a voicemail server 160 connected to a local exchange 120 might provide a voicemail facility for the customers whose telephony terminals 130 are directly connected to that local exchange 120. Of course it will be appreciated that a
5 multitude of other arrangements are possible.

The call data stores 170 each store a plurality of call data records representing some or all of the telephony calls made over the telephone network 100 for a given period of time. Each call data record will comprise the telephone number (or indeed any other suitable identifier) used by the calling party, the telephone number (or other suitable identifier) of the
10 called party, the time that the call started and the time that the call was terminated (or a time that the call started or ended and a duration of the call). The call data is provided periodically to the data stores by the one or more local exchanges 120 (and/or, in some embodiments, by the core exchanges 110) as calls are placed, connected and terminated in the telephone network 100. The provision of the call data to the call data stores 170 can be
15 achieved using any appropriate means of communication, such as by using a data network that is separate from the telephone network 100. As will be appreciated, each data store, may receive data from different sets of local exchanges 170, such that call data for the network 100 as a whole is spread across the data stores 170.

The analyst terminals 180 are computer systems which can access the data stored in the
20 data stores 170 (or at least, in some of the data stores 170). Programs may run on the analyst terminals 180 to analyse the call data stored in the data stores 170 including, for example, to classify whether particular callers are a source of nuisance (or unwanted) telephone calls, in accordance with embodiments of this invention.

As is well known, calls made by a customer telephony terminal 130 are initially handled
25 by the local exchange 120 to which the terminal 130 is connected via its respective communication link 150. If the destination of the call is another terminal 130 that is connected to the same local exchange 120, that local exchange 120 can route the call directly to its destination without involving any of the other components of the telephone network 100. Otherwise, if the destination terminal 130 is not on the same local exchange
30 120, the local exchange 120 routes the call to the respective core exchange 110 to which it is connected to handle the further routing of the call. If the call is destined for another terminal 130 on the network, the core exchange 110 routes the call, possibly via one of the other core exchanges 110, to the local exchange 120 to which that terminal 130 is connected. However, if the call is destined for a terminal on another network, the core
35 exchange 110 routes the call to one of the gateways for onward routing to that network. In

some cases, instead of routing a call to a customer's telephony terminal 160, the telephone network 100 can instead route a call to one of the voicemail servers 60 which provides a voicemail facility for that customer. The caller can then leave a message which will be recorded by the voicemail server 160 and can later be replayed by the customer at a time convenient to them. If a call is routed to the voicemail server a notification, such as a computer or smartphone notification, an SMS message and/or an email will be sent to the customer informing them of the presence of an unheard voicemail on the voicemail server 160. The decision to route a call to one of the voicemail servers 160 may be made if, for example, there is no answer from the customer's telephony terminal 160 after a predetermined number of rings or if a customer has specified that all call should be redirected to their voicemail.

It will be understood that the telephone network illustrated in figure 1 is merely exemplary and that various modifications may be made according to the needs of a specific telephone network. In some embodiments, various components described above may be absent from the telephone network 100. For example, the network may not include domestic gateways 140 and/or international gateways 150 if such connectivity to other networks is not required. Similarly, the telephone network 100 may not include voicemail servers 160 if no voicemail service is offered to customers of the network. Furthermore, a wide range of other components not illustrated in figure 1 may be present in the telephone network 100. Indeed, in general, it will be appreciated that there are many different forms that telephone network 100 may take using different combinations, numbers, types and/or arrangements of these components.

Figure 2 shows a flowchart illustrating how a method 200 according to the present invention can be implemented. The method 200 is operable to route calls from a network identity in a communications network, such as the telephone network 100 discussed in relation to figure 1. The method 200 starts with an operation 210.

At operation 210, the method 200 generates call duration distributions for nuisance callers.

A nuisance caller is an entity which is identified by a particular network identity, such as a Call Line Identification (CLI) number, which has already been determined to have placed nuisance calls via the network. For example, callers placing unusually large volumes or frequencies of calls, or who are the subject of numerous complaints, may be investigated and determined to have been making nuisance calls within the network and so are identified as being nuisance callers. A list of previously identified (or 'known') nuisance callers may be

maintained. As new nuisance callers are identified, they may be added to this known nuisance caller list. The list may include an indication of a period during which each nuisance caller is determined to have been making nuisance calls. That is to say, the list may indicate a point in time at which a particular network identity started being a source of nuisance calls and, if applicable, a point in time at which that network identity ceased being a source of nuisance calls (for example, when measures preventing that network identity placing nuisance calls via the network took effect or when the network identity becomes associated with a different entity that is not a source of nuisance calls). Alternatively, a separate list may be maintained to cover a particular period of time. For example, a separate list may be maintained for each day which indicates the network identities that were determined to have made nuisance calls on that particular day. Of course it will be appreciated that separate lists covering any other suitable period of time (e.g. week, month, quarter, year, etc.) could be used instead. In any case, the list, or lists, enable those callers which were known to be placing nuisance calls via the network 100 during a particular period of time, to be identified.

A call duration distribution for a known nuisance caller (which may be referred to as a nuisance call duration distribution) is generated from call data associated with the network identity (e.g. CLI) of the nuisance caller. The call duration distribution is a frequency distribution which represents the number (or frequency) of calls made by the nuisance caller which have a duration which lies in each of a plurality of ranges of durations. For example, the plurality of ranges of durations used to generate the call duration distribution may each span a particular time period, such as one minute. Accordingly, a first value in the distribution may represent the number of calls that were made by the nuisance caller which had a duration of up to 1 minute, whilst a second value in the distribution may represent the number of calls that were made by the nuisance caller which had a duration of at least 1 minute but less than 2 minutes, a third distribution may represent the number of calls that were made by the nuisance caller which had a duration of at least 2 minutes but less than 3 minutes, and so on. Of course it will be appreciated that the use of 1 minute intervals to define the ranges in the distribution is merely exemplary and any other suitable time period (whether longer or shorter) may be used to define the intervals for the ranges in the call duration distribution instead. Accordingly, these nuisance call duration distributions characterise how long customers spend on the phone when they are contacted by each of the nuisance callers.

Each call duration distribution may be associated with a respective period of time. That is to say, it may represent the duration distribution of the calls placed by a network identity

within a particular period of time. For example, each call duration distribution may be associated with a particular day and may reflect the durations of the calls made by the network identity on that particular day (although again, it will be appreciated that periods of time either shorter or longer than a day may be used instead).

5 To generate the nuisance call duration distributions, the method 200 may use the list of known nuisance callers to identify the network identities of nuisance callers that were active (i.e. making nuisance calls) during a particular time period. For example, the method 200 may refer to the list containing all known nuisance callers that were identified as placing nuisance calls on a given day. The method 200 may then obtain call data relating to those
10 network identities which covers that particular time period. For example, where the method 200 is being performed by an analyst terminal 180 in network 100, the analyst terminals may access the call data stored in the data stores 170 to retrieve call data for the network identities which covers that time period. The call data represents a plurality of calls that were made by the network identities during that time period and indicates the durations of
15 each call placed by the network identity via the network 100, either explicitly or implicitly (e.g. by providing a start time and an end time of the call, thereby allowing its duration to be determined). A distribution can then be formed for each nuisance caller by determining a number of calls that were made by that nuisance caller which fall within each of the plurality of ranges of duration intervals within the distribution, as discussed above.

20 In some cases, the method 200 may determine the nuisance call duration distributions for a single period of time. This means that a single set of call duration distributions is generated by generating a call duration distribution for each known nuisance caller that was active during that period of time. Accordingly, the call duration distribution for each known nuisance caller is generated based on call data covering the same single period of time.

25 However, in other cases, multiple sets of nuisance call duration distributions may be generated, wherein each set of nuisance call duration distributions relates to a different period of time. Whilst the nuisance call duration distributions in a particular set are all generated based on call data covering the same period of time, multiple nuisance call duration distributions may be generated for a particular network identity each covering a
30 different time period and being associated with a different set of nuisance call duration distributions that all relate to that same time period. It will be appreciated that nuisance call duration distributions may only be generated for a particular network identity during those periods of time in which that network identity was determined to have been making nuisance calls via the network 100. Accordingly, the set of nuisance callers that is used to generate
35 each set of nuisance call duration distributions may differ between different time periods

where different network identities have been identified as making nuisance calls during those time periods.

Where multiple time periods are considered, those time periods may be consecutive adjoining (i.e. non-overlapping) time periods. Accordingly, the multiple time periods may
5 subdivide a larger time period such that each point of the larger time period is associated with exactly one of the time periods. For example, a number k of sets of nuisance call duration distributions N_i may be generated to cover a period of k preceding days (i.e. $\{N_1, N_2, \dots, N_k\}$), each set containing nuisance call duration distributions for a respective one of the k preceding days (although, again, any other suitable time period could be used
10 instead). These sets may be continually updated over time. For example, as each day passes, a set of nuisance call duration distributions may be generated for all network identities that have been identified as placing nuisance calls during that day and the set of nuisance call duration distributions can be added to the sets of nuisance call duration distributions that are available for use with the invention. It will be appreciated that this need
15 not be done in real time. That is to say, the nuisance call duration distribution for a particular day may actually be determined sometime (such as a number of days) after that day has passed. This may allow time for investigations from other sources (such as in response to customer complaints) to identify (or confirm) nuisance callers that were active during that time period.

20 In some cases, rather than being consecutive adjoining (i.e. non-overlapping) time periods the time periods for which each set of nuisance call duration distributions are determined may be consecutive occurrences of a periodically occurring time period. That is to say, each of the time periods may be separated from each other by a regular pattern. Under this approach, there are periods of time between time periods for which no nuisance
25 call duration distributions are generated. For example the time periods may have a duration of a few hours and may recur daily. This means that each of the time periods covers a particular portion of the day, such that a respective set of nuisance call duration distributions will be produced to cover that same portion of each day. However, calls made outside of that portion of the day may be ignored. For example, the portion of the day for which the
30 nuisance call duration distributions are generated may be one of morning (e.g. 8am – 12pm), early afternoon (e.g. 12pm – 3pm), late-afternoon (e.g. 3pm – 6pm) and evening (e.g. 6pm onwards), although it will be appreciated that other time period recurring in different patterns may be used instead. As a further example, the time periods might be for a whole day and repeat every seven days such that the nuisance call duration distributions all relate to a
35 particular day of the week and so on. As will be discussed in more detail below, the

nuisance call duration distributions may exhibit more distinctive characteristics during certain periods of time compared to others.

In any case, having generated one or more sets of nuisance call duration distributions for one or more respective periods of time, the method 200 proceeds to an operation 220.

5 At operation 220, the method 200 determines common characteristics of the call duration distributions for the nuisance callers. This is achieved by using a dimensionality reduction technique, such as Principal Component Analysis (PCA), to analyse each set of nuisance call duration distributions that was generated in operation 210 (which in some cases may only be a single set) to determine the common characteristics of the call duration
10 distributions in each set. In some cases, the common characteristics may be represented by the principal components of the set of nuisance call duration distributions. Accordingly, the eigenvectors (which provide the principal components) of each set of nuisance call duration distributions can be determined. As will be appreciated by those skilled in the art, who will be familiar with PCA as well as other dimensionality reduction techniques, these
15 eigenvectors characterise the call duration distributions for the nuisance callers during the period of time associated with the set of nuisance call duration distributions from which they were determined. In some cases, the lower valued eigenvectors may be ignored, leaving only the higher valued eigenvectors which are more representative of the nuisance call duration distributions. Although this discussion of the invention mainly contemplates the use
20 of PCA, any other suitable technique for deriving the common characteristics of the sets of call duration distributions may be used instead.

In cases where multiple time periods are being considered, there are generally two approaches that may be taken to storing the common characteristics for the different sets of nuisance call duration distributions associated with each time period.

25 A first approach is to store the common characteristics for each time period so that they can be used individually. For example, the eigenvectors of the set of nuisance call duration distributions for each day may be stored in a database, such that there are k days of stored eigenvectors V_i for $i = 0, 1, \dots, k$ generated from historical data which are individually available for use (as will be discussed in relation to operation 230 below).

30 A second approach is to combine the common characteristics for each time period into a single set of common characteristics which represent all of the time periods. This single set of common characteristics may then be updated whenever a new set of common characteristics for a further time period are derived. This set of common characteristics can

then be updated whenever the common characteristics of another set of nuisance call duration distributions becomes available (e.g. when a new day's call data is processed). A suitable approach for combining the common characteristics into a single set is provided by the Recursive PCA technique that is described in in "*Perturbation-Based Eigenvector Updates for On-Line Principal Components Analysis and Canonical Correlation Analysis*" by Hedge et al published in the Journal of VLSI Signal Processing 45, 85-95, 2006 (DOI: 10.1007/s11265-006-9773-6), which is hereby incorporated by reference in its entirety (and referred to herein as "Hedge et al"). The combination of the common characteristics may be a straight forward combination where all of the time periods are treated equally (i.e. those eigenvectors obtained in a more recent time period, such as yesterday, are treated the same as those obtained in a less recent time period, such as the same day one year ago). This means that there is no difference in the contribution from the common characteristics of an older time period compared to those for a more recent time period. Alternatively, a forgetting factor may be introduced, which weights the eigenvectors for more recent time periods more strongly. This means that, eigenvectors from older time periods have less influence on the combined common characteristics. Hence, the impact of the common characteristics derived during a particular period of time may be considered to be "*forgotten*" once they have reached a certain age. These approaches are discussed, for example, in section 3.2 of Hedge et al., which introduces a memory depth parameter λ_k for a time period k that can be used to weigh all samples equally (by setting $\lambda_k = \frac{1}{k}$) or can provide a first-order dynamical forgetting strategy (by setting $\lambda_k = \lambda$ where $\lambda \in (0, 1)$ which will give an average memory depth of $\frac{1}{\lambda}$ samples).

The eigenvectors which are stored (and which represent the common characteristics of nuisance calls over a particular period of time) may be truncated. That is to say, only those eigenvectors associated with the largest eigenvalues may be stored, whilst those eigenvectors with smaller eigenvalues may be ignored. As will be appreciated by those skilled in the art, the eigenvectors with the larger eigenvalues will be more representative of the characteristics of nuisance calls than eigenvectors with smaller eigenvalues. Similarly, the eigenvectors with smaller eigenvalues may be more likely to represent characteristics that are also present in non-nuisance calls. Accordingly, only a predetermined proportion of the eigenvalues having the highest eigenvalues may be stored. Alternatively, a predetermined threshold may be defined such that only those eigenvectors having an eigenvalue higher than the threshold will be kept.

In any case, having determined common characteristics of the call duration distributions for the nuisance caller (whether by PCA or any other suitable technique), the method 200 proceeds to an operation 230.

At an operation 230, the method 200 classifies the calls made by a network identity. This network identity may be referred to in the following description as a candidate nuisance caller. The classification of the calls for the candidate nuisance caller provides an indication of whether the calls are determined to be nuisance calls (or not). The method 200 classifies the calls based on a similarity between a call duration distribution of the calls and the common characteristics of the nuisance call duration distributions (as determined at operation 220). Accordingly, at operation 230, the method 200 generates a call duration distribution for the calls made by the network identity. This call duration distribution can be generated in a similar manner to the generation of the nuisance call duration distributions for the nuisance callers described above and so will not be discussed again here.

As will be appreciated, the method 200 may be used to classify the calls made by multiple network identities, in which case, the operation 230 may be repeated in respect of the calls made by each network identity to produce a respective classification of the calls made by that network identity. Similarly, multiple sets of calls by the same network identity may be considered separately be a respective repetition of operation 230. For example, the method 200 may operate to classify the calls made by a network identity on each of a plurality of days.

In order to identify a network identity whose calls should be evaluated by operation 230, the method 200 may monitor a total number of calls Vol_{CLI} made by each of a plurality of network identities (such as all of the network identities in use in the network) over a particular period of time, such as a day (although other periods of time may be used instead). Any network identities that place more than a predetermined threshold number of calls (i.e. where $Vol_{CLI} > threshold$) may be considered as a candidate nuisance caller and operation 230 of method 200 may be performed in respect of that network identity's calls during that time period. This can be useful where limited computational resources are available for performing method 200 because it can help focus resources on those network identities that are more likely to be the source of nuisance calls (or at least whose nuisance calls are likely to be causing a greater amount of nuisance). However, it will be appreciated that this technique for selecting a nuisance caller need not be used with method 200 and that other approaches are possible. For example, if sufficient computation resources are available all network identities that placed calls in the network 100 could be evaluated.

Alternatively, a random sampling approach could be adopted. As a yet further example, the

network identity may be one that is identified in a complaint that is received by an operator of the network.

Any suitable technique for determining a similarity between the common characteristics of the nuisance call duration distributions and the call duration distribution of the candidate
5 nuisance caller may be used.

As an example, where the common characteristics of the nuisance call duration distributions are represented by a set of eigenvectors V that has been determined via principle component analysis, the similarity between the call duration distribution x for the candidate nuisance caller and the common characteristics can be determined by projecting
10 the call duration distribution onto a domain defined by the principal components that represent the common characteristics of the nuisance call duration distributions and then reconstructing the call duration distribution by projecting it back into the original domain. That is to say, a projected call duration distribution p can be determined in the domain defined by the principal components is generated by $p = V^T x$. A reconstructed call
15 distribution $x^{(r)}$ can then be determined by projecting the projected call duration distribution p back into the original domain by $x^{(r)} = Vp$. A measure of loss δ can then be determined between the original call duration distribution x for the candidate nuisance caller and the reconstructed call distribution $x^{(r)}$ by $\delta = \|x - x^{(r)}\|^2$. Where the call duration distribution of the candidate nuisance caller has the same common characteristics as the call duration
20 distributions of the known nuisance callers, there will be a minimal loss of information between the projections into the domain defined by the principal components and back again. Conversely, where the call duration distribution of the candidate nuisance caller does not share many of the same common characteristics as the call duration distributions of the known nuisance callers, there will likely be a much greater loss of information between the
25 projections into the domain defined by the principal components and back again. Accordingly, the measure of loss δ provides a measure of similarity between the call duration distribution of a candidate nuisance caller and the common characteristics represented by a set of eigenvectors V . The lower the measure of loss δ the more similar the call duration distribution of the candidate nuisance caller is to the common
30 characteristics of the nuisance call duration distributions (and vice-versa).

Therefore, where a single time period is considered when generating the nuisance call duration distributions for the known nuisance callers such that a single set of common characteristics is determined at operation 220, or where a single combined set of common characteristics is maintained from the common characteristics determined for multiple time

periods, the classification may be based on the similarity between the call duration distribution for the candidate nuisance caller and that single set of common characteristics. For example, following the approach outlined above, if the measure of loss δ is less than a predetermined threshold (i.e. where the similarity is greater than a predetermined threshold),
 5 the calls of the candidate nuisance caller may be classified as nuisance calls.

However, where multiple sets of nuisance call duration distributions are stored by operation 220 associated with a plurality of periods of time, the classification of the plurality of calls made by the network identity may be based on a similarity between the call duration distribution of the plurality of calls for the candidate nuisance caller and the common
 10 characteristics of each set of nuisance call duration distributions. That is to say, a respective measure of similarity may be determined between each of the sets and the call duration distribution for the candidate nuisance caller. For example, returning to the approach discussed above where the common characteristics of each of the sets of the nuisance call duration distributions are determined via principle component analysis, k sets of
 15 eigenvectors V_i may be available, each set of eigenvectors V_i being associated with a respective one of k periods of time. For example, if a period of time of one day is used and the common characteristics of nuisance call duration distributions on each day of the preceding week are stored, there will be seven sets of eigenvectors V_i available for use, each representing the calling characteristics of known nuisance callers on a respective one
 20 of the preceding seven days. Following the same approach, a set of k projections p_i of the call duration distribution x for the candidate nuisance caller can be created, one for each time period k , by applying each of the stored eigenvectors V_i to the call duration distribution, i.e. $\{p_1 = V_1^T x, p_2 = V_2^T x, \dots, p_k = V_k^T x\}$. Similarly, the k projections p_i of the call duration distribution x for the candidate nuisance caller can be projected back into the original
 25 domain by an inverse operation to generate a set of k reconstructed call duration distributions $X = \{x_1^{(r)}, x_2^{(r)}, \dots, x_k^{(r)}\}$, i.e. $\{x_1^{(r)} = V_1 V_1^T x, x_2^{(r)} = V_2 V_2^T x, \dots, x_k^{(r)} = V_k V_k^T x\}$. The set of k reconstructed call duration distributions X can then be used to generate a set Δ of k measures of loss, i.e. $\Delta = \{\|x - x_1^{(r)}\|^2, \|x - x_2^{(r)}\|^2, \dots, \|x - x_k^{(r)}\|^2\}$, whereby each of the measures of loss represents the similarity between the call duration distribution x and the
 30 common characteristics of the nuisance call duration distributions for nuisance calls made within a respective one of the k periods of time. Accordingly, if a period of time of one day is used and the common characteristics of nuisance call duration distributions on each day of the preceding week are stored, seven measures of similarity may be generated, one for each day. The set of calls for the candidate nuisance caller may be classified as nuisance
 35 calls if any of the measures of loss in the set Δ are less than a given threshold (i.e. if the call

duration distribution of the set of calls has similar characteristics to the call duration distributions of known nuisance callers during any of the k periods of time).

Having classified the calls made by a candidate nuisance caller at operation 230, the method proceeds to an operation 240.

- 5 At operation 240, the method 200 routes further calls from the network identity in accordance with the classification that was determined at operation 230.

Where the classification indicates that the network identity is considered to have been making nuisance calls, the network 100 may be configured to handle those calls appropriately. For example, a list may be maintained by the network 100 to store the
10 network identity of callers who have been classified as nuisance callers at operation 230. Any calls from network identities that are not on this list (i.e. which were not classified as making nuisance calls during operation 230) may be handled as normal, for example by routing them to the intended recipient. However, any calls from network identities that are on this list may be handled differently from normal. For example, the network 100 may be
15 configured to route calls from network identities on this list to a voicemail server so that customers are not disturbed by those calls. If a call is routed to a voicemail server, then a notification (such as an email, SMS message, smartphone or PC notification and so on) may be sent to the customer informing them that a calling party attempted to make a call to them which was blocked due to it being classified as a suspected nuisance call. The notification
20 may provide a link to the voicemail message to allow the customer to listen to the voicemail message if they so wish. Of course it will be appreciated that any other suitable method of handling further calls from the network identities that have been classified as making nuisance calls may be used. As a further example, calls from such numbers could simply be blocked instead. Alternatively or additionally, the network identity may be flagged for further
25 investigation to confirm whether it is actually placing nuisance calls (and, in some cases, such confirmation may be required before call routing is actually changed for such numbers).

If a particular network identity is added to this list (following the classification of a plurality of calls made by that network identity as being nuisance calls at operation 230), then the owner of that network identity may be informed by the network operator. It is known that
30 entities which place large numbers of nuisance calls typically lease telephone numbers from third party brokers and thus are likely to move from number to number as they are added to this list. Therefore, the network operator may also provide a mechanism by which such numbers can be removed from this list.

Having configured the network 100 to route further calls from the network identity based on the classification of the plurality of calls from that network identity, the method 200 ends. However, it will be appreciated that the method may be re-iterated to classify calls from a different network identity (or for the same network identity in a different period of time). Of course, it may not be necessary to repeat all operations of the method 200. In particular, operations 210 and 220 need not be repeated unless there is new data for known nuisance callers to be processed (e.g. where another period of time has passed when the method 200 is operated continuously). Furthermore, even though the method 200 ends, it will be appreciated that the routing for calls from the network identity, as configured in operation 240, may continue to have effect.

It will be understood that the present invention relates to a method of classifying a source of telephone calls to determine whether they can be classified as nuisance calls or non-nuisance calls. It will be further understood that the exact details of how these calls are then processed are not directly relevant to the present invention and that alternative schemes to those outlined above may be used. It will be understood that some large users of the telephony network (banks, government agencies, etc.) will make many calls each day but that these calls are not nuisance calls and are not expected to have call duration distributions which have the same common characteristics as nuisance calls. Nonetheless, the network operator may also establish a permitted list of telephone numbers used by such entities which will be generating large numbers of calls. Telephone numbers on this permitted list will not be analysed in accordance with the method described above and, in particular, will not be selected as candidate nuisance callers for operation 230.

The foregoing discussion has been focussed on the use of the present invention to detect, and then re-route, nuisance calls in a conventional PSTN network. It will be understood that the present invention could also be used to detect nuisance calls made using IP-based voice services, for example Voice over IP calls. In this case, the network identity may comprise an email address, a SIP address, a conventional telephone number or any other unique identifier.

Through the use of dimensionality reduction techniques to identify the common characteristics of known nuisance callers, the invention enables other (previously unknown) nuisance callers to be identified. In particular, the call duration distributions of nuisance callers, can be expected to differ from those of regular (i.e. non-nuisance) callers as they reflect the recipients' response to the incoming nuisance calls (e.g. as determined by the length of time for users to figure out that the call is a nuisance call and/or the length of time for a nuisance message to be completely relayed before terminating the call).

It may be expected that the characteristics of call duration distributions for nuisance callers change over time, for example, as different strategies for engaging the recipients of nuisance calls are employed or different techniques for trying to mask their calls are employed (such as by mixing nuisance and non-nuisance calls being made from the same number) or simply as the types of scams being employed by nuisance callers vary on a national level. However, these changing characteristics can be monitored and accounted for through the analysis of nuisance call duration distributions over multiple periods of time, possibly on an ongoing basis (e.g. daily). Therefore, a candidate nuisance caller's calling behaviour can be compared to the different types of behaviour that have been seen previously and the invention can adapt to reflect changes in nuisance caller behaviour over time.

It may further be expected that the recipient's response to nuisance calls may be especially distinctive at particular times of the day (or indeed on various other periodically occurring time periods, such as at weekends). Indeed, it is known from operations at legitimate call centres that customers tend to be willing to spend longer talking to operators at certain times of day. Accordingly, there may be a greater contrast between the call duration distributions of nuisance and non-nuisance calls at certain times of day (or during other periodically occurring time periods), which may allow nuisance callers to be more readily identified.

Ideally, where the common characteristics of the nuisance call duration distributions are derived from nuisance calls made during a plurality of periods of time (such as each day over the past week), the call duration distribution for the candidate nuisance caller will be determined from a time period of similar duration (such as over the course of a single day). This may help to provide a more accurate comparison. Similarly, where the plurality of periods of time from which the nuisance call duration distributions are drawn cover a periodically occurring time period, such as a particular portion of each day, the call duration distribution for the candidate nuisance caller should ideally be generated from the same time period, such as from the same portion of the day (e.g. if the nuisance call duration distributions are generated from calls that were placed during the morning, the calls from which the call duration distribution of the candidate nuisance caller should ideally also be generated from calls that were placed during the morning). However, it will be appreciated that this is not absolutely necessary, as even if the periods of time differ in duration, some of the common characteristics are likely to still be present (if the candidate nuisance caller is indeed a nuisance caller) allowing a comparison of characteristics between call duration distributions to be made.

Insofar as embodiments of the invention described are implementable, at least in part, using a software-controlled programmable processing device, such as a microprocessor, digital signal processor or other processing device, data processing apparatus or system, it will be appreciated that a computer program for configuring a programmable device,

5 apparatus or system to implement the foregoing described methods is envisaged as an aspect of the present invention. The computer program may be embodied as source code or undergo compilation for implementation on a processing device, apparatus or system or may be embodied as object code, for example. Suitably, the computer program is stored on a carrier medium in machine or device readable form, for example in solid-state memory, 10 magnetic memory such as disk or tape, optically or magneto-optically readable memory such as compact disk or digital versatile disk etc., and the processing device utilises the program or a part thereof to configure it for operation. The computer program may be supplied from a remote source embodied in a communications medium such as an electronic signal, radio frequency carrier wave or optical carrier wave. Such carrier media 15 are also envisaged as aspects of the present invention. It will be understood by those skilled in the art that, although the present invention has been described in relation to the above described example embodiments, the invention is not limited thereto and that there are many possible variations and modifications which fall within the scope of the invention. The scope of the present invention includes any novel features or combination of features disclosed 20 herein. The applicant hereby gives notice that new claims may be formulated to such features or combination of features during prosecution of this application or of any such further applications derived therefrom. In particular, with reference to the appended claims, features from dependent claims may be combined with those of the independent claims and features from respective independent claims may be combined in any appropriate manner 25 and not merely in the specific combinations enumerated in the claims.

CLAIMS

1. A method of routing calls from a network identity in a communications network, the method comprising:

5 using a dimensionality reduction technique to determine one or more common characteristics of a plurality of nuisance call duration distributions, each of the nuisance call duration distributions being generated from a plurality of calls placed by a respective nuisance caller;

classifying a plurality of calls made by the network identity based on a similarity

10 between a call duration distribution of the plurality of calls and the common characteristics of the nuisance call duration distributions; and

routing further calls from the network identity in accordance with the classification.

2. The method of claim 1, wherein:

15 the plurality of nuisance call duration distributions comprise a plurality of sets of nuisance call duration distributions, each set being associated with a respective period of time, wherein the nuisance call duration distributions in each set are generated from a plurality of calls placed by nuisance callers during the respective period of time; and

20 using the dimensionality reduction technique to determine one or more common characteristics of the plurality of nuisance call duration distributions comprises using the dimensionality reduction technique to determine one or more common characteristics of each set of nuisance call duration distributions.

3. The method of claim 2, wherein the classification of the plurality of calls made by the
25 network identity is based on a similarity between the call duration distribution of the plurality of calls and the common characteristics of each set of nuisance call duration distributions.

4. The method of claim 3, wherein the plurality of calls are classified as nuisance calls when the similarity between the call duration distribution and any of the nuisance call
30 duration distributions is greater than a predetermined threshold.

5. The method of claim 2, further comprising combining the common characteristics of each set of nuisance call duration distributions, wherein the classification of the plurality of calls is based on a similarity between the call duration distribution of the plurality of calls and
35 the combined common characteristics of the plurality of sets of nuisance call duration distributions.

6. The method of any one of claims 2 to 5, wherein the respective periods of time for the plurality of sets of nuisance call duration distributions are consecutive adjoining preceding time periods.

5

7. The method of claim 6, wherein the periods of time each have a duration of a day.

8. The method of any one of claims 2 to 5, wherein the respective periods of time for the plurality of sets of nuisance call duration distributions are consecutive occurrences of a periodically occurring time period.

10

9. The method of claim 8, wherein the periodically recurring time period is daily and the periods of time have a duration which covers a particular portion of the day.

15

10. The method of claim 9, wherein the particular portion of the day is one of: morning; early-afternoon; late-afternoon; and evening.

11. The method of any one of the preceding claims, wherein the common characteristics are represented by principal components obtained from carrying out a principal component analysis of the nuisance call duration distributions.

20

12. The method of claim 11, wherein the similarity between the call duration distribution of the plurality of calls and the common characteristics is determined by:

projecting the call duration distribution onto a domain defined by the principal components that represent the common characteristics to create a projected call duration distribution;

25

projecting the projected call duration distribution back into its original domain to create a reconstructed call duration distribution; and

determining a measure of loss between the reconstructed call duration distribution and the call duration distribution of the plurality of calls, the measure of loss indicating the similarity between the call duration distribution and the common characteristics.

30

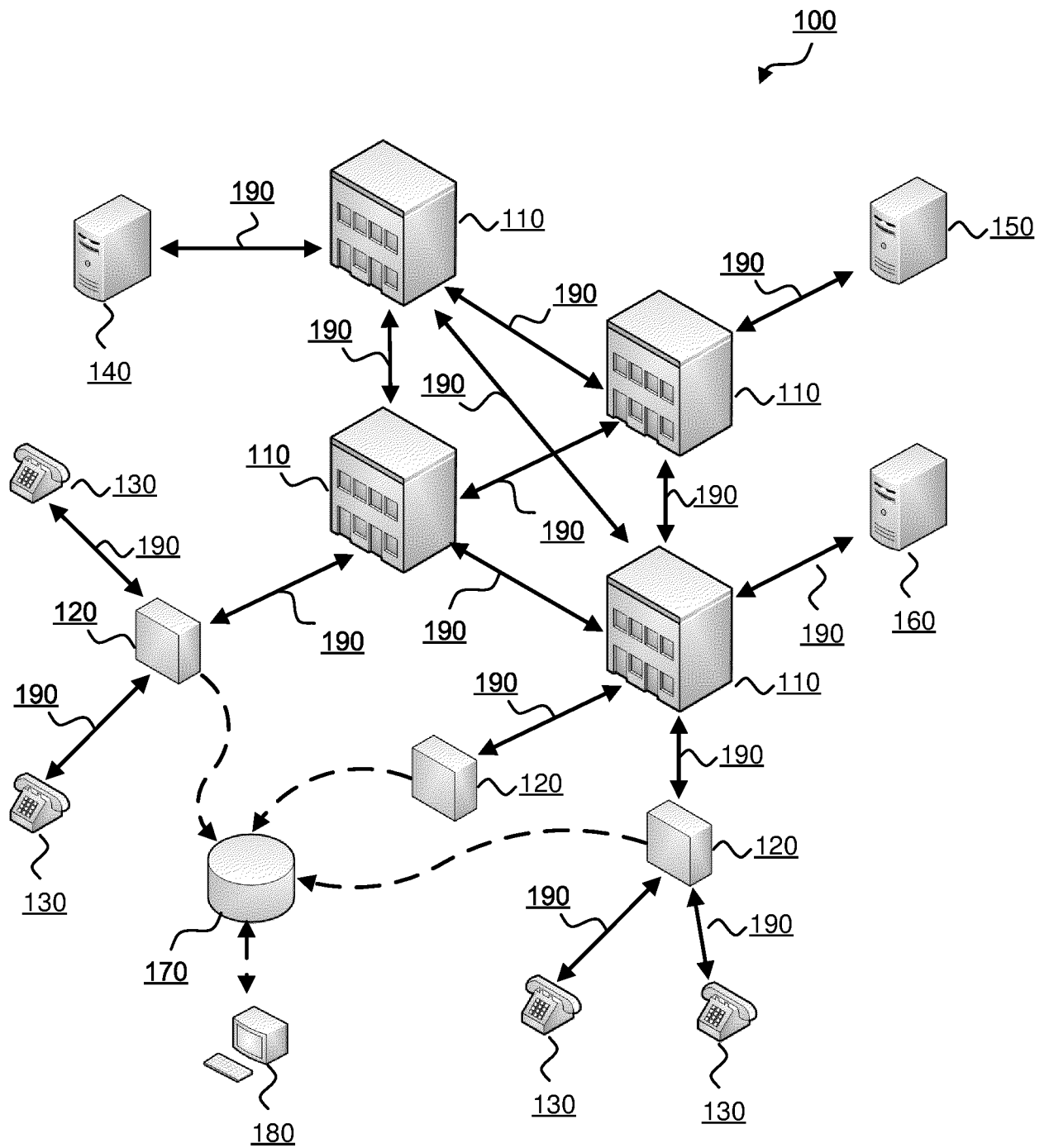
13. The method of any one of the preceding claims, wherein the network identity making the plurality of calls to be classified is selected for classification in response to a total number of calls made by the network identity in a particular period of time exceeding a predetermined threshold.

35

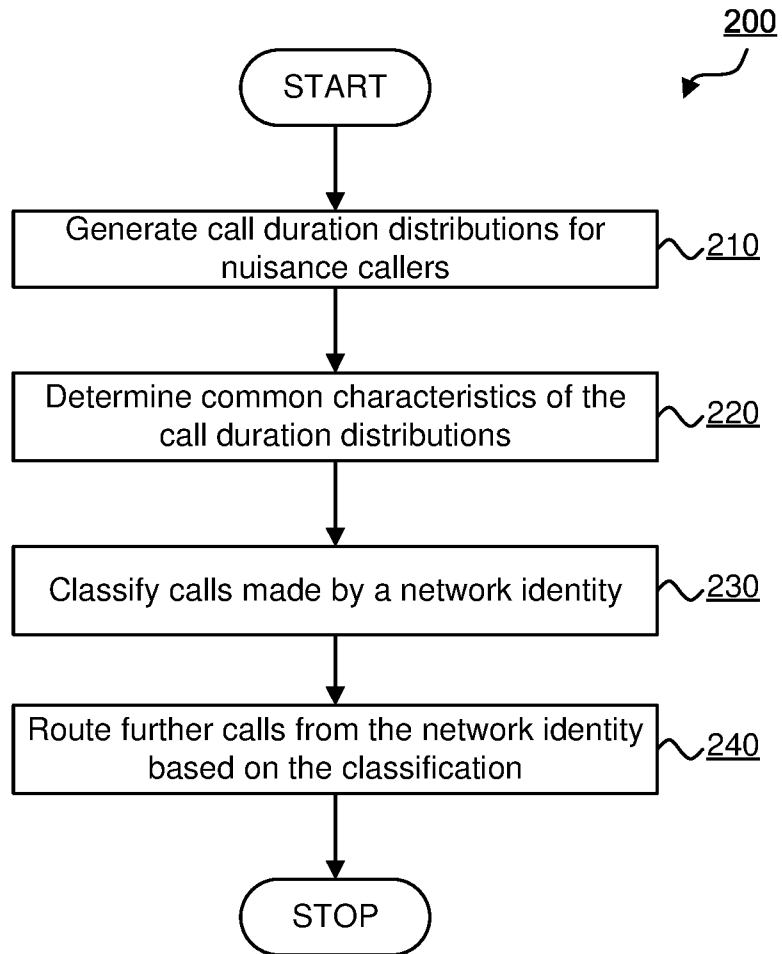
14. A computer system comprising a processor and a memory storing computer program code for performing the steps of any one of the preceding claims.

15. A computer program which, when executed by one or more processors, is arranged
5 to carry out a method according to any one of claims 1 to 13.

1/2

**FIGURE 1**

2/2

**FIGURE 2**

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2022/057916

A. CLASSIFICATION OF SUBJECT MATTER**INV. H04M3/436****ADD.**

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)

H04M

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 110 913 081 A (SHANGHAI GUANAN INFORMATION TECH CO LTD) 24 March 2020 (2020-03-24) abstract figure 1 paragraph [0014] - paragraph [0030] paragraph [0038] - paragraph [0039] paragraph [0068] - paragraph [0080] paragraph [0102] - paragraph [0117] -----	1-15
X	EP 3 580 920 A1 (BRITISH TELECOMM [GB]) 18 December 2019 (2019-12-18) abstract figures 2-4 paragraph [0010] - paragraph [0011] paragraph [0013] - paragraph [0018] paragraph [0019] - paragraph [0021] paragraph [0022] - paragraph [0023] -----	1-15



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

14 June 2022

Date of mailing of the international search report

23/06/2022

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2

NL - 2280 HV Rijswijk

Tel. (+31-70) 340-2040,

Fax: (+31-70) 340-3016

Authorized officer

Molinari, Fausto

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2022/057916

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
CN 110913081	A	24-03-2020	NONE

EP 3580920	A1	18-12-2019	EP 3580920 A1 18-12-2019
		US 2020045169 A1	06-02-2020
		WO 2018178393 A1	04-10-2018
