



(51) International Patent Classification:

H04L 29/08 (2006.01) *G06Q 30/02* (2012.01)
G06F 17/30 (2006.01) *H04W 4/18* (2009.01)

(21) International Application Number:

PCT/EP2017/025053

(22) International Filing Date:

20 March 2017 (20.03.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

15/073,717 18 March 2016 (18.03.2016) US
1604649.2 18 March 2016 (18.03.2016) GB

(71) Applicant: **ADBRAIN LTD** [GB/GB]; Albert House, 256-260 Old Street, London EC1V 9DD (GB).

(72) Inventors: **ATLAS, Jason**; 5 Brecon Mews, London N7 0BN (GB). **KALO, Fady**; Flat 3, Chaplin Apartments, 5 Sylver Path, London E8 1EN (GB). **MA, Jiefei**; 165 Seaforth Gardens, Stoneleigh, Surrey, KT19 0LW (GB). **SABER, Pourya**; 29 Applegate, 20 Victory Lane, London E20 1GF (GB). **CHUA, Seyiang**; Flat 2, Helier Court, Eleanor Close, London SE16 6PZ (GB). **ABRAHAM, Timothy Ashley**; 32 Heath Drive, London EN6 1EH (GB).

(74) Agent: **NORRIS, Timothy Sweyn**; BASCK LTD, 16 Sax-on Road, Cambridge, Cambridgeshire CB5 8HS (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, 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) 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).

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))
- of inventorship (Rule 4.17(iv))

Published:

- with international search report (Art. 21(3))

(54) Title: DATA COMMUNICATION SYSTEMS AND METHODS OF OPERATING DATA COMMUNICATION SYSTEMS

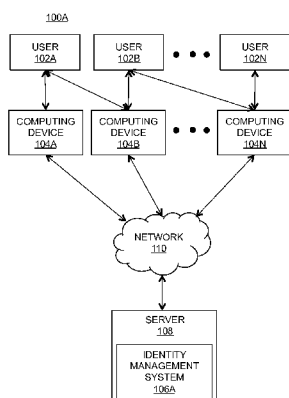


FIG. 1A

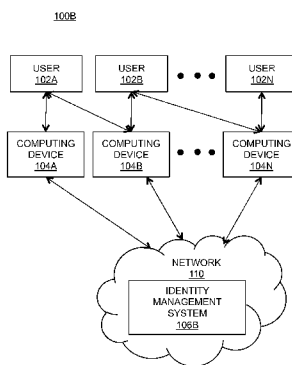


FIG. 1B

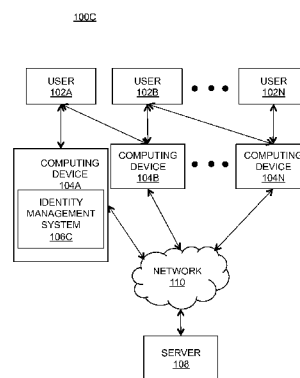


FIG. 1C

(57) Abstract: The present disclosure provides a data communication system and methods of controlling data flows within a data communication system. The data communication system includes a data communication network to which users are coupled when the data communication system is in operation. The data communication system also includes a data server. The users have associated user identifiers, and can interact with the system via user devices, or the user identifiers are derived in a transformed manner from information describing the users. The method includes: receiving the user identifiers; determining matched pairs of the user identifiers; comparing the matched pairs of the user identifiers with each of the received user identifiers, and determining a similarity score for the matched pairs and the user identifiers; mapping the user identifiers onto at least one single customer view identifier, each customer view identifier is distinctly (namely to within a mapping threshold) mapped to user identifier(s) for generating a mapping function; and using the mapping function for selectively controlling distribution of data content to the users.

DATA COMMUNICATION SYSTEMS AND METHODS OF OPERATING DATA COMMUNICATION SYSTEMS

TECHNICAL FIELD

[001] The present disclosure relates to data communication systems, and also to methods of operating aforesaid data communication systems; for example, the present disclosure relates to data communication systems employing data mining, wherein the data mining involves creating non-overlapping groups of user identifiers representing a single user. Moreover, the present disclosure discloses methods of operating data communication systems for controlling data flows to multiple users that are coupled via data communication networks of the aforesaid data communication systems. Furthermore, the present disclosure relates to computer program products comprising non-transitory computer-readable storage media having computer-readable instructions stored thereon, the computer-readable instructions being executable by a computerized device comprising processing hardware to execute the aforesaid methods.

BACKGROUND

[002] Recently, there has been a rapid increase in the number of users connecting to data communications networks such as, for example, the Internet®, for performing various user activities. The user activities include, for example, social networking, executing e-commerce transactions, communicating data, providing and/or receiving entertainment, and so forth. For example, a given user can connect to a given data communication network and access information therefrom, for example data content, using multiple devices coupled to the given data communication network, wherein the multiple devices are located at mutually different spatial locations. Nowadays, many data mining companies analyze and utilize data pertaining to activities of users coupled to the Internet® in order to derive substantial information relating to personalities of the users, spatial locations

of the user, shopping preferences of the users, personal interests of the users, trends associated with the users, and so forth. Such analyzed data, namely information, assists the data mining companies to plan customer-centric marketing strategies, such as targeted advertising campaigns, targeting a specific demographic area, targeting a group of users, providing improvements in data communication network data delivery performance, providing technical data processing improvements, and so forth. As a result of a given same user accessing the Internet® using multiple devices, there may result redundant user activity data in respect of mutually different user devices; the size of user activity data increases substantially, though it is not of much relevance to data mining companies for identifying trends, and user preferences in comparison to multiple devices' data is of a single user only. An ability to identify and present to individuals or organizations a probabilistic graph of user profile is a nascent field, with considerable activity presently being expended to improve such ability. Both technology and approaches for generating such probabilistic graphs pertaining to these very well understood problems, from marketing and advertisement technology ("*ad tech*"), to media and entertainment/connected homes, to the Internet® of Things, all depend on having an unambiguous understanding of what devices coupled via use of Internet Protocol are associated with a given user, in order to create a unified identity for the given user. The concept of identity on the Internet® has largely been established through an explicit process of authentication, which allows applications and services to have explicit identification relationships as a given user moves between multiple devices that are couplable to the Internet®.

[003] Presently, there exist many techniques for analyzing operating data of mutually different devices and data related to network usage by one or more users to identify identification relationships between the one or more users and the mutually different devices. The techniques for analyzing and/or determining the identification relationships are based primarily around authentication of user credentials such as, but not limited to, a login identifier, a login password, and so forth. Every device connected to the network has an associated Internet Protocol (IP) address. Other than the IP address, every device also has a unique device

identifier such as, an International Mobile Equipment Identity (IMEI).

[004] For accessing the Internet®, a given user may use a web browsing application such as, but not limiting to, the Internet Explorer®, Google Chrome®, and the like for accessing various websites. Some websites save cookies or browsing history of the given user. The cookies may include a fragment of text or information sent to the web browsing application while the given user is accessing a given user-selected website. Various models to calculate a probability of a same given user accessing the Internet® through multiple devices are based on IP address identification and use of cookies. Many existing techniques use cookies to identify various relationships such as, a relationship between various activities of the user, a relationship of the user with multiple devices, and so forth. However, determination of such relationships across multiple websites using cookies may present privacy concerns for the users.

[005] Furthermore, every user connected to a data communication network such as the Internet® has a digital identity including online information related to the user. Certain methods of analyzing the digital identity also involve scrutinizing Internet® usage data for all the devices using a same given IP address. The techniques that adopt the analysis of Internet® usage frequency count across multiple devices may be highly inaccurate, since a given user may not make use of the given user's devices equally. There also exist methods that utilize the browsing history, including consistency or frequency of access to similar websites across multiple devices by the users, to determine user identity.

[006] In view of the preceding, there is a problem that contemporary techniques for identifying users, or mapping users onto multiple devices in a given data communication network, are not sufficiently developed and versatile. Moreover, there arises a further problem with accuracy of such aforementioned contemporary techniques.

[007] Various approaches to associate user identities (IDs) with user activities in a data communication network, for example the Internet®, has been described in the following patent documents:

- (i) US2012/041939A1 (Amsterdamski);
- (ii) US2015/235258A1 (Shah);
- (iii) US2014/033317A1 (Barber); and
- (iv) US011/191179A1 (Linetsky).

SUMMARY

[0008] The present disclosure seeks to provide improved methods, systems, and computer program products involving data mining.

[0009] Moreover, the present disclosure further seeks to provide improved systems and methods that utilize improved techniques for probabilistic true mapping of unique users to IP-based devices.

[0010] The present disclosure seeks to provide an improved device graph clustering process for disambiguating differentiated algorithms into a normalized process for visualization and data mining purposes.

[0011] In an aspect of the present disclosure, a method of controlling data flows within a data communication system is disclosed.

[0012] In another aspect of the present disclosure, a data communication system for controlling data flows is disclosed.

[0013] In further aspect of the present disclosure, a non-transitory tangible computer readable medium comprising instructions for controlling data flows is disclosed.

[0014] In another aspect of the present disclosure, a method of identifying relationships between a number of users and a number of devices connected to a network is disclosed.

[0015] In yet another embodiment, a system for identifying relationships between a plurality of users and a plurality of devices connected to a network is disclosed.

[0016] According to an aspect of the present disclosure, there is provided a method of controlling data flows within a data communication system, wherein the data communication system includes a data communication network to which a plurality of users are coupled when the data communication system is in operation, wherein the data communication system includes at least one data server for use in implementing the method, wherein the plurality of users have associated therewith corresponding user identifiers (IDs), and wherein the plurality of users interact with the data communication system by using a plurality of user devices, wherein the method comprises steps of:

- (i) receiving the plurality of user identifiers (IDs);
- (ii) determining a plurality of matched pairs (P) of the user identifiers (IDs);
- (iii) comparing the determined plurality of the matched pairs (P) of the user identifiers (IDs) with each of the received plurality of user identifiers (IDs), and determining a similarity score for the matched pairs and the user identifiers (IDs);
- (iv) mapping the plurality of user identifiers (IDs) onto at least one single customer view identifier (SCV_ID), wherein each single customer view identifier (SCV_ID) is distinctly mapped, namely mapped to within a mapping threshold, to at least one user identifier (ID) of the plurality of user identifiers (IDs), to generate a mapping function (F); and
- (v) using the mapping function for selectively controlling distribution of data content to the plurality of users.

[0017] According to another aspect of the present disclosure, there is provided a data communication system for controlling data flows, wherein the data communication system includes a data communication network to which a plurality of users are coupled when the data communication system is in operation, wherein the data communication system includes at least one data server for use in implementing the method, wherein the plurality of users have associated therewith corresponding user identifiers, and wherein the plurality of users interact with the data communication system by using a plurality of user devices, wherein the data communication system is operable to:

- (i) receive the plurality of user identifiers (IDs);
- (ii) determine a plurality of matched pairs of the user identifiers (IDs);
- (iii) compare the determined plurality of the matched pairs of the user identifiers (IDs) with each of the received plurality of user identifiers , and determining a similarity score for the matched pairs and the user identifiers (IDs);
- (iv) map the plurality of user identifiers (IDs) onto at least one single customer view identifier, wherein each customer view identifier is distinctly mapped, namely mapped to within a mapping threshold, to at least one user identifier of the plurality of user identifiers, to generate a mapping function; and
- (v) use the mapping function for selectively controlling distribution of data content to the plurality of users.

[0018] According to another aspect of the present disclosure, there is provided a method of identifying relationships between a plurality of users and a plurality of devices connected to a network, the method comprising:

at a server:

receiving, by a transceiving module, a plurality of user identifiers (IDs) associated with the plurality of users present in the network;

determining, by a data retrieving module, a plurality of matched pairs of user identifiers;

comparing, by an identity mapping module, the determined plurality of matched pairs of user identifiers (IDs) with each of the received plurality of user identifiers (IDs) to create a first graph, wherein the first graph comprises one or more of the plurality of matched pairs of user identifiers (IDs) overlapped with the received plurality of user identifiers and a similarity score, wherein the similarity score is determined based on the comparison of the matched pairs of user identifiers (IDs) with the plurality of user identifiers;

mapping, by the identity mapping module, the plurality of user identifiers (IDs) of the first graph onto at least one single customer view

identifier, wherein each distinct single customer view identifier is mapped to at least one user identifier of the plurality of user identifiers (IDs);

creating, by the identity mapping module, a second graph based on the mapping of the plurality of user identifiers (IDs) of the first graph onto the at least one single customer view identifier, the second graph comprises a plurality of single customer view identifiers, wherein each single customer view identifier of the plurality of single customer view identifiers in the second graph represent a single user and one or more user identifiers mapped to the single customer view identifier belongs to same user of the plurality of users; and

identifying, by the identity mapping module, at least one user associated with one or more of the plurality of computing devices based on the plurality of single customer view identifiers in the second graph, wherein the at least one single customer view identifier comprises browsing event including at least one of browsing history, timestamp information, images, uniform resource locators, images, cookie information, and device identifier.

[0019] According to another aspect of the present disclosure, there is provided a system for identifying relationship between a plurality of users and a plurality of devices connected to a network, the system comprising:

a server comprising:

a transceiving module configured to receive a plurality of user identifiers (IDs) associated with the plurality of users present in the network;

a data retrieving module configured to determine a plurality of matched pairs of user identifiers; and

an identity mapping module configured to:

compare the determined the plurality of matched pairs of user identifiers with each of the received plurality of user identifiers to create a first graph, wherein the first graph comprises one or more of

the plurality of matched pairs of user identifiers overlapped with the received plurality of user identifiers and a similarity score, wherein the similarity score is determined based on the comparison of the matched pairs of user identifiers with the plurality of user identifiers;

map the plurality of user identifiers of the first graph onto at least one single customer view identifier, wherein each distinct, namely mapped to within a mapping threshold, single customer view identifier is mapped to at least one user identifier of the plurality of user identifiers;

create a second graph based on the mapping of the plurality of user identifiers of the first graph onto the at least one single customer view identifier, the second graph comprises a plurality of single customer view identifiers, wherein each single customer view identifier of the plurality of single customer view identifiers in the second graph represent a single user and one or more user identifiers mapped to the single customer view identifier belongs to same user of the plurality of users; and

identify, by the identity mapping module, at least one user associated with one or more of the plurality of computing devices based on the plurality of single customer view identifiers in the second graph, wherein the at least one single customer view identifier comprises browsing event including at least one of browsing history, timestamp information, images, uniform resource locators, images, cookie information, and device identifier.

According to another aspect of the present disclosure, there is provided a computer program product comprising non-transitory computer-readable storage media having computer-readable instructions stored thereon, the computer-readable instructions being executable by a computerized device comprising processing hardware to execute the aforesaid methods.

BRIEF DESCRIPTION OF THE FIGURES

[0020] The foregoing summary, as well as the following detailed description of preferred embodiments, is better understood when read in conjunction with the appended drawings. For the purposes of illustration, there are shown in the drawings exemplary embodiments; however, the present disclosure is not limited to the specific methods and instrumentalities disclosed. In the drawings:

[0021] FIGS. 1A-1C are illustrations of various exemplary data communication systems, in accordance with various embodiments of the present disclosure, wherein:

[0022] FIG. 2 is an illustration of a block diagram of a server (or data server), in accordance with an embodiment of the present disclosure;

[0023] FIG. 3 is an illustration of a block diagram of a computing device (or a user device), in accordance with an embodiment of the present disclosure;

[0024] FIG. 4 is a flowchart illustrating an exemplary method for creating non-overlapping groups of user identifiers representing a single user, in accordance with an embodiment of the present disclosure; and

[0025] FIG. 5 is a flowchart illustrating an exemplary method for selectively controlling distribution of data to a number of users in a data communication network, in accordance with an embodiment of the present disclosure.

[0026] In the accompanying diagrams, an underlined number is employed to represent an item over which the underlined number is positioned or an item to which the underlined number is adjacent. A non-underlined number relates to an item identified by a line linking the non-underlined number to the item. When a number is non-underlined and accompanied by an associated arrow, the non-underlined number is used to identify a general item at which the arrow is pointing.

[0027]

DETAILED DESCRIPTION

[0028] Description of embodiments of the present disclosure is not intended to limit the scope of claims of the present disclosure. Rather, the inventors have contemplated that the claimed subject matter might also be embodied in other ways, to include different steps or elements similar to the ones described in this document, in conjunction with other present or future technologies. Moreover, although the term “*step*” may be used herein to connote different aspects of methods employed, the term should not be interpreted as implying any particular order among or between various steps herein disclosed unless and except when the order of individual steps is explicitly described.

[0029] Functional units described in this disclosure have been labeled as systems or devices. A module, device, or a system may be implemented in programmable hardware devices such as, processors, digital signal processors, central processing units, field programmable gate arrays, programmable array logic, programmable logic devices, cloud processing systems, or the like. The devices/modules may also be implemented in software for execution by various types of processors. An identified device/module may include executable code and may, for example, comprise one or more physical or logical blocks of computer instructions, which may, for example, be organized as an object, procedure, function, or other construct. Nevertheless, the executable of an identified device / module need not be physically located together, but may comprise disparate instructions stored in different locations which, when joined logically together, comprise the device and achieve the stated purpose of the device.

[0030] An executable code of a device could be a single instruction, or many instructions, and may even be distributed over several different code segments, among different applications, and across several memory devices. Similarly, operational data may be identified and illustrated herein within the device, and may be embodied in any suitable form and organized within any suitable type of data structure. The operational data may be collected as a single data set, or may be

distributed over different locations including over different storage devices, and may exist, at least partially, as electronic signals on a system or network.

[0031] Reference throughout this specification to "*a select embodiment*", "one embodiment", or "an embodiment" means that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment of the disclosed subject matter. Thus, appearances of the phrases "a select embodiment", "in one embodiment", or "in an embodiment" in various places throughout this specification are not necessarily referring to the same embodiment.

[0032] Furthermore, the described features, structures, or characteristics may be combined in any suitable manner in one or more embodiments. In the following description, numerous specific details are provided, to provide a thorough understanding of embodiments of the disclosed subject matter. One skilled in the relevant art will recognize, however, that the disclosed subject matter can be practiced without one or more of the specific details, or with other methods, components, materials, etc. In other instances, well-known structures, materials, or operations are not shown or described in detail to avoid obscuring aspects of the disclosed subject matter.

[0033] In various embodiments of the present disclosure, definitions of one or more terms that will be used as provided below.

[0034] As used herein, a "computing device" or user device includes a single device or a combination of multiple devices, which may be capable of communicating, and exchanging one or messages with other devices present in a network.

[0035] As used herein, a "User Interface" or a "Graphical User Interface" (GUI) can include an interface on a display, such as a screen, of the computing device enabling a user to interact with the device or computing device.

[0036] Furthermore, as used herein, a "database" refers to a single or multiple modules or devices including hardware, software, firmware, or combination of these that can be configured to store user identifiers, matched pairs of the user

identifiers, device identifiers, browsing history of the users, browsing events including browsing history of the users on various devices, and so forth.

[0037] Furthermore, as used herein, an "Input / Output module" refers to a single or multiple modules or devices including hardware, software, firmware, or combination of these that can be configured to receive an input from a user or to present an output to the user.

[0038] Furthermore, as used herein, a "memory" refers to a single or multiple modules or devices including hardware, software, firmware, or combination of these that can be configured to store instructions that can be executed by other modules / devices.

[0039] As used herein, a "central processing unit" refers to a single or multiple modules or devices including a software, hardware, firmware or combination of these, that is configured to execute instructions stored in the memory.

[0040] As used herein, a "data communication system" refers to a network of devices and/or users. The data communication system may include a data communication network to which a number of users are coupled when the data communication system is in operation. The data communication network may include a data server. Each of the users has an associated user device to connect to the data communication network.

[0041] It will be appreciated that the terms "first", "second", and the like, herein do not denote any order, quantity, or importance, but rather are used to distinguish one element from another. Further, the terms "a" and "an" herein do not denote a limitation of quantity, but rather denote the presence of at least one of the referenced item.

[0042] As used herein, a "user identifier" refers to one or more elements of information that are capable of defining a user or a plurality of users (for example a group of users). In a simple example, a "user identifier" (ID) may include familiar subject matter such as, but not limited to, a login identifier (ID), a password, a phone number, an image identifier, a name, a date of birth of a user, and so forth.

However, the user identifier (ID) is alternatively a hybrid identifier including one or more elements of information that are subjected to a grouping or clustering transformation. Optionally, the transformation is determined by applying one or more pre-determined rules. Alternatively, optionally, the transformations are implemented in an adaptive manner, for example in a recursive manner, as a function of matching results generated by employing methods of user identification pursuant to the present disclosure. Certain categories of user identifier (ID) relate to user-defined information, for example a password, an access code, a user-selected bio-identification (e.g. choice of fingerprint of a user-selected finger); these are referred to as "*variant categories*" of ID. Other categories relate to historical information that cannot be altered by the user, for example date-of-birth, an original family name at birth; these are referred to as "*invariant categories*" of ID.

[0043] The concept of an User Identity (ID) on the Internet®, for purposes of describing embodiments of the present disclosure, fall into the following definitions:

1. IP to ID: this is the most general use of User Identity. For the vast majority of cases, an IP does not resolve to a single individual, but the collection of ID's serving one purpose of category;
2. HouseHold (HH): a Household is identified using a variety of techniques, including behavior, and event timing, there can thereby be created a picture of a House Hold; such categorization gives much more granularity in respect of a given customer;
3. Family - this grouping is a more of a meta-group, as opposed to a physical one. A household can be a collection of adults living together. A Family is meant to be a demographic identity linked to marketing and advertising needs. For purposes of describing embodiments of the present disclosure, such a definition of Family is beneficially employed. A group consisting of parents and children living together in a household could also be regarded as being a Family;
4. Person - this is an actual individual view. It is a single human being.

[0044] In FIGS. 1A-1C, there are illustrated data communication systems

100A-100C, in accordance with various embodiments of the present disclosure. As shown in FIG. 1A, the data communication system **100A** primarily includes a plurality of computing devices (or user devices) **104A-104N** associated with a plurality of users **102A-102N**. In describing embodiments of the present disclosure, the terms "user device" and "computing device" are used interchangeably on account of mutual similarities in their functionality and structure. Moreover, each of the users **102A-102N** may connect to a network **110** (also referred as a data communication network **110**) for accessing information using their associated computing devices **104A-104N**. In an embodiment, the users are coupled to the data communication network **110** when the data communication system is in operation. Each of the computing devices **104A-104N** also has an associated device identifier. An example of the device identifier may include an International Mobile Equipment Identity (IMEI).

[0045] In an embodiment, a given user of the plurality of users **102A-102N** uses more than one computing device for accessing the network **110** by using same or different user IDs. For example, the user **102A** can use the computing device **104A** and the computing device **104B** for accessing the network **110**. Similarly, the user **102B** can use computing device **104B** and **104N** for accessing the network **110**. Moreover, each of the users **102A-102N** may have more than one associated user IDs. As aforementioned, examples of the user identifiers may include such as, but not limited to, a login identifier (ID), a password, a phone number, an image identifier, a name, date of birth of a user, and so forth.

[0046] Examples of the computing devices **102A-102N** include cell phones, phablet computers, tablet computers, desktop computers, a personal digital assistant (PDA), and so forth. A typical example of the computing device is a wireless data access-enabled device, for example, an iPhone® smart phone, a BlackBerry® smart phone, a NEXUS ONE™ smart phone, an iPad® device, and so forth, that is capable of sending and receiving data in a wireless manner using protocols like the Internet Protocol, or IP, and the wireless application protocol, or WAP. Hereinafter, the terms network and data communication network are used interchangeably without change in their meaning due to similarity in their

functionality and structure.

[0047] Furthermore, the computing devices **102A-102N** may be connected to the network **110** using wired or wireless technologies. Examples of the network **110** may include, such as, but are not limited to, Local Area Networks (LANs), Wide Area Networks (WANs), Metropolitan Area Networks (MANs), Wireless LANs (WLANs), Wireless WANs (WWANs), Wireless MANs (WMANs), the Internet®, second generation (2G) telecommunication networks, third generation (3G®) telecommunication networks, fourth generation (4G®) telecommunication networks, and Worldwide Interoperability for Microwave Access (WiMAX®) networks. Furthermore, the terms network and data communication network are used interchangeably throughout this draft due to similarity in their structure and functionality.

[0048] The data communication system **100A** also includes a server **108** (or a data server **108**) connected to the network **110**. Furthermore, the terms server and data server are used interchangeably without change in their meaning due to similarity in their functionality and structure. The server **108** further includes an identity management system **106A**. The identity management system **106A** is operable to identify and establish relationship between the users **102A-102N** and the computing devices **104A-104N** connected to or present in the network **110**. In an embodiment, the identity management system **106A** is operable to identify and establish relationship between the users **102A-102N** and the computing devices **104A-104N** based on browsing events including such as, but not limiting to, browsing history, timestamp information, website information, uniform resource locators, device identifiers, images, and so forth.

[0049] The identity management system **106A** is operable, namely configured, to receive a number of user identifiers (IDs) associated with the plurality of users **102A-102N** present in the data communication network **110**. The identity management system **106A** is further configured to determine a number of matched pairs of user identifiers. In an embodiment, the matched pairs of user identifiers are pre-defined and pre-stored in a database of the server **108**. In an embodiment, the user identifiers are received from one of the computing

devices **104A-104N**. In an embodiment, the matched pairs of the user identifiers are pre-stored in a database using a Parquet format or a file system. The storing of the matched pairs in Parquet format may maintain integrity of the associations and meta-information in information, for example a first graph or a second graph, displayed to the users **102A-102N**. The Parquet format is a column oriented binary file format, which is considered to be very efficient for storing large scale data and executing queries.

[0050] Furthermore, the data communication system and primarily the identity management system **106A** are operable, namely are configured, to compare the determined matched pairs (P) of the user identifiers (IDs) with each of the received user identifiers (IDs). The identity management system **106A** is also configured to determine a similarity score for the matched pairs and the user identifiers. In an embodiment, the similarity score for the matched pairs and the user identifiers may have a numeric value for example, in a range of 0 to 1. Furthermore, the similarity score may indicate a probability of two or more user IDs being associated with one user based on a mapping of the matched pairs and the user identifiers. For example, the value of the similarity score is "1" if there is 100 percent probability and the value of the similarity score is "0" if probability is zero. In another embodiment, the similarity score value may be selected from the following: "High", "Low", "Very Low", and "Very High". An identity management system **106A** may further set a pre-defined minimum and maximum probability threshold. For example, the minimum probability threshold is 0.1, the maximum probability threshold is 0.9, then if the probability is below 0.1 then the similarity score is "Very Low", and when the probability is more than 0.9, then the similarity score while comparing the user IDs is "Very High". The similarity score may be determined by using suitable methods such as, but not limited to, methods based on correlation, neural network methods, multi-dimensional correlation, clustering methods, Kalman Filtering, regression based methods, association rule learning methods, dependency modeling, anomaly detection, outlier/change/deviation detection methods, summarization methods, classification methods, and so forth. In an embodiment, in of cluster analysis, a similarity score between two vertices x

and y is determined by using the following equation: $s(x, y) = -\|x - y\|_2^2$ wherein $\|x - y\|_2^2$ is the Euclidean distance.

[0051] In another embodiment, similarity score may be determined using cosine similarity for real-valued vectors, as used for retrieving information for scoring similarity between documents in a vector space model. Depending on the stage on the graph creation, the cosine method is used, and may be enhanced by using cluster analysis methods. This way, the similarity scores, and the associations and their respective "strengths" may be determined. The following exemplary cosine equation may be used to determine the similarity score:

$$D_C(A, B) = 1 - S_C(A, B)$$

Furthermore, the cosine equation can be normalized with cluster analysis methods to handle lack of triangulation in the cosine.

The identity management system **106A** is also operable, namely configured, to map the user identifiers onto at least one single customer view identifier (SCV_ID). A single customer view ID may be based on browsing event including browsing history, timestamp information, uniform resource locators, website information, device ID and so forth. The browsing event may include information monitored by or collected from various devices or software applications such as, but not limiting to, routers, gateways, computing devices **104A-104N**, data monitoring software, in the network **110**. Such collection of information is achieved by analyzing DNS (digital network signaling) routing information of data packets conveying a data payload, irrespective of whether or not the payload is encrypted. The SCV_ID may also indicate a linking of a user identity or user identifier to a device (or device identifier). Furthermore, the identity can be an internet protocol (IP) address or any other unique Data/networking element or device, that can be mapped to all the other devices and systems associated with the user, for example, user **102B**. Furthermore, multiple users may be mapped onto a single SCV_ID. The single SCV_ID may represent and indicate that the multiple users IDs that are being mapped onto the same SCV_ID belong to a same user. Furthermore, an SCV_ID represent a single user and one or more user identifiers mapped to the SCV_ID

belongs to same user of the users **102A-102N**. Each single customer view identifier is distinctly mapped, namely has a matching score that is greater than a pre-defined threshold, to at least one user identifier of the user identifiers, to generate a mapping function "F". In other words, each SCV ID is mapped, with a score greater than a pre-defined threshold, to at least one user identifier of the user identifiers, to generate a distinct mapping function "F", such that each user identifier can be mapped to at most one SCV ID. Optionally, the threshold is adaptively varied, for example when performing computations recursively to find a best mapping is user identification (IDs) to user activities. Optionally, the threshold is set, at least in part, depending upon a nature of activities against which the user identifications are being matched. Optionally, when performing computations, it is feasible to employ a one-time uniquely generated ID instead of distinct ID, for example a pseudo-constant ID, for example when implementing iterative recursive computations.

[0052] A distinct SCV_ID represents a single unique user and may be based on the browsing event, or a plurality of browsing events, across multiple devices **104A-104N**. The mapping function "F" may be determined using suitable analysis methods and/or computational methods based on correlation, neural network methods, multi-dimensional correlation, clustering methods, Kalman Filtering, regression based methods, association rule learning methods, dependency modeling, anomaly detection, outlier/change/deviation detection methods, summarization methods, classification methods, and so forth, as aforementioned.

[0053] In an exemplary scenario, the mapping function "F" may be determined using multidimensional correlation method and model to identify relationships between the user devices **104A-104N**, single customer view IDs and the user identifiers of the users **102A-102N**. The identity management system **106A** is also operable, namely configured, to use the mapping function for selectively controlling distribution of data content to the users **102A-102N**. In an example embodiment, the data content is presented on associated user devices **104A-104N** of the users **102A-102N**. Examples of the data content may include, such as, but not limited to, advertisements, surveys, entertainment data like movies, songs, news, and so forth. In an embodiment, the identity management

system **106A** is configured to repeat, for example iteratively, the process of comparing the determined matched pairs of the user identifiers with each of the received user identifiers and determining the similarity scores, and mapping the user identifiers onto at least one single customer view identifier for achieving a distinct mapping in the mapping function "F". The mapping function "F" may indicate relationship of the user IDs and various device identifiers of the user devices that are being used for accessing the data communication network by multiple users.

[0054] In an example embodiment, the identity management system **106A** is operable, namely configured, to compare the determined matched pairs of user identifiers with each of the received user identifiers to create a first graph. The first graph may include one or more of the matched pairs "P" of user identifiers overlapped with the received user identifiers and a similarity score. The similarity score may be determined based on the comparison of the matched pairs of user identifiers with the received user identifiers.

[0055] The identity management system **106A** is further configured to map the user identifiers of the first graph onto at least one single customer view identifier. Each distinct SCV_ID is mapped to at least one user identifier of the user identifiers. The identity management system **106A** is further operable, namely configured, to create a second graph based on the mapping of the user identifiers of the first graph onto the at least one single customer view identifier. The second graph may include a number of SCV_IDs. In some embodiments, one or more of the single customer view identifiers are in the received user identifiers. In alternative embodiment, one or more of the user identifiers mapping to a same single customer view identifier are absent in the second graph. In an embodiment, the second graph is empty when there is not enough information about the user identifiers, or the user identifiers are found not to have a mapping to another device. In another embodiment, the second graph is an empty graph when the identity management system **106A** determines that some of the user IDs are bots, or suspicious. This may avoid giving false positives in the second graph. In an alternative embodiment, the single graph includes at least one singleton match,

thereby eliminating empty graphs.

[0056] Furthermore, the identity management system **106A** is configured to present or display the second graph on a computing device such as the computing device **104B** of the user **102B**. For example, the second graph or an output graph may be presented on the computing device **104A** of the user **102A**. In an embodiment, an SCV_ID in the second graph represent a single user and one or more user identifiers mapped to the SCV_ID belongs to same user of the users **102A-102N**. In an embodiment, one or more of the user identifiers mapping to a same SCV_ID are absent in the second graph.

[0057] Furthermore, the identity management system **106A** is operable, namely configured, to identifying at least one user associated with one or more of the computing devices **104A-104N** based on the single customer view identifiers in the second graph. A single customer view identifier of the multiple single customer view identifiers in the second graph may represent a single user and one or more user identifiers mapped to the single customer view identifier belongs to a same user of the multiple users **102A-102N**.

[0058] As shown in FIG. 1B, an identity management system **106B** may be present in a cloud network **110** or on any network device in the network **110**.

[0059] Furthermore, in an embodiment, an identity management system **106C** may be present on a computing device itself. For example, the identity management system **106C** may be present on the computing device **104A** as shown in FIG. 1C. The functionality of the identity management systems **106B-106C** is similar to the identity management device **106A** as described in the foregoing with reference to FIG. 1A.

[0060] In FIG. 2, there is shown a block diagram **200** of a server **202**, in accordance with an embodiment of the present disclosure. As shown, the server **202** primarily includes an identity management system **204**, an Input / Output module **214**, a memory **216**, a central processing module or unit (CPU) **218**, and a network interfacing module **220**. The server **202** may be a single device or may include more than one devices including software, hardware, firmware, or

combination of these. The system or modules of the server **202** may be individual device or combination of devices capable of performing one or more functions.

[0061] The identity management system **204** is operable, namely configured to identify a relationship between a number of users (such as the users **102A-102N**) and a number of devices (such as the computing devices **104A-104N**) connected to the network **110** (or a data communication network **110**). The identity management system **204** includes a transceiving module **206**, a data retrieving module **208**, an identity mapping module **210**, and a database **212**. The transceiving module **206** is configured to receive a number of user identifiers (IDs) associated with a number of users (such as the users **102A-102N**) present in a network (such as the network **110**).

[0062] The data retrieving module **208** is configured to determine a number of matched pairs of user identifiers, for example to determine a plurality of such matched pairs. In an embodiment, the matched pairs of user identifiers are pre-defined and pre-stored in a database of the server **202** or of a remotely located device in the network **110**. In an embodiment, the data retrieving module **208** is operable, namely configured, to retrieve the matched pairs of user identifiers from the database **212**.

[0063] The identity mapping module **210** is operable, namely configured, to compare the determined matched pairs of user identifiers with each of the received user identifiers and determine a similarity score for the matched pairs and the user identifiers (IDs or UIIDs). The identity mapping module **210** is also operable, namely configured, to map the user identifiers onto at least one single customer view identifier (SCV_ID). Each single customer view identifier is distinctly mapped to at least one user identifier of the received user identifiers to generate a mapping function "F".

[0064] The identity mapping module **210** is operable, namely configured, to use the mapping function for selectively controlling distribution of data content to the users **102A-102N**. In an embodiment, the transceiving module **206** of the server **202** may send the mapping to another device or server in the network **110**. Furthermore, the another server may control distribution of data content to the

users in the network based on the received mapping function and SCV_IDs. For example, if a user John is attached to four computing devices like laptop, phones and so forth. then the server may decide to send content, such as a survey or an advertisement, to only one of the devices associated with John, for example to conserve data communication bandwidth, network energy consumption, to reduce data communication network response latency, and so forth.

[0065] In an embodiment, the identity mapping module **210** is operable, namely configured, to compare the determined matched pairs of user identifiers with each of the received user identifiers to create a first graph. The first graph may include one or more of the matched pairs of user identifiers overlapped with the received user identifiers and a similarity score. The similarity score may be determined based on the comparison of the matched pairs of user identifiers with the received user identifiers. In an embodiment, the user identifiers are received from one of the computing devices **104A-104N**.

[0066] The identity mapping module **210** is further operable, namely configured, to map the user identifiers of the first graph onto at least one single customer view identifier. Each distinct SCV_ID is mapped to at least one user identifier of the user identifiers. The identity mapping module **210** is further operable, namely configured, to create a second graph based on the mapping of the user identifiers of the first graph onto the at least one single customer view identifier (SCV_ID). The second graph may include a number of SCV_IDs. The single customer view ID may be based on browsing event including browsing history, timestamp information, images, uniform resource locators, images, cookie information, and device identifier monitored from devices such as routers, monitoring applications in the network 110. In an embodiment, the second graph includes the SCV_IDs such that one or more of the SCV_IDs are in the received user identifiers. In another embodiment, one or more of the user identifiers that are mapping to a same SCV_ID are not in the second graph. A single customer view identifier of the plurality of single customer view identifiers in the second graph represent a single user, and one or more user identifiers mapped to the single customer view identifier belong to a same given user of the users **102A-**

102N.

[0067] The identity mapping module **210** is also operable, namely configured, to identify at least one user associated with one or more of the computing devices **104A-104N** based on the single customer view identifiers in the second graph. For example, if there are four SCV_IDs "A", "B", "C", and "D" and there are eight user IDs i.e., "John", "Tim", "Ross", "Sara", "Rachel", "Joey", "Monica", and "Jack". Based on the mapping function "F" and in the second graph if "John", "Tim", "Sara" are mapped to the SCV_ID "A", this means that the user IDs "John", "Tim", "Sara" belongs to same user. This may help the server **202** to control distribution of content such as advertisements, surveys, movies, songs, etc. to the one or more of the users **102A-102N**. For example, the server **202** may send content to only one computing or user device associated with the three user IDs "John", "Tim", and "Sara". However, it will be appreciated from the foregoing that the ID's themselves may be hybrid user identifications derived from a plurality of elements by applying a transformation, wherein the transformation performs clustering or grouping of the elements, and the transformation itself can be made adaptive depending upon earlier matching results achieved, for example in a recursive manner, to obtain a best mapping of IDs to various user performed events (for example Internet® surfing activities). This maps back to the way we categorize individuals - from IP, to HH, to Family, to User. The ability to achieve finer and finer grained views allows us to make the most pivots off the data.

[0068] Moreover, the transceiving module **206** is operable, namely configured, to send to or present the second graph on a computing device of a user. For example, the second graph or an output graph may be presented at the computing device **104A** of the user **102A**. In an embodiment, an SCV_ID of the plurality of SCV_IDs in the second graph represent a single user. Furthermore, one or more user identifiers mapped to the SCV_ID belongs to same user of the plurality of users **102A-102N**. Furthermore, as used herein, a "memory" refers to a single or multiple modules or devices including hardware, software, firmware, or combination of these that can be configured to store instructions that can be executed by other modules/devices.

[0069] The central processing module **218** may include a single or multiple modules or devices including a software, hardware, firmware or combination of these, that is configured to execute instructions stored in the memory **216**. The network interfacing module **220** may enable the server **202** to establish connection with the network **110** or/and with other network devices such as the computing devices **104A-104N** present in the network **110**. The network **110** may include Internet®.

[0070] In FIG. 3, there is provided an illustration of a block diagram **300** that represents a computing device **302**, in accordance with an embodiment of the present disclosure. As shown, the computing device **302** primarily includes an identity management system **304**, an input/output module **314**, a memory **316**, a central processing module or unit (CPU) **318**, and a network interfacing module **320**. Furthermore, the computing device **302** may include software, hardware, firmware, or combination of these. The system or modules of the computing device **302** may be individual device or combination of devices capable of performing one or more functions as described below.

[0071] The identity management system **304** is operable, namely configured, to identify a relationship between a number of users (such as the users **102A-102N**) and a number of devices (such as the computing devices **104A-104N**) connected to the network **110**. The identity management system **304** includes a transceiving module **306**, a data retrieving module **308**, an identity mapping module **310**, and a database **312**. The transceiving module **306** is operable, namely configured, to receive a number of user identifiers (IDs) associated with a number of users (such as the users **102A-102N**) present in a network (such as the network **110**). The data retrieving module **308** is operable, namely configured, to determine a plurality of matched pairs of user identifiers. In an embodiment, the matched pairs of user identifiers are pre-defined and pre-stored in a database of the computing device **302** or of a remotely located device in the network **110**. In an embodiment, the data retrieving module **308** is configured to retrieve the matched pairs of user identifiers from the database **312**.

[0072] The identity mapping module **310** is operable, namely configured, to

compare the determined plurality of matched pairs of user identifiers with each of the received plurality of user identifiers to create a first graph. The first graph may include one or more of the plurality of matched pairs of user identifiers overlapped with the received user identifiers and a similarity score. The similarity score may be determined based on the comparison of the matched pairs of user identifiers with the received user identifiers. In an embodiment, the user identifiers are received from one of the computing devices **104A-104N**.

[0073] The identity mapping module **310** is further operable, namely configured, to map the user identifiers of the first graph onto at least one single customer view identifier (SCV_ID). Each distinct SCV_ID is mapped to at least one user identifier of the user identifiers. The identity mapping module **210** is further operable, namely configured, to create a second graph based on the mapping of the plurality of user identifiers of the first graph onto the at least one single customer view identifier (SCV_ID). The second graph may include a number of SCV_IDs. In an embodiment, the second graph includes the SCV_IDs such that one or more of the SCV_IDs are in the received user identifiers. In another embodiment, one or more of the user identifiers mapping to a same SCV_ID are not in the second graph.

[0074] The identity mapping module **310** is also operable, namely configured, to identify at least one user associated with one or more of the computing devices **104A-104N** based on the single customer view identifiers in the second graph. As aforementioned, the identity mapping module **310** is also operable to apply one or more threshold to determine whether or not there is an association between the at least one user with one or more of the computing devices **104A-104N**; whether or not the threshold is satisfied determines whether or not there is a distinct association.

[0075] The functionality and structure of the identity mapping module **310** is similar to the identity mapping module **210** of FIG. 2.

[0076] Furthermore, the Input / Output module **314** is operable configured to present or display the second graph on the computing device **302** of a user. For example, the second graph or an output graph may be presented at the computing

device **104A** of the user **102A**. In an embodiment, an SCV_ID of the plurality of SCV_IDs in the second graph represent a single user. Furthermore, one or more user identifiers mapped to the SCV_ID belongs to same user of the users **102A-102N**. Furthermore, as used herein, a “memory” refers to a single or multiple modules or devices including hardware, software, firmware, or combination of these that can be configured to store instructions that can be executed by other modules/devices.

[0077] The central processing module **318** may include a single or multiple modules or devices including a software, hardware, firmware or combination of these, that is configured to execute instructions stored in the memory **316**. The network interfacing module **320** may enable the computing device **302** to establish connection with the network **110** or/and with other network devices such as the computing devices **104A-104N** present in the network **110**.

[0078] In FIG. 4, there is shown a flowchart illustrating an exemplary method **400** of creating non-overlapping groups of user identifiers representing a single user, in accordance with an embodiment of the present disclosure. The method **400** may be implemented or performed by an identity management system such as, the identity management systems **106A-106C**, **204**, and **304** as described with reference to FIGs. 1A-1C, 2 and 3. The identity management system may be a single device or may be a combination of multiple devices including software, hardware, firmware, or combination of these. The identity management system may further include following modules or devices i.e., a transceiving module, a database, a data retrieving module, and an identity mapping module. Each of these modules can be a device or a combination of multiple devices. The identity management system may be present or implemented on a computing device or a server or on any network device in the network.

[0079] At a step **402**, a plurality of user identifiers are received. The user identifiers may be associated with a plurality of users. In an embodiment, the transceiving module receives the user identifiers. Then at a step **404**, a number of matched pairs of user identifiers are determined. In an embodiment, the matched pairs of user identifiers are determined from a database such as a database of a

server.

[0080] At a step **406**, the matched pairs of the user identifiers are compared with received user identifiers to create a first graph (or a relevant graph). In an embodiment, the identity mapping module compares the matched pairs of the user identifiers with the received user identifiers to create the first graph and determine similarity score. At a step **408**, map the user identifiers of the first graph onto at least one single customer view identifier (SCV_ID). In an embodiment, the identity mapping module maps the user identifiers of the first graph onto at least one SCV_ID. At a step **410**, create a second graph based on the mapping done at the step **408**. In an embodiment, the identity mapping module creates a second graph based on the mapping. At a step **412**, at least one user associated with one or more computing devices is identified based on the SCV_IDs present in the second graph. In an embodiment, the identity mapping module of the identity management system identifies the at least one user associated with the one or more computing devices based on the SCV_IDs present in the second graph.

[0081] In FIG. 5, there is shown a flowchart illustrating an exemplary method **500** of creating non-overlapping groups of user identifiers representing a single user, in accordance with an embodiment of the present disclosure. In an embodiment, the user identifiers are associated with one or more users, for example the users **102A-102N** as described in the foregoing with reference to FIG. 1A, who are either accessing or have accessed a data communication system. The data communication system includes a data communication network (or the network **110**). The users may be coupled or connected to the data communication network when the data communication system is in operation. Furthermore, the data communication system may include at least one data server, such as the server **108**, for use in implementing the method **500**. The users have associated therewith corresponding user identifiers (IDs or UIIDs). The users can interact with the data communication system by using a number of associated user devices, for example computing devices **104A-104N**. The method **500** may be implemented or performed by an identity management system such as, the identity management systems **106A-106C**, **204**, and **304** as discussed with reference to FIGs. 1A-1C, 2

and 3. The identity management system may be a single device or may be a combination of multiple devices including software, hardware, firmware, or combination of these. The identity management system may further include following modules or devices i.e., a transceiving module, a database, a data retrieving module, and an identity mapping module. Each of these modules can be a device or a combination of multiple devices. The identity management system may be present or implemented on a user device or the data server or on any network device in the data communication network.

[0082] At a step **502**, a number of user identifiers are received. In an embodiment, the transceiving module receives the user identifiers. In an embodiment, the user IDs are received from a computing device/user device. In an alternative embodiment, the user IDs are retrieved from a database of the data server. Then at a step **504**, a number of matched pairs of user identifiers are determined. In an embodiment, the matched pairs of user IDs are determined from a database such as the database of the data server.

[0083] At a step **506**, the matched pairs of the user IDs are compared with each of the received user IDs to determine a similarity score for the matched pairs and the user IDs. In an embodiment, the identity mapping module compares the matched pairs of the user identifiers with each of the received user identifiers to determine one or more similarity scores, for example plurality of similarity scores.

[0084] At a step **508**, the user IDs are mapped onto at least one single customer view identifier (SCV_ID). A single customer view ID may be based on browsing event including browsing history, timestamp information, uniform resource locators, website information, device ID, images, and such like in a device. Furthermore, multiple users may be mapped onto a single SCV_ID. The single SCV_ID may represent and indicate that the multiple users IDs that are being mapped onto the same SCV_ID belong to a same user. Furthermore, an SCV_ID represent a single user and one or more user identifiers mapped to the SCV_ID belongs to same user of the users.

[0085] At a step **510**, a mapping function "F" is generated. Furthermore, each customer view identifier is distinctly mapped to at least one user identifier of

the plurality of user identifiers to generate a mapping function "F". Furthermore, in an embodiment, the identity management system maps each single customer view identifier to at least one user identifier of the plurality of user identifiers, to generate a mapping function "F". A distinct SCV_ID may represent a single unique user and may be based on the browsing event across multiple devices. Multiple user IDs may be mapped to a single distinct SCV_ID. The mapping function "F" may be determined using suitable analysis methods and/or computational methods based on correlation, neural network based methods, multi-dimensional correlation, clustering methods, and so forth. In an exemplary scenario, the mapping function "F" may be determined using one or more multidimensional correlation methods and models to identify relationships between the user devices, single customer view IDs and the user identifiers of the users.

[0086] Furthermore, the steps **508** and **510** may be iteratively performed for achieving a distinct mapping in the mapping function "F". The mapping function "F" may indicate relationship of the user IDs and various device identifiers of the user devices that are being used for accessing the data communication network by multiple users. For example, Kalman Filtering, also known as linear quadratic estimation (LQE), is an algorithm that uses a series of measurements observed over time, containing statistical noise and other inaccuracies, and produces estimates of unknown variables that tend to be more precise than those based on a single measurement alone. The Kalman filter is named after Rudolf Kalman, one of the primary developers of its theory.

[0087] An Expectation-maximization algorithm is an iterative method to find a maximum likelihood of estimates of parameters in statistical models using unobserved or hidden variables. Moreover, Expectation-maximization has numerous applications in technology, including but not limited to, data clustering in machine learning, segmentation and object recognition in computer vision, natural language processing. Furthermore, Expectation-maximization is capable of estimating the best values for parameters based on observed data via unobserved data. Optionally, it can be employed in an iterative manner, until the parameter values converge for obtaining most stable results for mappings determining

relationship and associations for the user identifications (ID's).

[0088] The expectation-maximization algorithm operates on a statistical model with a set of observed data (for example, the scores between IDs), a set of assumed unobserved data and a vector of parameters (for example, the user-identifier to SCV ID mappings) with a likelihood function. The algorithm works in a two-step process. In the expectation step, the algorithm computes the expected value of the log likelihood function with respect to the conditional distribution of unobserved data given the observed data under the current estimate of the parameters. In the maximization step, it finds the parameters that maximizes the log likelihood. The two steps iterates until convergence.

[0089] The mapping function "F" may be determined using suitable analysis methods and/or computational methods based on correlation, neural network methods, multi-dimensional correlation, clustering methods, Kalman Filtering, regression based methods, association rule learning methods, dependency modeling, anomaly detection, outlier/change/deviation detection methods, summarization methods, classification methods, and so forth, as aforementioned. As an alternative or addition to employing Kalman Filtering, Expectation Maximization techniques can be employed in embodiments of the present disclosure. In Expectation Maximization, in a given iteration, a postulate that a given mapping of user ID's to browsing events is tested against data; in an event that the postulate is not found to be satisfied, further recursive iterations are performed with other postulates to be tested. Optionally, rules are employed to define postulates to be tested recursively.

[0090] Kalman filters have numerous applications in technology. A common application for Kalman filters is for guidance, navigation and control of vehicles, particularly aircraft and spacecraft. Furthermore, Kalman filters are a widely applied concept in time series analysis used in fields such as signal processing and econometrics. Moreover, Kalman filters are also employed for robotic motion planning and control, and they are sometimes included in trajectory optimization. Kalman filter are capable of providing a model needed for making estimates of the

current state of a given system and issuing updated commands. Optionally, the Kalman filters are employed in a recursive iterative manner, until parameters of the Kalman filters converge for obtaining most stable results for mappings determining relationships and associations for the user identifications (ID's). Similar consideration pertain for other types of filter than just Kalman-type filters.

[0091] The Kalman filter algorithm works in a two-step process. In a prediction step, the Kalman filter produces estimates of the current state variables, along with their uncertainties. Once the outcome of the next measurement (necessarily corrupted with some amount of error, including random noise) is observed, these estimates are updated using a weighted average, with more weight being given to estimates with higher certainty. The algorithm is recursive, for example as aforementioned. It can run in real time, using only the present input measurements and the previously calculated state and its uncertainty matrix; no additional past information is required.

[0092] At a step **512**, the mapping function "F" may be used for selectively controlling distribution of data, such as advertising content, surveys, movies, and so forth, to the users. For example, the server may send the advertising content to users **102B**, **102C**, and **102D** or to their associated devices **104B**, **104C**, and **104D**, respectively. In an embodiment, the mapping function and/or the second graph is sent to another device in a network such as another server device connected to the data communication network. Thereafter, the another device may control distribution of data to the users **102A-102N** in the network **110**. For example, another server device may select few of the users **102A-102N** to send a survey on their associated user devices.

[0093] In an exemplary scenario, a customer or a client user may want to create non-overlapping groups of user identifiers, which may represent a single user. The customer may want to create this to understand reach of his/her advertising campaign. As a customer of cortex single view, the customer should be able to count the number of individuals who have seen an advertisement. Therefore to identify a single user from a list of user identifiers (UIID) or received

user identifiers i.e.

|UIID| |A| |D| |C| |G| |X| |Y| |Z|

[0094] Then, based on the input user identifiers, a first graph or a relevant graph including matched pairs overlapped with input user identifiers is created. For example, "#IsSeed true" means a user identifier (or a UIID or ID) is in the received user identifiers set. Similarly, similarity scores for the matched pairs of user identifiers etc. are determined using the following: "| ID(IsSeed) | ID (IsSeed) | similarity_score |". For example, | A(true) | B(false) | High | , | B(false) | C(true) | High | , | A(true) | D(true) | High | , | A(true) | E(false) | High | , | D(true) | E(false) | High | , | F(false) | G(true) | Low | , | F(False) | H(False) | Low | , | F(false) | A(false) | Very Low |

[0095] Furthermore, the similarity score may have a numeric value like 0.1, 0.2, 0.5, and so forth based on a probability of the matched pairs and the user identifiers. Such numerical values are suitable for feeding into algorithms, for example the aforementioned Kalman filter, as parameter values.

[0096] Based on the first graph a second graph (or an output map or an SCV_Map) may be created or requested. An exemplary output map or second map format may look like "| SCV_ID | Set { UIID 1, UIID 2,... UIID N} |".

[0097] If a user identifier provided is in the matched pairs, then it should be in the output map or the second graph. In an embodiment, a user identifier (of the first graph) should only appear against one SCV_ID. Moreover, each distinct SCV_ID should be mapped to at least one user identifier from the received user identifiers. However, it will be appreciated that user identifiers with the same SCV_ID must be in a connected graph in the determined matched pairs of user identifiers. Moreover, however, it will be appreciated that user identifiers with the same SCV_ID do not necessarily have to be a connected graph in the output graph. In addition, singletons, an SCV_ID mapped to a single user identifier mapped to a single user identifier can be returned. Moreover, there may not be any user identifiers which are not in the matched pairs of user identifiers.

[0098] In another exemplary scenario, | Type | | cookie | |

device ids | | device ids | @NegativeTest Scenario: Provided UIIDs are not in device graph

[0099] In an exemplary scenario, when some data is received from the customer:

| UUID | | H | | I | | J | | K |

And the following matched pairs:

ID (isSeed)	ID (isSeed)	similarity_score
A (true)	B (false)	High
B (false)	C (true)	High
A (true)	D (true)	High
A (true)	E (false)	High
D (true)	E (false)	High
F (false)	G (true)	Low
F (false)	H (false)	Low
F (false)	A (false)	Very Low

When the disclosed method is executed, then following output is presented on the computing device: | SCV_ID | UIIDs | | empty | empty | #log file,

[00100] The disclosed method may be implemented when an input is a mixture of IDs or UIIDs, some of which are the matched pairs some are not.

[00101] The disclosed method may also be used to return a same single customer view if UUID is in the same cluster multiple times.

[00102] The present disclosure provides multi-staged, clustering methods to first derive data aggregate data from one or more data sources, and then via clustering, group those together into a unified output. Whether the output is for Single Consumer View, or to provide a means to provide aggregate views for demographics, time, geographies, and other critical data elements is now also possible, via a multistage approach to extract the core data (beyond ETL, this is to put into use for disclosed methods), getting one result set from the extraction, and using the extraction, or combination of extractions to provide a secondary result

that would not be possible without first examining and providing contextual understanding to the initial data result.

[00103] Various companies may use the disclosed methods and systems for frequency capping or planning their marketing campaigns.

[00104] While the disclosure has been presented with respect to certain specific embodiments, it will be appreciated that many modifications and changes may be made by those skilled in the art without departing from the spirit and scope of the disclosure. It is intended, therefore, by the appended claims to cover all such modifications and changes as fall within the true spirit and scope of the disclosure.

CLAIMS

What is claimed is:

1. A method of controlling data flows within a data communication system, wherein the data communication system includes a data communication network to which a plurality of users are coupled when the data communication system is in operation, wherein the data communication system includes at least one data server for use in implementing the method, wherein the plurality of users have associated therewith corresponding user identifiers (IDs), and wherein the plurality of users interact with the data communication system by using a plurality of user devices, wherein the method comprising steps of:

(i) receiving the plurality of user identifiers (IDs);

(ii) determining a plurality of matched pairs (P) of the user identifiers (IDs);

(iii) comparing the determined plurality of the matched pairs (P) of the user identifiers (IDs) with each of the received plurality of user identifiers (IDs), and determining a similarity score for the matched pairs and the user identifiers (IDs);

(iv) mapping the plurality of user identifiers (IDs) onto at least one single customer view identifier (SCV_ID), wherein each single customer view identifier (SCV_ID) is mapped to within a mapping threshold to at least one user identifier (ID) of the plurality of user identifiers (IDs), to generate a mapping function (F); and

(v) using the mapping function for selectively controlling distribution of data content to the plurality of users.

2. A method of claim 1, wherein the method includes generating the user identifies (IDs) in a recursive iterative manner until the mapping function of (iv) converges.

3. A method of claim 1, wherein the method includes adaptively varying the mapping threshold depending upon a nature of content associated with the at least

one single customer view identifier (SCV_ID).

4. A method of claim 1, wherein the steps (iii) and (iv) are performed iteratively for achieving a distinct mapping in the mapping function.

5. A method of claim 1, wherein comparing the determined plurality of matched pairs of user identifiers with each of the received plurality of user identifiers includes creating a first graph, wherein the first graph comprises one or more of the plurality of matched pairs of user identifiers overlapped with the received plurality of user identifiers and a similarity score, wherein the similarity score is determined based on the comparison of the matched pairs of user identifiers with the plurality of user identifiers.

6. A method of claim 1 or 5 further comprising:

mapping the plurality of user identifiers of the first graph onto at least one single customer view identifier, wherein each distinct single customer view identifier is mapped to at least one user identifier of the plurality of user identifiers; and

creating a second graph based on the mapping of the plurality of user identifiers of the first graph onto the at least one single customer view identifier, the second graph comprises a plurality of single customer view identifiers, wherein each single customer view identifier of the plurality of single customer view identifiers in the second graph represent a single user and one or more user identifiers mapped to the single customer view identifier belongs to same user of the plurality of users.

7. A method of claim 1, 5, or 6, wherein the method further comprises presenting the second graph on a computing device of a user, wherein a single customer view identifier of the plurality of single customer view identifiers in the

second graph represent a single user and one or more user identifiers mapped to the single customer view identifier belongs to same user of the plurality of users.

8. A method of claim 1, 5, 6, or 7, wherein, in the method, the second graph comprises a plurality of single customer view identifiers, wherein one or more of the plurality of single customer view identifiers are in the received plurality of user identifiers.

9. A method of claim 1, 5, 6, 7, or 8, wherein, in the method, one or more of the plurality of user identifiers mapping to a same single customer view identifier are absent in the second graph.

10. A method of claim 1, 5, 6, 7, 8, or 9, wherein, in the method, the plurality of matched pairs of user identifiers are pre-defined and pre-stored in a database of the server.

11. A method of claim 1, 5, 6, 7, 8, 9, or 10 wherein the method further comprises arranging for the data communication system to employ the Internet as a part of its data communication network.

12. A data communication system for controlling data flows, wherein the data communication system includes a data communication network to which a plurality of users are coupled when the data communication system is in operation, wherein the data communication system includes at least one data server for use in implementing the method, wherein the plurality of users have associated therewith corresponding user identifiers, and wherein the plurality of users interact with the data communication system by using a plurality of user devices, wherein the data communication system is operable to:

- (i) receive the plurality of user identifiers;

- (ii) determine a plurality of matched pairs of the user identifiers;
- (iii) compare the determined plurality of the matched pairs of the user identifiers with each of the received plurality of user identifiers, and determining a similarity score for the matched pairs and the user identifiers;
- (iv) map the plurality of user identifiers onto at least one single customer view identifier, wherein each customer view identifier is mapped, to within a mapping threshold, to at least one user identifier of the plurality of user identifiers, to generate a mapping function; and
- (v) use the mapping function for selectively controlling distribution of data content to the plurality of users.

13. A data communication system of claim 12, wherein the data communication system is operable to perform (iii) and (iv) iteratively for achieving a distinct mapping in the mapping function.

14. A data communication system of claim 12, wherein the data communication system, when comparing the determined plurality of matched pairs of user identifiers with each of the received plurality of user identifiers, is operable to create a first graph, wherein the first graph comprises one or more of the plurality of matched pairs of user identifiers overlapped with the received plurality of user identifiers and a similarity score, wherein the similarity score is determined based on the comparison of the matched pairs of user identifiers with the plurality of user identifiers.

15. A data communication system of claim 12 or 14, wherein the data communication system is operable to:

map the plurality of user identifiers of the first graph onto at least one single customer view identifier, wherein each distinct single customer view identifier is mapped to at least one user identifier of the plurality of user

identifiers; and

create a second graph based on the mapping of the plurality of user identifiers of the first graph onto the at least one single customer view identifier, the second graph comprises a plurality of single customer view identifiers, wherein each single customer view identifier of the plurality of single customer view identifiers in the second graph represent a single user and one or more user identifiers mapped to the single customer view identifier belongs to same user of the plurality of users.

16 A data communication system of claim 12, 14, or 15, wherein the data communication system is operable to present the second graph on a computing device of a user, wherein a single customer view identifier of the plurality of single customer view identifiers in the second graph represent a single user and one or more users mapped to the single customer view identifier belongs to same user of the plurality of users.

17. A data communication system of claim 12, 14, 15, or 16, wherein the second graph comprises a plurality of single customer view identifiers, wherein one or more of the plurality of single customer view identifiers are in the received plurality of user identifiers.

18. A data communication system of claim 12, 14, 15, 16, or 17, wherein one or more users of the plurality of user identifiers mapping to a same single customer view identifier are absent in the second graph.

19. A data communication system of any one of claims 12 to 18, wherein the data communication system is operable to present the second graph to the user on a computing device associated with the user.

20. A data communication system of any one of claims 12 to 19, wherein the plurality of matched pairs of user identifiers are pre-defined and pre-stored in a database of the server.

21. The data communication system of any one of claims 12 to 20, wherein the data communication system is operable to employ the Internet as a part of its data communication network.

22. A non-transitory tangible computer readable medium comprising instructions for the execution of the method of claim 1.

23. A computer program product comprising non-transitory computer-readable storage media having computer-readable instructions stored thereon, the computer-readable instructions being executable by a computerized device comprising processing hardware to execute the method of claim 1.

24. A method of identifying relationship between a plurality of users and a plurality of devices connected to a network, the method comprising:

at a server:

receiving, by a transceiving module, a plurality of user identifiers (IDs) associated with the plurality of users present in the network;

determining, by a data retrieving module, a plurality of matched pairs of user identifiers;

comparing, by an identity mapping module, the determined plurality of matched pairs of user identifiers with each of the received plurality of user identifiers to create a first graph, wherein the first graph comprises one or more of the plurality of matched pairs of user identifiers overlapped with the received plurality of user identifiers and a similarity score, wherein the similarity score is determined based on the comparison of the matched pairs of user identifiers with the plurality of user identifiers;

mapping, by the identity mapping module, the plurality of user identifiers of the first graph onto at least one single customer view identifier, wherein each single customer view identifier is mapped, to within a mapping threshold, to at least one user identifier of the plurality of user identifiers;

creating, by the identity mapping module, a second graph based on the mapping of the plurality of user identifiers of the first graph onto the at least one single customer view identifier, the second graph comprises a plurality of single customer view identifiers, wherein each single customer view identifier of the plurality of single customer view identifiers in the second graph represent a single user and one or more user identifiers mapped to the single customer view identifier belongs to same user of the plurality of users; and

identifying, by the identity mapping module, at least one user associated with one or more of the plurality of computing devices based on the plurality of single customer view identifiers in the second graph, wherein the at least one single customer view identifier comprises browsing event including at least one of browsing history, timestamp information, images, uniform resource locators, images, cookie information, and device identifier.

25. A method of claim 24, wherein the second graph comprises a plurality of single customer view identifiers, wherein one or more of the plurality of single customer view identifiers are in the received plurality of user identifiers.

26. A method of claim 24, wherein one or more of the plurality of user identifiers mapping to a same single customer view identifier are absent in the second graph.

27. A method of any one of claims 24 to 26, wherein the second graph is presented to the user on an associated computing device.

28. A method of any one of claims 24 to 27, wherein the plurality of matched pairs of user identifiers are pre-defined and pre-stored in a database of the server.

29. A system for identifying relationship between a plurality of users and a plurality of devices connected to a network, the system comprising:

a server comprising:

a transceiving module configured to receive a plurality of user identifiers (IDs) associated with the plurality of users present in the network;

a data retrieving module configured to determine a plurality of matched pairs of user identifiers; and

an identity mapping module configured to:

compare the determined the plurality of matched pairs of user identifiers with each of the received plurality of user identifiers to create a first graph, wherein the first graph comprises one or more of the plurality of matched pairs of user identifiers overlapped with the received plurality of user identifiers and a similarity score, wherein the similarity score is determined based on the comparison of the matched pairs of user identifiers with the plurality of user identifiers;

map the plurality of user identifiers of the first graph onto at least one single customer view identifier, wherein each single customer view identifier is mapped, to within a mapping threshold, to at least one user identifier of the plurality of user identifiers;

create a second graph based on the mapping of the plurality of user identifiers of the first graph onto the at least one single customer view identifier, the second graph comprises a plurality of single customer view identifiers, wherein each single customer view identifier of the plurality of single customer view identifiers in the second graph represent a single user and one or more user identifiers mapped to the single customer view identifier belongs to same user of the plurality of

users; and

identify, by the identity mapping module, at least one user associated with one or more of the plurality of computing devices based on the plurality of single customer view identifiers in the second graph, wherein the at least one single customer view identifier comprises browsing event including at least one of browsing history, timestamp information, images, uniform resource locators, images, cookie information, and device identifier.

1/7

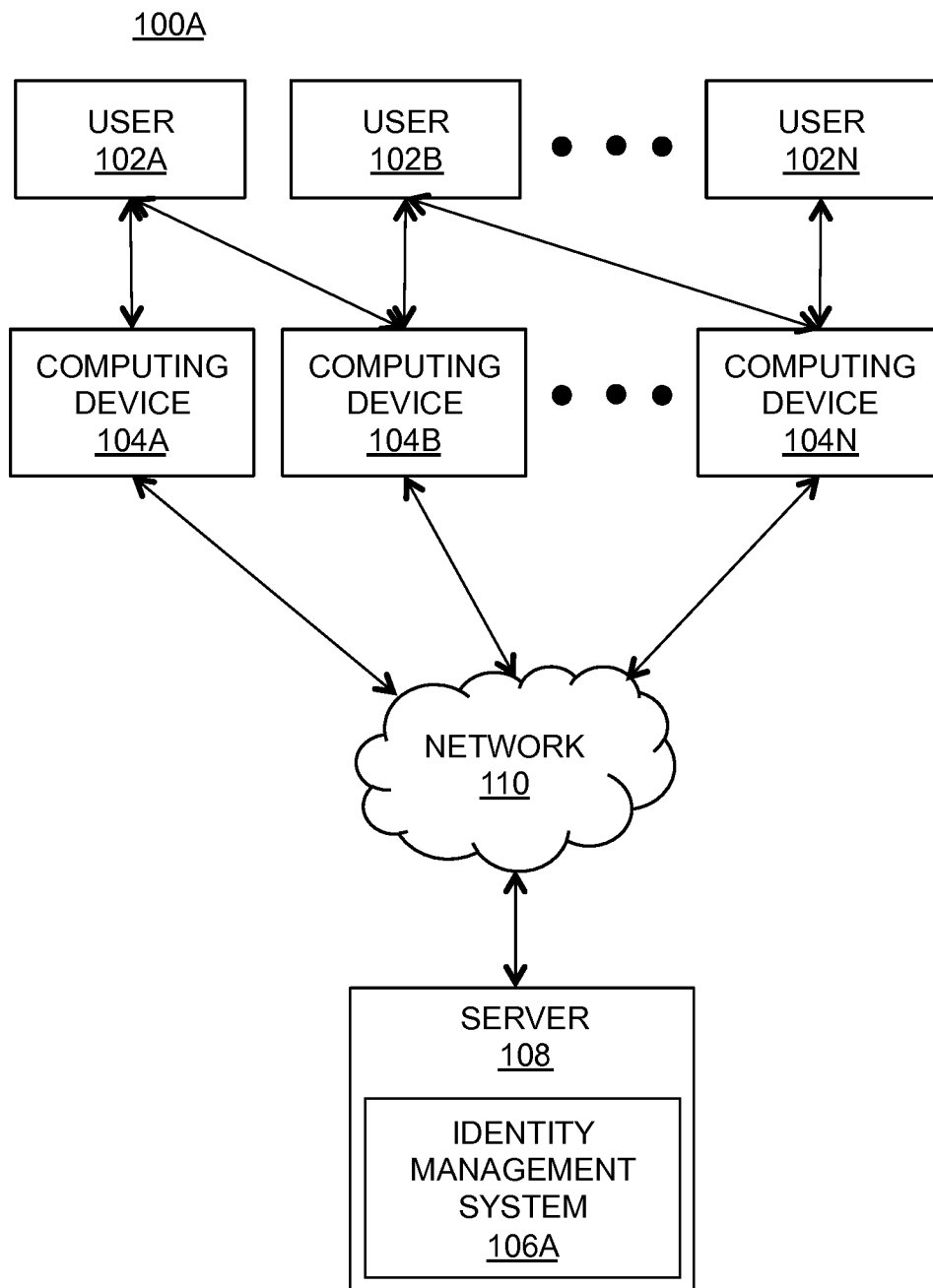


FIG. 1A

2/7

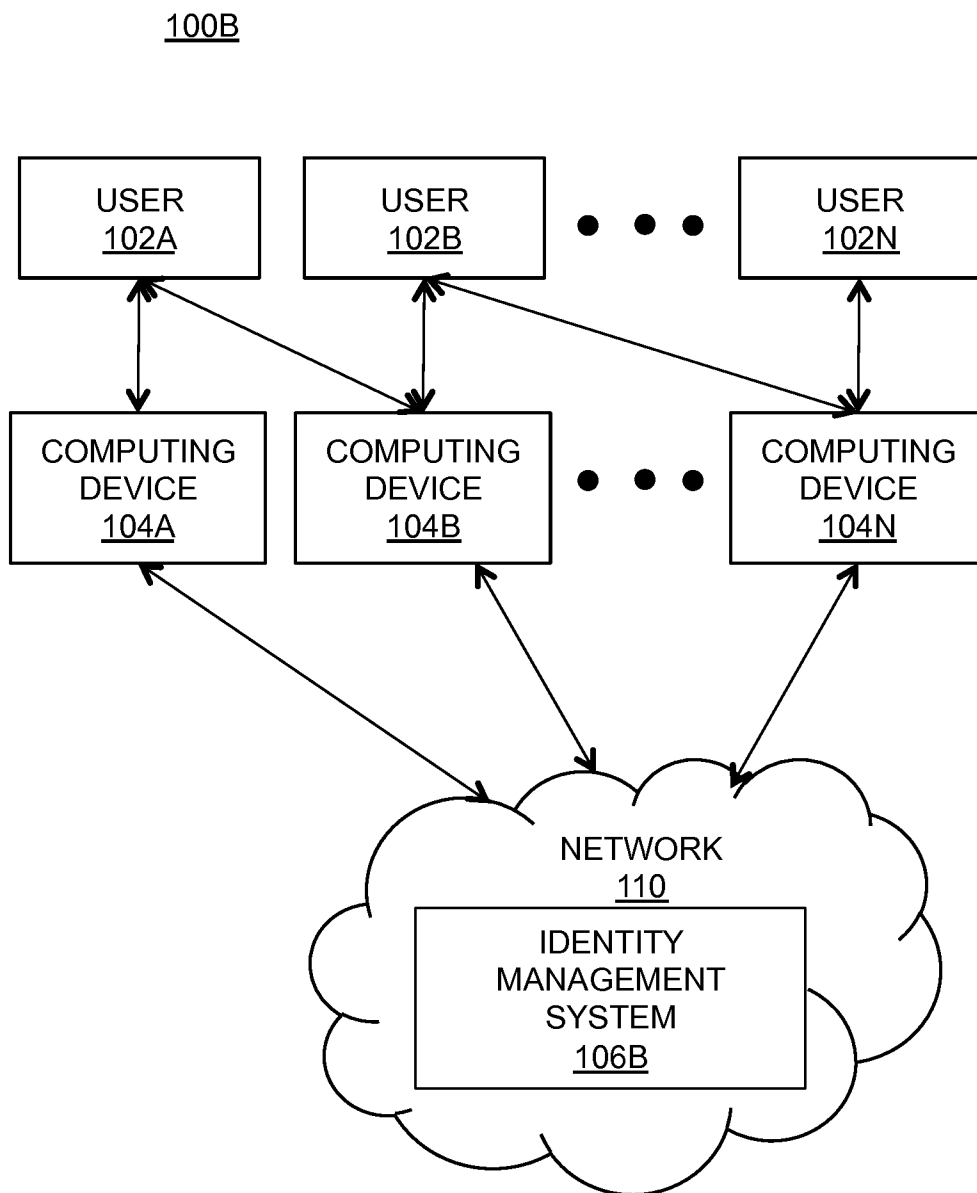


FIG. 1B

3/7

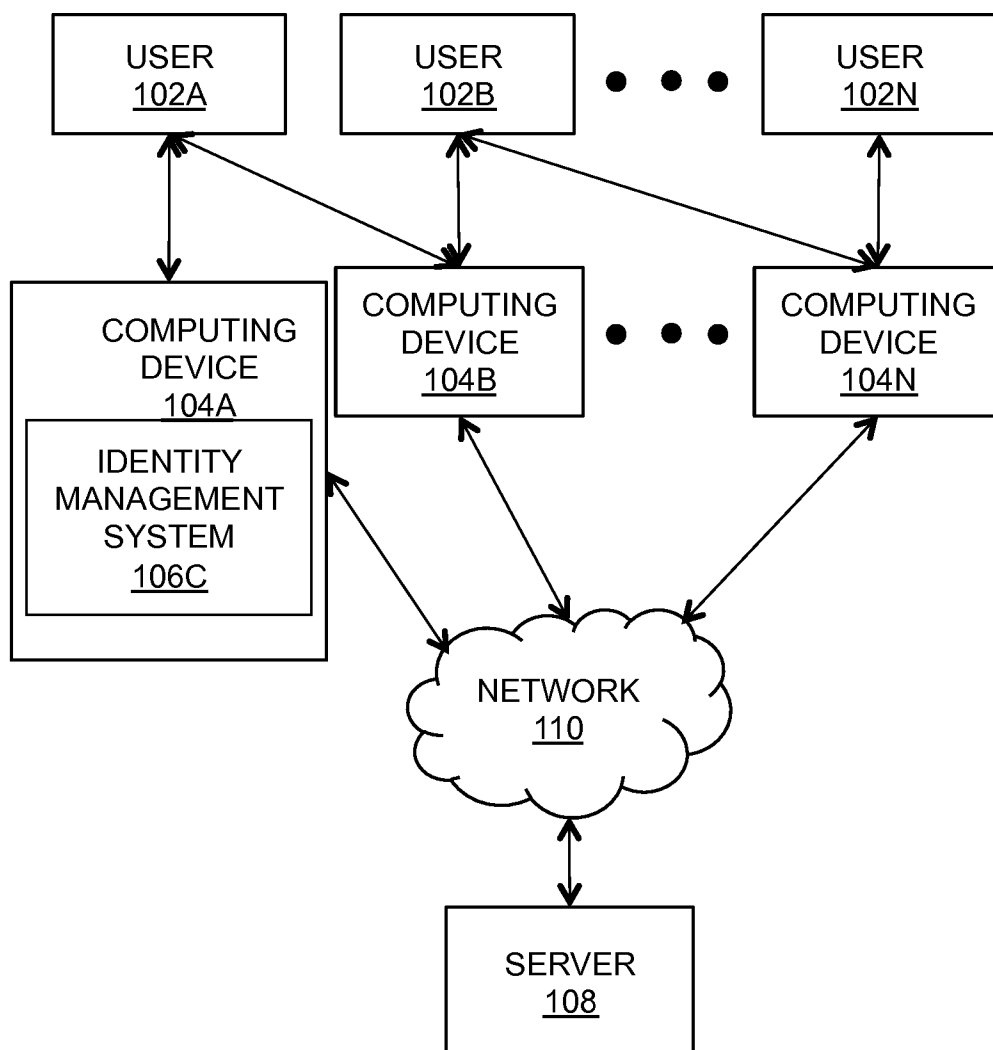
100C

FIG. 1C

4/7

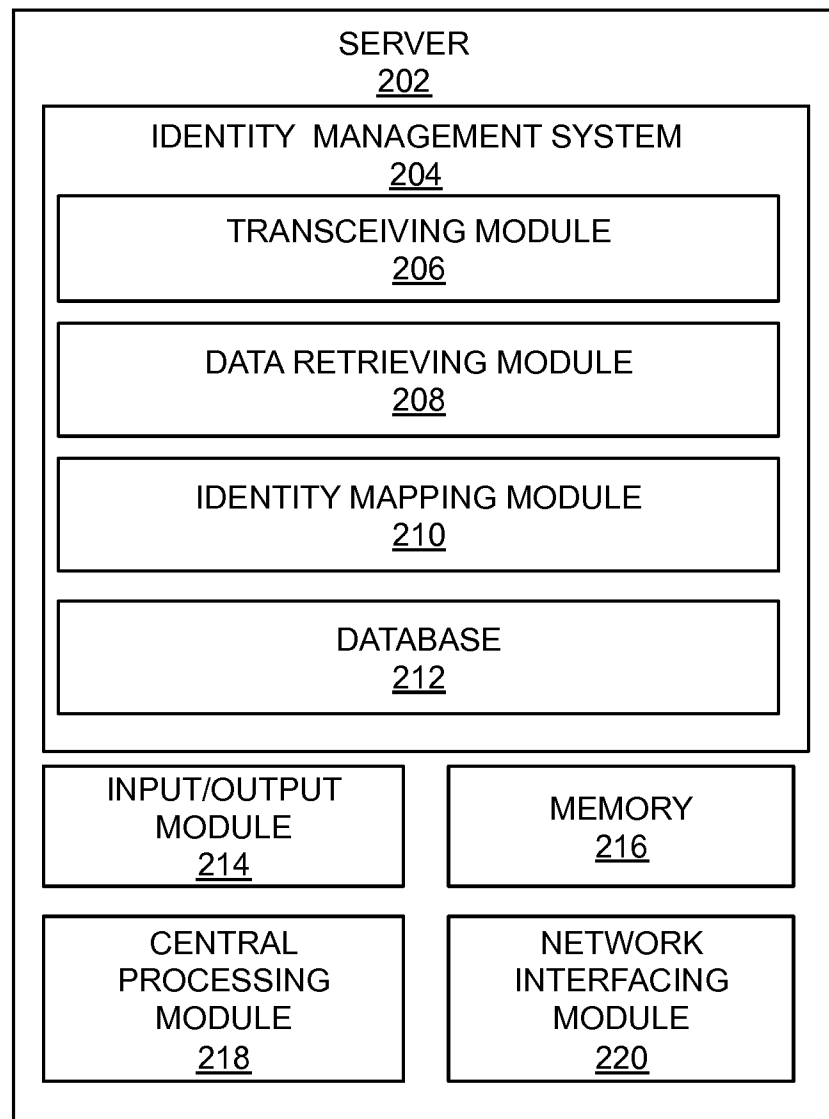
200

FIG. 2

5/7

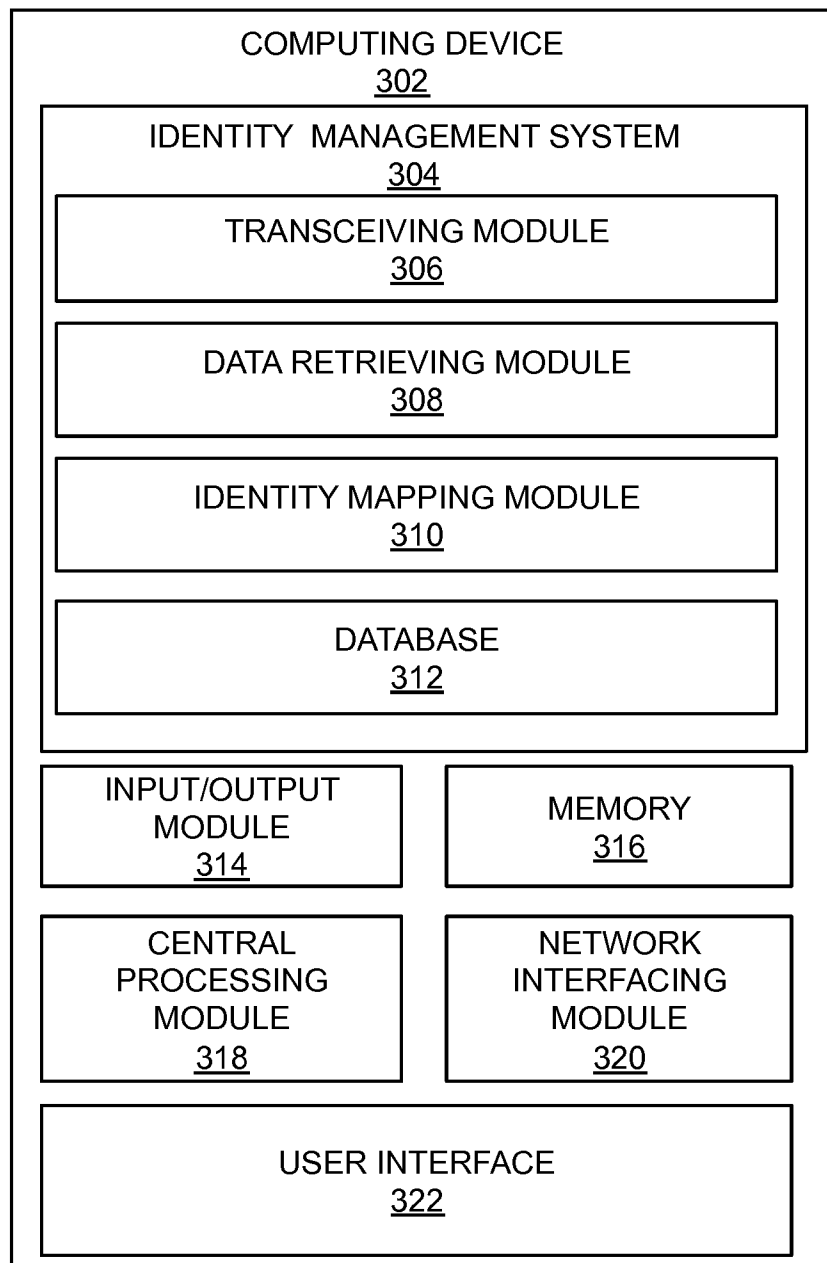
300

FIG. 3

6/7

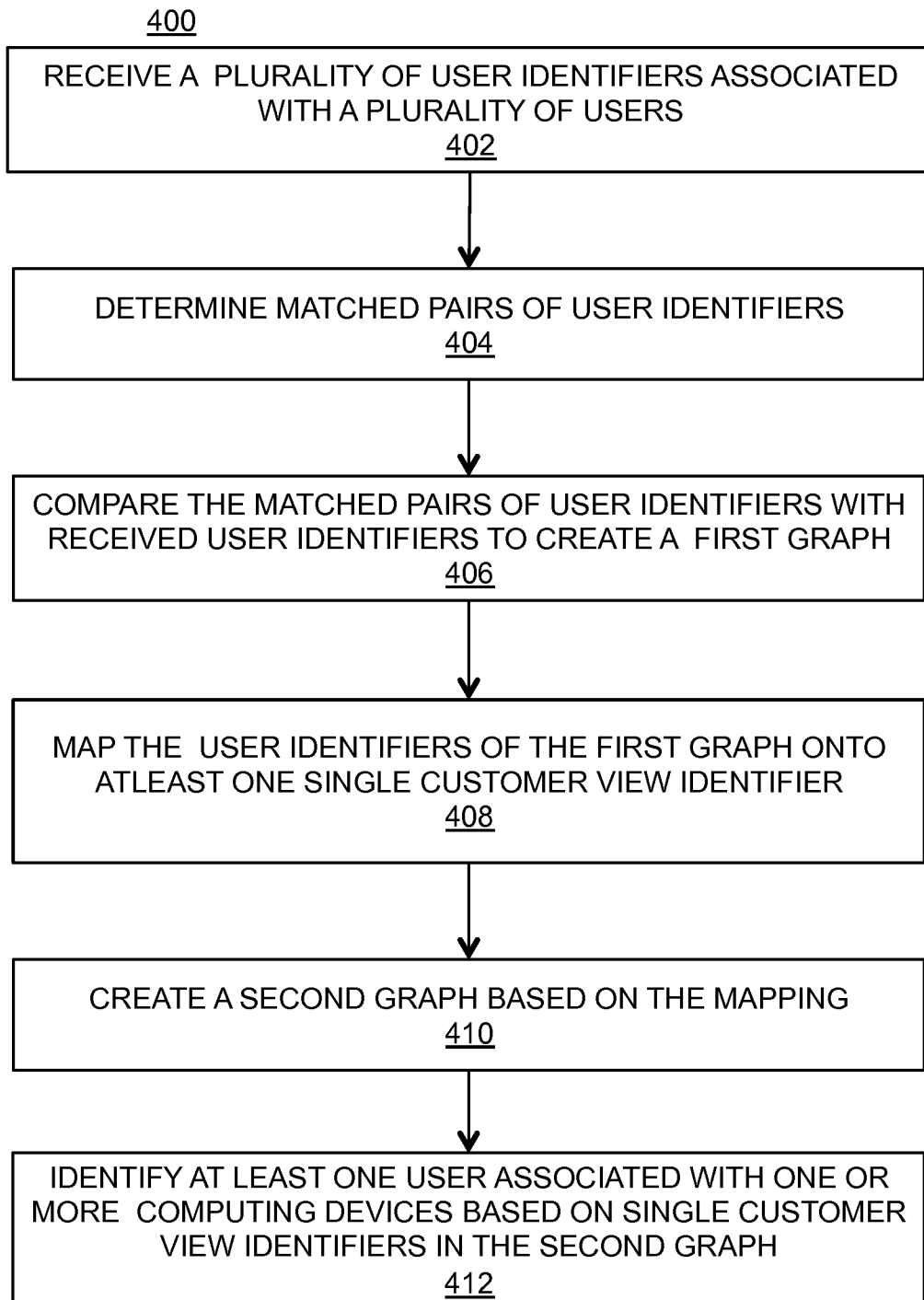


FIG. 4

7/7

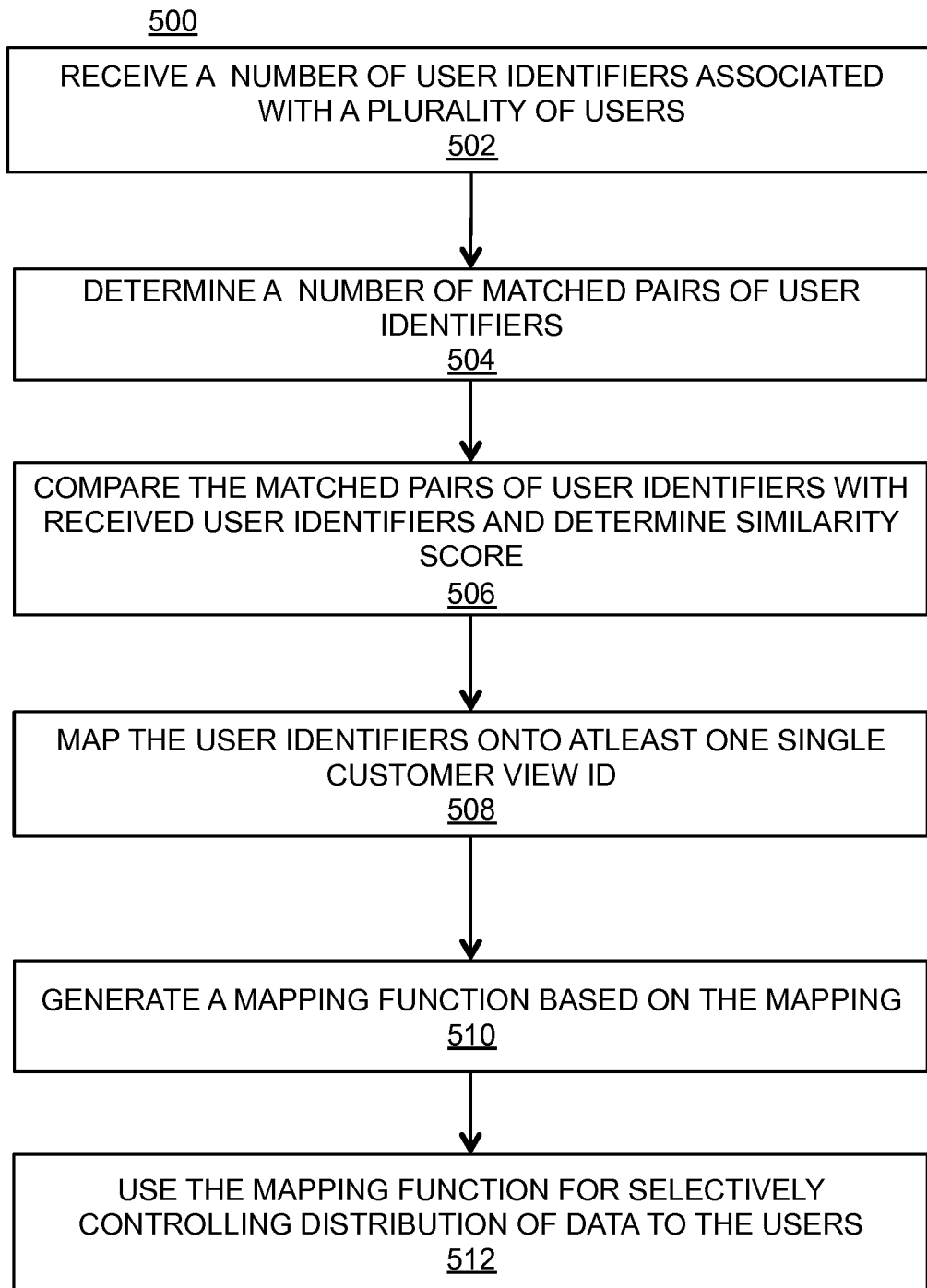


FIG. 5

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2017/025053

A. CLASSIFICATION OF SUBJECT MATTER

INV. H04L29/08 G06F17/30 G06Q30/02 H04W4/18
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)

H04L G06F G06Q H04W

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, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 8 438 184 B1 (WANG CHANGFENG C [US] ET AL) 7 May 2013 (2013-05-07)	1-5, 10-14, 20-23
A	column 1, line 20 - line 31 column 2, line 64 - column 3, line 54 column 5, line 64 - column 8, line 27 column 10, line 24 - column 12, line 13 column 13, line 10 - line 15	6-9, 15-19, 24-29
Y	US 2013/124309 A1 (TRAASDAHL ARE HELGE [US] ET AL) 16 May 2013 (2013-05-16) paragraphs [0005], [0007], [0021] - [0025], [0032] - [0034], [0060] - [0062]; figures 2,5,6,7 ----- -/-	1,5,12



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

7 June 2017

Date of mailing of the international search report

14/06/2017

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

Lefebvre, Laurent

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2017/025053

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2012/041939 A1 (AMSTERDAMSKI LIOR [IL]) 16 February 2012 (2012-02-16) paragraphs [0008], [0009], [0020], [0028], [0031], [0036] - [0038], [0042], [0044], [0048] -----	1,5,12

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2017/025053

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 8438184	B1	07-05-2013	CN 105190595 A 23-12-2015
			EP 2880553 A1 10-06-2015
			US 8438184 B1 07-05-2013
			WO 2014022327 A1 06-02-2014

US 2013124309	A1	16-05-2013	CA 2890627 A1 23-05-2013
			CN 104054055 A 17-09-2014
			EP 2780805 A1 24-09-2014
			HK 1202173 A1 18-09-2015
			JP 6125519 B2 10-05-2017
			JP 2015512066 A 23-04-2015
			US 2013124309 A1 16-05-2013
			WO 2013074750 A1 23-05-2013

US 2012041939	A1	16-02-2012	IL 207123 A 30-04-2015
			US 2012041939 A1 16-02-2012
