(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 August 2013 (01.08.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/112911 A1

(51) International Patent Classification:
G06Q 50/30 (2012.01) *G06F 19/00* (2011.01)

(21) International Application Number:
PCT/US2013/023259

(22) International Filing Date:
25 January 2013 (25.01.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/591,263 26 January 2012 (26.01.2012) US

(71) Applicant *(for all designated States except US)*: **THE NIELSEN COMPANY (US), LLC** [US/US]; 150 North Martingale Road, Schaumburg, Illinois 60173 (US).

(72) Inventors; and
(71) Applicants *(for US only)*: **OLIVER, James R.** [US/US]; 2320 Howard Ave, San Carlos, California 94070 (US). **STACKHOUSE, Herbert F.** [US/US]; 295 Oxford Street, San Francisco, California 94134 (US). **DOE, Peter C.** [US/US]; Ridgewood, New York (US). **TANG, Candy** [US/US]; Sunnyvale, California (US). **HA, Michelle** [US/US]; 2115 Marques Ave, San Jose, California 95125 (US).

(74) Agent: **ZIMMERMAN, Michael W.**; Hanley, Flight & Zimmerman, LLC, 150 S. Wacker Drive, Suite 2100, Chicago, Illinois 60606 (US).

(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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, 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, 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,

[Continued on next page]

(54) Title: SYSTEMS, METHODS, AND ARTICLES OF MANUFACTURE TO MEASURE ONLINE AUDIENCES

(57) Abstract: Methods and apparatus to monitor media content at a content display site are described. An example method includes obtaining panelist data corresponding to a plurality of panelists accessing web pages at measured locations, classifying the panelist data according to demographic information of the panelists, generating a virtual panel based on an estimate of web page access at unmeasured locations, and classifying page view data associated with the unmeasured locations based on data corresponding to the virtual panel.

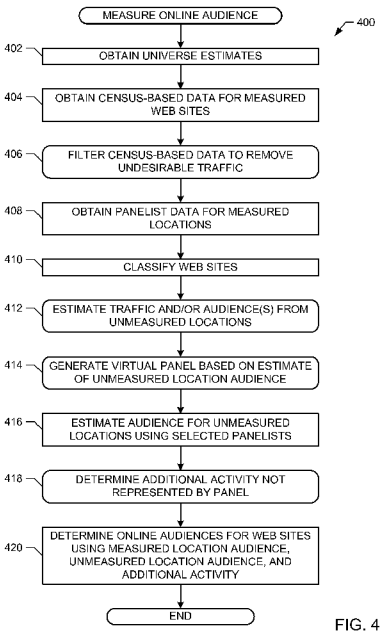


FIG. 4

WO 2013/112911 A1

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, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

SYSTEMS, METHODS, AND ARTICLES OF MANUFACTURE TO MEASURE ONLINE AUDIENCES

RELATED APPLICATIONS

[0001] This patent claims priority to U.S. Provisional Patent Application No. 61/591,263, filed January 26, 2012, the entirety of which is hereby incorporated by reference.

FIELD OF THE DISCLOSURE

[0002] The present disclosure relates generally to audience measurement and, more particularly, to systems, methods, and articles of manufacture to measure online audiences.

BACKGROUND

[0003] Online audience measurement based on panelist device metering and online audience measurement based on web site/web page tagging share the goal of measuring web traffic. In each case, the objective is to count or estimate the number of occasions when a person has the opportunity to see an element of online media (e.g., content, advertising, etc.). The objective may also include estimating a total unique audience exposed to a particular web site.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] FIG. 1 is a block diagram illustrating a disclosed example system constructed in accordance with the teachings of this disclosure to measure online audiences.

[0005] FIG. 2 is a block diagram of an example system that may be used to implement the hybrid online audience measurement system of FIG. 1.

[0006] FIG. 3 is a block diagram of an example virtual panel generator to implement the virtual panel generator of FIG. 2.

[0007] FIG. 4 is a flowchart representative of example computer readable instructions that may be executed to implement the hybrid online audience measurement system of FIGS. 1 and/or 2 to measure an online audience of a web site.

[0008] FIG. 5 is a flowchart representative of example computer readable instructions which, when executed, cause a processor to filter non-human traffic from a set of traffic data.

[0009] FIG. 6 is a flowchart representative of example computer readable instructions which, when executed, cause a processor to estimate access to a web site from unmeasured locations.

[0010] FIG. 7 is a flowchart representative of example computer readable instructions which, when executed, cause a processor to calculate negative binomial distribution parameters.

[0011] FIG. 8 is a flowchart representative of example computer readable instructions which, when executed, cause a processor to calculate a negative binomial distribution.

[0012] FIG. 9 is a flowchart representative of example computer readable instructions which, when executed, cause a processor to select panelists to represent an unmeasured location audience.

[0013] FIGS. 10A and 10B collectively comprise a flowchart representative of example computer readable instructions which, when executed, cause a processor to estimate an audience for unmeasured locations using selected panelists.

[0014] FIGS. 11A and 11B collectively comprise a flowchart representative of example computer readable instructions which, when executed, cause a processor to estimate an audience for unmeasured locations using selected panelists.

[0015] FIG. 12 is a flowchart representative of example computer readable instructions which, when executed, cause a processor to smooth a volume metric.

[0016] FIGS. 13A and 13B collectively comprise a flowchart representative of example computer readable instructions which, when executed, cause a processor to calculate an online audience.

[0017] FIG. 14 illustrates an example panel stream distribution percentage among demographic groups for a day part of a day of a week.

[0018] FIG. 15 illustrates a distribution of a number of streams for the demographic groups, the day part, and the day of the week of FIG. 14.

[0019] FIG. 16 illustrates a calculation of a number of census streams for an example one of the demographic groups of FIG. 14 for a reporting period.

[0020] FIG. 17 illustrates a calculation of scaled census sessions for the example demographic groups of FIG. 14.

[0021] FIG. 18 is a block diagram of an example processor platform capable of executing the instructions of FIGS. 4-13B to implement the systems of FIGS. 1, 2, and/or 3.

DETAILED DESCRIPTION

[0022] Example systems, methods, apparatus, and articles of manufacture disclosed herein generate hybrid online audience measurement data via observational user-centric approaches using panel data (e.g., data obtained from a panel of participants such as panelists who have agreed to have their online browsing activity monitored) as core source data. In some examples, the panel data is calibrated using information derived from web site-centric census data, such as a web site server's counts of web page hits. Example observational user-centric approaches disclosed herein effectively correct for the weaknesses of known online audience measurement systems, enables reporting on any web site, and/or enable determination of additional analytics that are available from panel data that could not previously be determined from census-based data.

[0023] Example systems, methods, apparatus, and articles of manufacture disclosed herein provide a measurement of audience exposures and behaviors at previously unmeasured locations (e.g., locations outside of home and work environments). Examples of such previously unmeasured locations include usage on work computers shared by multiple users, secondary computers at home, public access locations (e.g., public library computers), mobile device usage, and/or other previously unmeasured devices. The audience from unmeasured locations is combined with the audience from measured locations to obtain a total unique audience. The combined measured and unmeasured audiences are more accurate than total unique online audiences obtained under previous methods.

[0024] In some examples, participating publishers and/or web sites insert or embed a tag within the source code (e.g., HTML code) of their respective content. The tag may include Java, Javascript, and/or other executable instructions, which cause the page view to be recorded by a data collection facility when the tag executes on the browser.

[0025] Tags are well known from Blumenau, US Patent 6,108,637, which is hereby incorporated by reference in its entirety. Because a tag is embedded in the HTML defining a webpage and/or referenced by a pointer in the HTML of a web page, the tag is executed whenever a browser renders the corresponding webpage. Typically, a

tag will cause the browser to send a request for content to the data collection facility. The request may be thought of as a "dummy request," in that, unlike a conventional Internet request, the dummy request is typically not seeking to download content. Instead, the dummy request is actually being used to convey audience measurement information to the data collection facility. To this end, the payload of the dummy request carries identification information to be collected, compiled and/or analyzed by the data collection facility. The identification information may identify the webpage with which the tag is associated, the user device on which the webpage is accessed, the browser on which the webpage is accessed, the user, etc. In some examples, the data collection facility responds to the receipt of a dummy request by requesting a cookie from the requesting user device. In some such examples, if a cookie is not received (i.e., no cookie in the domain of the data collection facility is presently set on the user device), a cookie is set to facilitate identification of the device from webpage visit to webpage visit.

[0026] Tagging such as that described above is advantageous in that it enables the collection of census like data. In other words, because every browser that accesses the tagged webpage will respond to the tag by sending the dummy request, every access to the webpage will be known to the data collection facility. Moreover, the collection of this data does not require the use of a special browser, or of special metering software at the user devices. Rather, because a dummy request appears to a conventional commercially available browser (e.g., Firefox, Microsoft Explorer, etc) as any other request to retrieve Internet media (e.g., as a request to obtain content or advertisement material to be displayed as part of the webpage), any such browser will participate in the audience measurement process without requiring modification. As a result, tagging enables collection of this audience measurement data from panelists and non-panelists alike. Therefore, data collected via a tagging approach such as that described above, is described herein as census data.

[0027] As mentioned above, panelists are persons that have agreed to be monitored by, for example, an audience measurement entity such as The Nielsen Company (U.S.), LLC. Typically, such panelists provide detailed demographic information (e.g., race, age, income, home location, education level, gender, etc) when they register to participate in the panel. Additionally, panelists are provided with a meter to collect their media exposure. For example, a software meter, such as

that disclosed in Coffey, U.S. Patent 5,675,510, which is incorporated by reference in its entirety, may be downloaded onto the panelist's computer, smart phone, tablet and/or other browsing device. The meter will collect data indicating media access activities (e.g., web site names, dates/times of access, clickstream data, and/or other information identifying media (e.g., webpage content, advertisements, etc) to which the panelist is exposed. This data is uploaded, periodically or aperiodically, to the data collection facility. The data collected by a meter is referred to herein as panelist data. Panelist data is advantageous in that it is married to detailed demographic information since the panelist has provided their demographics as part of the registration and the activity data collected by the meter can, thus, be associated with that demographic information. When a panelist user device accesses a tagged page, the access will be logged by the meter and by the data collection facility via the tagging mechanism mentioned above. Thus, the panelist who accesses a tagged webpage provides a bridge between panelist data and census data.

[0028] Based on the panelist information received via the meters and the census information received via the tags, example systems and methods disclosed herein generate online audience measurement information (e.g., exposure statistics, demographics, etc.) using the following technique(s): (1) apply filtration techniques to the census-based data to remove undesirable traffic (e.g., automatic/programmatic refreshes of web pages, which causes additional hits on the web page, robot traffic, traffic originating from out-of-market geographic locations, etc.); (2) apply dictionary definitions to classify observed behaviors, web sites, brands, and/or channels; (3) determine the size and/or demographics of the population accessing the web pages from unmeasured locations; and (4) weight the measured and unmeasured location behaviors to represent the observed traffic.

[0029] Some example methods disclosed herein include obtaining panelist data corresponding to a plurality of panelists accessing web pages at measured locations, classifying the panelist data according to demographic information of the panelists, generating a virtual panel based on an estimate of web page access at unmeasured locations, and classifying page view data associated with the unmeasured locations based on data corresponding to the virtual panel.

[0030] Some example methods disclosed herein include assigning weights to a plurality of panelists based on first estimated characteristics, selecting a subset of the

panelists based on the weights, re-weighting the selected panelists based on second estimated characteristics, and generating a virtual panel including the selected panelists as virtual panelists, wherein data collected from the selected panelists for the virtual panel are to be assigned weights based on the re-weighting.

[0031] Example apparatus disclosed herein include a panel data collector to collect panelist data corresponding to a plurality of panelists accessing web pages at measured locations, a virtual panel generator to generate a virtual panel based on an estimate of web page access at unmeasured locations, and an audience classifier to classify the panelist data according to demographic information of the panelists and to classify page view data associated with the unmeasured locations based on data corresponding to the virtual panel.

[0032] Example apparatus disclosed herein include a sample weighter to assign weights to a plurality of panelists based on first estimated characteristics, a sample selector to select a subset of the panelists based on the weights, and a sample re-weighter to re-weight the selected panelists based on second estimated characteristics, and to generate a virtual panel including the selected panelists as virtual panelists, wherein data collected from the selected panelists for the virtual panel are to be assigned weights based on the re-weighting.

[0033] As used herein, the following terms are defined to mean:

[0034] Uniform resource locator (URL) pattern – a set of similar URL instances that are classified together due to similarity of content and/or purpose.

[0035] URL instance – a unique URL string. A URL instance may differ from other URL instances belonging to a URL pattern based on particulars of the URL string, arguments in the URL string, and/or any other variation on the URL string that makes the URL unique.

[0036] Page view – used interchangeably herein with "exposure," refers to a web page or other online media (e.g., content, advertisements, video, image, audio, etc.) being provided (e.g., successfully transmitted) to a consumer or requester or device, with or without any indication that the provided web page was actually viewed by the consumer or requester.

[0037] Stream – also used interchangeably herein with "exposure," refers to providing an instance of streaming video and/or audio, similar to a page view except referring to different types of media.

[0038] Session – time spent on the internet from log on to log off, or a continuous surfing of a particular web site or multiple web sites by an individual.

[0039] Universe/population – totality of individuals/devices/households of interest, may be limited to only online individuals/devices/households.

[0040] Census-based data – data collected based on tags or another mechanism not limited to panel measurement.

[0041] Panelist - a person or group of persons that have agreed to have one or more aspects of their behavior (e.g., browsing activity, television viewing, etc) monitored.

[0042] Panelist data - data collected by a meter associated with a panelist.

[0043] Meter - a tool of any type (e.g., software and/or hardware) which collects data reflecting (or enabling the determination of) the identity of a user and/or the identity of media to which a person (e.g. a panelist) is exposed.

[0044] Media - any type of content and/or advertisement delivered via any type of delivery mechanism (e.g., webpages, television, video, audio, etc).

[0045] Measured site - a user device or physical location (e.g. a room) at which a meter is installed for collecting panelist data.

[0046] Unmeasured site - a user device or physical location which is not monitored by a meter.

[0047] Measured traffic - one or more accesses to media from one or more measured sites.

[0048] Unmetered traffic - one or more accesses to media from unmeasured sites.

[0049] Some examples below are described with reference only to page views for brevity. However, some examples disclosed herein are applicable to other types of media such as video streams, audio streams, and/or any other type of measureable online traffic.

[0050] FIG. 1 is a block diagram illustrating a disclosed example system 100 constructed in accordance with the teachings of this disclosure to measure online audiences. The example system 100 of FIG. 1 obtains measurements of online audience traffic from measured sites based on a panel of respondents or panelists, estimates online audience traffic from unmeasured sites based on census data, and estimates a total online audience for a web site based on the measured traffic from measured sites and the estimate of unmeasured traffic. The example system 100 may

further provide demographic data regarding the estimated total online audience (e.g., across measured and unmeasured sites) for a web site.

[0051] The example system 100 of FIG. 1 includes a hybrid online audience measurement system 102, a universe estimator 104, measured web servers 106, 108, a network 110, and an online audience panel 112.

[0052] The example hybrid online audience measurement system 102 of FIG. 1 obtains (e.g., receives, retrieves, collects, etc.) panel-based online audience measurement data, census-based measurement data, and survey-based audience data, and determines online audiences (size and/or demographic composition) for specific media. The example hybrid online audience measurement system 102 of FIG. 1 may determine the online audiences for a designated reporting period for web sites of interest, such as web sites served by the example web servers 106, 108, and/or for aggregations of web sites belonging to channel, brand, and/or parent entities. The online audience measurement information generated by the example hybrid online audience measurement system 102 may be used to improve web site traffic, analyze web sites for purposes of purchasing advertising space, price advertisement placement, and/or any other use of online audience measurement information.

[0053] As described below, the example hybrid online audience measurement system 102 of FIG. 1 provides a more accurate measurement of online audiences than known audience measurement systems by combining the advantages of panel-based audience measurement with the advantages of or census-based audience measurement. For example, panel-based measurement has the advantage of more accurately representing a population or universe to be measured with respect to demographics and other useful statistics. In contrast, census-based audience measurement has the advantage of accurately measuring a total quantity of online traffic.

[0054] The example universe estimator 104 of FIG. 1 generates an estimate of online audiences, including the demographics, locations, and/or behaviors of online audiences. In some examples, the universe estimator 104 provides an independent and/or continuous enumeration study to provide current estimates of an online population. The example universe estimator 104 of FIG. 1 generates the estimates from an ongoing quarterly Internet universe enumeration survey. The example survey uses a sample of panelists recruited by the remote digit dial (RDD) methodology to

collect Internet access information (e.g., web pages visited, time spent online, etc.) and/or demographic profiles (e.g., age, gender, etc.) of Internet users. The example universe estimator 104 of FIG. 1 collects Internet access information including estimates of online behavior in locations other than measured (e.g., work and/or home) environments of the panelists (e.g., at unmeasured locations such as coverage of work environment usage on computers shared by multiple users, secondary computers in home environments, public access locations, mobile usage, and previously unmeasured devices). Estimating the behavior corresponding to unmeasured locations improves overall response rates and reduces response biases that can be associated with non-proprietary, omnibus surveys.

[0055] The example web servers 106, 108 of FIG. 1 are communicatively coupled to the network 110 (e.g., the Internet) to serve web pages, video streams, and/or other web traffic to requesting devices. In the example system 100 of FIG. 1, the web servers 106, 108 serve web pages and/or video streams that have been tagged for measurement in accordance with the Blumenau methodology explained above. For example, the web servers 106, 108 may tag served web pages and/or video streams by including one or more monitoring instructions in each served web page and/or video stream. The example tag code may be active content (e.g., Javascript instructions, Java instructions, HTML5 instructions, etc.), which causes the device receiving the tagged web page and/or video stream to execute the instructions to browser information to the example hybrid online audience measurement system 102 (e.g., to a daemon 136 to store the browser information), to the web servers 106, 108, and/or to a different logging server.

[0056] The example online audience panel 112 of FIG. 1 provides measurements of online activities of panelists 114, such as web pages visited, video streams downloaded, and/or lengths of time spent browsing the web pages and/or playing video streams. The example panelists 114 are associated with respective computers 116, each of which is provided with an online monitor application (i.e., a meter 118) when the panelist 114 becomes a member of the online audience panel 112. The online monitor applications 118 are uniquely associated with respective user devices 116 and, thus, the activity data they collect can be mapped to the demographic characteristics of the panelists 114. Thus, the measured activities of the panelists 114 are correlated to the demographic characteristics of the panelists 114. The example

online monitor applications 118 records the uniform resource locators (URLs) of the web pages and/or video streams received at the computers 116, keystrokes entered, items clicked on via a cursor, and/or any other interaction(s) performed by the panelists 114 with the computers 116. The example computers 118 also execute tags embedded in monitored web pages (e.g., monitoring instructions). In some examples, the tags are recognized by the meter 118. In other examples, the tags are logged by a daemon 136 at the hybrid online audience measurement system 102 and not recognized by the meter 118. The online monitor applications 118 transmit logs of online activities of the panelists to the example hybrid online audience measurement system 102. The logs may be transmitted at regular intervals (e.g., daily, weekly, biweekly, monthly, etc.), on demand, in response to an event or request, at predetermined times, and/or according to any other schedule(s) and/or condition(s).

[0057] User devices and/or locations monitored by one or more meters 118 are referred to as measured locations. Measured locations may include home environments 120 (e.g., computers located at the panelists' homes) and work environments 122 (e.g., computers located at the panelists' places of employment). The activities of a given panelist 114 may be measured in the home environments 120, the work environments 122, and/or both the home and work environment. As some businesses preclude installation of meters 118, some panelists are only monitored at their home environment and not at their work environments.

[0058] Online activities also occur outside of the home and work environments 120, 122. The example system 100 of FIG. 1 does not directly measure online activities performed in unmeasured locations, but these activities result in additional page views and/or video streams being served by the example servers 106, 108. Such example unmeasured locations that can generate web traffic include traffic generated in the work environment 122 via computers 124 that do not include a meter 118 (e.g., computers that are shared by multiple users 126 including non-panelists, panelists and non-panelists, etc.), secondary (e.g., non-primary) computers 128 in the home environments 120, computers 130 in public access locations (e.g., libraries, universities, etc.), mobile devices 132 (e.g., mobile phones, smartphones, tablet computers, etc.), and/or any other type(s) of unmeasured devices (e.g., Internet-connected appliances such as smart televisions, digital media players, game consoles,

etc.). Activities (e.g., media exposures) in unmeasured locations can be performed by panelists and/or non-panelists.

[0059] The example web servers 106, 108 of FIG. 1 include server monitors 134 that measure the web traffic (e.g., web pages served, video streams served, etc.) served by the respective web servers 106, 108. The example server monitors 134 of FIG. 1 collect information such as the details of browsers or other applications requesting web pages and/or video streams from the servers 106, 108, the IP addresses of requesters, lengths of browsing sessions of persons on the servers 106, 108, and/or any other information that can be determined by the servers 106, 108 (e.g., via logging and/or analyzing requests for web pages and/or via cookies). The data collected via the server monitors 134 are considered to be server-based, server-centric data. Server-centric data is considered to have limitations. For example, such server side data is subject to manipulation (e.g., by the proprietor of the server and/or by robots or other devices programmed to repeatedly request data from servers). Thus, server-side data can over count page views. As web pages are often cached in user devices, a second or later access to a webpage may not involve a request to a server. Instead, the webpage may simply be retrieved from a local cache of the user device or served by an intervening proxy server. As such, server side data may additionally undercount page views.

[0060] In contrast to these potential overcounting and undercounting problems of server side data, the census data collected by the tagging system is accurate as every access to a tagged web page (whether from a cache or not), will cause the tag to fire, resulting in issuance of a dummy request and logging of an exposure to the tagged webpage.

[0061] FIG. 2 is a block diagram of an example implementation of the hybrid online audience measurement system 102 of FIG. 1. The example hybrid online audience measurement system 102 of FIG. 2 obtains input data including panel-based online activity data, an estimate of total and/or subsets of an online population, and/or census data measurements of traffic for particular web site(s). Based on the input data, the example hybrid online audience measurement system 102 of FIG. 2 classifies and/or outputs data reflecting an online audience for web site(s), channel(s), brand(s), parent(s), and/or any other organization unit of interest. In some examples, the hybrid

online audience measurement system 102 classifies data reflecting an online audience for a particular reporting period, for day(s) of the week, and/or part(s) of the day.

[0062] The example hybrid online audience measurement system 102 of FIG. 2 obtains data from a census-based data collector 202, a survey-based data collector 204, and a panel data collector 206. The example hybrid online audience measurement system 102 may obtain the data via a network (e.g., the network 110 of FIG. 1), via manual data entry (e.g., entry of survey responses), and/or using any other method of receiving data.

[0063] The example census-based data collector 202 of FIG. 2 receives census-based traffic information. Census-based data may be obtained from, for example, server logs generated by the server monitors 134, tag-based data collected by the daemon 136, and/or any other source of census data. The census-based traffic information may include data collected from dummy requests made to the census-based data collector 202 resulting from the execution of tags in tagged web pages and/or statistics collected via the server monitors 134 based on server logs reflecting requests for web pages, session measurements, and/or other traffic information that can be collected via the server monitors 134. In some examples, the census-based data collector 202 implements the example daemon 136 to collect, parse, and/or store data received from the devices 116, 124, 128, 130, 132 of FIG. 1 in response to execute of tag instructions.

[0064] The example survey-based data collector 204 of FIG. 2 receives survey-based behavior information, such as "universe estimates" of the total audience and/or subsets of an online audience (e.g., from the example universe estimator 104 of FIG. 1). In the examples of FIGS. 1 and 2, the universe estimates are obtained from personal interviews, which may be conducted via telephone. The interviews provide Internet access information and demographic profiles of Internet users, including estimates for measured locations such as work and home, and for unmeasured locations such as outside of work and home. Survey data is based on the respondent's ability to accurately recall their activities and willingness to truthfully report.

[0065] The example panel data collector 206 of FIG. 2 receives panelist data including page view data representative of panel activities (e.g., collected via the monitoring applications 118 such as meters). The panelist data may include web sites visited (e.g., URLs), sessions including multiple URLs, timestamps reflecting

time/date of occurrence of web site requests and/or sessions, demographic characteristics of the panelists, and/or any other information that may be collected via the online monitor applications of FIG. 1. The example panel data collector 206 of the illustrated example filters the received panelist data according to rules. Additionally or alternatively, the example panel data collector 206 may sort the received panelist data according to demographic categories, day-of-week information, and/or time-of-day information to obtain more precise data.

[0066] The example panel data collector 206 of FIG. 2 weights panelist web site requests and/or sessions based on a determined representation of the panelists relative to the universe. For example, the behavior of a first panelist who represents a larger portion of the universe than a second panelist will be weighted more heavily (e.g., multiplied by a larger factor) than the activities of the second panelist. The example panel data collector 206 of FIG. 2 determines estimated activities for individual measured locations and for the measured locations in general (e.g., by removing overlapping or duplicated audience entries). An individual panelist may be counted in multiple measured locations (e.g., at home and at work), so the panelist's presence in the multiple locations is counted for by removing an apparent duplicate audience member.

[0067] The example hybrid online audience measurement system 102 of FIG. 2 processes the census-based traffic information to clean and/or classify the data. To this end, the example hybrid online audience measurement system 102 includes a traffic filter 208 and a site classifier 210.

[0068] Many web sites (e.g., web servers 106, 108) receive traffic (e.g., page views) that is generated by non-human and/or indirectly human activities (e.g., robots, web crawlers web spiders, automatic page refreshes, and/or other traffic not generated by a person consciously or intentionally requesting a web site). The example traffic filter 208 cleans the census-based information (e.g., server logs and/or tag-based counts) to avoid counting irrelevant data and/or other non-human activity. For example, the traffic filter 208 of FIG. 2 applies a list of known user agents, known IP addresses, and/or activity duration thresholds to the census-based information to identify and remove non-human traffic. The list of known user agents and/or known IP addresses may be obtained from the Interactive Advertising Bureau (IAB)/Audit Bureau of

Circulations Electronic (ABCe) International Spider & Robot List. In some examples, the traffic filter 208 further applies activity-based filters to detect non-human traffic.

[0069] The example traffic filter 208 of FIG. 2 categorizes the census-based data by geographical region (e.g., state, region, nation, continent, etc.) using the IP addresses of requesting devices. Thus, international traffic to the web servers 106, 108 may be removed if not of interest for a particular use. The example traffic filter 208 of FIG. 2 also removes automatic web page refreshes (e.g., web page code that causes a request for an updated version of the same web page to be sent, typically at some interval after the initial web page is received). Because the example monitor applications 118 recognize tags, the example panel data collector 206 may estimate the auto-refresh page view activity by identifying the human-requested page views and distinguish the page views resulting from auto-refreshing. For example, the monitor applications 118 (e.g., meters) may identify web page refreshes (e.g., web page requests) by tracking browser activity and determining whether corresponding a panelist action (e.g., button clicks, keystrokes, etc.) occurred to cause the web page refresh and/or whether browser execution of instructions in the web page source code causes the web page refresh.

[0070] The human page views and the auto-refresh page views may be converted to a rate or ratio, which the example traffic filter 208 of FIG. 2 applies to the census-based traffic (e.g., page views) to reduce or remove page views attributable to auto-refreshes. For example, the auto-refresh rate may be determined from panel-based data for a URL pattern by a day of the week and a day part, as the ratio of number of auto-refresh generated page views to all page views. The auto-refresh rate or ratio is then applied to the census-based page views for a selected URL pattern (URL page views) (e.g., which have been cleaned of other non-human and/or non-geographic market traffic) to determine an adjusted or cleaned number of census-based URL page views (adjusted URL page views). An example calculation is shown in the following equation: Adjusted URL page views = URL page views * (1 – auto-refresh rate).

[0071] The example site classifier 210 of FIG. 2 receives the filtered census-based traffic information (e.g., the adjusted URL page views for the URLs of interest) and classifies the URLs (e.g., web pages and/or video streams) into categories (e.g., sports, retail, etc.). For example, the site classifier 210 of FIG. 2 applies a dictionary of classifications to assist in classifying and/or modeling panelist activities by the

categories. The example site classifier 210 classifies in multiple ways, such as brands, parent entities, channels, Internet domains, Internet sub-domains, and/or in any other way. For example, a parent may include multiple brands, each of which may include multiple channels.

[0072] The example hybrid online audience measurement system 102 of FIG. 2 further includes a virtual panel generator 212. FIG. 3 is a block diagram of an example implementation of the example virtual panel generator 212 of FIG. 2. While the example panelist data and the example census data may be used to effectively estimate and/or classify online audiences in measured locations (e.g., the work and home environments 120, 122 of FIG. 1), census-based traffic often indicates that increasing amounts of traffic originate from unmeasured locations. The example virtual panel generator 212 generates a virtual panel to enable more accurate estimation of traffic from unmeasured locations than was previously possible. Example unmeasured locations may include, but are not limited to, shared computers in work environments, secondary or tertiary computers in home environments, computers in public access locations, mobile devices, and/or other devices not measured and/or not measurable via panel-based methods.

[0073] In the example of FIG. 2, a negative binomial distribution (NBD) calculator 214 determines a number of people who access the web site(s) from unmeasured locations. The negative binomial distribution is a discrete probability distribution of a number of successes r in a sequence of n independent Bernoulli trials (the outcome of each Bernoulli trial being defined as a success or a failure). The example NBD calculator 214 may consider a success to be a page view of a URL by a particular panelist in a trial of whether a panelist visits the URL. The example NBD calculator 214 of FIG.2 determines characteristics (e.g., size and/or demographic composition) of an audience who access web pages only from unmeasured locations (e.g., are not represented by the panelist data measuring measured locations). To determine the audience at unmeasured locations, the example NBD calculator 214 of FIG. 2 receives weighted audience characteristic data for each of multiple demographic groups determined from the panelist data, weighted session data for each of the demographic groups determined from the panelist data, numbers of cleaned sessions for each of the demographic groups determined from the census-based data after processing by the traffic filter 208, and a universe estimate of the audience size determined from survey

data. The demographic groups may define subsets of a universe or population. Accordingly, the example NBD calculator 214 of the illustrated example performs multiple NBD calculations corresponding to multiple demographic groups. The example NBD calculator 214 of FIG. 2 determines variables to be used in an NBD process and calculates the NBDs for the demographic groups based on the variables. Example processes (e.g., computer readable instructions) to calculate the NBD and the input variables to the NBD are described below.

[0074] Once the number of persons who access the web sites solely from unmeasured locations is determined (e.g., from the NBD calculator 214), the example virtual panel generator generates a virtual panel to represent the audience for the unmeasured locations. The example virtual panel generator 212 of FIG. 2 generates the virtual panel (e.g., the unmeasured location sample) by selecting a subset of panelists of measured locations (e.g., home and work environments), duplicating the selected panelists (and their corresponding online behaviors) to form an unmeasured location sample. In the example of FIG. 2, the duplicated versions of the panelists are provided with separate identifiers to simulate or represent the duplicated panelists as actual panelists. The example virtual panel generator 212 selects the subset of the panelists to have a demographic composition similar to an estimated demographic distribution of the unmeasured location audiences, and so that the activities of the duplicated panelists represent the surveyed behavior of the unmeasured location audiences. The example virtual panel generator 212 may select all or a portion of the activities of the duplicated panelists to meet volume targets (numbers of page views to the web sites calculated based on differences between panelist data and census-based data).

[0075] The example virtual panel generator 212 of FIG. 3 includes a sample weighter 302, a sample selector 304, and a sample re-weighter 306. The example virtual panel generator 212 receives a selected sample size 308 of the resulting virtual panel (e.g., a value input by an operation reflecting a desired panel size such as 10,000 members), audience target(s) 310 (e.g., a difference between a census-based audience and a panel based audience and/or an estimated number of page views resulting from the unmeasured locations) for one or more demographic groups, panel behavior data 312 (e.g., online activities), NBD output 314 (e.g., calculations from the

NBD calculator 214 of FIG. 2), and an estimated universe size for unmeasured locations 316.

[0076] The selected sample size 308 of FIG. 3 represents a number of panelists to be selected (e.g., by the sample selector 304) to form the virtual panel. The selected sample size 308 may be at least as large as a number of panelists needed to represent the demographics and/or activities of an audience that only accesses web pages from unmeasured locations (e.g., persons who are not active online in measured locations such as home or work environments).

[0077] The example panelist behavior data 312 (e.g., a subset of panelist data) includes data representing the activities for each of the panelists to be weighted and/or considered for the virtual panel. The example sample weighter 302 and/or the example sample re-weighter 306 of FIG. 3 compare the activities reflected in the panel behavior data 312 to activities in the unmeasured locations to determine weights for the panelists.

[0078] The example audience target(s) 310 of FIG. 3 (e.g., an expected or estimated number of audience members) are calculated as a difference between a census-based audience estimate (e.g., an estimate of a total audience for one or more web pages based on the cleaned census-based page views) and a panel-based audience estimate (e.g., an estimate of an audience at measured locations). The audience target 310 of FIG. 3 is an estimated or expected number of audience members with online activity occurring only in unmeasured locations. The panel-based audience estimate is calculated by the NBD calculator 214 of FIG. 2. The census-based audience estimate is calculated based on the census-based web traffic (e.g., a number of page requests or impressions, a number of video streams, etc.) and a panel-based page view volume calculation (e.g., web page requests or impressions and panelist sessions). For example, the number of page views determined for a URL pattern from the cleaned census-based data may be divided by the average number of page views per panelist (e.g., audience member) to determine an estimated number of audience members for the census-based data. In some examples, the audience target(s) 310 include a total audience target for an entire population, and audience targets for individual subsets of the population corresponding to demographic subgroups.

[0079] The universe size for unmeasured locations 316 is determined in the illustrated example based on a total universe size (e.g., a total online population

determined from surveys) and a total audience for measured locations (e.g., an audience from all measured locations determined from the panelist data). In the example of FIG. 3, the universe size for unmeasured locations 316 is the difference between the total universe size and the total audience for the measured locations.

[0080] The example survey-based data collector 204 of FIG 2 provides the virtual panel generator 212 with an estimated size and demographic composition of the population accessing the web sites from unmeasured locations. The census-based data collector 202 provides census-based page view information and the panel data collector 206 provides panel-based information to the virtual panel generator 212. The panelist data and the census data provide guidance on the page views from the unmeasured locations on a site-by-site basis, such as page view volume discrepancies between estimates of page views from measured locations and page views measured by the census-based data collector 202 (e.g., via the server monitors 134 and/or the daemon 136).

[0081] To create the virtual panel (e.g., the unmeasured location sample), the example virtual panel generator 212 (e.g., via the sample weighter 302) applies or assigns a weight to each of the panelists. The sample weighter 302 generates the weights to represent the desirability or suitability of each of the example panelists for representing an audience in unmeasured locations. For example, the sample weighter 302 may determine desirability based on the audience targets 310, including demographic targets (e.g., a estimated demography of the unmeasured locations) and/or entity targets (e.g., an estimated unique audience 314 in the unmeasured locations determined by the NBD calculator 214, an estimated difference in page views between the census-based data and the panelist data). Using Calmar weighting, the example sample weighter 302 of FIG. 2 assigns a weight to each panelist based on the panelist's individual online activities and/or demographic characteristics compared to the demographic and/or entity targets. A weight is determined for each panelist when the Calmar weighting converges. The weight designates how closely the panelist represents a typical person in an unmeasured location. Panelists that are similar to estimated demographic profiles and/or behavioral profiles of persons accessing a web site from an unmeasured location are given higher weights.

[0082] After weighting, the example sample selector 304 of FIG. 3 selects a number of the panelists for inclusion in the virtual panel using random selection. The weights

of each of the panelists, calculated by the sample weighter 302, are used as probabilities of selection (or inclusion). A panelist having a high weight will have a higher likelihood of selection by the sample selector 304. The example virtual panel generator 212 may randomly select a number of the panelists (e.g., a fixed number) using, for example, the FastCube method. Selected panelists and their corresponding demographic, behavioral, and/or other data are duplicated to create a virtual panelist while retaining the original panelist in the panel.

[0083] After selecting (e.g., generating) the virtual panelists, the example sample re-weighter 306 of FIG. 3 re-weights the selected virtual panelists against the demographic targets and/or audience targets. The example targets used in the re-weighting can be the same as during the first weighting or different from those used in the first weighting, depending on how much is known about the audience and/or behaviors in the unmeasured locations. The example virtual panel generator 212 outputs a virtual panel including the set of selected panelists and corresponding weights to represent the audience and/or behavior targets.

[0084] The example panelist data (e.g., the behavior information of the panelists), the census-based data, and the virtual panel (e.g., panelist data for panelists selected and/or duplicated for the virtual panel) are provided to an activity fulfiller 216. The example activity fulfiller 216 of FIG. 2 fulfills (estimates the origin of) any remaining activity between measured and unmeasured panel volume and the measured census-based volume. To fulfill the virtual panel activity, the example activity fulfiller 216 determines an amount of activity to fulfill (e.g., estimate, match) the activities collected by the census-based data collector 202, but not accounted for by the panelist data and the virtual panel, for each of the demographic groups for each day part. The discrepancy in activity may result from unmeasured activities in measured environments (e.g., activities on devices that are located in home and/or work environments but are not metered). The example activity fulfiller 216 then randomly and/or probabilistically duplicates instances of panelist activity (e.g., sets of page views and/or video streams logged for panelists, online sessions of panelists, etc.) to compensate for the differences in activity.

[0085] The example activity fulfiller 216 of FIG. 2 generates and/or outputs a report of the audience, the demographics of the audience, and/or the determined activities of the audience for measured URLs, measured sets of similar URLs, measured brands,

measured channels, measured parent entities, and/or for any other measured entity or sub-entity.

[0086] The example hybrid online audience measurement system 102 of FIG. 2 further includes an audience classifier 218. The example audience classifier 218 receives the determined audiences for measured locations (e.g., from the panel data collector 206) and for unmeasured locations (e.g., from the virtual panel generator 212), and additional activity not represented by the panels (e.g., from the activity fulfiller 216). The example audience classifier 218 further receives page view data from the example census-based data collector 202 and/or cleaned page view data from the site classifier 210. The audience classifier 218 of FIG. 2 classifies a first portion of the page view data based on the panelist data (e.g., for measured locations) and classifies a second portion of the page view data based on the virtual panel (e.g., for unmeasured locations). Classifying the page view data may include generating statistics and/or reports to classify audiences and/or traffic for combinations of URLs and/or entities, day parts, days of week, and/or any other classification. The example audience classifier 218 of FIG. 2 estimates an audience for one or more web sites during a reporting period. For example, the audience classifier 218 of the illustrated example estimates numbers of unique audience members of web sites and the demographics of web site audiences.

[0087] While an example manner of implementing the hybrid online audience measurement system 102 of FIG. 1 has been illustrated in FIG. 2 and an example manner of implementing the virtual panel generator 212 of FIG. 2 has been illustrated in FIG. 3, one or more of the elements, processes and/or devices illustrated in FIGS. 1-3 may be combined, divided, re-arranged, omitted, eliminated and/or implemented in any other way. Further, the example hybrid online audience measurement system 102, the example universe estimator 104, the example web servers 106, 108, the example monitoring applications 118, the example server monitors 134, the example census-based data collector 202, the example survey-based data collector 204, the example panel data collector 206, the example traffic filter 208, the example site classifier 210, the example virtual panel generator 212, the example NBD calculator 214, the example activity fulfiller 216, the example audience classifier 218, the example sample weighter 302, the example sample selector 304, the example sample re-weighter 306 and/or, more generally, the example system 100, the example hybrid

online audience measurement system 102, and/or the example virtual panel generator 212 of FIGS. 1, 2, and/or 3 may be implemented by hardware, software, firmware and/or any combination of hardware, software and/or firmware. Thus, for example, any of the example hybrid online audience measurement system 102, the example universe estimator 104, the example web servers 106, 108, the example monitoring applications 118, the example server monitors 134, the example census-based data collector 202, the example survey-based data collector 204, the example panel data collector 206, the example traffic filter 208, the example site classifier 210, the example virtual panel generator 212, the example NBD calculator 214, the example activity fulfiller 216, the example audience classifier 218, the example sample weighter 302, the example sample selector 304, the example sample re-weighter 306 and/or, more generally, the example system 100, the example hybrid online audience measurement system 102, and/or the example virtual panel generator 212 could be implemented by one or more circuit(s), programmable processor(s), application specific integrated circuit(s) (ASIC(s)), programmable logic device(s) (PLD(s)) and/or field programmable logic device(s) (FPLD(s)), etc. When reading any of the apparatus or system claims of this patent to cover a purely software and/or firmware implementation, at least one of the example hybrid online audience measurement system 102, the example universe estimator 104, the example web servers 106, 108, the example monitoring applications 118, the example server monitors 134, the example census-based data collector 202, the example survey-based data collector 204, the example panel data collector 206, the example traffic filter 208, the example site classifier 210, the example virtual panel generator 212, the example NBD calculator 214, the example activity fulfiller 216, the example audience classifier 218, the example sample weighter 302, the example sample selector 304, and/or the example sample re-weighter 306 are hereby expressly defined to include a tangible computer readable storage medium such as a memory, DVD, CD, Blu-ray, etc. storing the software and/or firmware. Further still, the example system 100, the example hybrid online audience measurement system 102, and/or the example virtual panel generator 212 of FIGS. 1, 2, and/or 3 may include one or more elements, processes and/or devices in addition to, or instead of, those illustrated in FIGS. 1, 2, and/or 3, and/or may include more than one of any or all of the illustrated elements, processes and devices.

[0088] Flowcharts representative of example machine readable instructions for implementing the hybrid online audience measurement system 102 of FIGS. 1-3 are shown in FIGS. 4-13B. In these examples, the machine readable instructions comprise program(s) for execution by a processor such as the processor 1812 shown in the example processing platform 1800 discussed below in connection with FIG. 18. The program may be embodied in software stored on a tangible computer readable storage medium such as a CD-ROM, a floppy disk, a hard drive, a digital versatile disk (DVD), a Blu-ray disk, or a memory associated with the processor 1812, but the entire program and/or parts thereof could alternatively be executed by a device other than the processor 1812 and/or embodied in firmware or dedicated hardware. Further, although the example program is described with reference to the flowcharts illustrated in FIGS. 4-13B, many other methods of implementing the example hybrid online audience measurement system 102 and/or the example virtual panel generator 212 may alternatively be used. For example, the order of execution of the blocks may be changed, and/or some of the blocks described may be changed, eliminated, or combined.

[0089] As mentioned above, the example processes of FIGS. 4-13B may be implemented using coded instructions (e.g., computer readable instructions) stored on a tangible computer readable storage medium such as a storage drive, a storage disc, a hard disk drive, a flash memory, a read-only memory (ROM), a compact disk (CD), a digital versatile disk (DVD), a Blu-ray disc, a cache, a random-access memory (RAM) and/or any other storage device or storage disc in which information is stored for any duration (e.g., for extended time periods, permanently, brief instances, for temporarily buffering, and/or for caching of the information). As used herein, the term tangible computer readable storage medium is expressly defined to include any type of computer readable storage device and/or storage disc and to exclude propagating signals. Additionally or alternatively, the example processes of FIGS. 4-13B may be implemented using coded instructions (e.g., computer readable instructions) stored on a non-transitory computer readable medium such as a hard disk drive, a flash memory, a read-only memory, a compact disk, a digital versatile disk, a cache, a random-access memory and/or any other storage device or storage disc in which information is stored for any duration (e.g., for extended time periods, permanently, brief instances, for temporarily buffering, and/or for caching of the information). As

used herein, the term non-transitory computer readable medium is expressly defined to include any type of computer readable storage device and/or storage disc irrespective of the duration of storage and to exclude propagating signals. As used herein, when the phrase "at least" is used as the transition term in a preamble of a claim, it is open-ended in the same manner as the term "comprising" is open ended.

[0090] FIG. 4 is a flowchart representative of example computer readable instructions 400 that may be executed to implement the hybrid online audience measurement system 102 of FIGS. 1 and/or 2 to measure an online audience of a web site.

[0091] The example instructions 400 of FIG. 4 includes obtaining (e.g., receiving, collecting) universe estimates (e.g., via the survey-based data collector 204 of FIG. 2) (block 402). The example universe estimates include an estimate of an online population (e.g., a number of persons who are capable of accessing web sites or another definition of an online population). The universe estimates may further include surveyed behaviors of the online population.

[0092] The example hybrid online audience measurement system 102 (e.g., via the census-based data collector 202) obtains server-centric or census-based data for web sites (block 404). The census-based data may include tag-based data and/or measurements of web site traffic performed by the example server monitors 134 of FIG. 1. The example hybrid online audience measurement system 102 filters (e.g., via the traffic filter 208) the census-based data to remove undesirable traffic (block 406). Examples of undesirable (or non-representative) traffic include non-human traffic such as robots or spiders, traffic from non-representative geographical locations, and/or traffic resulting from automatic page refreshes. Example computer readable instructions to implement block 406 are described below with reference to FIG. 5. The example hybrid online audience measurement system 102 obtains panelist data (e.g., via the panel data collector 206 of FIG. 2) for measured locations (block 408). Example panelist data includes demographic and online behavior information for individual panelists.

[0093] The example hybrid online audience measurement system 102 classifies the web site(s) in the panelist data, the census-based data, and/or the survey-based data (e.g., via the site classifier 210 of FIG. 2) (block 410). For example, the site classifier 210 may classify URLs, sets of similar URLs, channels, brands, parent entities, and/or

any other organization of web sites using a dictionary of classifications (e.g., sports, retail, etc.).

[0094] The example hybrid online audience measurement system 102 estimates traffic and/or audiences from unmeasured locations (e.g., via the NBD calculator 214 of FIG. 2) (block 412). The example NBD calculator 214 estimates traffic and/or audiences from unmeasured locations based on the census-based data, the panelist data, and the survey-based data. Example computer readable instructions to implement block 412 are described below with reference to FIGS 6-8.

[0095] The example hybrid online audience measurement system 102 generates a virtual panel based on the estimate of the unmeasured location audience (e.g., via the virtual panel generator 212 of FIG. 2) (block 414). For example, the virtual panel generator 212 of FIG. 2 may weight the panelists included in the panelist data based on demographic targets, audience targets, and/or activity targets, and based on the demographic characteristics and/or online behaviors of the panelists. The example virtual panel generator 212 selects a number of the panelists based on the weights for inclusion in a virtual panel. Example computer instructions to implement block 414 are described below with reference to FIG. 9.

[0096] The example hybrid online audience measurement system 102 estimates an audience for unmeasured locations using the selected panelists (e.g., via the virtual panel generator 212 of FIG. 2) (block 416). For example, the virtual panel generator 212 may re-weight the panelists selected in block 414 to represent the demographic targets, audience targets, and/or activity targets. The combination of re-weighted panelists in the virtual panel may provide an estimated audience for unmeasured locations including demographic characteristics of the estimated audience.

[0097] The example hybrid online audience measurement system 102 determines additional activity not represented by the panel measurements (e.g., via the activity fulfiller 216 of FIG. 2 (block 418). For example, the activity fulfiller 216 may determine a difference between a census-based measurement of web site traffic and a panel-based estimation of web site traffic (e.g., measured panel and virtual panel). The example activity fulfiller 216 generates and/or classifies additional audience activity based on actual panelist activity to represent the difference. The difference in activity may include activity performed in measured locations (e.g., home environments, work environments) but not measured or represented by the panel.

Example computer readable instructions to implement block 418 are described below with reference to FIGS. 10A and 10B and/or 11A and 11B.

[0098] The example hybrid online audience measurement system 102 determines online audience(s) for web sites using the measured location audience, the unmeasured location audience, and the additional activity (block 420). In some examples, the hybrid online audience measurement system 102 classifies the online audiences and/or determines portions of the online audience corresponding to particular criteria, such as demographic groups, geographic locations, day parts, days of the week, and/or other criteria. In some examples, the hybrid online audience measurement system 102 classifies page view data associated with the unmeasured locations based on data corresponding to the virtual panel. The example instructions 400 of FIG. 4 may then end and/or iterate to determine additional online audiences.

[0099] FIG. 5 is a flowchart representative of example computer readable instructions 500 which, when executed, cause a processor to filter traffic (e.g., page views) from a set of traffic data (e.g., page view data). The example instructions 500 of FIG. 5 may be executed by the example traffic filter 208 of FIG. 2 to perform block 406 of FIG. 4. The example instructions 500 are performed subsequent to obtaining census-based data for measured web sites (e.g., traffic logs generated by the server monitors 134 for the web servers 106, 108 of FIG. 1).

[00100] The example traffic filter 208 obtains a list of known user agents and/or IP addresses representing non-human traffic (block 502). For example, the traffic filter 208 may obtain the IAB/ABCe International Spider & Robot List. The example traffic filter 208 selects a URL in the census-based data (block 504). The census-based data may have multiple entries (e.g., page views, video streams) for the URL and/or one or more variations of the URL that correspond to a same URL pattern. The example traffic filter 208 selects an entry in the census-based data for the selected URL (block 506). Each example entry includes information about the page view or video stream, such as the IP address of the requesting device, a user agent used to request the device, a time at which the page view was requested, and/or other information.

[00101] The example traffic filter 208 determines whether the IP address and/or the user agent of the selected entry match the list of user agents and/or IP addresses (block 508). If there is not a match, the example traffic filter 208 determines whether the IP address corresponds to a geographical region to be excluded from the online

audience measurement (e.g., traffic from another country) (block 510). For example, the traffic filter 208 may compare the IP address of the selected entry to a mapping of IP addresses to geographical regions. If the IP address and/or the user agent of the selected entry corresponds to the list of non-human IP addresses and/or user agents (block 508) of if the IP address of the selected entry corresponds to an excluded geographical region (block 510), the example traffic filter 208 removes the selected entry from the census-based data (block 512). In some examples, the traffic filter 208 archives the selected entry or otherwise marks the selected entry to not be used for determining the online audience.

[00102] After removing the selected entry (block 512), or if the selected entry is not matched to the list of IP address and/or user agents (block 508) and does not correspond to an excluded geographical region (block 510), the example traffic filter determines whether there are additional entries for the selected URL (or URL pattern) (block 514). If there are additional entries (block 514), control returns to block 506 to select another entry in the census-based data. When there are no additional entries for the selected URL (e.g., the non-human and/or excluded geographical region entries have been removed) (block 514), the example traffic filter 208 determines whether there is an auto-refresh rate available for the URL (or URL pattern) (block 516). For example, an auto-refresh rate may be determined for the URL based on panel-based observations of an average number of automatic refreshes of the URL.

[00103] If there is an auto-refresh rate available (block 516), the example traffic filter 208 removes a number of entries for the selected URL based on the auto-refresh rate (block 518). For example, the traffic filter 208 may remove a number of the entries for the URL proportional to the average number of automatic refreshes per human request. After removing the entries (block 518), or if there is no auto-refresh rate data available for the URL (block 516), the example traffic filter 208 determines whether there are additional URLs in the census-based data (block 520). If there are additional URLs (or URL patterns) (block 520), control returns to block 504 to select another URL from the census-based data. When there are no more URLs (block 520), the example instructions 500 of FIG. 5 end and control returns to block 408 of FIG. 4.

[00104] FIG. 6 is a flowchart representative of example computer readable instructions 600 which, when executed, cause a processor to estimate access to a web site from unmeasured locations. The example instructions 600 of FIG. 6 may be

performed by the example NBD calculator 214 of FIG. 2 to implement block 412 of FIG. 4. In some examples, the instructions 600 cause the NBD calculator 214 to estimate an audience of persons with access only from unmeasured locations.

[00105] The example NBD calculator 214 of FIG. 2 calculates NBD parameters (block 602). For example, the NBD calculator 214 may determine whether a Poisson condition exists and/or calculate variables to be used in the NBD process for each of the demographic groups to be measured. Example instructions for calculating the NBD parameters are described below with reference to FIG. 7. Using the calculated parameters, the example NBD calculator 214 calculates an NBD for the demographic groups (block 604). Example instructions for calculating the NBD are described below with reference to FIG. 8.

[00106] The example instructions 600 of FIG. 6 may then end and return a total audience estimate, including an audience from unmeasured locations, to the virtual panel generator 212 and/or to the activity fulfiller 216 of FIG. 2. Control returns to block 414 of FIG. 4.

[00107] FIG. 7 is a flowchart representative of example computer readable instructions 700 which, when executed, cause a processor to calculate negative binomial distribution parameters. The example instructions 700 of FIG. 7 may be performed by the example NBD calculator 214 of FIG. 2 to implement block 602 of FIG. 6.

[00108] The example NBD calculator 214 obtains inputs for determining the NBD parameters (block 702). Example inputs to the NBD calculator 214 include a weighted panel audience for a demographic group (Ui), weighted panel sessions for the demographic group (Vi), cleaned server sessions for the demographic group (Xi), and the estimated universe for the demographic group (Yi).

[00109] The weighted panel audience for the demographic group Ui is the estimated number of persons in the selected demographic group having at least one page view for a URL of interest, scaled to represent the universe for measured locations. An example determination of the weighted audience for the demographic group is described below. The weighted panel sessions for the demographic group Vi is the estimated number of sessions of the panelists, scaled to represent the universe for measured locations. Determination of the cleaned server sessions for the demographic group Xi is described below with reference to FIGS. 13A and 13B. The estimated

universe for the demographic group Yi may be determined from survey data and represents the number of persons in the total population of interest who are in the demographic group.

[00110] Based on the inputs, the example NBD calculator 214 can calculate additional information including panel sessions per person in the universe for the demographic group (e.g., Gr = 100 * Vi/Yi), adjusted panel sessions per person in the universe for the demographic group (e.g., Gp = 100 * Xi/Yi), and a weighted proportion of persons with zero page views (e.g., fr(0) = 1-Ui/Yi).

[00111] The example NBD calculator 214 then amends the variables (if necessary) from indicating 100% reach for the demographic group (e.g., all members of the demographic group universe have visited the web site during the reporting period) and/or 0% reach for the demographic group (e.g., no members of the demographic group universe have visited the web site) (block 704). For example, if fr(0) = 1, then fr(0) is changed to a number slightly less than 1 (e.g., 0.999). Conversely, if fr(0) = 0, then fr(0) is changed to a number slightly greater than 0 (e.g., 0.001).

[00112] The example NBD calculator 214 further determine a value for a constant "c" to be used in calculating the NBD parameters (block 706). In the example of FIG. 7, the constant "c" is calculated to be c = Gr/(100*ln(fr(0))). The example NBD calculator 214 determines whether the value of the constant "c" is greater than or equal to -1 to determine whether a Poisson condition is present (block 708). The Poisson condition represents a scenario in which members of a demographic group have less than a threshold likelihood of visiting a web site or genre of web site.

[00113] If the Poisson condition is not present (e.g., the constant "c" is less than -1) (block 708), the example NBD calculator 214 estimates the NBD parameter "A." To estimate the parameter "A," the example, NBD calculator 214 sets A = -2*(1+c) (block 710). The example NBD calculator 214 sets a placeholder variable "B" equal to "A" (block 712). The NBD calculator 214 calculates an updated value of A based on the previous value of "A" and based on the constant "C" (e.g., A = C*(A-(1+A)*LN(1+A))/(1+A+C)) (block 714).

[00114] The example NBD calculator 214 determines whether the value of "A" has converged (e.g., determines whether the updated value of A is within a threshold amount of B, or the previous value of A) (block 716). If the value of "A" has not converged (block 716), control returns to block 712 to iterate the calculation. When

the value of "A" has converged (block 716), the example NBD calculator 214 sets a second NBD parameter "k" (block 718). In the example of FIG. 7, the NBD calculator 214 sets the NBD parameter k = Gr/(100*A).

[00115] The example NBD calculator 214 scales the NBD parameter A to be consistent with the cleaned page views and/or video streams by calculating a parameter A' = A * (Xi/Vi) (block 720) and calculating an adjusted NBD parameter α = 1/A' (block 722).

[00116] If the Poisson condition is present (block 708), the NBD is treated as a Poisson distribution having one parameter (λ). Thus, the example NBD calculator 214 does not calculate the NBD parameters α and k and, instead, calculates a Poisson parameter λ = Gr/100 (block 724). When the NBD parameters α and k are calculated (block 722) or when the Poisson parameter is calculated (block 724), the example instructions 700 of FIG. 7 end and control returns to block 604 of FIG. 6.

[00117] FIG. 8 is a flowchart representative of example computer readable instructions 800 which, when executed, cause a processor to calculate a negative binomial distribution. The example instructions 800 of FIG. 8 may be performed by the example NBD calculator 214 of FIG. 2 to implement block 604 of FIG. 6. The example NBD calculations of FIG. 8 are performed for a demographic. Accordingly, the example NBD calculator 214 may iterate the instructions 800 for different demographic groups for a reporting period.

[00118] The example NBD calculator 214 of FIG. 2 determines whether the Poisson condition is present (block 802). For example, the NBD calculator 214 may determine which parameter(s) were calculated for the NBD calculation (e.g., λ if the Poisson condition is present, α and k if the Poisson condition is not present). If the Poisson condition is not present (block 802), the example NBD calculator 214 calculates the scaled reach for the demographic group (block 804). For example, the NBD calculator 214 of FIG. 2 calculates the scaled reach = $100*(1 - (\alpha/(\alpha + t))^k)$, where the time variable "t" is a unit of time such as a reporting period (e.g., t days). The time variable "t" may be set to 1 to scale unique audience members to match the cleaned page views and/or video streams of the illustrated example (e.g., daily page views, daily reporting, etc.).

[00119] The example NBD calculator 214 of FIG. 2 calculates the probability of zero exposures (e.g., page views) (block 806). For example, the probability of zero page

views, fp(0), may be determined according to $fp(0) = (\alpha/(\alpha + t))^k$, where the time variable "t" is set to 1 to scale unique audience members to match the cleaned page views and/or video streams. The example NBD calculator 214 of FIG. 2 calculates the probability of a number n of page views for $n \geq 1$ (block 808). For example, the probability of n page views, fp(n), may be determined according to $fp(n) = ((k + n - 1)/n)*((t/(\alpha + t))*fp(n-1)$, where the time variable "t" is set to 1 to scale unique audience members to match the cleaned page views and/or video streams. Thus, the probability of a number of page views fp(n) is based on probabilities of lower numbers of page views (fp(n-1), fp(n-x)).

[00120] The example NBD calculator 214 of FIG. 2 calculates an average number of page views (or average frequency) AveF (block 810). The average number of page views may be determined by $AveF = kt/\alpha$, where the time variable "t" is set to 1 to scale unique audience members to match the cleaned page views and/or video streams.

[00121] The example NBD calculator 214 calculates a number of panel sessions per person, representing gross rating points for a time "t" (GRP(t)), for the demographic group universe (block 812). The number of panel sessions per person may be determined by GRP(t) = t*Gr, where the time variable "t" is set to 1 to scale unique audience members to match the cleaned page views and/or video streams.

[00122] If the Poisson condition is present (block 802), the example NBD calculator 214 determines a frequency distribution for a time t, fp(i), where "i" is a number of page views (block 814). For example, the NBD calculator 214 may determine the frequency distribution fp(i) for a time t (e.g., the distribution of the numbers of page views) according to $fp(i) = (\lambda t)^i * e^{-\lambda t}/i!$ (where ! indicates the factorial operator).

[00123] The example NBD calculator 214 of FIG. 2 determines the schedule reach (expressed as a percentage) over t days (block 816). The scaled reach may be determined according to $reach = 100*(1- e^{-\lambda t})$. The example NBD calculator 214 of FIG. 2 determines a number of panel sessions per person (Gp(t)) (block 818). The number of panel sessions per person Gp(t) may be determined by Gp(t) = t*Gr.

[00124] After performing the NBD for the Poisson condition (blocks 814-818) or for no Poisson condition (blocks 804-812), the example instructions 800 end and control returns to block 414 of FIG. 4. The example instructions 800 may return the calculations determined from the NBD, such as the input variables (e.g., Xi, Vi, Yi,

Ui), the scaled reach for the demographic group(s), the probabilities of $n \geq 0$ page views for the demographic group(s), the average number(s) of page views for the demographic group(s), the panel sessions per person for the demographic group(s), and/or the frequency distribution(s) for one or more time periods for the demographic group(s).

[00125] FIG. 9 is a flowchart representative of example computer readable instructions 900 which, when executed, cause a processor to select panelists to represent an unmeasured location audience. The example instructions 900 may be executed by the example virtual panel generator 212 of FIG. 2 to implement block 414 of FIG. 4.

[00126] The example virtual panel generator 212 of FIG. 2 obtains demographic targets for an audience corresponding to unmeasured locations (block 902). The demographic targets may be received from the survey-based data collector 204 and provide an estimated demography of the unmeasured locations. The example sample weighter 302 obtains audience targets and volume targets (block 904). The example audience target is the difference between a unique audience reported for the measured locations (e.g., home and work environments) and an estimated unique audience determined by the NBD calculator 214 (e.g., home, work, and unmeasured environments or locations). The example volume target is the difference between a reported volume of page views (e.g., received from the census-based data collector 202) and/or video streams and a smoothed volume of page views and/or video streams, and discounted by the traffic from unmeasured locations. The virtual panel generator 212 of FIG. 2 assumes that volume metrics, such as page views and video streams, follow a Pareto distribution. Example instructions to generate the smoothed volume are described below with reference to FIG. 12. Given audience and volume targets of a specific entity, volume targets for sample selection may be generated by computing selected percentiles of the Pareto distribution. In some examples, the volume metrics are categorized into finite levels prior to calculating the computing the percentiles.

[00127] The example sample weighter 302 selects a panelist from the set of panelists for measured location(s) (block 906). The set of panelists may be obtained from the example panel data collector 206. The example sample weighter 302 assigns a weight to the selected panelist based on a conformance of the selected panelist's demographic

information and/or measured activities to target demographics and/or target behaviors of the unmeasured location population (block 908). The example sample weighter 302 determines whether there are additional panelists to be weighted (block 910). If there are additional panelists (block 906), control returns to block 906 to select another panelist.

[00128] When there are no additional panelists to be weighted (block 910), the example sample selector 304 converts the panelist weights to selection probabilities (block 912). For example, a higher panelist weight results in a higher probability of selection for the corresponding panelist. Based on the selection probabilities, the example sample selector 304 selects a number of the panelists (block 914). In some examples, the sample selector 304 selects the panelists randomly using the selection probabilities to determine the likelihood of randomly selecting any given panelist. The number of panelists selected may be predetermined (e.g., the selected sample size 308 of FIG. 3) based on a number of panelists to represent the unmeasured location audience and/or may be dynamically determined.

[00129] The sample re-weighter 306 re-weights the selected panelists (block 916). The re-weighting may be similar or identical to the weighting performed in block 908 and/or may be based on similar or identical factors. The re-weighting causes the selected panelists to more closely represent the demographic, audience, and/or volume targets of the unmeasured location audience. The example instructions 900 may then end and control returns to block 416 of FIG. 4.

[00130] FIGS. 10A and 10B show a flowchart representative of example computer readable instructions 1000 which, when executed, cause a processor to fulfill additional activities using a probability-based method. The example instructions 1000 may be executed to implement the example activity fulfiller 216 of FIG. 2 to perform block 418 of FIG. 4.

[00131] The example activity fulfiller 216 of FIG. 2 compares a panel-based volume (e.g., web page views, video streams) with a census-based volume to determine how much activity is needed to fill a gap in volume between the census and panel-based data (block 1002). The example panel-based volume is based on panel activity from measured locations and unmeasured locations (e.g., actual panelist data and virtual panel data). The example activity fulfiller 216 selects a URL pattern from a list of URL patterns (e.g., from a census-based list of URLs and/or URL patterns requested

from and/or provided by the servers 106, 108) (block 1004). In some examples, the activity fulfiller 216 and/or the server monitors 134 of FIGS. 1 and/or 2 aggregate instances of URLs into URL patterns.

[00132] The example activity fulfiller 216 selects a combination of day part and demographic group (block 1006). The example day part categories are illustrated below in Table 1. The example demographic groups (e.g., gender/age categories) are illustrated below in Table 2. Demographic groups may include additional and/or alternative distinctions. The example activity fulfiller 216 of FIG. 2 selects one of the example day part categories and one of the example demographic groups.

Day part	Definition
1	midnight-6am
2	6am-9am
3	9am-5pm
4	5pm-8pm
5	8pm-midnight

Table 1 – Day Part Categories

Gender/Age	Definition
1	Male 2-11
2	Female 2-11
3	Male 12-17
4	Female 12-17
5	Male 18-24
6	Female 18-24
7	Male 25-34
8	Female 25-34
9	Male 35-44
10	Female 35-44
11	Male 45-54
12	Female 45-54
13	Male 55+
14	Female 55+

Table 2 – Demographic Groups

[00133] The example activity fulfiller 216 of FIG. 2 calculates a weighted measured panel volume (e.g., for measured locations) for the selected group as variable A (block 1008). The example variable A may be determined by estimating the traffic during the day part and by the selected gender and age group from the panelist data. This data is weighted against the virtual panel data to be representative of the audience from measured locations, and can predict a volume of page views from the selected group from measured locations. The activity fulfiller 216 calculates the weighted virtual panel volume (e.g., for unmeasured locations) for the selected group as variable B (block 1010). The example variable B may be determined by estimating the traffic during the day part and by the selected gender group from the virtual panel data. The example virtual panel data is weighted against the panelist data to be representative of the audience from unmeasured location and can predict a volume of page views from the selected group from unmeasured locations. The activity fulfiller 216 calculates the total server page view volume for the selected group as variable C (block 1012). The example variable C may be an estimated portion of a total census-based page view volume during the selected day part that is attributable to the selected demographic group.

[00134] The example activity fulfiller 216 of FIG. 2 calculates an activity duplication probability for the selected day part, demographic group, and URL pattern (block 1014). The example activity duplication probability may be determined from the variables A, B, and C (e.g., determined from blocks 1008-1012) as (C-A-B)/A*100%.

[00135] The example activity fulfiller 216 determines whether there are additional categories to be processed (block 1016). If there are additional categories, control returns to block 1006 to select another day part and demographic group. When there are no additional categories for the selected URL pattern (block 1016), the example activity fulfiller 216 determine whether there are additional URL patterns to be processed (block 1018). If there are additional URL patterns (block 1018), control returns to block 1004 to select another URL pattern.

[00136] When there are no additional URL patterns (block 1018), the example audience classifier generates a probability lookup table specifying a duplication probability for each combination of URL pattern, day part category, and demographic group (block 1020).

[00137] Turning to FIG. 10B, the example activity fulfiller 216 of FIG. 2 selects a URL instance (e.g., a page view URL, a video stream URL, etc.) (block 1022). The URL instance may be selected from a table of panel and/or virtual panel activities. The activity fulfiller 216 retrieves the duplication probability from the lookup table for the URL pattern to which the URL instance belongs based on the day part and demographic group of the selected URL instance (block 1024). The activity fulfiller 216 determines whether the duplication probability is less than 1 (e.g., 100%) (block 1026). For example, the duplication probability may be less than 1 if panel activities constitute a large portion of the total server volume for the day part and demographic group.

[00138] If the duplication probability is 1 or more (block 1026), the example activity fulfiller 216 duplicates the selected instance and reduces the duplication probability of the URL pattern in the table by 1 (block 1028). Duplicating the selected instances results in the generation of an identical instance. If the duplication probability is less than 1 (block 1026), the example activity fulfiller 216 generates a random number having a uniform distribution between 0 and 1 (block 1030). If the generated number is less than or equal to the duplication probability (block 1032), the example activity fulfiller 216 duplicates the selected instance (block 1034).

[00139] If the activity fulfiller 216 duplicates the instance (block 1028 or block 1034), the example audience classifier duplicates the entire session of which the instance is a part (block 1036). Duplication of the session causes all of the instances in the session to be duplicated (without duplicating the selected instance twice). After duplicating the session (block 1036), or if the instance is not duplicated (block 1032), the example audience classifier determines whether there are additional instances in the table of panel and/or virtual panel activities (block 1038). If there are additional instances (block 1038), control returns to block 1022 to select another instance.

[00140] When there are no additional instances (block 1038), the example instructions 1000 of FIGS. 10A-10B may end and control returns to block 420 of FIG. 4. The example activity fulfiller 216 may return a listing of additional activity generated via the instructions 1000 and/or an updated list of activity including panel-based activity, virtual panel-based activity, and additional activity generated via the instructions 1000.

[00141] FIG. 11 is a flowchart representative of example computer readable instructions 1100 which, when executed, cause a processor to fulfill additional activities using a scaling-based method. The example scaling-based method of FIG. 11 differs from the probability-based method of FIGS. 10A-10B by computing a scaling factor for "missing" activities instead of adding more rows to the table by duplicating panel sessions. The example instructions 1100 may be executed to implement the example activity fulfiller 216 of FIG. 2 to perform block 418 of FIG. 4.

[00142] The example activity fulfiller 216 of FIG. 2 compares a panel-based volume (e.g., web page views, video streams) with a census-based volume to determine how much activity is needed to fill a gap in volume between the census and panel-based data (block 1102). The example panel-based volume is based on panel activity from measured locations and unmeasured locations (e.g., actual panelist data and virtual panel data). The example activity fulfiller 216 selects a URL pattern from a list of URL patterns (e.g., from a census-based list of URLs and/or URL patterns requested from and/or provided by the servers 106, 108) (block 1104). In some examples, the activity fulfiller 216 and/or the server monitors 134 of FIGS. 1 and/or 2 aggregate instances of URLs into URL patterns.

[00143] The example activity fulfiller 216 selects a combination of day part and demographic group (block 1106). The example day part categories are illustrated above in Table 1. The example demographic groups (e.g., gender/age categories) are illustrated above in FIG. 2. The example activity fulfiller 216 of FIG. 2 selects one of the example day part categories and one of the example demographic groups.

[00144] The example activity fulfiller 216 of FIG. 2 calculates a weighted measured panel volume (e.g., for measured locations) for the selected group as variable A (block 1108). The example variable A may be determined by estimating the traffic during the day part and by the selected gender and age group from the panelist data. This data is weighted against the virtual panel data to be representative of the audience from measured locations, and can predict a volume of page views from the selected group from measured locations. The activity fulfiller 216 calculates the weighted virtual panel volume (e.g., for unmeasured locations) for the selected group as variable B (block 1110). The example variable B may be determined by estimating the traffic during the day part and by the selected gender group from the virtual panel data. The example virtual panel data is weighted against the panelist data to be

representative of the audience from unmeasured location and can predict a volume of page views from the selected group from unmeasured locations. The activity fulfiller 216 calculates the total census page view volume for the selected group as variable C (block 1112). The example variable C may be an estimated portion of a total census-based page view volume during the selected day part that is attributable to the selected demographic group.

[00145] The example activity fulfiller 216 of FIG. 2 calculates a scaling factor for the selected day part, demographic group, and URL pattern (block 1114). The example activity duplication probability may be determined from the variables A, B, and C (e.g., determined from blocks 1108-1112) as (C-A-B)/A*100%.

[00146] The example activity fulfiller 216 determines whether there are additional categories to be processed (block 1116). If there are additional categories, control returns to block 1106 to select another day part and demographic group. When there are no additional categories for the selected URL pattern (block 1116), the example activity fulfiller 216 determine whether there are additional URL patterns to be processed (block 1118). If there are additional URL patterns (block 1118), control returns to block 1004 to select another URL pattern.

[00147] When there are no additional URL patterns (block 1118), the example audience classifier generates a probability lookup table specifying a duplication probability for each combination of URL pattern, day part category, and demographic group (block 1120).

[00148] Turning to FIG. 11B, the example activity fulfiller 216 of FIG. 2 generates a scaling factor lookup table including a scaling factor for each combination of URL pattern, day part category, and demographic group (block 1122). The example activity fulfiller 216 of FIG. 2 selects a URL instance (e.g., a page view URL, a video stream URL, etc.) (block 1124). Based on the URL pattern to which the selected instance belongs, the example activity fulfiller 216 retrieves the scaling factor from the lookup table for the day part and the demographic group of the URL instance (block 1126).

[00149] The activity fulfiller 216 applies the scaling factor to the instance (block 1128). For example, the activity fulfiller 216 may apply a scaling value to the instance to scale a number of page views or streams corresponding to the instance. The scaled value of the selected instance is not necessarily an integer. The example activity fulfiller 216 determines whether there are additional instances (block 1130). If there

are additional instances (block 1130), control returns to block 1124 to select another instance.

[00150] When there are no additional instances (block 1130), the instructions 1100 may end and control returns to block 420 of FIG. 4. The example activity fulfiller 216 may return a listing of additional activity generated via the instructions 1100 and/or an updated list of activity including panel-based activity, virtual panel-based activity, and additional activity generated via the instructions 1100.

[00151] FIG. 12 is a flowchart representative of example computer readable instructions which, when executed, cause a processor to calculate a smoothed volume metric. The example instructions 1200 of FIG. 12 may be executed as part of determining a volume target (e.g., block 904 of FIG. 9).

[00152] The example sample weighter 302 of FIG. 3 determines a number of census-measured page views for each combination of URL pattern, day part, and day of week (block 1202). For example, the server monitor 134 and/or the daemon 136 may provide data showing that a URL pattern had 1.2 million page views on Monday at 1:23:45 PM. The example sample weighter 302 calculates average census-measured page views for each combination of URL pattern, day part, day of week, and demographic group (block 1204). For example, the sample weighter 302 may determine from census data (e.g., via the census-based data collector 202) that average census page views for a URL pattern for the day part including 1:23:45 PM on Mondays during the previous 4 weeks is 800,000 page views.

[00153] The sample weighter 302 determines average panel-based page views for each combination of URL pattern, day part, day of week, and demographic group (block 1206). For example, a Men, age 18-24, demographic group may be determined from panel-based data for the day part including 1:23:45 PM on Mondays during the previous 4 weeks as 60,000 page views. The sample weighter 302 determines an unmeasured location factor for the selected demographic group (block 1208). For example, the unmeasured location factor for Men, age 18-24, for an example genre related to a URL pattern is 1.1.

[00154] The example sample weighter 302 calculates a smoothed volume metric for each combination of demographic group, day part, and day of week (block 1210). The sample weighter 302 determines the smoothed volume metric according to: smoothed volume metric = Fi*Pi*S/T. The example smoothed page views for Men, age 18-24,

for the day part of Monday including 1:23:45 PM is (1.1*60,000*1,200,000)/800,000. The example instructions 1200 may then end and control returns to block 906 of FIG. 9.

[00155] FIG. 13 is a flowchart representative of example computer readable instructions 1300 which, when executed, cause a processor to calculate inputs for an NBD calculation. The example instructions 1300 may be executed by the example NBD calculator 214 of FIG. 2 to perform block 702 of FIG. 7. To perform the instructions 1300, the NBD calculator 214 receives as inputs weighted panel activity (e.g., from the panel data collector 206 of FIG. 2), census page view and/or video stream measurements (e.g., from the census-based data collector 202 of FIG. 2), and a listing of URL patterns, channels, brands, and/or parents.

[00156] The NBD calculator 214 selects a URL pattern (e.g., from a list of URL patterns to be measured) (block 1302). The NBD calculator 214 computes a page view demographic distribution for the selected URL pattern by day of week and/or day part (block 1304). The page view demographic distribution determines a percentage of page views for the selected URL pattern per demographic group during the day of week and/or day part. The page view demographic distribution may be determined using the following equation: panel stream distribution % = (weighted panel-measured page views for demographic group for day part and/or day of week)/(weighted panel-measured page views for all demographic groups for the same day part and/or day of week).

[00157] FIG. 14 illustrates an example demographic distribution 1400 for an example selected URL pattern, day of week 1402, and day part 1404. The example demographic distribution may be generated for the URL pattern for the remaining combinations of day part and day of week. The example demographic distribution illustrates the page views 1406 for the selected combination of URL pattern, day of week 1402, and day part 1404, for each of a set of demographic groups 1408. The page views 1406 are determined from measurement of a weighted panel (e.g., collected by the panel data collector 206). A demographic distribution, or panel page view distribution 1410, percentage represents the percentage of page views from the corresponding ones of the demographic group 1408 for the corresponding day of week 1402 and day part 1404.

[00158] Returning to FIG. 13A, the NBD calculator 214 distributes the adjusted (e.g., cleaned) census page views into demographic groups by multiplying the adjusted census page views by the demographic distribution for each day of week and day part (block 1306). The adjusted census page views are obtained via the census-based data collector 202 of FIG. 2. The demographic distribution is obtained from the panel stream distributions 1410 determined in block 1304. FIG. 15 illustrates a distribution 1500 of a number of streams for the demographic groups, the day part, and the day of the week of FIG. 14. The example distribution 1500 of FIG. 15 includes the day of week 1402, the day part 1404, the demographic groups 1408, and the panel page view distribution 1410 of FIG. 14. The distribution 1500 further includes a number of adjusted census page views 1502 for the day of week 1402 and day part 1404. The example NBD calculator 214 determines calculated numbers of census-measured page views 1504 for each of the demographic groups 1408 for the day of week 1402 and the day part 1404 by multiplying the corresponding adjusted census page views 1502 by the corresponding panel page view distribution 1410.

[00159] Returning to FIG. 13A, the example NBD calculator 214 determines whether there are additional URL patterns (block 1308). If there are additional patterns (block 1308), control returns to block 1302 to select another URL pattern. When there are no additional URL patterns for which the adjusted census-measured page views are to be determined (block 1308), the example NBD calculator 214 selects a channel, brand, or parent (block 1310). The selection of a channel, brand, or parent may be based on an entity for which measurement is to be performed. The example NBD calculator 214 selects a demographic group (e.g., one of the demographic groups 1408 of FIGS. 14 and/or 15) (block 1312).

[00160] The example NBD calculator 214 aggregates URL patterns into the selected channel, brand, and/or parent for the selected demographic group (block 1314). For example, the NBD calculator 214 may sum the census-measured page views from multiple URL patterns belonging to the selected entity. The NBD calculator 214 aggregates page views from a full reporting period into day parts and/or days of the week for the selected group (block 1316). The reporting period may be a designated or requested reporting period, such as page views for a selected hour, day, week, month, or any other reporting period. FIG. 16 illustrates numbers of census-measured aggregated page views for an example one of the demographic groups 1408 of FIG.

14 for a reporting period. The example number of census-measured aggregated page views may represent the aggregation performed in blocks 1314 and 1316 of FIG. 13A.

[00161] Example numbers of aggregated page views 1602 are illustrated for each of the example days of week 1402 and day parts 1404 of FIGS. 14 and/or 15, as well as the days of the week and day parts not illustrated in FIGS. 14 and/or 15. In the example of FIG. 16, a first demographic group (e.g., group 1 of the 14 example groups of FIGS. 14 and/or 15) is credited with 21,305 page views for the selected channel, brand, and/or parent (e.g., the URL patterns belonging to the selected channel, brand, and/or parent) during a first day part of the third day of the week during a designated reporting period.

[00162] Returning to FIG. 13A, the example NBD calculator 214 aggregates weighted panel sessions for the selected group and selected channel, brand, and/or parent (block 1318). The example weighted panel sessions are obtained from the panel data collector 206 of FIG. 2. The sessions may be aggregated to represent sessions occurring during a selected reporting period, including all day parts and days of week, for the selected demographic group.

[00163] Turning to FIG. 13B, the example NBD calculator 214 computes a number of scaled census sessions (block 1320). For example, the NBD calculator 214 may compute the number of scaled census sessions using the ratio of calculated census page views and weighted panel page views, and the aggregated number of weighted panel sessions. The following equation may be used to compute scaled census sessions for a selected demographic group and channel, brand, and/or parent:

$$scaled\ server\ sessions = aggregated\ weighted\ panel\ sessions \times \frac{calculated\ server\ page\ views}{weighted\ panel\ page\ views}$$

[00164] Other methods to determine scaled census sessions may be used. The example equation uses the ratio of panel-based page views to census-based page views to determine a number of census sessions from a panel-based number of sessions. FIG. 17 illustrates example calculated scaled census sessions 1702 for the example demographic groups 1408 of FIGS. 14-16. The example scaled census sessions 1702 are calculated for the example demographic groups based on numbers of weighted panel page views 1704 (e.g., obtained from the panel-based data collector 206), a number of calculated census page views 1706 (e.g., from the calculated page

views 1602 of FIG. 16, calculated at block 1316 of FIG. 13A), and a number of weighted panel sessions 1708 (e.g., from the panel-based data collector 206). It should be noted that while the example total number of calculated census page views 1602 of FIG. 16 is different than the calculated census page views 1706 of FIG. 16 for demographic group 1 for illustration, the example calculated census page views 1706 may be obtained as the total number of calculated census page views 1602 determined from block 1316. As illustrated in FIG. 17, the scaled census sessions 1702 for demographic group 2 is calculated as 18,130,627 = 6,989,826*(72,732,430/28,040,233) (e.g., truncating the fraction).

[00165] Returning to FIG. 13B, the example NBD calculator 214 retrieves a panel combination audience (e.g., an audience for measured locations) for URL patterns that belong to the selected channel, brand, and/or parent and that match between the census-based data and the panel-based data (block 1322). A URL pattern is considered to match in the example of FIGS. 13A-13B if the URL pattern is both measured by the census (e.g., the server monitors 134 and/or the daemon 136) and visited by a member of the panel. The example panel combination audience, which is the total audience for all measured locations (e.g., the home and work environments 120, 122 of FIG. 1) and including overlap in audience, may be retrieved from the example panel-based data collector 206. The panel-based data collector 206 determines the panel combination audience by, for example, weighting panelist data to represent a population represented by the panel, and extrapolating the panel audience to identify an estimate of an audience from the population.

[00166] The example NBD calculator 214 determines whether there is a duplication factor for the selected brand, channel, and/or parent (block 1324). The duplication factor represents an overlap (e.g., a percentage, a ratio, etc.) in audience between the different measured locations (e.g., audience members measured at both home and work environments, etc., who are counted as multiple unique audience members). The duplication factor may be obtained from the example panel-based data collector 206. If there is no duplication factor available (block 1324), the example NBD calculator 214 uses the retrieved panel combination audience as a panel combination audience for the demographic group (block 1326).

[00167] If there is a duplication factor (block 1324), the example NBD calculator retrieves the duplication factor for the selected brand, channel, and/or parent (block

1328) and computes a weighted audience for the measured locations and the selected group (block 1330). The weighted audience calculated in block 1330 uses the audience for any URLs for the selected channel, brand, and/or parent regardless of whether the URLs match the census-based data. The NBD calculator calculates an adjusted combination audience for the selected channel, brand, and/or parent and the selected group (block 1332). An example equation to calculate the adjusted combination audience is shown below:

$$Adjusted\ Combination\ Audience = (audience\ for\ measured\ locations) * (1 - duplication) * \frac{matched\ combo\ audience\ (block\ 1322)}{weighted\ audience\ (block\ 1330)}$$

where the audience for measured locations includes the audience determined via the panel for each measured location (e.g., home environment and work environment). This rate is adjusted by the duplication factor (e.g., multiplied by (1-duplication factor)) to remove overlapping audience members among the measured locations. The adjusted combination audience is the estimated number of persons in the selected demographic group having at least one page view for URLs of the selected channel, brand, and/or parent, and is used in the calculation of NBD parameters as the parameter Ui.

[00168] The example NBD calculator 214 determines whether there are additional demographic groups (block 1334). If there are additional demographic groups (block 1334), control returns to block 1312 of FIG. 13A. If there are no additional demographic groups for calculation for the selected channel, brand, and/or parent (block 1334), the example NBD calculator 214 determines whether there are additional channels, brands, and/or parents for which an NBD is to be calculated (block 1336). If there are additional channels, brands, and/or parents (block 1336), control returns to block 1310 of FIG. 13A. When there are no additional channels, brands, or parents (block 1336), the example instructions 1300 end and control returns to block 704 of FIG. 7. The example instructions 1300 return the NBD inputs for scaled census sessions (Xi) and adjusted combination audience (Ui) to the NBD calculation. The instructions 1300 may further return the weighted panel sessions (Vi) if the weighted panel sessions are not provided independently.

[00169] FIG. 18 is a block diagram of an example processing platform 1800 capable of executing the instructions of FIGS. 4-13B to implement the hybrid online audience

measurement system 102 and/or the virtual panel generator 212 of FIGS. 1-3. The processing platform 1800 can be, for example, a server, a personal computer, and/or any other type of computing device.

[00170] The system 1800 of the instant example includes a processor 1812. For example, the processor 1812 can be implemented by one or more microprocessors or controllers from any desired family or manufacturer.

[00171] The processor 1812 includes a local memory 1813 (e.g., a cache) and is in communication with a main memory including a volatile memory 1814 and a non-volatile memory 1816 via a bus 1818. The volatile memory 1814 may be implemented by Synchronous Dynamic Random Access Memory (SDRAM), Dynamic Random Access Memory (DRAM), RAMBUS Dynamic Random Access Memory (RDRAM) and/or any other type of random access memory device. The non-volatile memory 1816 may be implemented by flash memory and/or any other desired type of memory device. Access to the main memory 1814, 1816 is controlled by a memory controller.

[00172] The processing platform 1800 also includes an interface circuit 1820. The interface circuit 1820 may be implemented by any type of interface standard, such as an Ethernet interface, a universal serial bus (USB), and/or a PCI express interface.

[00173] One or more input devices 1822 are connected to the interface circuit 1820. The input device(s) 1822 permit a user to enter data and commands into the processor 1812. The input device(s) can be implemented by, for example, a keyboard, a mouse, a touchscreen, a track-pad, a trackball, isopoint and/or a voice recognition system.

[00174] One or more output devices 1824 are also connected to the interface circuit 1820. The output devices 1824 can be implemented, for example, by display devices (e.g., a liquid crystal display, a cathode ray tube display (CRT), a printer and/or speakers). The interface circuit 1820, thus, typically includes a graphics driver card.

[00175] The interface circuit 1820 also includes a communication device such as a modem or network interface card to facilitate exchange of data with external computers via a network 1826 (e.g., an Ethernet connection, a digital subscriber line (DSL), a telephone line, coaxial cable, a cellular telephone system, etc.).

[00176] The processing platform 1800 also includes one or more mass storage devices 1828 for storing software and data. Examples of such mass storage devices 1828

include floppy disk drives, hard drive disks, compact disk drives and digital versatile disk (DVD) drives.

[00177] The coded instructions 1832 of FIGS. 4-13B may be stored in the mass storage device 1828, in the volatile memory 1814, in the non-volatile memory 1816, and/or on a removable storage medium such as a CD or DVD.

[00178] Although certain example systems, methods, apparatus and articles of manufacture have been described herein, the scope of coverage of this patent is not limited thereto. On the contrary, this patent covers all systems, methods, apparatus and articles of manufacture fairly falling within the scope of the claims of this patent.

What is claimed is:

1. A method to classify audience data, comprising:

obtaining panelist data corresponding to a plurality of panelists accessing web pages at measured locations;

classifying the panelist data according to demographic information of the panelists;

generating a virtual panel based on an estimate of web page access at unmeasured locations; and

classifying page view data associated with the unmeasured locations based on data corresponding to the virtual panel.

2. A method as defined in claim 1, wherein the virtual panel comprises a subset of the plurality of panelists.

3. A method as defined in claim 1, wherein generating the virtual panel comprises estimating a total audience.

4. A method as defined in claim 1, wherein generating the virtual panel is based on a census-based count of page views, the panelist data, and an estimated universe.

5. A method as defined in claim 4, wherein generating the virtual panel comprises estimating an audience at the unmeasured locations using a negative binomial distribution.

6. A method as defined in claim 1, wherein generating the virtual panel is based on at least one of a demographic target or an activity target.

7. A method as defined in claim 1, wherein classifying the panelist data comprises estimating a first portion of an online audience based on the panelist data.

8. A method as defined in claim 7, wherein classifying the page view data comprises estimating a second portion of the online audience, the first portion and the second portion comprising numbers of unique audience members.

9. A method as defined in claim 1, further comprising:

determining a difference between a census-based count of page views and a number of page views associated with the first and second portions of the page view data; and

estimating additional audience activity not represented by the plurality of panelists and the virtual panel based on the difference.

10. A method to generate a virtual panel, comprising:

assigning weights to a plurality of panelists based on first estimated characteristics;

selecting a subset of the panelists based on the weights;

re-weighting the selected panelists based on second estimated characteristics; and

generating a virtual panel including the selected panelists as virtual panelists, wherein data collected from the selected panelists for the virtual panel are to be assigned weights based on the re-weighting.

11. A method as defined in claim 10, further comprising estimating an audience, the first estimated characteristics of the panelists being based on the estimated audience.

12. A method as defined in claim 11, wherein the first estimated characteristics are based on a difference between the estimated audience and a measured audience.

13. A method as defined in claim 10, wherein selecting the subset of the panelists comprises randomly selecting the panelists using the weights assigned to corresponding panelists as selection probabilities for the corresponding panelist.

14. A method as defined in claim 10, further comprising calculating the first estimated characteristics based on at least one of a demographic target or an activity target.

15. A method as defined in claim 14, wherein calculating the first estimated characteristics comprises comparing online activities of the panelists to census-based page view data.

16. An apparatus, comprising:

a panel data collector to collect panelist data corresponding to a plurality of panelists accessing web pages at measured locations;

a virtual panel generator to generate a virtual panel based on an estimate of web page access at unmeasured locations; and

an audience classifier to classify the panelist data according to demographic information of the panelists and to classify page view data associated with the unmeasured locations based on data corresponding to the virtual panel.

17. An apparatus as defined in claim 16, further comprising a negative binomial distribution calculator to estimate an audience at unmeasured locations based on the plurality of panelists and a universe estimate.

18. An apparatus as defined in claim 16, further comprising an activity fulfiller to determine online activity based on the plurality of panelists, the audience classifier to classify a third portion of the page view data based on the online activity.

19. An apparatus as defined in claim 16, further comprising a census-based data collector to receive the page view data representing a number of page views of a web page from measured and unmeasured locations.

20. An apparatus as defined in claim 16, further comprising a survey-based data collector to collect a universe estimate, the virtual panel generator to generate the virtual panel based on the universe estimate.

21. An apparatus, comprising:

a sample weighter to assign weights to a plurality of panelists based on first estimated characteristics;

a sample selector to select a subset of the panelists based on the weights; and

a sample re-weighter to re-weight the selected panelists based on second estimated characteristics, and to generate a virtual panel including the selected panelists as virtual panelists, wherein data collected from the selected panelists for the virtual panel are to be assigned weights based on the re-weighting.

22. An apparatus as defined in claim 21, wherein the sample weighter is to:

obtain an estimated audience; and

calculate the first estimated characteristics of the panelists based on the estimated audience.

23. An apparatus as defined in claim 22, wherein the sample weighter is to calculate the first estimated characteristics based on a difference between the estimated audience and a measured audience.

24. An apparatus as defined in claim 21, wherein the sample selector is to select the panelists in the subset by randomly selecting the panelists using the weights assigned to corresponding panelists as selection probabilities for the corresponding panelists.

25. An apparatus as defined in claim 21, wherein the sample weighter is to calculate the first estimated characteristics based on at least one of a demographic target or an activity target.

26. An apparatus as defined in claim 25, wherein the sample weighter is to calculate the first estimated characteristics by comparing online activities of the panelists to measured page view data.

27. A tangible computer readable storage medium comprising machine readable instructions which, when executed, cause a processor to at least:

obtain panelist data corresponding to a plurality of panelists accessing web pages at measured locations;

classifying the panelist data according to demographic information of the panelists;

generating a virtual panel based on an estimate of web page access at unmeasured locations; and

classifying page view data associated with the unmeasured locations based on data corresponding to the virtual panel.

28. A storage medium as defined in claim 27, wherein the virtual panel comprises a subset of the plurality of panelists.

29. A storage medium as defined in claim 27, wherein the instructions are to cause the processor to generate the virtual panel by estimating a total audience.

30. A storage medium as defined in claim 27, wherein the instructions are to cause the processor to generate the virtual panel based on a census-based count of page views, the panelist data, and an estimated universe.

31. A storage medium as defined in claim 27, wherein the instructions are to cause the processor to generate the virtual panel by estimating an audience at the unmeasured locations using a negative binomial distribution.

32. A storage medium as defined in claim 27, wherein the instructions are to cause the processor to generate the virtual panel based on at least one of a demographic target or an activity target.

33. A storage medium as defined in claim 27, wherein the instructions are to cause the processor to classify the panelist data by estimating a first portion of an online audience based on the panelist data.

34. A storage medium as defined in claim 33, wherein the instructions are to cause the processor to classifying the page view data by estimating a second portion of the online audience, the first portion and the second portion comprising numbers of unique audience members.

35. A storage medium as defined in claim 27, wherein the instructions are further to cause the processor to:

determine a difference between a census-based count of page views and a number of page views associated with the first and second portions of the page view data; and

estimate additional audience activity not represented by the plurality of panelists and the virtual panel based on the difference.

36. A computer readable storage medium comprising machine readable instructions which, when executed, cause a processor to:

assign weights to a plurality of panelists based on first estimated characteristics;

select a subset of the panelists based on the weights;

re-weight the selected panelists based on second estimated characteristics; and

generate a virtual panel including the selected panelists as virtual panelists, wherein data collected from the selected panelists for the virtual panel are to be assigned weights based on the re-weighting.

37. A computer readable storage medium as defined in claim 36, wherein the instructions are further to cause the processor to estimate an audience, the first estimated characteristics of the panelists being based on the estimated audience.

38. A computer readable storage medium as defined in claim 37, wherein the first estimated characteristics are based on a difference between the estimated audience and a measured audience.

39. A computer readable storage medium as defined in claim 36, wherein the instructions are to cause the processor to select the subset of the panelists by randomly selecting the panelists using the weights assigned to corresponding panelists as selection probabilities for the corresponding panelist.

40. A computer readable storage medium as defined in claim 36, wherein the instructions are further to cause the processor to calculate the first estimated characteristics based on at least one of a demographic target or an activity target.

41. A computer readable storage medium as defined in claim 40, wherein the instructions are to cause the processor to calculate the first estimated characteristics by comparing online activities of the panelists to census-based page view data.

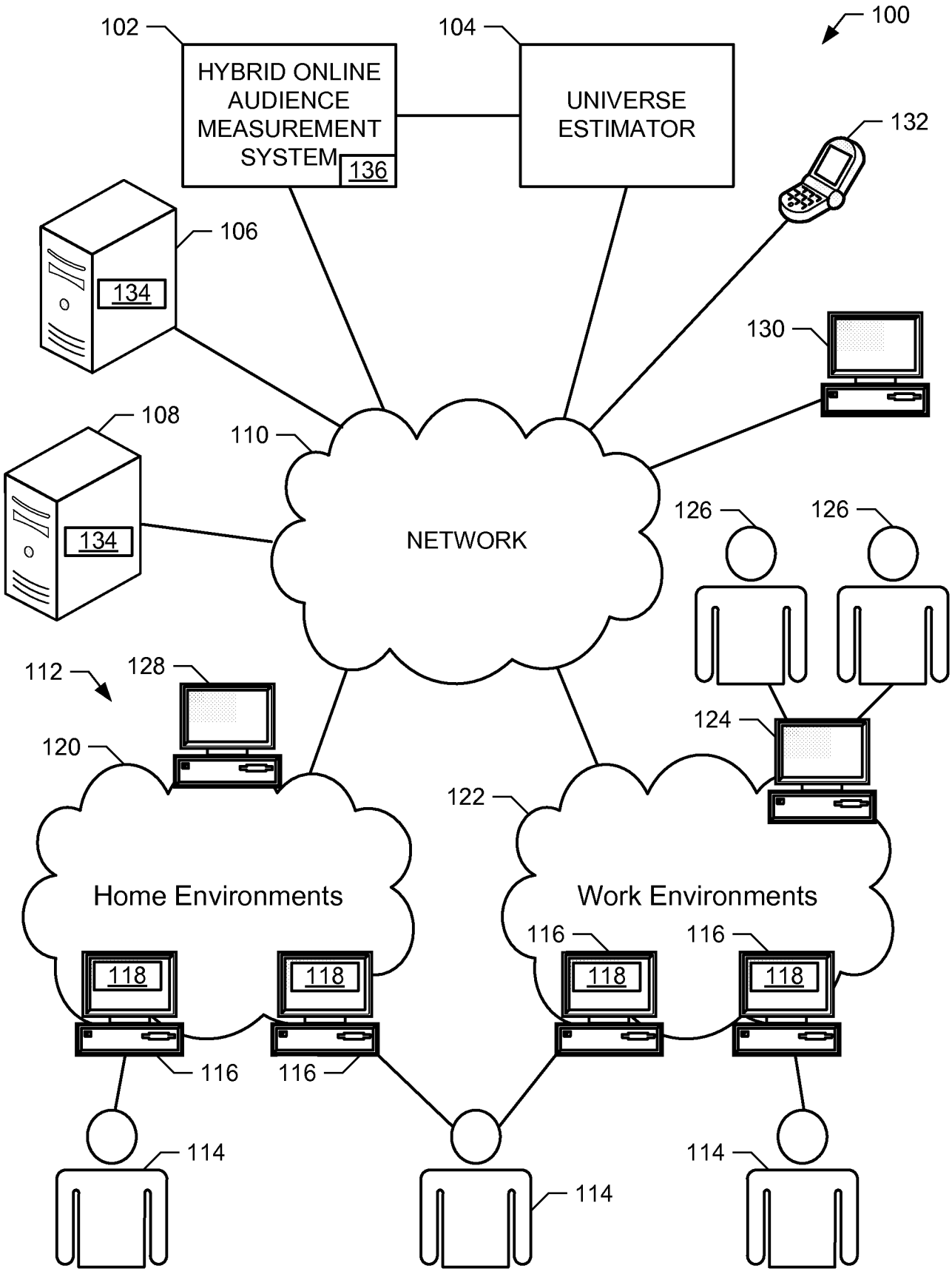

100
102
HYBRID ONLINE AUDIENCE MEASUREMENT SYSTEM
136
104
UNIVERSE ESTIMATOR
132
106
134
130
108
110
134
NETWORK
126
126
112
128
120
124
122
Home Environments
Work Environments
116
116
118
118
118
118
116
116
114
114
114

FIG. 1

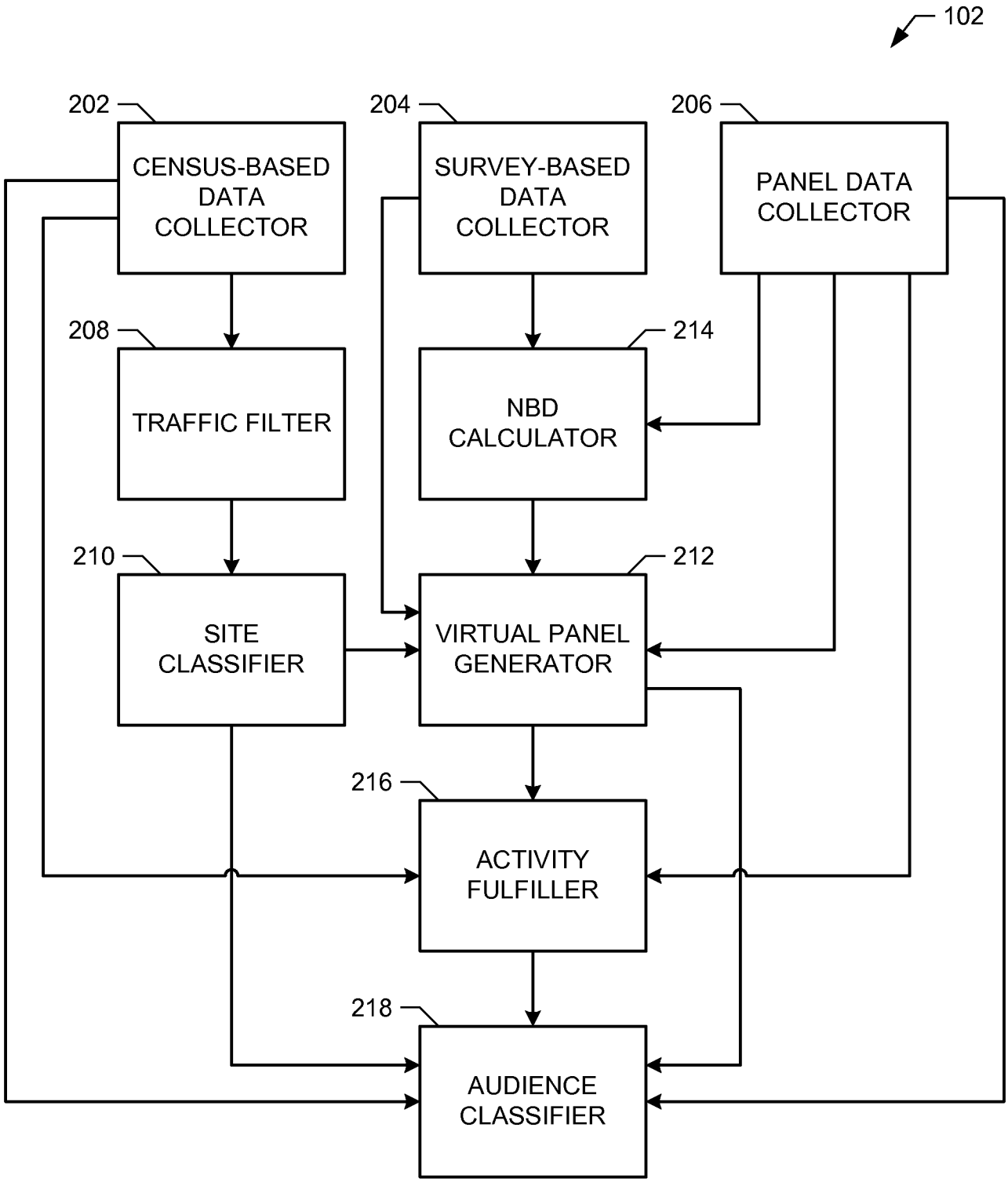

102
202
CENSUS-BASED DATA COLLECTOR
204
SURVEY-BASED DATA COLLECTOR
206
PANEL DATA COLLECTOR
208
TRAFFIC FILTER
214
NBD CALCULATOR
210
SITE CLASSIFIER
212
VIRTUAL PANEL GENERATOR
216
ACTIVITY FULFILLER
218
AUDIENCE CLASSIFIER

FIG. 2

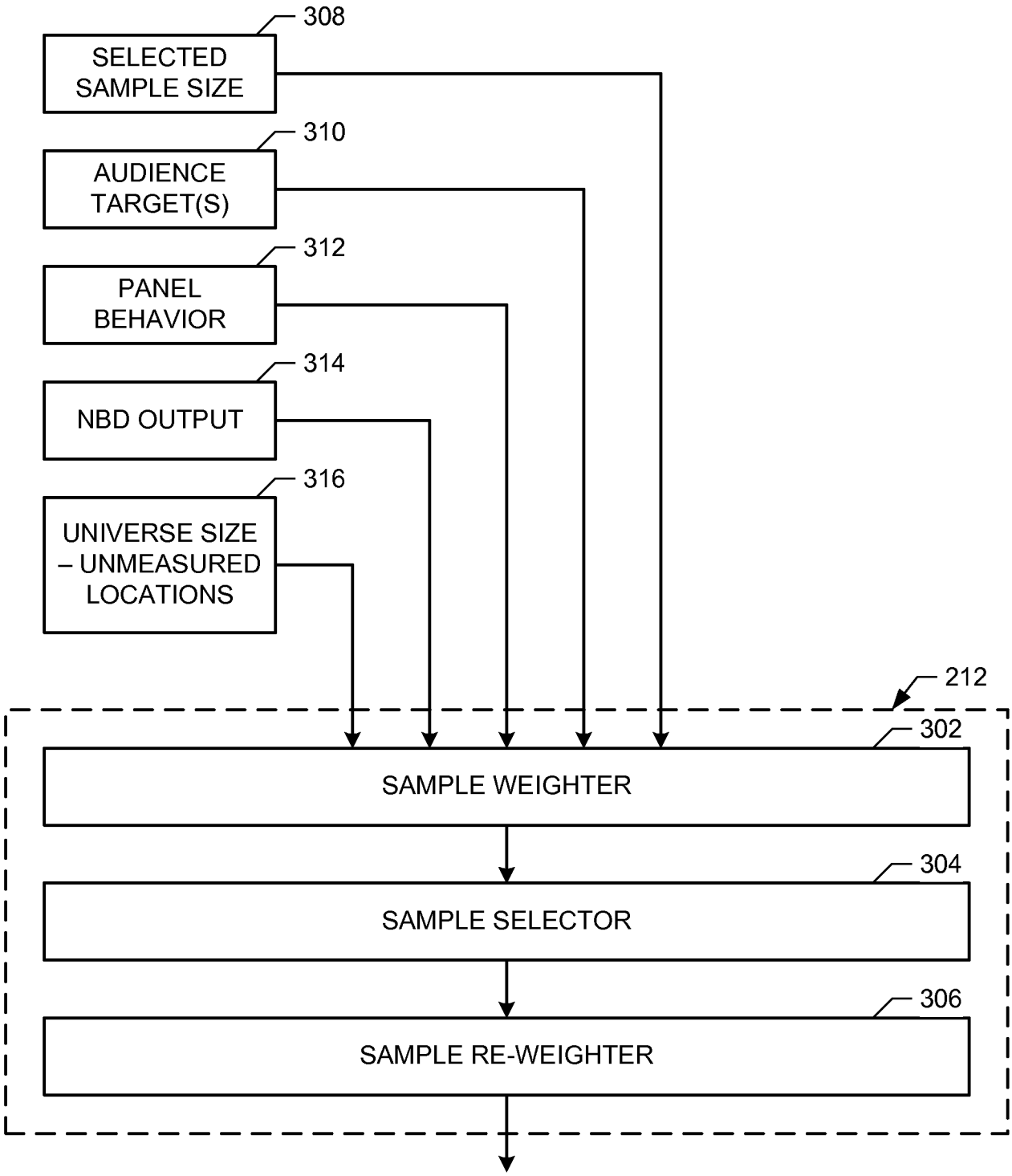

308
SELECTED SAMPLE SIZE
310
AUDIENCE TARGET(S)
312
PANEL BEHAVIOR
314
NBD OUTPUT
316
UNIVERSE SIZE – UNMEASURED LOCATIONS
212
302
SAMPLE WEIGHTER
304
SAMPLE SELECTOR
306
SAMPLE RE-WEIGHTER

FIG. 3

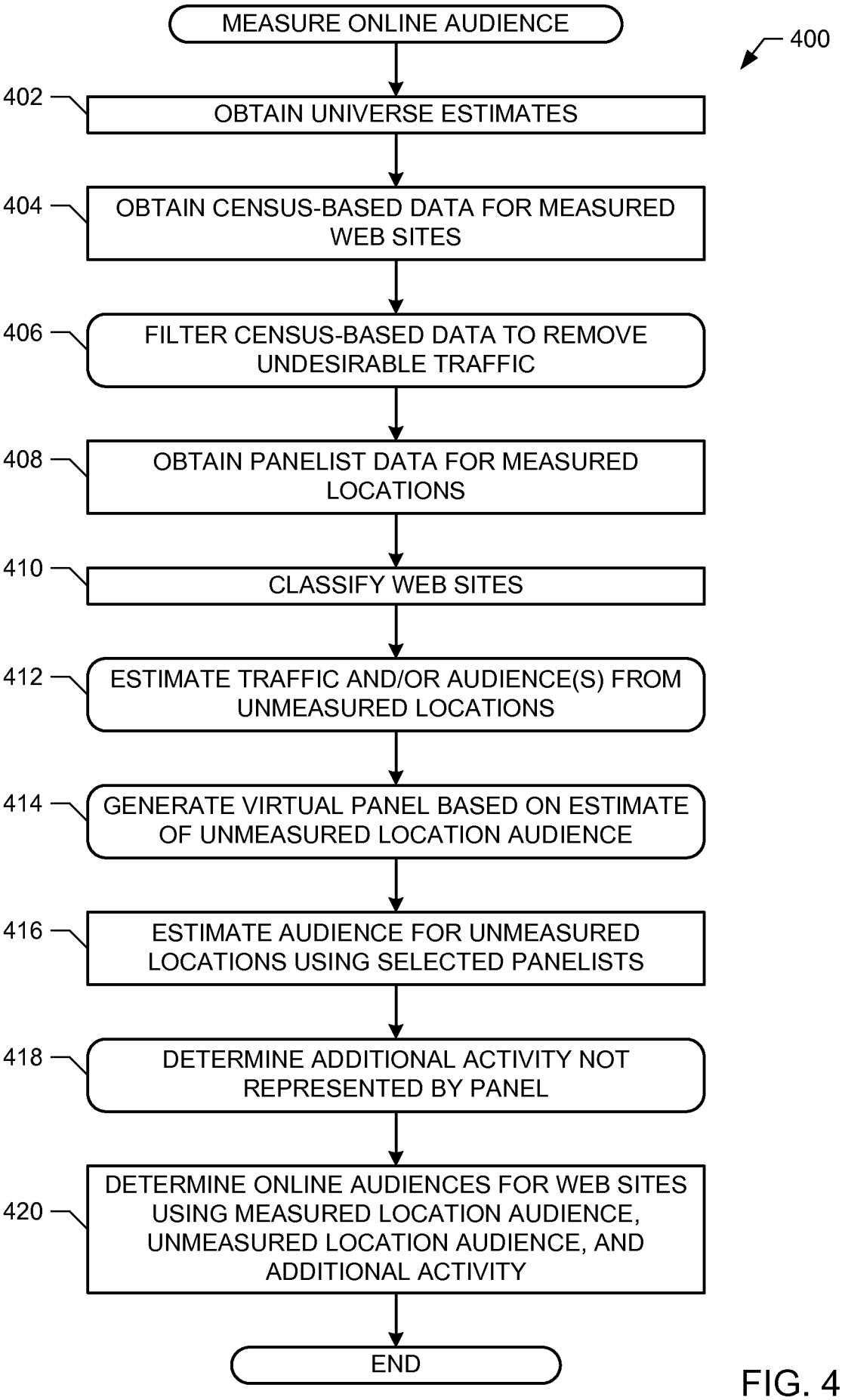

MEASURE ONLINE AUDIENCE
400
402
OBTAIN UNIVERSE ESTIMATES
404
OBTAIN CENSUS-BASED DATA FOR MEASURED WEB SITES
406
FILTER CENSUS-BASED DATA TO REMOVE UNDESIRABLE TRAFFIC
408
OBTAIN PANELIST DATA FOR MEASURED LOCATIONS
410
CLASSIFY WEB SITES
412
ESTIMATE TRAFFIC AND/OR AUDIENCE(S) FROM UNMEASURED LOCATIONS
414
GENERATE VIRTUAL PANEL BASED ON ESTIMATE OF UNMEASURED LOCATION AUDIENCE
416
ESTIMATE AUDIENCE FOR UNMEASURED LOCATIONS USING SELECTED PANELISTS
418
DETERMINE ADDITIONAL ACTIVITY NOT REPRESENTED BY PANEL
420
DETERMINE ONLINE AUDIENCES FOR WEB SITES USING MEASURED LOCATION AUDIENCE, UNMEASURED LOCATION AUDIENCE, AND ADDITIONAL ACTIVITY
END

FIG. 4

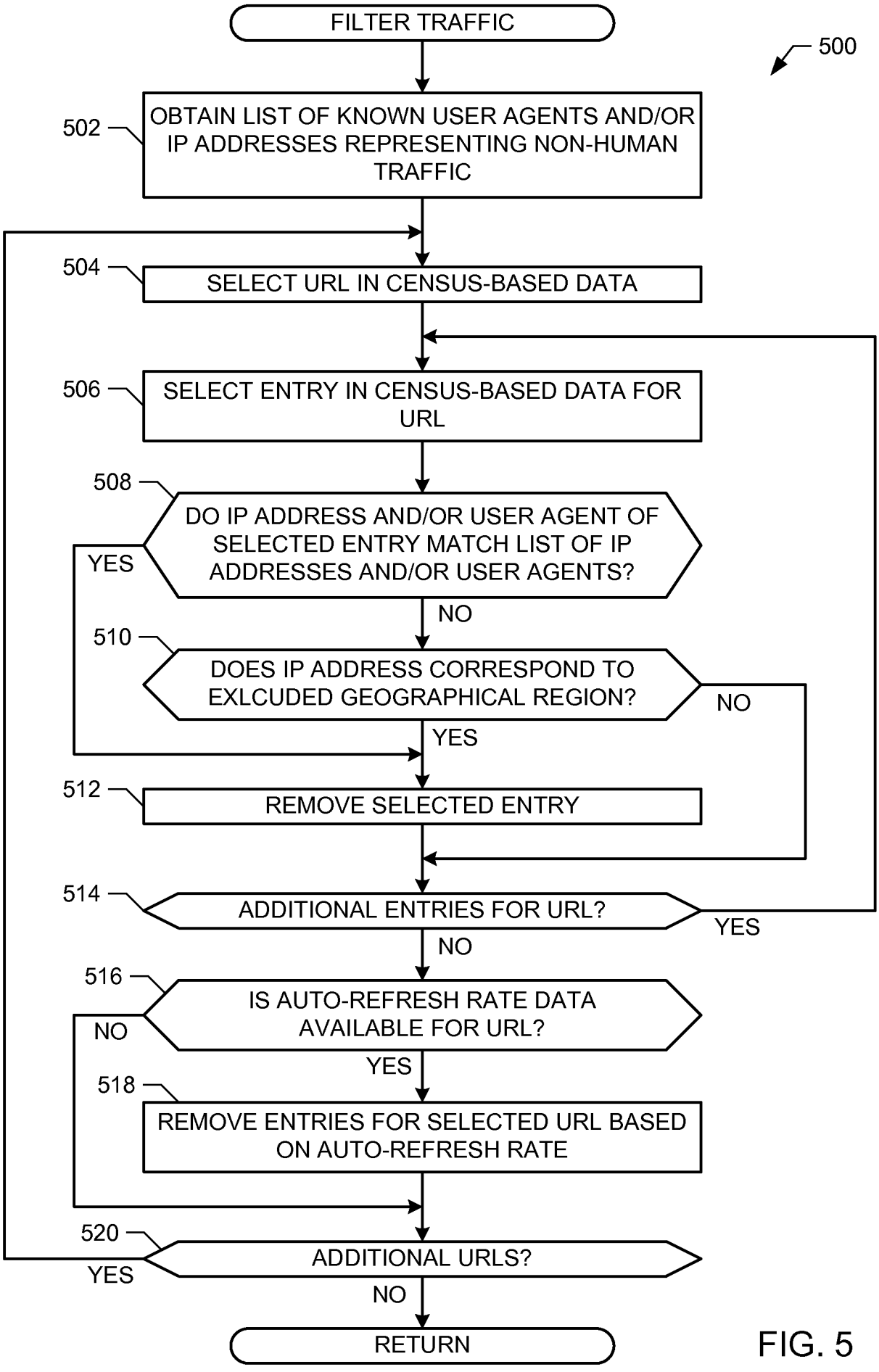

FILTER TRAFFIC
500
502
OBTAIN LIST OF KNOWN USER AGENTS AND/OR IP ADDRESSES REPRESENTING NON-HUMAN TRAFFIC
504
SELECT URL IN CENSUS-BASED DATA
506
SELECT ENTRY IN CENSUS-BASED DATA FOR URL
508
DO IP ADDRESS AND/OR USER AGENT OF SELECTED ENTRY MATCH LIST OF IP ADDRESSES AND/OR USER AGENTS?
YES
NO
510
DOES IP ADDRESS CORRESPOND TO EXLCUDED GEOGRAPHICAL REGION?
NO
YES
512
REMOVE SELECTED ENTRY
514
ADDITIONAL ENTRIES FOR URL?
YES
NO
516
IS AUTO-REFRESH RATE DATA AVAILABLE FOR URL?
NO
YES
518
REMOVE ENTRIES FOR SELECTED URL BASED ON AUTO-REFRESH RATE
520
ADDITIONAL URLS?
YES
NO
RETURN

FIG. 5

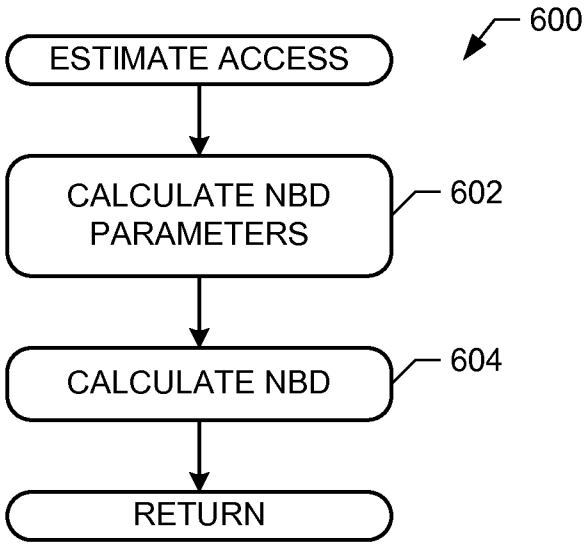

600
ESTIMATE ACCESS
CALCULATE NBD PARAMETERS
602
CALCULATE NBD
604
RETURN

FIG. 6

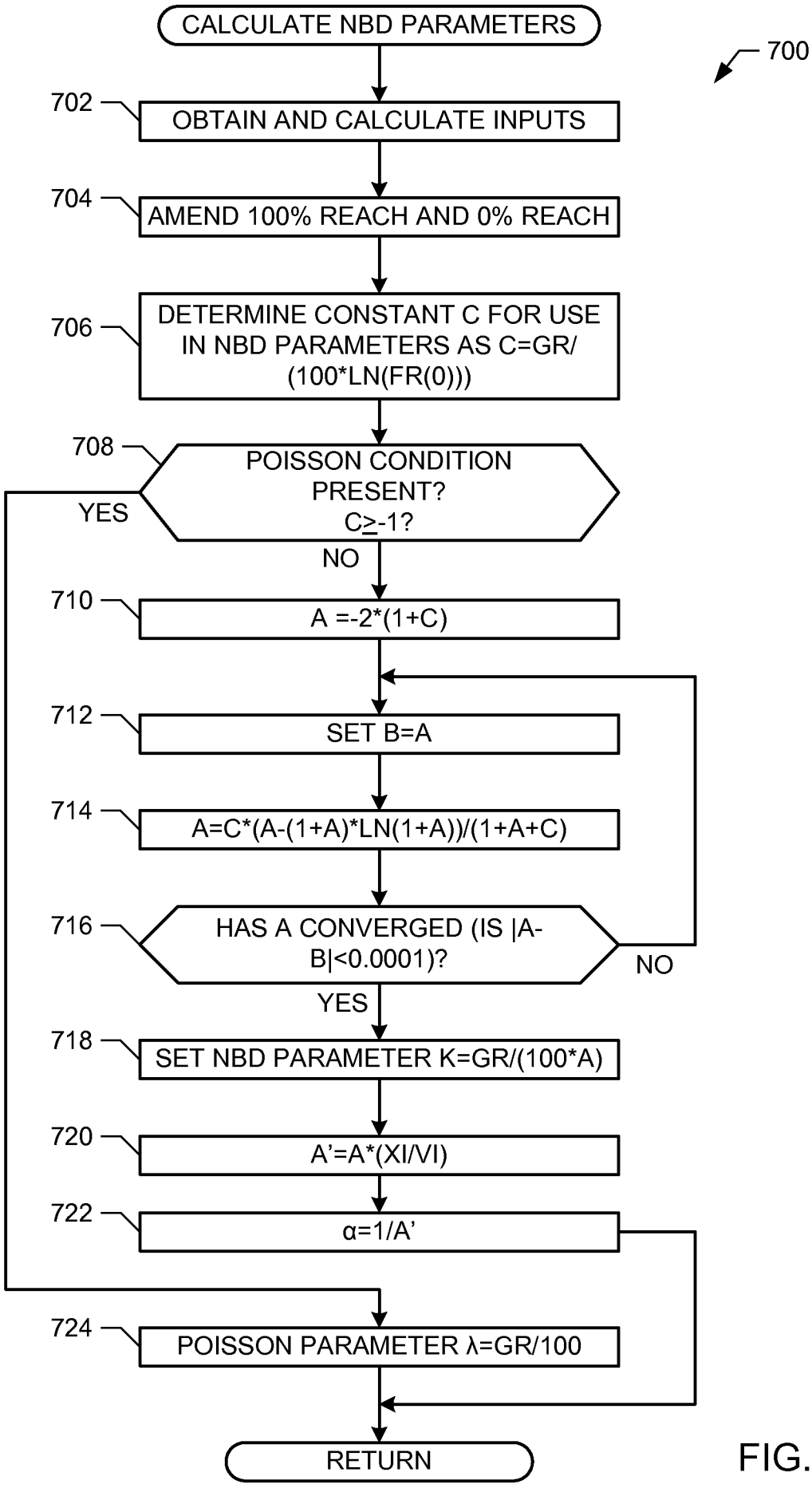

CALCULATE NBD PARAMETERS
700
702
OBTAIN AND CALCULATE INPUTS
704
AMEND 100% REACH AND 0% REACH
706
DETERMINE CONSTANT C FOR USE IN NBD PARAMETERS AS C=GR/(100*LN(FR(0)))
708
POISSON CONDITION PRESENT? C≥-1?
YES
NO
710
A =-2*(1+C)
712
SET B=A
714
A=C*(A-(1+A)*LN(1+A))/(1+A+C)
716
HAS A CONVERGED (IS |A-B|<0.0001)?
NO
YES
718
SET NBD PARAMETER K=GR/(100*A)
720
A'=A*(XI/VI)
722
α=1/A'
724
POISSON PARAMETER λ=GR/100
RETURN

FIG. 7

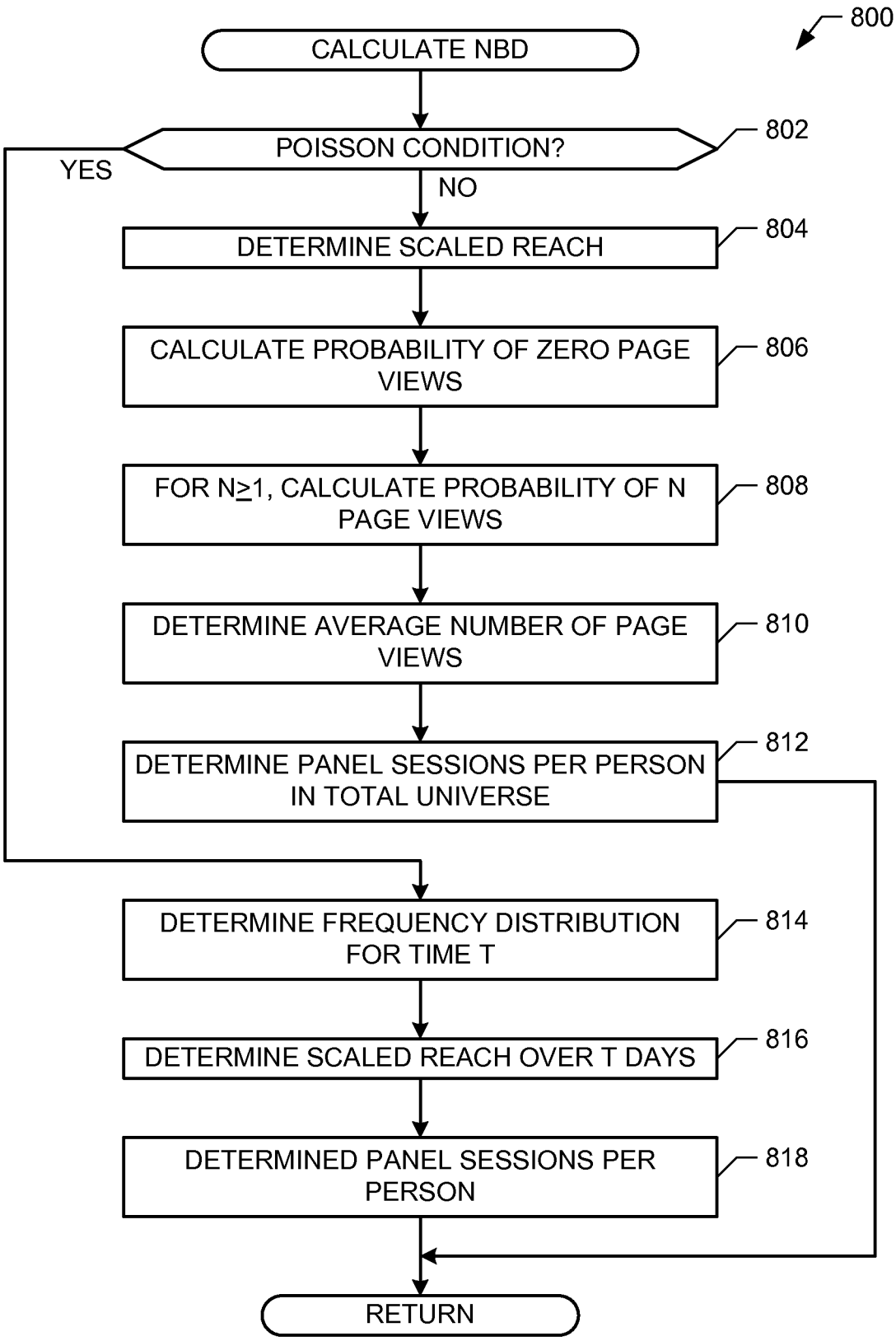

800
CALCULATE NBD
POISSON CONDITION?
802
YES
NO
DETERMINE SCALED REACH
804
CALCULATE PROBABILITY OF ZERO PAGE VIEWS
806
FOR N≥1, CALCULATE PROBABILITY OF N PAGE VIEWS
808
DETERMINE AVERAGE NUMBER OF PAGE VIEWS
810
DETERMINE PANEL SESSIONS PER PERSON IN TOTAL UNIVERSE
812
DETERMINE FREQUENCY DISTRIBUTION FOR TIME T
814
DETERMINE SCALED REACH OVER T DAYS
816
DETERMINED PANEL SESSIONS PER PERSON
818
RETURN

FIG. 8

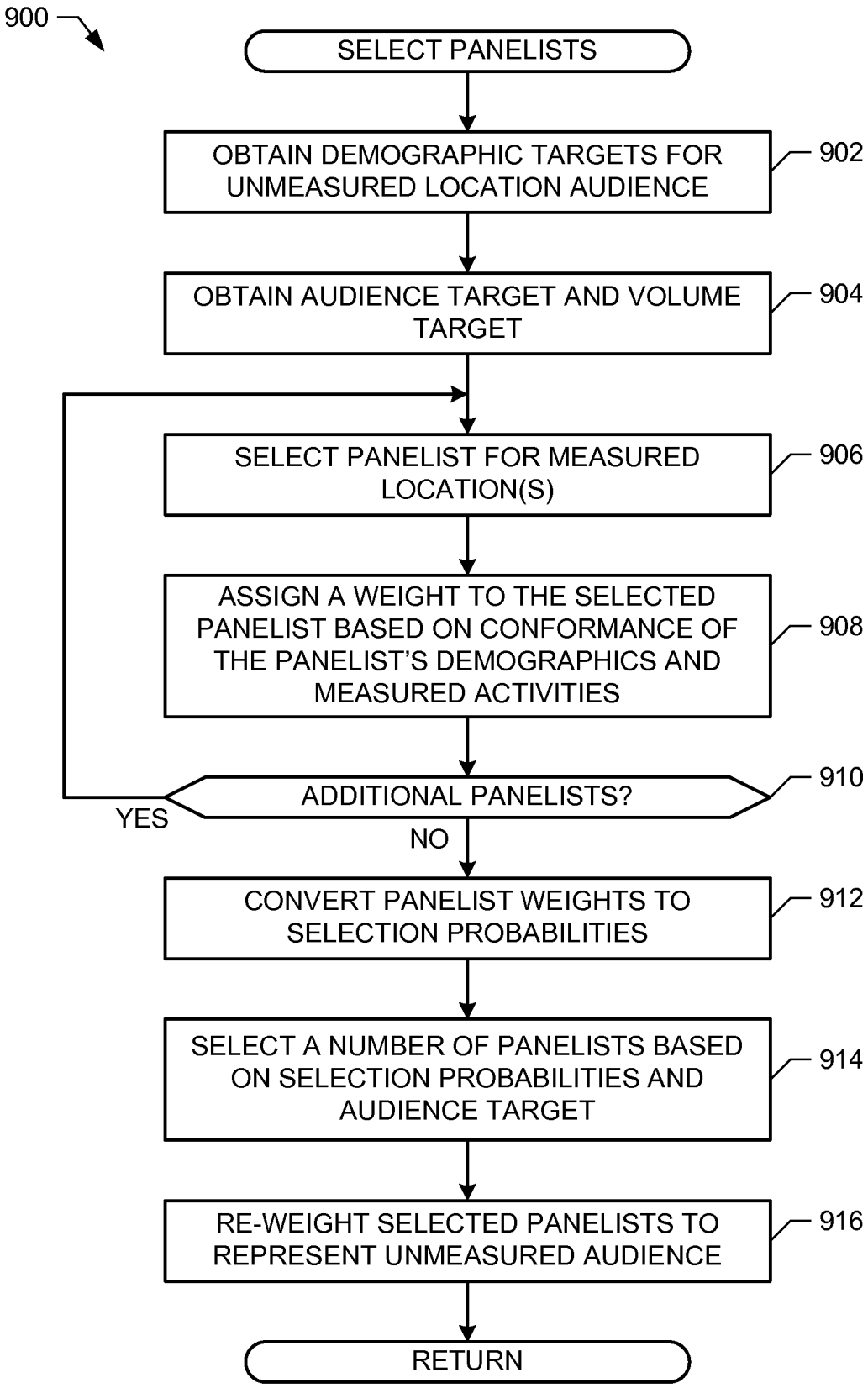

900
SELECT PANELISTS
OBTAIN DEMOGRAPHIC TARGETS FOR UNMEASURED LOCATION AUDIENCE
902
OBTAIN AUDIENCE TARGET AND VOLUME TARGET
904
SELECT PANELIST FOR MEASURED LOCATION(S)
906
ASSIGN A WEIGHT TO THE SELECTED PANELIST BASED ON CONFORMANCE OF THE PANELIST'S DEMOGRAPHICS AND MEASURED ACTIVITIES
908
ADDITIONAL PANELISTS?
910
YES
NO
CONVERT PANELIST WEIGHTS TO SELECTION PROBABILITIES
912
SELECT A NUMBER OF PANELISTS BASED ON SELECTION PROBABILITIES AND AUDIENCE TARGET
914
RE-WEIGHT SELECTED PANELISTS TO REPRESENT UNMEASURED AUDIENCE
916
RETURN

FIG. 9

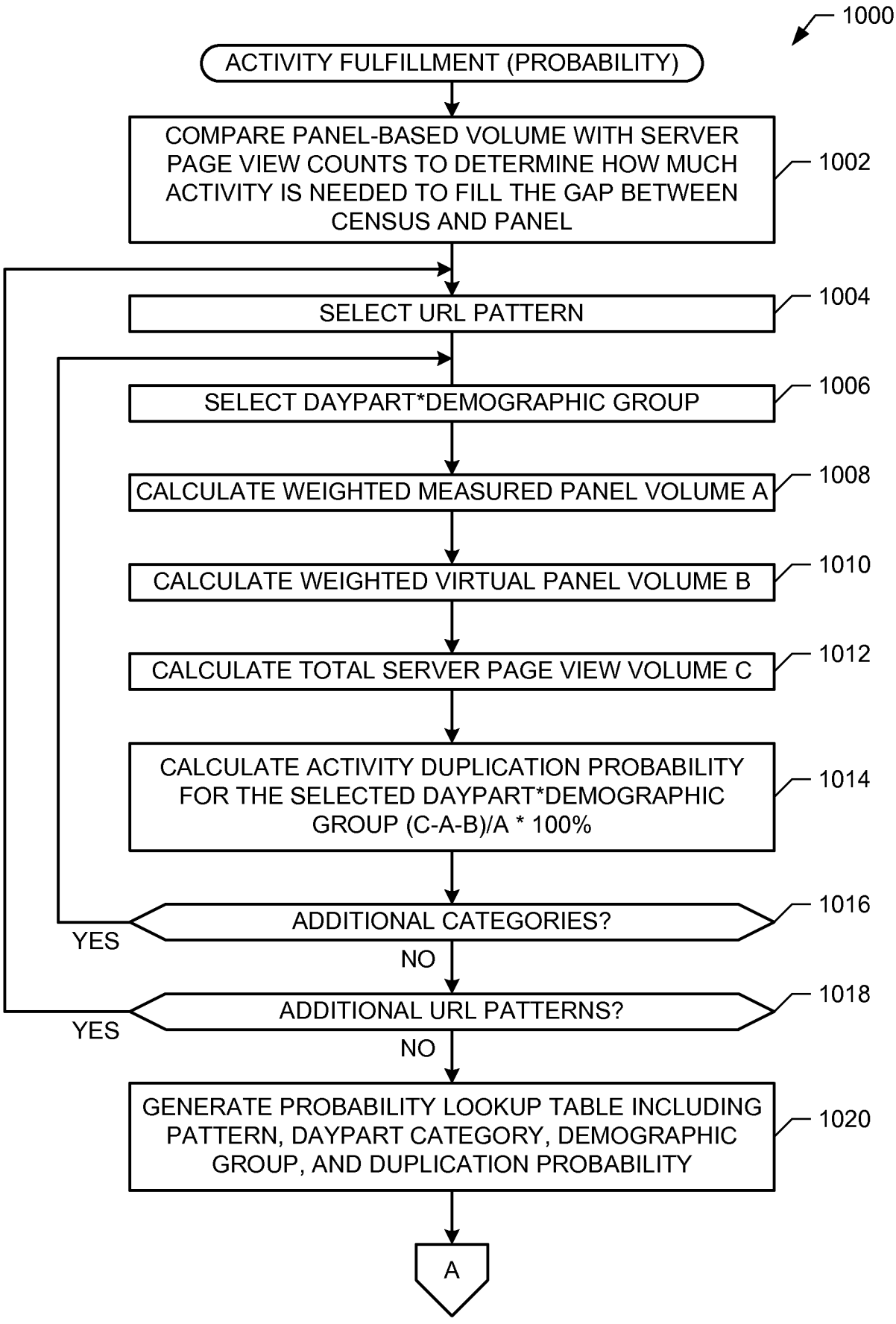

1000
ACTIVITY FULFILLMENT (PROBABILITY)
COMPARE PANEL-BASED VOLUME WITH SERVER PAGE VIEW COUNTS TO DETERMINE HOW MUCH ACTIVITY IS NEEDED TO FILL THE GAP BETWEEN CENSUS AND PANEL
1002
SELECT URL PATTERN
1004
SELECT DAYPART*DEMOGRAPHIC GROUP
1006
CALCULATE WEIGHTED MEASURED PANEL VOLUME A
1008
CALCULATE WEIGHTED VIRTUAL PANEL VOLUME B
1010
CALCULATE TOTAL SERVER PAGE VIEW VOLUME C
1012
CALCULATE ACTIVITY DUPLICATION PROBABILITY FOR THE SELECTED DAYPART*DEMOGRAPHIC GROUP (C-A-B)/A * 100%
1014
ADDITIONAL CATEGORIES?
1016
YES
NO
ADDITIONAL URL PATTERNS?
1018
YES
NO
GENERATE PROBABILITY LOOKUP TABLE INCLUDING PATTERN, DAYPART CATEGORY, DEMOGRAPHIC GROUP, AND DUPLICATION PROBABILITY
1020
A

FIG. 10A

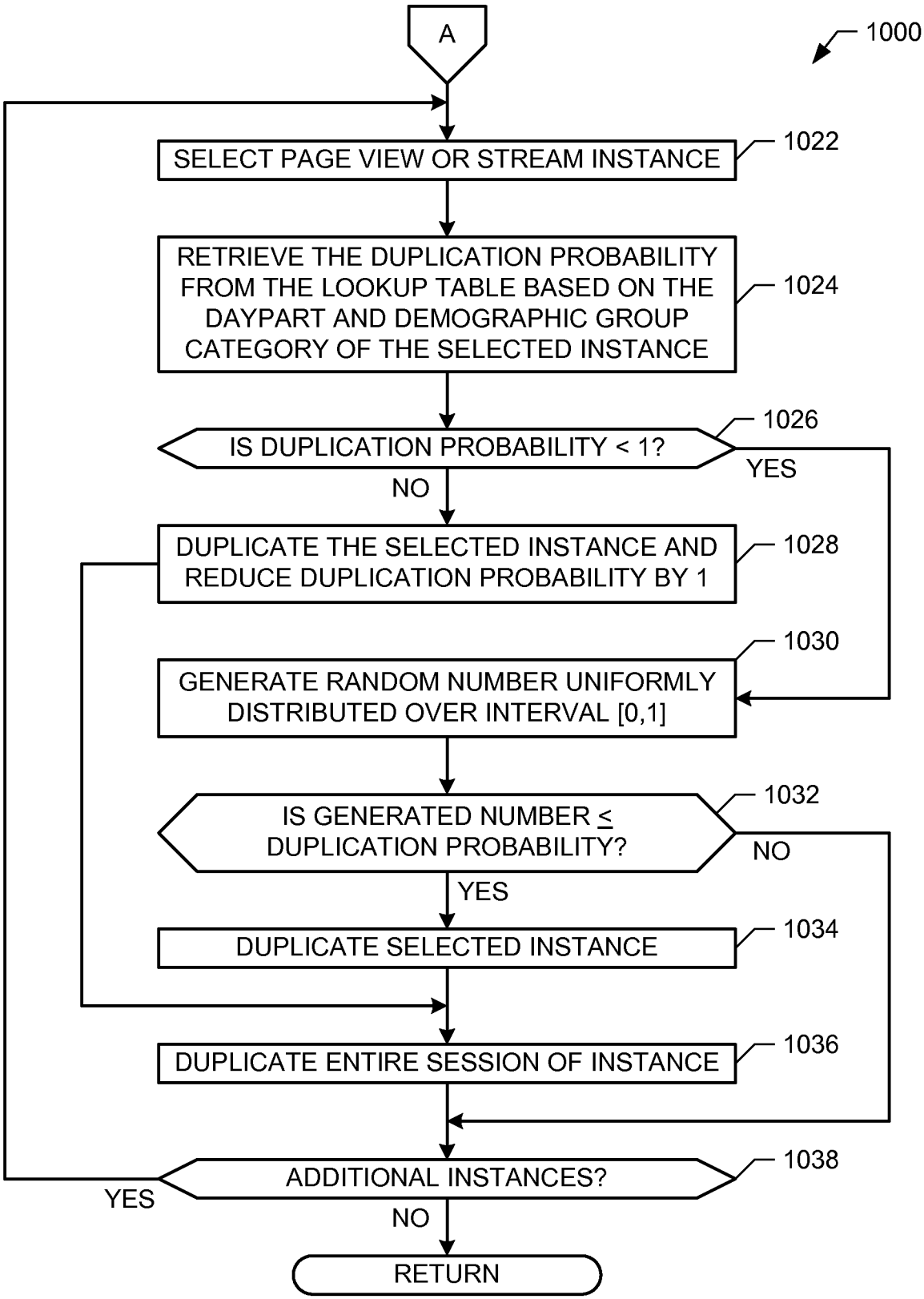

A
1000
SELECT PAGE VIEW OR STREAM INSTANCE
1022
RETRIEVE THE DUPLICATION PROBABILITY FROM THE LOOKUP TABLE BASED ON THE DAYPART AND DEMOGRAPHIC GROUP CATEGORY OF THE SELECTED INSTANCE
1024
IS DUPLICATION PROBABILITY < 1?
1026
YES
NO
DUPLICATE THE SELECTED INSTANCE AND REDUCE DUPLICATION PROBABILITY BY 1
1028
GENERATE RANDOM NUMBER UNIFORMLY DISTRIBUTED OVER INTERVAL [0,1]
1030
IS GENERATED NUMBER ≤ DUPLICATION PROBABILITY?
1032
NO
YES
DUPLICATE SELECTED INSTANCE
1034
DUPLICATE ENTIRE SESSION OF INSTANCE
1036
ADDITIONAL INSTANCES?
1038
YES
NO
RETURN

FIG. 10B

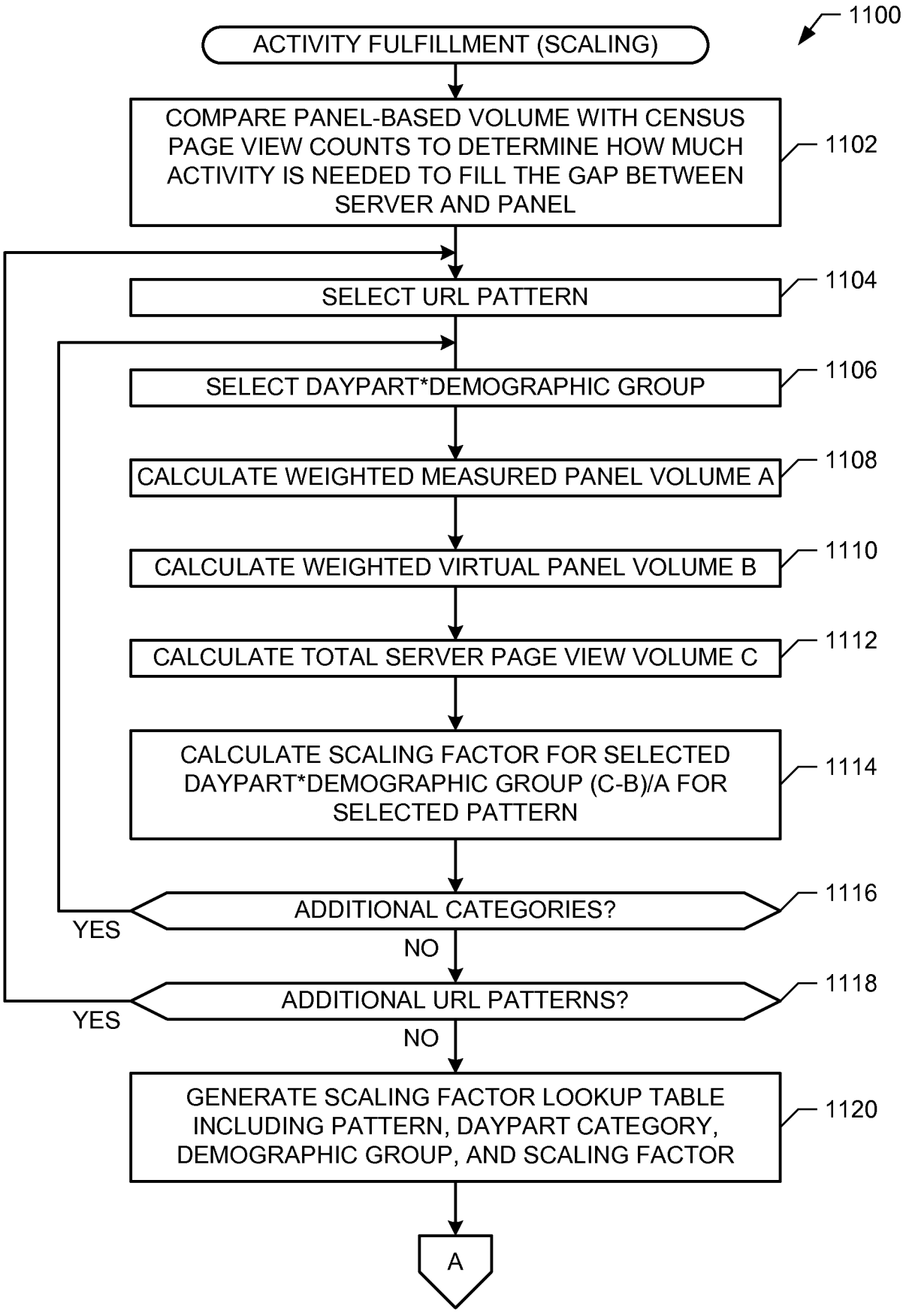

1100
ACTIVITY FULFILLMENT (SCALING)
COMPARE PANEL-BASED VOLUME WITH CENSUS PAGE VIEW COUNTS TO DETERMINE HOW MUCH ACTIVITY IS NEEDED TO FILL THE GAP BETWEEN SERVER AND PANEL
1102
SELECT URL PATTERN
1104
SELECT DAYPART*DEMOGRAPHIC GROUP
1106
CALCULATE WEIGHTED MEASURED PANEL VOLUME A
1108
CALCULATE WEIGHTED VIRTUAL PANEL VOLUME B
1110
CALCULATE TOTAL SERVER PAGE VIEW VOLUME C
1112
CALCULATE SCALING FACTOR FOR SELECTED DAYPART*DEMOGRAPHIC GROUP (C-B)/A FOR SELECTED PATTERN
1114
ADDITIONAL CATEGORIES?
1116
YES
NO
ADDITIONAL URL PATTERNS?
1118
YES
NO
GENERATE SCALING FACTOR LOOKUP TABLE INCLUDING PATTERN, DAYPART CATEGORY, DEMOGRAPHIC GROUP, AND SCALING FACTOR
1120
A

FIG. 11A

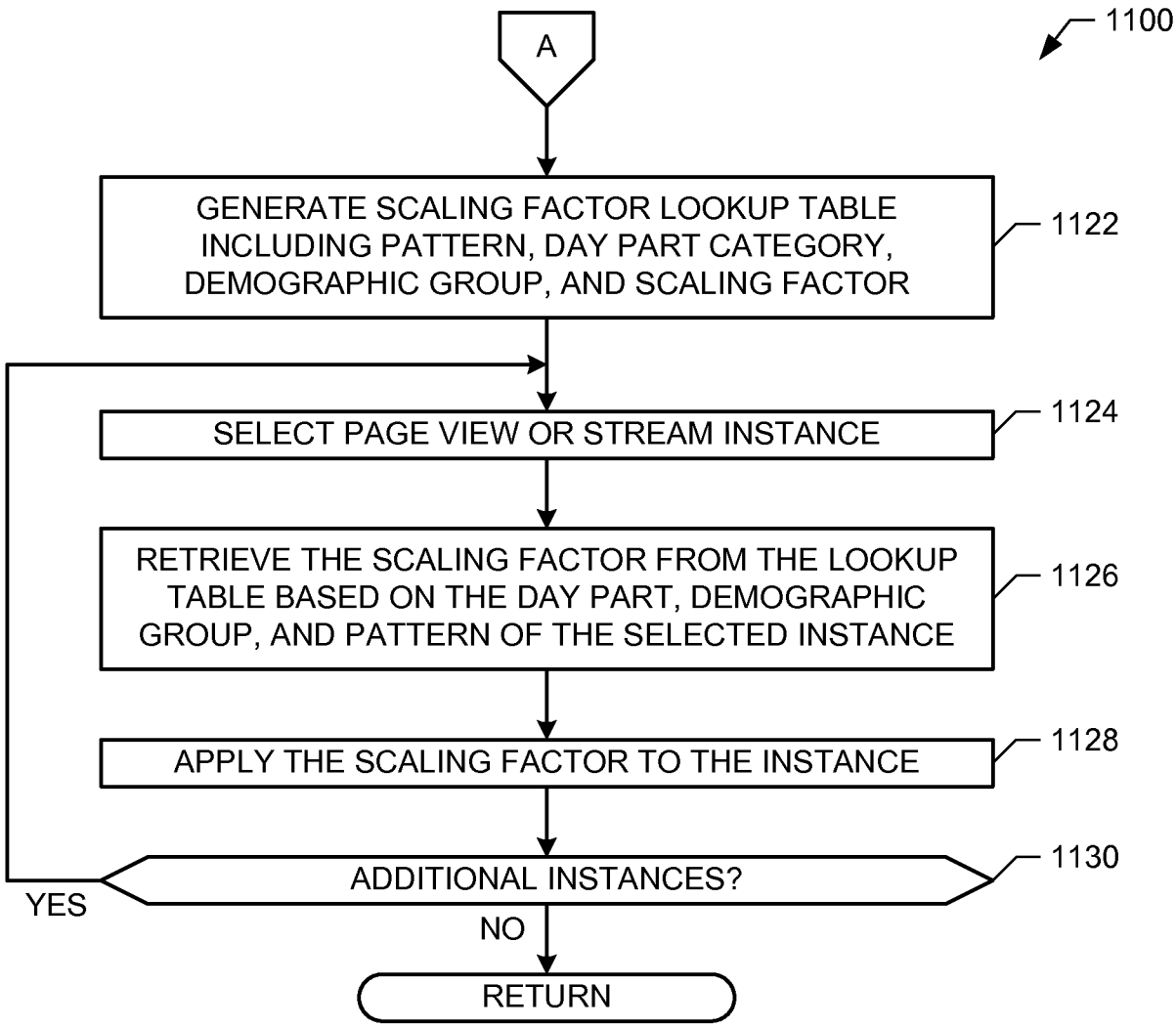

1100
A
GENERATE SCALING FACTOR LOOKUP TABLE INCLUDING PATTERN, DAY PART CATEGORY, DEMOGRAPHIC GROUP, AND SCALING FACTOR
1122
SELECT PAGE VIEW OR STREAM INSTANCE
1124
RETRIEVE THE SCALING FACTOR FROM THE LOOKUP TABLE BASED ON THE DAY PART, DEMOGRAPHIC GROUP, AND PATTERN OF THE SELECTED INSTANCE
1126
APPLY THE SCALING FACTOR TO THE INSTANCE
1128
ADDITIONAL INSTANCES?
1130
YES
NO
RETURN

FIG. 11B

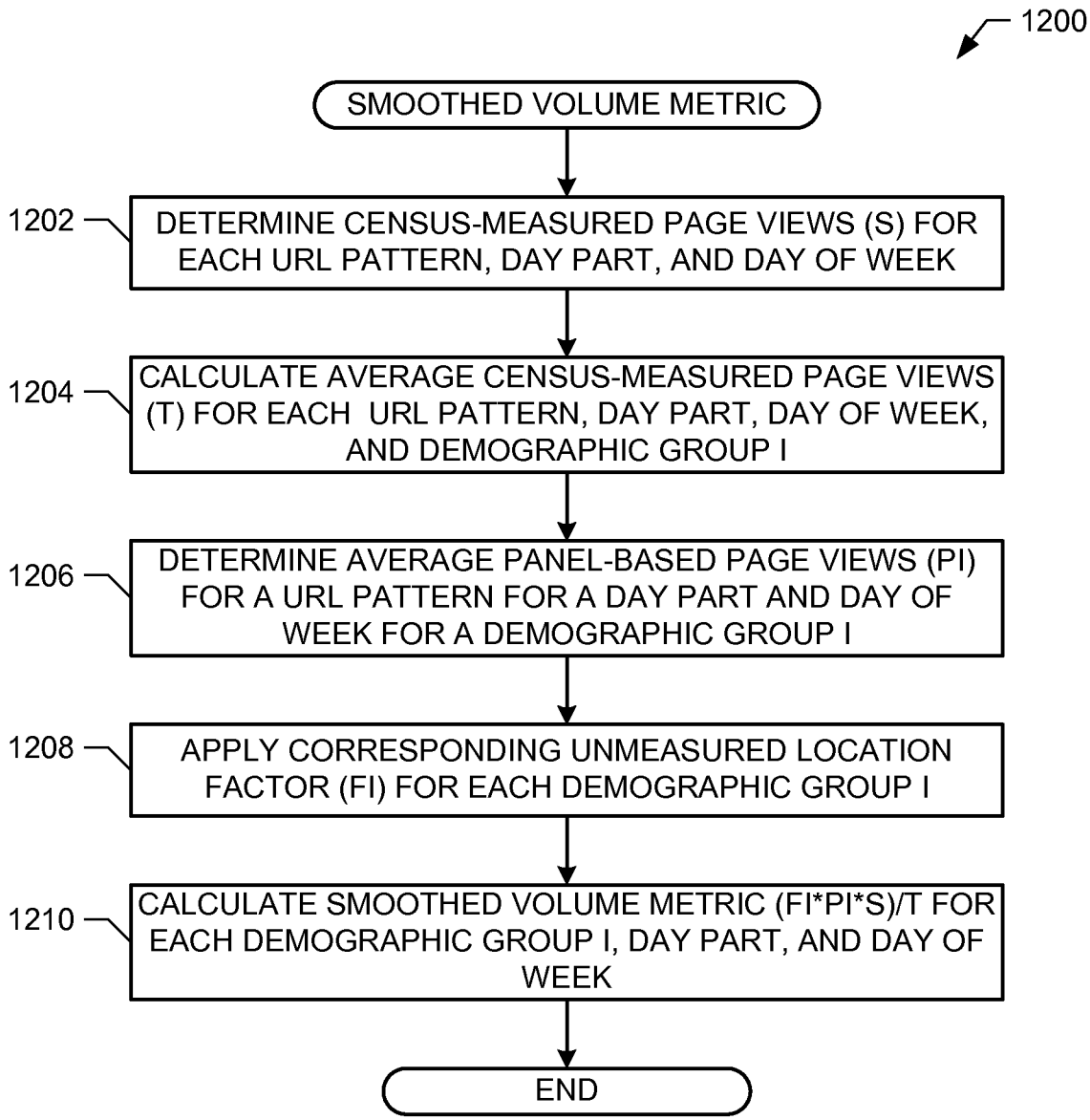

1200
SMOOTHED VOLUME METRIC
1202
DETERMINE CENSUS-MEASURED PAGE VIEWS (S) FOR EACH URL PATTERN, DAY PART, AND DAY OF WEEK
1204
CALCULATE AVERAGE CENSUS-MEASURED PAGE VIEWS (T) FOR EACH URL PATTERN, DAY PART, DAY OF WEEK, AND DEMOGRAPHIC GROUP I
1206
DETERMINE AVERAGE PANEL-BASED PAGE VIEWS (PI) FOR A URL PATTERN FOR A DAY PART AND DAY OF WEEK FOR A DEMOGRAPHIC GROUP I
1208
APPLY CORRESPONDING UNMEASURED LOCATION FACTOR (FI) FOR EACH DEMOGRAPHIC GROUP I
1210
CALCULATE SMOOTHED VOLUME METRIC (FI*PI*S)/T FOR EACH DEMOGRAPHIC GROUP I, DAY PART, AND DAY OF WEEK
END

FIG. 12

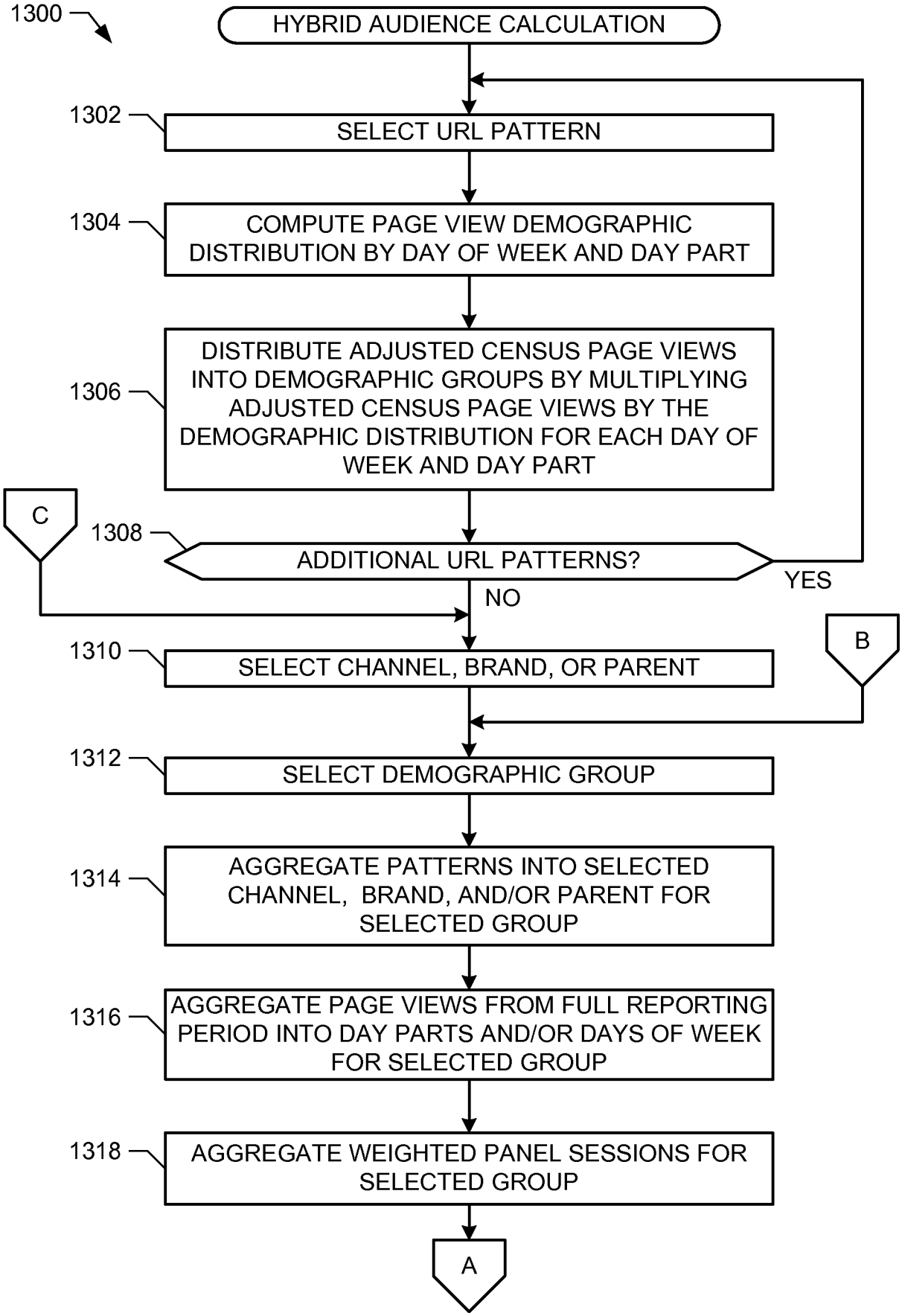

1300
HYBRID AUDIENCE CALCULATION
1302
SELECT URL PATTERN
1304
COMPUTE PAGE VIEW DEMOGRAPHIC DISTRIBUTION BY DAY OF WEEK AND DAY PART
1306
DISTRIBUTE ADJUSTED CENSUS PAGE VIEWS INTO DEMOGRAPHIC GROUPS BY MULTIPLYING ADJUSTED CENSUS PAGE VIEWS BY THE DEMOGRAPHIC DISTRIBUTION FOR EACH DAY OF WEEK AND DAY PART
C
1308
ADDITIONAL URL PATTERNS?
YES
NO
B
1310
SELECT CHANNEL, BRAND, OR PARENT
1312
SELECT DEMOGRAPHIC GROUP
1314
AGGREGATE PATTERNS INTO SELECTED CHANNEL, BRAND, AND/OR PARENT FOR SELECTED GROUP
1316
AGGREGATE PAGE VIEWS FROM FULL REPORTING PERIOD INTO DAY PARTS AND/OR DAYS OF WEEK FOR SELECTED GROUP
1318
AGGREGATE WEIGHTED PANEL SESSIONS FOR SELECTED GROUP
A

FIG. 13A

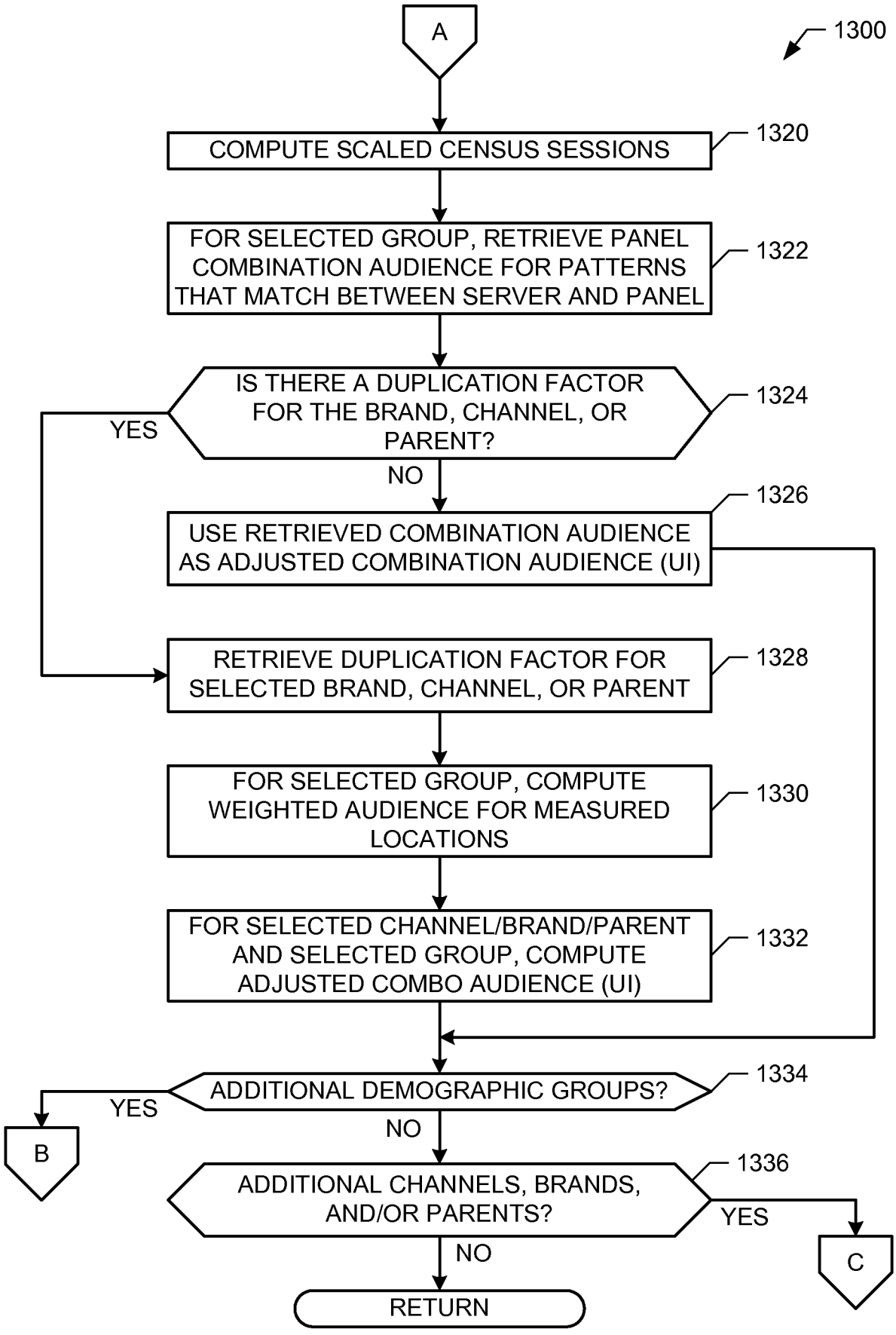

A
1300
COMPUTE SCALED CENSUS SESSIONS
1320
FOR SELECTED GROUP, RETRIEVE PANEL COMBINATION AUDIENCE FOR PATTERNS THAT MATCH BETWEEN SERVER AND PANEL
1322
IS THERE A DUPLICATION FACTOR FOR THE BRAND, CHANNEL, OR PARENT?
1324
YES
NO
USE RETRIEVED COMBINATION AUDIENCE AS ADJUSTED COMBINATION AUDIENCE (UI)
1326
RETRIEVE DUPLICATION FACTOR FOR SELECTED BRAND, CHANNEL, OR PARENT
1328
FOR SELECTED GROUP, COMPUTE WEIGHTED AUDIENCE FOR MEASURED LOCATIONS
1330
FOR SELECTED CHANNEL/BRAND/PARENT AND SELECTED GROUP, COMPUTE ADJUSTED COMBO AUDIENCE (UI)
1332
ADDITIONAL DEMOGRAPHIC GROUPS?
1334
YES
B
NO
ADDITIONAL CHANNELS, BRANDS, AND/OR PARENTS?
1336
YES
C
NO
RETURN

FIG. 13B

1400

1402 1404 1408 1406 1410

day of week	Day part	demo group	weighted panel page views	panel page view distribution %
1	1	1	6738.647	0.25%
1	1	2	23096.261	0.86%
1	1	3	9347.1714	0.35%
1	1	4	6155.4873	0.23%
1	1	5	280235.0921	10.41%
1	1	6	1091639.805	40.55%
1	1	7	73995.9797	2.75%
1	1	8	340683.8341	12.66%
1	1	9	237807.3983	8.83%
1	1	10	312455.8553	11.61%
1	1	11	170626.8178	6.34%
1	1	12	90809.9909	3.37%
1	1	13	14589.1937	0.54%
1	1	14	33629.1696	1.25%

FIG. 14

1500

1402 1404 1408 1410 1502 1504

day of week	Day part	demo group	panel page view distribution %	Adjusted Census page views	calculated page views
1	1	1	0.25%	5,413,047	13,551
1	1	2	0.86%	5,413,047	46,445
1	1	3	0.35%	5,413,047	18,797
1	1	4	0.23%	5,413,047	12,378
1	1	5	10.41%	5,413,047	563,534
1	1	6	40.55%	5,413,047	2,195,213
1	1	7	2.75%	5,413,047	148,801
1	1	8	12.66%	5,413,047	685,092
1	1	9	8.83%	5,413,047	478,214
1	1	10	11.61%	5,413,047	628,327
1	1	11	6.34%	5,413,047	343,119
1	1	12	3.37%	5,413,047	182,613
1	1	13	0.54%	5,413,047	29,338
1	1	14	1.25%	5,413,047	67,626

FIG. 15

1402 day of week	1404 day part	1408 demo group	1602 calculated page views
1	1	1	33,551
1	2	1	0
1	3	1	94,640
1	4	1	15,003
1	5	1	303,185
2	1	1	53,277
2	2	1	3,456
2	3	1	45,670
2	4	1	0
2	5	1	0
3	1	1	21,305
3	2	1	0
3	3	1	0
3	4	1	0
3	5	1	57,987
4	1	1	23,476
4	2	1	0
4	3	1	16,155
4	4	1	0
4	5	1	5,488
5	1	1	25,411
5	2	1	0
5	3	1	0
5	4	1	3,453
5	5	1	9,666
6	1	1	31,721
6	2	1	0
6	3	1	7,763
6	4	1	0
6	5	1	1,917
7	1	1	33,978
7	2	1	3,853
7	3	1	250,049
7	4	1	9,414
7	5	1	0
Total for demo group 1 :			**1,050,418**

FIG. 16

1408	1704	1706	1708	1702
Demographic Group	Weighted Panel hits	Calculated Census Hits	Weighted Panel Sessions	Scaled Census Sessions
1	3,829,438	10,659,387	1,501,075	4,178,299
2	28,040,233	72,732,430	6,989,826	18,130,627
3	23,793,017	58,664,788	6,454,412	15,914,195
4	47,605,566	105,049,652	8,783,620	19,382,528
5	24,197,316	61,571,824	7,193,873	18,305,330
6	24,286,131	61,252,859	7,066,847	17,823,531
7	7,528,919	17,803,181	1,597,329	3,777,107
8	29,962,546	71,000,489	6,811,842	16,141,622
9	31,874,593	77,764,534	6,459,959	15,760,381
10	16,211,006	42,416,337	5,565,773	14,562,928
11	17,002,289	43,714,488	4,306,656	11,072,819
12	13,798,564	36,655,649	2,888,567	7,673,428
13	112,449	356,843	60,461	191,865
14	232,375	706,181	71,048	215,913
15	529,139	1,431,329	138,368	374,288
16	1,461,985	4,297,373	231,656	680,933

FIG. 17

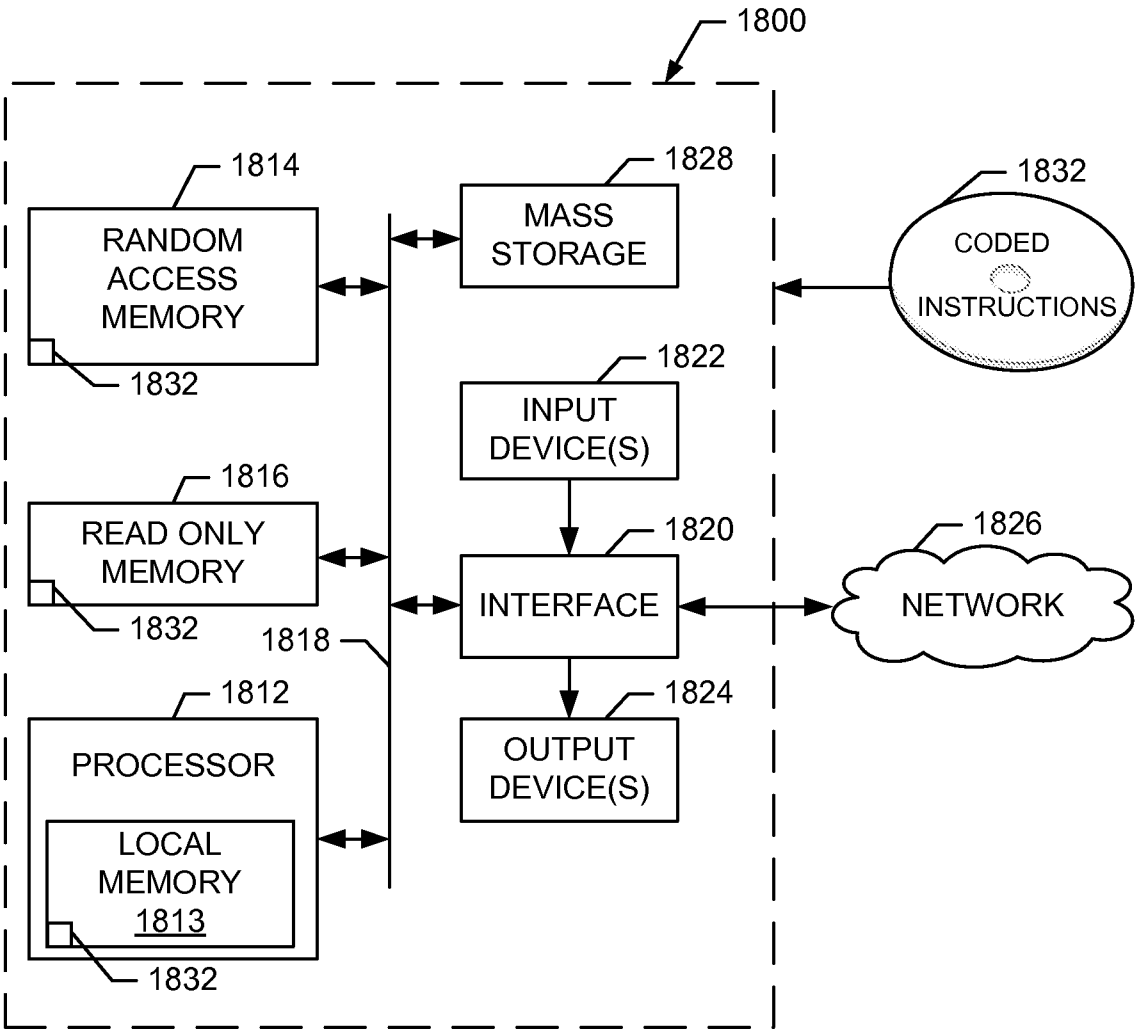

1800
1814
RANDOM ACCESS MEMORY
1832
1828
MASS STORAGE
1832
CODED INSTRUCTIONS
1822
INPUT DEVICE(S)
1816
READ ONLY MEMORY
1832
1820
INTERFACE
1826
NETWORK
1818
1812
PROCESSOR
LOCAL MEMORY 1813
1832
1824
OUTPUT DEVICE(S)

FIG. 18

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2013/023259

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 50/30(2012.01)i, G06F 19/00(2011.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06Q 50/30; G06F 15/16; H04H 60/45; G06F 15/173; H04H 9/00; G06F 17/30; G06F 17/60

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

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

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

eKOMPASS(KIPO internal) & Keywords: online audience measurement, panel, unmeasured location, estimating online population, virtual panel

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2011-0246641 A1 (PUGH, BRIAN et al.) 06 October 2011	1-3,6-16,18-19
A	See abstract, paragraphs [0012]-[0015], [0025], [0035]-[0041], [0048]-[0049], [0071], [0084], [0104] and claims 1, 4, 15.	,21-29,32-41 4-5,17,20,30-31
Y	US 2012-0005213 A1 (HANNAN, KEVIN MICHAEL et al.) 05 January 2012	1-3,6-16,18-19
A	See abstract, paragraphs [0038]-[0039], [0047]-[0048], claims 1, 4 and figures 5-6.	,21-29,32-41 4-5,17,20,30-31
A	US 2011-0208860 A1 (SIM, LIM OR et al.) 25 August 2011 See abstract, paragraphs [0057], [0059]-[0072], [0080]-[0081] and figures 4a-4b.	1-41
A	US 2004-0088212 A1 (HILL, CLARKE R.) 06 May 2004 See abstract, paragraphs [0030]-[0038] and figures 1-2.	1-41
A	US 2008-0086741 A1 (FELDMAN, KONRAD et al.) 10 April 2008 See abstract, paragraphs [0015]-[0026] and figure 1.	1-41

☐ 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 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	Date of mailing of the international search report
03 May 2013 (03.05.2013)	**03 May 2013 (03.05.2013)**

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon Metropolitan City, 302-701, Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

OH, Eung Gie

Telephone No. 82-42-481-8744

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2013/023259

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2011-0246641 A1	06.10.2011	WO 2011-127027 A1	13.10.2011
US 2012-0005213 A1	05.01.2012	CN 102314488 A	11.01.2012
		EP 2402899 A1	04.01.2012
		US 8307006 B2	06.11.2012
US 2011-0208860 A1	25.08.2011	AU 1999-06399 D0	26.08.1999
		AU 2000-62543 A1	05.03.2001
		AU 2000-62543 B2	28.04.2005
		AU 6254300 A	05.03.2001
		AU 781008 B2	28.04.2005
		EP 1226525 A1	31.07.2002
		EP 1226525 A4	09.08.2006
		EP 2293495 A2	09.03.2011
		JP 2003-506790 A	18.02.2003
		JP 2011-018384 A	27.01.2011
		JP 4799788 B2	26.10.2011
		US 2008-0256235 A1	16.10.2008
		US 2008-0263200 A1	23.10.2008
		US 2010-0228855 A1	09.09.2010
		US 2012-0158959 A1	21.06.2012
		US 7376722 B1	20.05.2008
		US 7953791 B2	31.05.2011
		US 7953839 B2	31.05.2011
		US 8112511 B2	07.02.2012
		WO 01-11506 A1	15.02.2001
US 2004-0088212 A1	06.05.2004	None	
US 2008-0086741 A1	10.04.2008	WO 2008-045899 A1	17.04.2008