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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 December 2016 (01.12.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/190856 A1

(51) International Patent Classification:
G06F 11/08 (2006.01) *G06N 3/08* (2006.01)

(21) International Application Number:
PCT/US2015/032570

(22) International Filing Date:
27 May 2015 (27.05.2015)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **HEWLETT PACKARD ENTERPRISE DEVELOPMENT LP** [US/US]; 11445 Compaq Center Drive West, Houston, TX 77070 (US).

(72) Inventors: **KAUR, Satwant**; 2350 Charleston Rd, Mountain View, California 94043 (US). **MAKKINEJAD, Babak**; 585 South Boulevard, Pontiac, Michigan 48341 (US). **DOSHI, Parag**; 5555 Windward Pkwy, Alpharetta, Georgia 30004 (US). **WICK, Corey**; 5400 Legacy, Plano, Texas 75024 (US).

(74) Agent: **LEMMON, Marcus B,**; Hewlett Packard Enterprise, 3404 E. Harmony Road, Mail Stop 79, Fort Collins, CO 80528 (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, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,

[Continued on next page]

(54) Title: DATA VALIDATION

(57) Abstract: Examples associated with data validation are disclosed. One example includes a sensor attribute data store having information describing attributes of a set of distributed sensors. A pattern data store stores information describing patterns indicating anomalous sensor activity. A data aggregation module flags data received from a tested sensor as anomalous data when the anomalous data exceeds a variance level described by an attribute of the tested sensor. A data validation module validates the anomalous data by comparing the anomalous data to the patterns indicating anomalous sensor activity. A learning module updates the pattern indicating anomalous sensor activity based on a result received from the validation logic after the validation logic validates data received from the tested sensor.

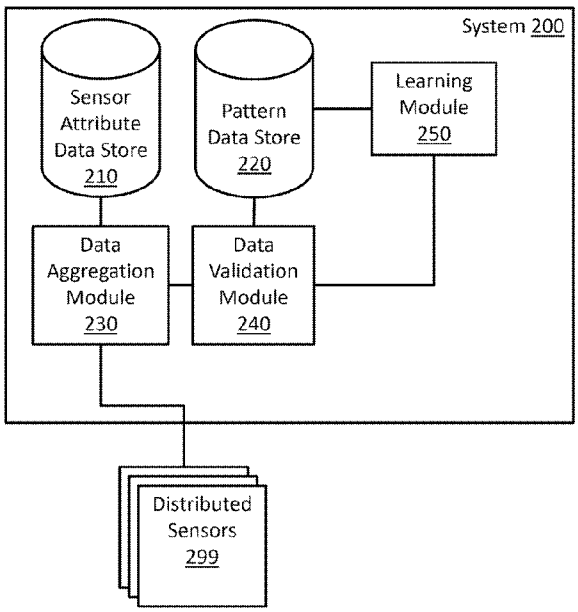

Figure 2

WO 2016/190856 A1

LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to the identity of the inventor (Rule 4.17(i))*

Published:

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

DATA VALIDATION

BACKGROUND

[0001] As more and more devices are embedded with electronics and networked applications, these devices are being connected to the internet, creating what is becoming known as the "internet of things". Manufacturers, data aggregators, other devices, and so forth receive data from the devices describing their use, environment, and so forth. Applications range from environmental monitoring (e.g., disaster early warning systems), to health care (e.g., remote patient monitoring), and even automating and/or remotely controlling home appliances (e.g., air conditioning). When the devices transmit data, the data itself may valuable, especially when aggregated with data from other devices including other distributed devices of the same type, other local devices, and so forth.

BRIEF DESCRIPTION OF THE DRAWINGS

[0002] The present application may be more fully appreciated in connection with the following detailed description taken in conjunction with the accompanying drawings, in which like reference characters refer to like parts throughout, and in which:

[0003] FIG. 1 illustrates an example network in which example systems, methods, and equivalents, may operate.

[0004] FIG. 2 illustrates an example system associated with data validation.

[0005] FIG. 3 illustrates another example system associated with data validation.

[0006] FIG. 4 illustrates a flowchart of example operations associated with data validation.

[0007] FIG. 5 illustrates another flowchart of example operations associated with data validation.

[0008] FIG. 6 illustrates another flowchart of example operations associated with data validation.

[0009] FIG. 7 illustrates another flowchart of example operations associated with data validation.

[0010] FIG. 8 illustrates an example computing device in which example systems, and methods, and equivalents, may operate.

DETAILED DESCRIPTION

[0011] Systems, methods, and equivalents associated with data validation are described. As mentioned above, many devices that are connected to the internet transmit data regarding device status, environmental states, and so forth. This data is collected by various sensors in the devices and transmitted to, e.g., an operator of the device, a manufacturer of the device, and so forth. In some cases, after data is collected, valuable information may be derived from the data. However, over time, devices may begin to malfunction due to wear and tear. Other factors (e.g., network error) may also contribute to situations where a device transmits data that is incorrect. Consequently, it may be valuable to distinguish when anomalous data received from a sensor is a result of a sensor malfunction, or from an actual event of significance that the sensor was designed to monitor.

[0012] By way of illustration, an earthquake early warning system may be made up of many distributed sensors. If first sensor malfunctions and begins "detecting" seismic activity, the fact that other sensors are not detecting the seismic activity may indicate the first sensor is malfunctioning. On the other hand, if many sensors are detecting seismic activity, it is much more likely an event of significance has occurred, as it is unlikely for many of the sensors to be malfunctioning simultaneously. In another example, if a sensor measuring temperature in a server

room, suddenly jumps a very high temperature reading, but then returns to a prior value, it is likely the high temperature reading was an anomaly resulting from a malfunction. If the high temperature is preceded by a steady increase in temperature of the server room, it is likely the sensor is functioning properly and an event of significance has occurred.

[0013] Consequently, when data is received from a sensor, first the data may be checked to see if it falls within typical behavior patterns for that sensor as defined by, for example, specifications associated with that sensor based on sensor type, manufacturer, and so forth. If the data falls outside the typical behavior patterns, the data may be categorized as anomalous and flagged for further analysis. Anomalous data may then be verified by comparing the data to a set of patterns associated with anomalous data that facilitate identifying whether the data is a result of a malfunction or whether the data is a legitimate reading from the sensor. In some cases, the patterns may be updated as data is classified, and anomalous data may be logged, so that the anomalous data can be re-verified at a later time.

[0014] "Module", as used herein, includes but is not limited to hardware, firmware, software stored on a computer-readable medium or in execution on a machine, and/or combinations of each to perform a function(s) or an action(s), and/or to cause a function or action from another module, method, and/or system. A module may include a software controlled microprocessor, a discrete module, an analog circuit, a digital circuit, a programmed module device, a memory device containing instructions, and so on. Modules may include one or more gates, combinations of gates, or other circuit components. Where multiple logical modules are described, it may be possible to incorporate the multiple logical modules into one physical module. Similarly, where a single logical module is described, it may be possible to distribute that single logical module between multiple physical modules.

[0015] Figure 1 illustrates an example network 100 in which example systems and methods, and equivalents, may operate. It should be appreciated that the items depicted in figure 1 are illustrative examples and many different features and implementations are possible.

[0016] Figure 1 illustrates an example network 100. Network 100 includes a numerous sensors 198. In some examples the sensors may feed data into network hubs 199 who serve as owners or administrators of different networks of sensors 198. In these examples, network hubs 199 may provide aggregated and/or disaggregated data to a data validation system 105. In other examples sensors 198 may directly connect to data validation system 105, causing sensor data to be provided directly to data validation system 105 without an network hub 199 affecting the data.

[0017] Whether the sensor data is received directly from sensors 198 or from network hubs 199, the sensor data may be initially processed by data aggregation module 110. Data aggregation module 110 may process the sensor data based on sensor attributes 120. Sensor attributes 120 may include, for example, make, model, and specifications of sensors from which sensor data is retrieved. Sensor attributes 120 may also include, for example, known data variations, malfunctions, and causes. In various examples, sensor attributes 120 may state variance levels and/or variance levels may be derived from sensor attributes 120 to facilitate detecting when sensors are behaving according to expectations or producing anomalous data.

[0018] By way of illustration, a sensor that measures air temperature in a home may typically read between 65 and 85 degrees Fahrenheit depending on the time of year and climate control usage (e.g., heat, air conditioning), among other factors. Various companies may be interested in this data. For example, an energy company may use the data to facilitate planning energy production, a company that sells energy efficiency products may use the data for targeting advertisements to an owner of the home, and so forth. For the sensor, a reading above 100 degrees may be considered to be outside normal operation of the sensor. This may be because the reading over 100 degrees may exceed a variance level associated with that sensor as defined by or derived from sensor attributes 120. When the reading exceeds the variance level, that data may be flagged as anomalous by the data aggregation module 110. Data that falls within the variance level may be considered validated data 160 that is treated as accurate which may subsequently be provided to data consumers (e.g., the energy company, the energy efficiency company) 195 via distribution module 170.

[0019] Data flagged as anomalous, however, may not necessarily be related to a malfunction of, for example, the sensor, the network, and so forth. By way of illustration, the reading above 100 degrees could also be a result of a significant event such as, for example, a heat wave and a broken air conditioning system, a house fire, or another legitimate reason. Events of significance, though outside normal operation of the sensor, if they can be properly identified may be useful for triggering certain events (e.g., repair the broken air conditioner, call the fire department) in response to the significant events.

[0020] Consequently, when data is flagged as anomalous, a data validation module 130 may perform further processing on the anomalous data. In some examples, the data may be compared to various patterns 140. The patterns may include, for example, predefined patterns input by a user or administrator, patterns learned from analyzing data received from sensors 198 over time, and so forth. A pattern 140 may take more information into account when validating anomalous data than data aggregation module does at 110. For example, patterns 140 may account for past data of the sensor, data of nearby sensors, and other factors that may indicate whether data flagged as anomalous is a result of a sensor malfunction or an event of significance.

[0021] As used herein a sensor malfunction is intended to encompass any technological error, glitch, or otherwise that may contribute to a sensor providing data considered anomalous. These may include issues arising directly from the sensor, issues arising from transmitting and/or storing data received from the sensor, and so forth. An event of significance is intended to encompass any real event that produces real data that is accurate, even though the data may be considered anomalous as a result of being outside the normal behavior pattern of the sensor (e.g., several standard deviations from normal operation). Though sensors may be configured to track, monitor, and so forth, important events while operating normally, here, events of significance is a term intended to encompass actual events outside of expected behaviors of the sensor.

[0022] The patterns may be associated with weights that indicate a level of confidence of whether data matching the pattern is a result of a malfunction or a significant event. Consequently, the more and/or more strongly weighted patterns that a piece of anomalous data matches that indicate the data is a result of a significant event, the more confident data validation module 130 may be that that data should be included with validated data 160. Similarly, the more and/or more strongly weighted patterns that a piece of anomalous data matches that indicate the data is a result of an malfunction, the more confident data validation module 130 may be that the data should not be included with validated data 160.

[0023] By way of illustration, when the home temperature sensor detects the 100 degree temperature and that measurement is flagged by data aggregation module 110 as anomalous, data validation module 130 may compare that measurement to patterns 140. One pattern, for example, may compare the 100 degree measurement to measurements of other temperature sensors in the home. If many temperature sensors have similar readings, it may be more likely that the 100 degree measurement is valid. On the other hand, if the measurement when compared to other preceding and subsequent measurements of the home temperature sensor, is an outlier, data validation module 130 may be more confident that the reading was a one-time malfunction. In some examples, data validation module 130 may not be tasked with actually identifying what type of malfunction or significant event has occurred. Diagnosing malfunctions and/or significant events may be performed by other modules (not shown), data consumers, and so forth. Patterns of sensor behavior involving other types of sensors, external data, data histories, and so forth may also be taken into consideration by data validation module 130.

[0024] When data validation module makes a decision regarding whether data is a result of an event of significance or a result of a malfunction, data validation module 130 may communicate this decision to a learning module 150. Learning module 150 may use the decision to update the patterns 140. Updating patterns 140 may include modifying patterns 140, creating patterns 140, removing patterns 140, and so forth. Modifying patterns 140 may also include updating weights associated with patterns 140 to increase or decrease confidence as to whether data matching

patterns 140 is a result of an anomaly or an event of significance. Consequently, over time, the learning module 150 may increase reliability of data validation module 130 by increasing the likelihood that anomalous data is appropriately categorized.

[0025] Data validation system 105 also includes an error log 180. In some examples, error log 180 may store data marked as anomalous in addition to notations as to whether that data was validated as an event of significance or found to be a result of a malfunction. Logging anomalous data may facilitate periodically revalidating the anomalous data by, for example, data validation module 130. Revalidating anomalous data may be occasionally appropriate due to the updates to patterns 140 by learning module 150. In various examples, revalidating data may cause some data that was previously rejected as a malfunction to be subsequently added to validated data 160 and/or for data that was previously validated to be removed from the validated data 160. Error log 180 may also include for a given piece of anomalous data, for example, notations related to a pattern which was found to be strongly determinative as to whether the piece of anomalous data was considered a result of an event of significance or a malfunction, notations indicating a decision strength regarding whether the piece of anomalous data was considered a result of an event of significance or a malfunction (e.g., close to 50%, close to 100% or 0%), and so forth.

[0026] As mentioned above data validation system 105 also includes a distribution module 170. Distribution module 170 may facilitate distribution of validated data 160 to data consumers 195. How distribution module 170 operates may depend on who data consumers 195 are, their relationship to an operator of data validation system 105, privacy and/or security concerns, and so forth. For example, a lower security solution may be to provide raw validated data 160 to data consumers 195. In other examples, distribution module 170 may operate as an interface (e.g., a structured query language interpreter) to provide control over what queries can be executed on validated data 160. Distribution module 170 could also be designed to obtain specific results from the validated data 160 and provide these results to data consumers 195 without granting access to the underlying validated data 160. Other modes of operation of distribution module 170 may also be possible.

[0027] It is appreciated that, in the following description, numerous specific details are set forth to provide a thorough understanding of the examples. However, it is appreciated that the examples may be practiced without limitation to these specific details. In other instances, methods and structures may not be described in detail to avoid unnecessarily obscuring the description of the examples. Also, the examples may be used in combination with each other.

[0028] Figure 2 illustrates an example system 200 associated with data validation. System 200 includes a sensor attribute data store 210. Sensor attribute data store may store information describing attributes of a set of distributed sensors 299. Attributes of the distributed sensors may include, for example, attributes describing model information of the sensors (e.g., manufacturer, standards the sensor adheres to), attributes describing expected data to be received from the sensors (e.g., ranges), attributes describing expected anomalous behaviors of the sensors, and so forth (e.g., errors the sensor may raise, known sensor bugs).

[0029] System 200 also includes a pattern data store 220. Pattern data store 220 may store information describing patterns indicating anomalous sensor activity. Patterns indicating anomalous sensor activity may include, for example, specific (e.g., pre-defined) scenarios that describe a sensor behavior to be treated as anomalous, learned patterns of sensor behaviors to be treated as anomalous, and so forth. Patterns may facilitate comparing sensor data to sensor data from other types of sensors, other nearby sensors, historical and/or future data, and so forth. In some examples, pattern data store 220 may also include patterns that indicate non-anomalous sensor activity. These patterns of non-anomalous activity may also be, for example, pre-defined scenarios, learned scenarios, and so forth. In various examples, patterns may be associated with weights. Consequently, when a sensor behavior matches a pattern, the weight may indicate a confidence level of whether the sensor behavior is a result of, for example, a sensor malfunction, an event of significance, and so forth.

[0030] System 200 also includes a data aggregation module 230. Data aggregation module 230 may flag data received from a tested sensor as anomalous

data. The tested sensor may be a member of the set of distributed sensors 299. Data may be flagged as anomalous data by data aggregation module 230 when the anomalous data exceeds a variance level described by an attribute of the tested sensor (e.g., from sensor attribute data store 210). The variance level may be an actual attribute stored in sensor attribute data store 210, derived from an attribute in sensor attribute data store 210, and so forth.

[0031] System 200 also includes a data validation module 240. Data validation module 240 may validate the anomalous data by comparing the anomalous data to the patterns indicating anomalous sensor activity (e.g., in pattern data store 220. In some examples, validation module 240 may validate the data received from the tested sensor by comparing the data received from the tested sensor to data received from sensors located within a specified physical proximity to the tested sensor. The specified physical proximity may be defined by a pattern, and may depend on sensor attributes (e.g., what the sensor was designed to measure, sensor location), and so forth.

[0032] System 200 also includes a learning module 250. Learning module 250 may update a pattern indicating anomalous sensor activity in pattern data store 220 based on a result received from validation module 240. The signal may be received when validation module 240 validates data received from the tested sensor. In various examples, learning module 250 may update the pattern by modifying a weight associated with the pattern. As described above, the weight may indicate a likelihood that anomalous data satisfying the pattern is a result of a sensor malfunction, an event of significance, and so forth.

[0033] Figure 3 illustrates another example system 300 associated with data validation. System 300 includes many items similar to those described above with reference to system 200 (figure 2). For example, system 300 includes a sensor attribute data store 310, a pattern data store 320, a data aggregation module 330 to flag anomalous data received from distributed sensors 399, a data validation module 340, and a learning module 350. System 300 also includes an error logging data store 360. Error logging data store 360 may store anomalous data that fails validation.

[0034] System 300 also includes an error checking module 370. Error checking module 370 may revalidate data in error logging data store 360 when learning module 350 updates a pattern indicating anomalous sensor activity in pattern data store 320. Though system 300 illustrates error checking module 370 as a separate component from data validation module 340, in some examples, data validation module 340 may also serve as error checking module 370.

[0035] System 300 also includes a validated data store 380. Validated data store 380 may store data received from the tested sensor that falls within the variance level. Validated data store 380 may also store anomalous data received from the tested sensor that passes validation from data validation module 340.

[0036] System 300 also includes a distribution module 390. Distribution module 390 may facilitate distribution of data in validated data store 380. Distribution module 390 may distribute data from validated data store 380 by providing raw data associated with a portion of the verified data, providing query access to the verified data, providing a result of a query of the verified data, and so forth.

[0037] Figure 4 illustrates an example Method 400 associated with data validation. Method 400 may be embodied on a non-transitory computer-readable medium storing computer-executable instructions. The instructions, when executed by a computer, may cause the computer to perform method 400. In other examples, method 400 may exist within logic gates and/or RAM of an application specific integrated circuit (ASIC).

[0038] Method 400 includes receiving sensor data at 410. The sensor data may be received from a set of distributed sensors. Method 400 also includes marking sensor data as anomalous data at 420. The sensor data marked as anomalous data may be so marked when the anomalous data exceeds a variance level associated with a sensor from which the anomalous data was received. The variance levels for sensors may be generated based on device specifications of the sensors. The device specifications may describe, for example, known sensor malfunction scenarios for respective sensors. Consequently, the variance levels may be derived from the known sensor malfunction scenarios, and so forth.

[0039] Method 400 also includes validating the anomalous data at 430. The anomalous data may be validated based on a set of patterns describing anomalous sensor activity. Validating the anomalous data at 430 may indicate whether the anomalous data is a result of a sensor malfunction, an event of significance, and so forth.

[0040] Method 400 also includes updating the set of patterns at 440. The patterns may be updated based on whether the anomalous data is validated as the sensor malfunction and the event of significance. As described above, the set of patterns may include weights that indicate a likelihood that anomalous data is a result of a sensor malfunction. In this example, updating a pattern may include modifying a weight associated with the pattern to increase or decrease confidence in whether a sensor is malfunctioning when that sensor's behavior matches the pattern.

[0041] Method 400 also includes distributing verified data at 470. The verified data may include sensor data within the variance level and validated anomalous data. In various examples, the verified data may be distributed by providing raw data associated with a portion of the verified data, providing query access to the verified data, providing a result of a query of the verified data, and so forth. The technique used for distributing the verified data may depend on, for example, data security concerns, privacy concerns, processing availability, and so forth, and different distribution techniques may be appropriate depending on who the is receiving the distributed data.

[0042] Figure 5 illustrates another example method associated with data validation. Method 500 includes several actions similar to those described above with reference to method 400 (figure 4). For example method 500 includes receiving sensor data at 510, marking sensor data as anomalous data at 520, validating anomalous sensor data at 530, updating patterns at 540, and distributing verified data at 570.

[0043] Method 500 also includes logging anomalous data indicated as a sensor malfunction at 550. Method 500 also includes revalidating logged anomalous data at 560. The revalidation may occur, for example, after an update to the set of

patterns. Revalidating anomalous data indicated as a sensor malfunction may facilitate making accurate data available when distributing the verified data 570.

[0044] Figure 6 illustrates a method 600 associated with data validation. Method 600 includes receiving sensor data at 610. The sensor data may be received from a set of distributed sensors.

[0045] Method 600 also includes marking sensor data as anomalous data at 620. Sensor data may be marked as anomalous data when the anomalous data exceeds a variance level associated with a sensor from the anomalous data was received. The variance level for a sensor may be generated on accuracy information associated with the sensor, how recently the sensor was calibrated, and so forth.

[0046] Method 600 also includes storing non-anomalous data in a validated data store at 630. Method 600 also includes storing anomalous data found to be a result of an event of significance at 640. The anomalous data found to be resulting from significant events may also be stored to the validated data store. To determine whether anomalous data is a result of an event of significance, the anomalous data may be compared to a set of patterns describing anomalous sensor activity.

[0047] Method 600 also includes updating the set of patterns at 660. As described above, updating the set of patterns may include modifying a weight associated with a pattern that reflects a confidence level of whether data matching the pattern is a result of a sensor malfunction, an event of significance, and so forth. Method 600 also includes distributing data from the validated data store at 680. The data may be distributed to consumers of the data from the validated data store, who may, for example, pay for access to the validated data.

[0048] Figure 7 illustrates a method 700 associated with data validation. Method 700 includes several actions similar to those described above with reference to method 600 (figure 6). For example, method 700 includes, receiving sensor data at 710, marking sensor data as anomalous data at 720, storing non-anomalous data at 730, storing anomalous data resulting from events of significance at 740, updating patterns at 760, and distributing validated data at 780.

[0049] Method 700 also includes storing anomalous data found to be a result of a sensor malfunction at 750. Anomalous data resulting from sensor malfunctions may be stored in an error logging data store. Whether data results from a sensor function may also be determined based on the set of patterns.

[0050] Method 700 also includes, at 770, periodically evaluating whether data in the error logging data store is a result of an event of significance based on updated patterns.

[0051] Figure 8 illustrates an example computing device in which example systems and methods, and equivalents, may operate. The example computing device may be a computer 800 that includes a processor 810 and a memory 820 connected by a bus 830. The computer 800 includes a data validation module 840. Data validation module 840 may perform, alone or in combination, various functions described above with reference to the example systems, methods, apparatuses, and so forth. In different examples, data validation module 840 may be implemented as a non-transitory computer-readable medium storing computer-executable instructions, in hardware, software, firmware, an application specific integrated circuit, and/or combinations thereof.

[0052] The instructions may also be presented to computer 800 as data 850 and/or process 860 that are temporarily stored in memory 820 and then executed by processor 810. The processor 810 may be a variety of various processors including dual microprocessor and other multi-processor architectures. Memory 820 may include non-volatile memory (e.g., read only memory) and/or volatile memory (e.g., random access memory). Memory 820 may also be, for example, a magnetic disk drive, a solid state disk drive, a floppy disk drive, a tape drive, a flash memory card, an optical disk, and so on. Thus, memory 820 may store process 860 and/or data 850. Computer 800 may also be associated with other devices including other computers, peripherals, and so forth in numerous configurations (not shown).

[0053] It is appreciated that the previous description of the disclosed examples is provided to enable a person skilled in the art to make or use the present disclosure. Various modifications to these examples may be apparent, and the generic principles

defined herein may be applied to other examples without departing from the spirit or scope of the disclosure. Thus, the present disclosure is not intended to be limited to the examples shown herein but is to be accorded the widest scope consistent with the principles and features disclosed herein.

WHAT IS CLAIMED IS:

1. A system, comprising:

a sensor attribute data store to store information describing attributes of a set of distributed sensors;

a pattern data store to store information describing patterns indicating anomalous sensor activity;

a data aggregation module to flag data received from a tested sensor as anomalous data when the anomalous data exceeds a variance level described by an attribute of the tested sensor;

a data validation module to validate the anomalous data by comparing the anomalous data to the patterns indicating anomalous sensor activity;

a learning module to update a pattern indicating anomalous sensor activity based on a result received from the validation logic after the validation logic validates data received from the tested sensor.

2. The system of claim 1, further comprising:

an error logging data store to store anomalous data that fails validation; and

an error checking module to re-validate data in the error logging data store when the learning module updates a pattern indicating anomalous sensor activity.

3. The system of claim 1, further comprising:

a validated data store to store data received from the tested sensor that falls within the variance level and to store anomalous data received from the tested sensor that passes validation by the data validation module; and

a distribution module to facilitate distribution of data in the validated data store.

4. The system of claim 1, where the data validation module validates the data received from the tested sensor by comparing the data received from the tested

sensor to data received from sensors located within a specified physical proximity to the tested sensor, the specified physical proximity defined by a pattern.

5. The system of claim 1, where the attributes of the set of distributed sensors include:

attributes describing model information of the sensors;

attributes describing expected data to be received from the sensors; and

attributes describing expected anomalous behaviors of the sensors.

6. The system of claim 1, where the patterns indicating anomalous sensor activity include:

specific scenarios that describe a sensor behavior to be treated as anomalous; and

learned patterns of sensor behaviors to be treated as anomalous.

7. The system of claim 1, where the learning logic updates the pattern by modifying a weight associated with the pattern, the weight indicating a likelihood that anomalous data satisfying the pattern is a result of a sensor malfunction.

8. A method, comprising:

receiving sensor data from a set of distributed sensors;

marking sensor data as anomalous data when the anomalous data exceeds a variance level associated with a sensor from which the anomalous data was received;

validating, based on a set of patterns describing anomalous sensor activity, whether the anomalous data is a result of one of a sensor malfunction and an event of significance;

updating the set of patterns based on whether the anomalous data is validated as the sensor malfunction and the event of significance; and

distributing verified data comprising sensor data within the variance level and validated anomalous data.

9. The method of claim 7, further comprising:
logging anomalous data indicated as a sensor malfunction; and
revalidating logged anomalous data after an update to the set of patterns.

10. The method of claim 7, where distributing verified data comprises one or more of, providing raw data associated with a portion of the verified data, providing query access to the verified data, and providing a result of a query of the verified data.

11. The method of claim 7, where variance levels for sensors are generated based on device specifications describing known sensor malfunctions for respective sensors.

12. The method of claim 7, where the set of patterns includes weights that indicate a likelihood that anomalous data is a result of a sensor malfunction and where updating the set of patterns includes modifying the weights.

13. A non-transitory computer-readable medium storing computer-executable instructions that when executed by a computer cause the computer to:
receive sensor data from a set of distributed sensors;
mark sensor data as anomalous data when the anomalous data exceeds a variance level associated with a sensor from which the anomalous data was received;
store non-anomalous data in a validated data store;
store in the validated data store, anomalous data found, based on a set of patterns describing anomalous sensor activity, to be a result of an event of significance;

update the set of patterns; and

distribute data from the validated data store.

14. The non-transitory computer-readable medium of claim 13, where the instructions further cause the computer to:

store in an error logging data store, anomalous data found, based on the set of patterns, to be a result of a sensor malfunction; and

periodically evaluate whether data in the error logging data store is a result of an event of significance based on updated patterns.

15. The non-transitory computer-readable medium of claim 13, where the variance level for a sensor is based on accuracy information associated with the sensor and how recently the sensor was calibrated.

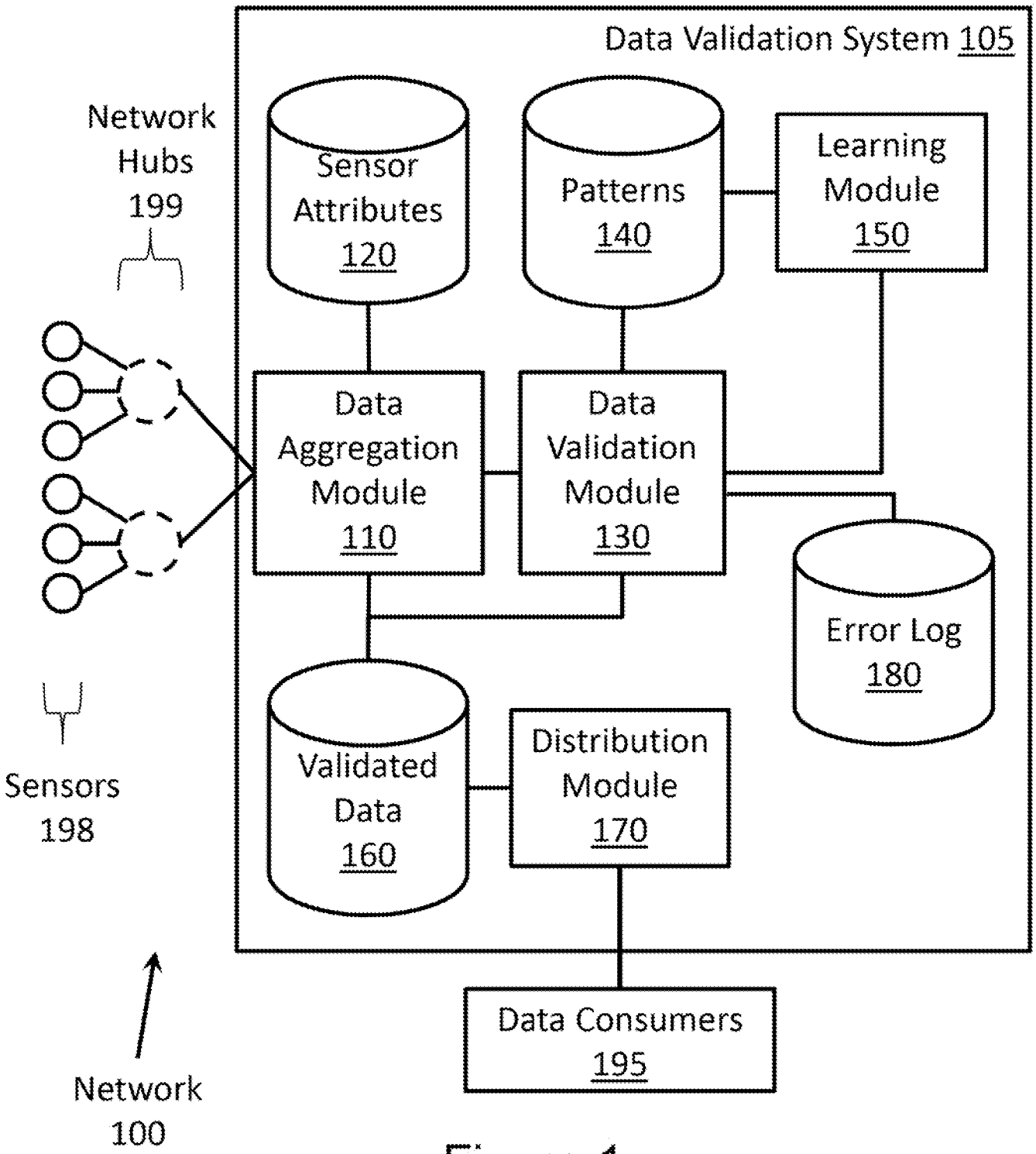
Data Validation System 105
Network Hubs 199
Sensor Attributes 120
Patterns 140
Learning Module 150
Data Aggregation Module 110
Data Validation Module 130
Error Log 180
Validated Data 160
Distribution Module 170
Data Consumers 195
Sensors 198
Network 100

Figure 1

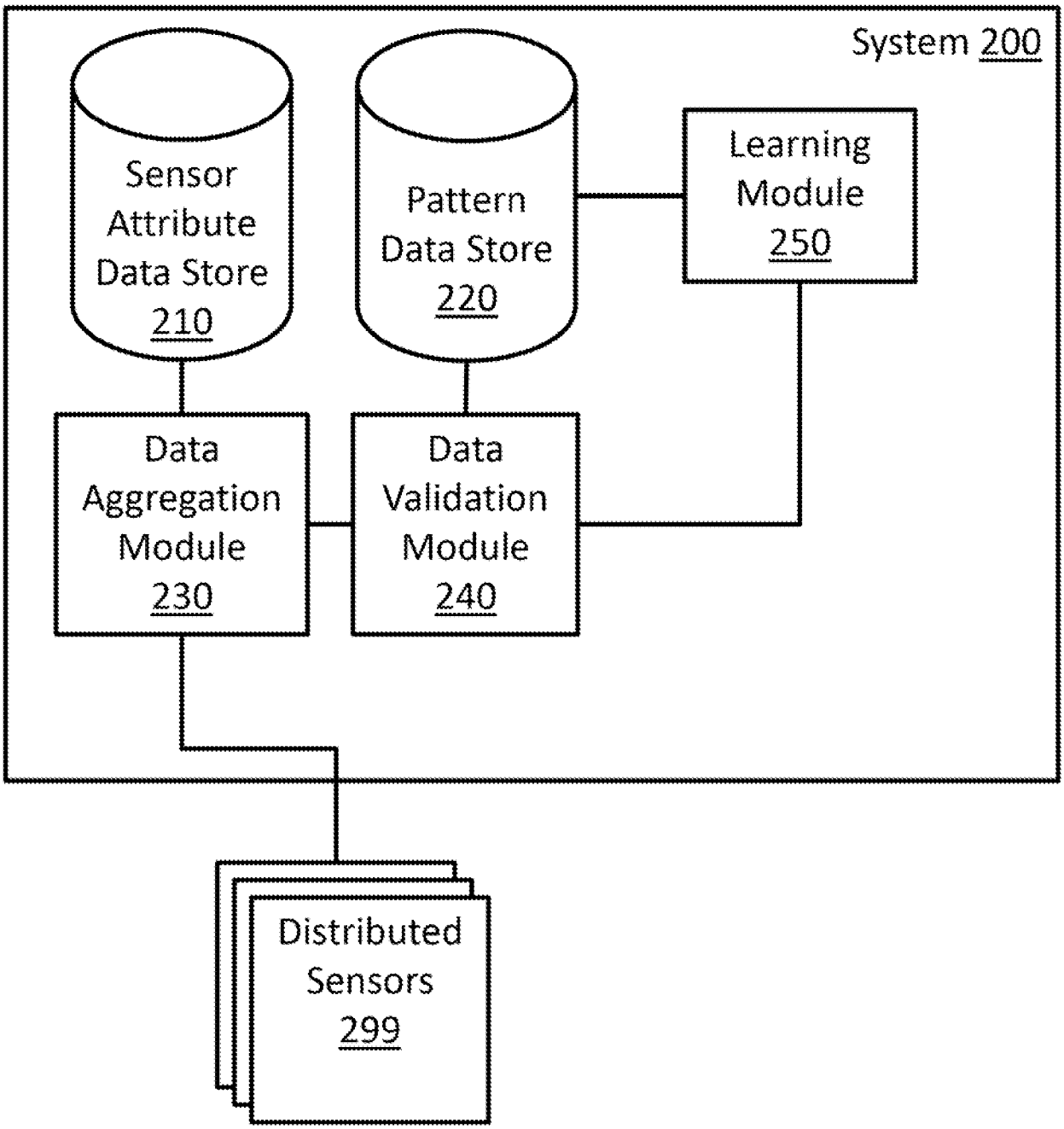
System 200
Sensor Attribute Data Store 210
Pattern Data Store 220
Learning Module 250
Data Aggregation Module 230
Data Validation Module 240
Distributed Sensors 299

Figure 2

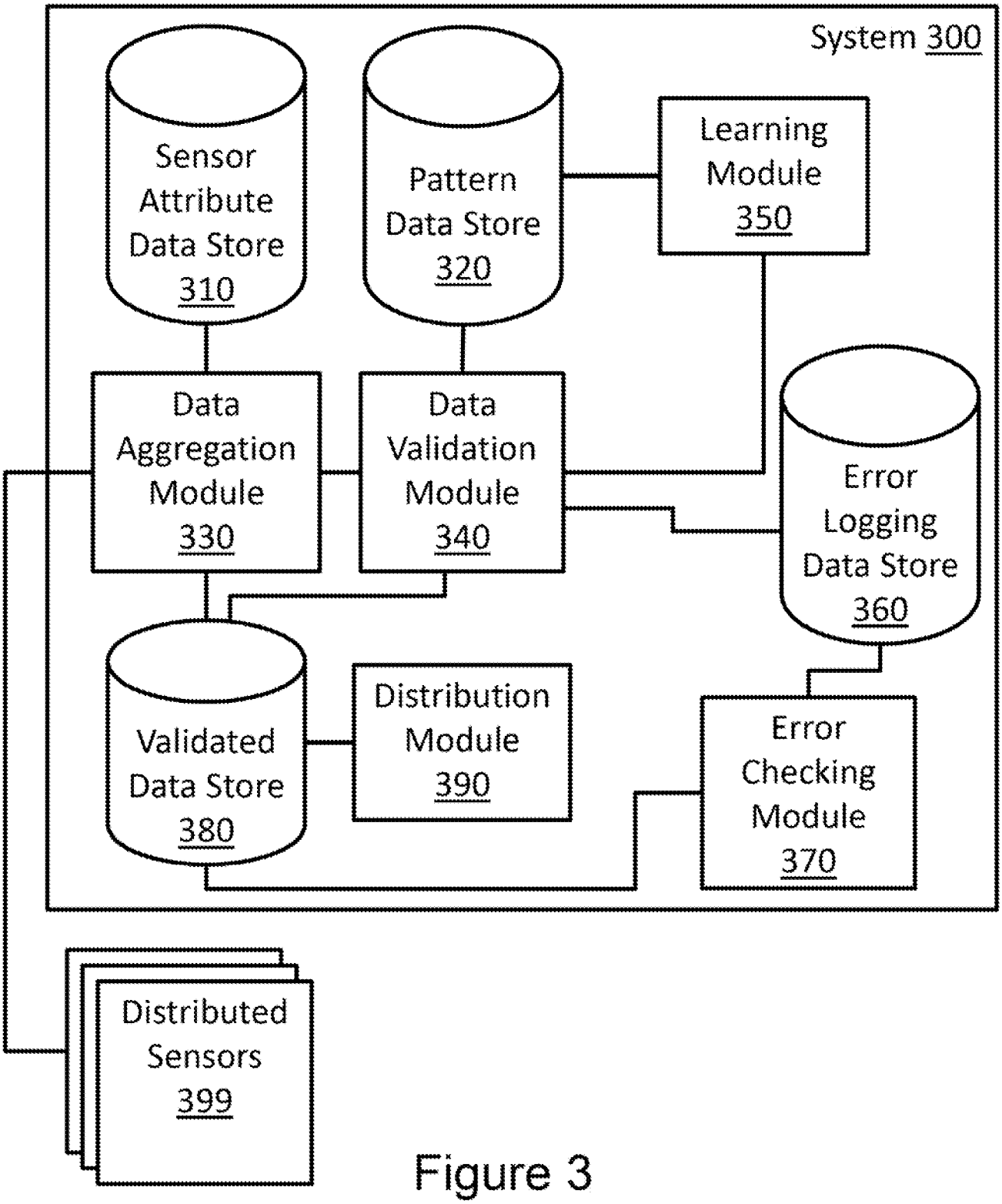
System 300
Sensor Attribute Data Store 310
Pattern Data Store 320
Learning Module 350
Data Aggregation Module 330
Data Validation Module 340
Error Logging Data Store 360
Validated Data Store 380
Distribution Module 390
Error Checking Module 370
Distributed Sensors 399

Figure 3

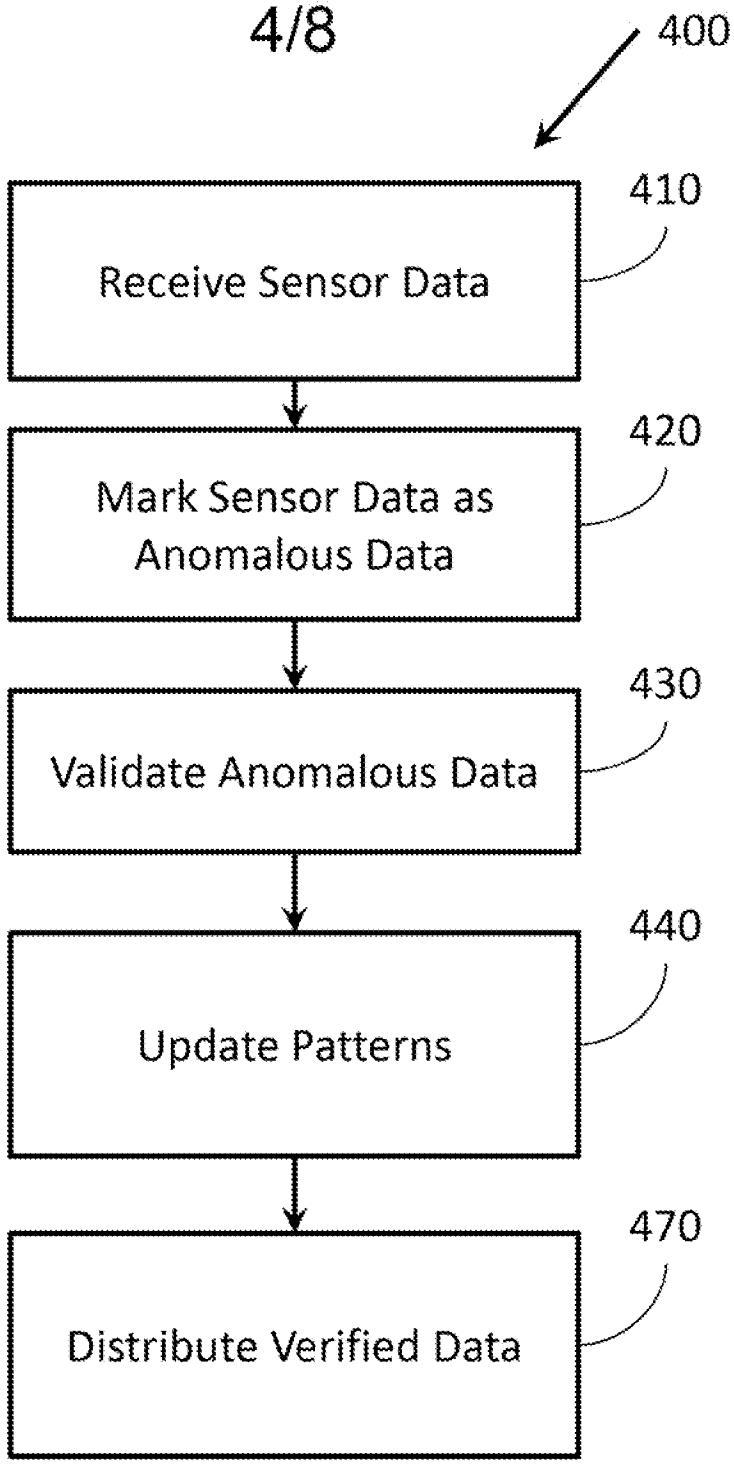
400
410
Receive Sensor Data
420
Mark Sensor Data as Anomalous Data
430
Validate Anomalous Data
440
Update Patterns
470
Distribute Verified Data

Figure 4

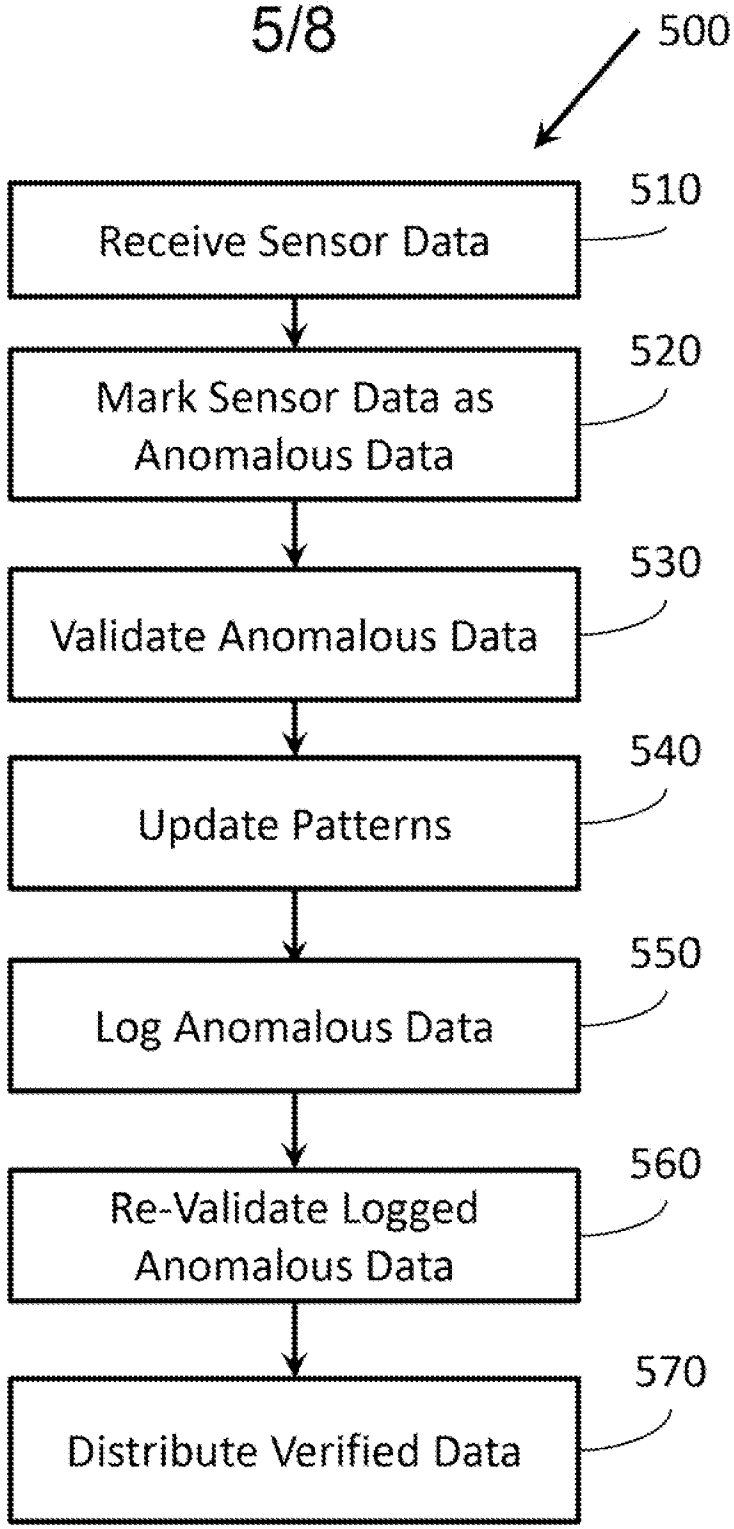
500
Receive Sensor Data
510
Mark Sensor Data as Anomalous Data
520
Validate Anomalous Data
530
Update Patterns
540
Log Anomalous Data
550
Re-Validate Logged Anomalous Data
560
Distribute Verified Data
570

Figure 5

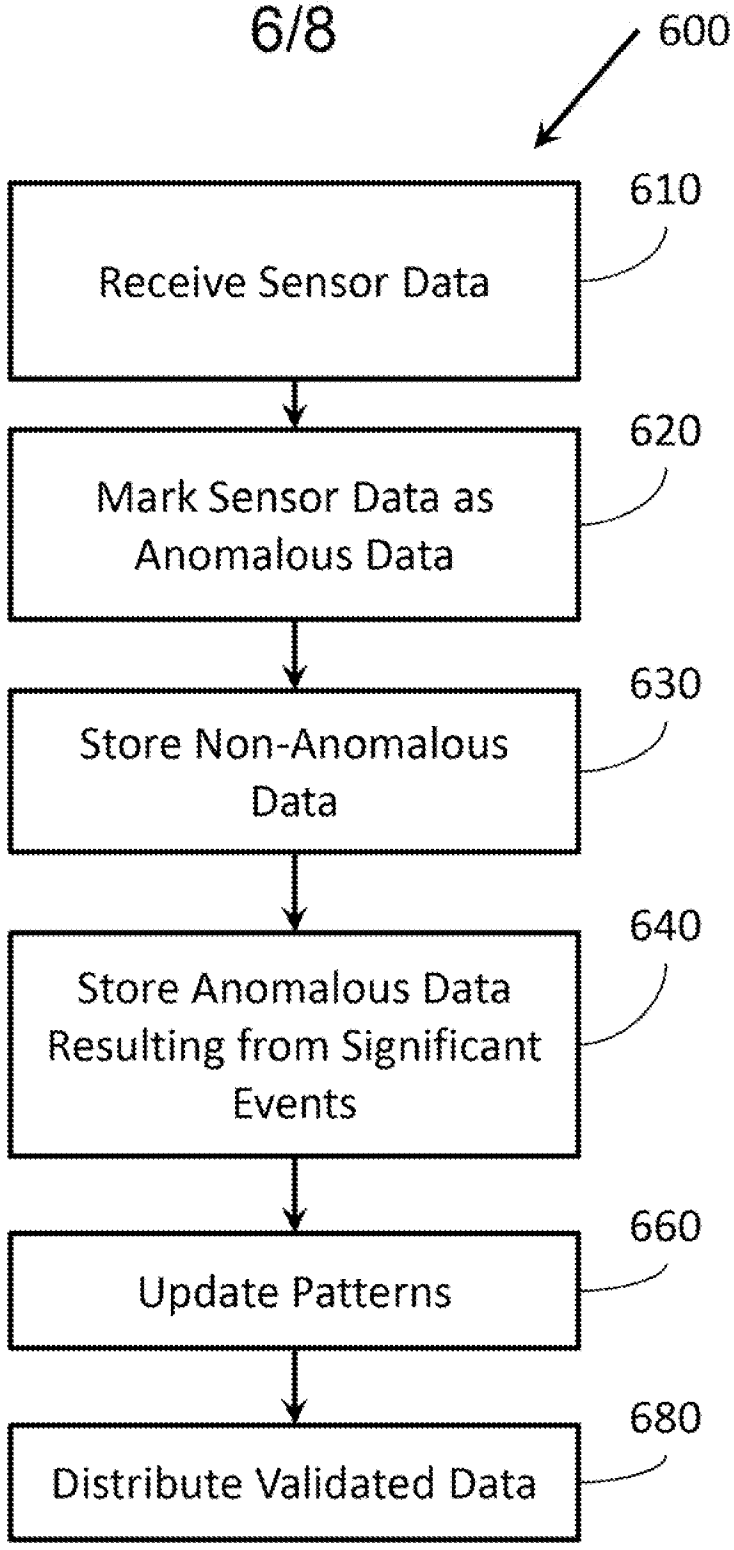
600
610
Receive Sensor Data
620
Mark Sensor Data as Anomalous Data
630
Store Non-Anomalous Data
640
Store Anomalous Data Resulting from Significant Events
660
Update Patterns
680
Distribute Validated Data

Figure 6

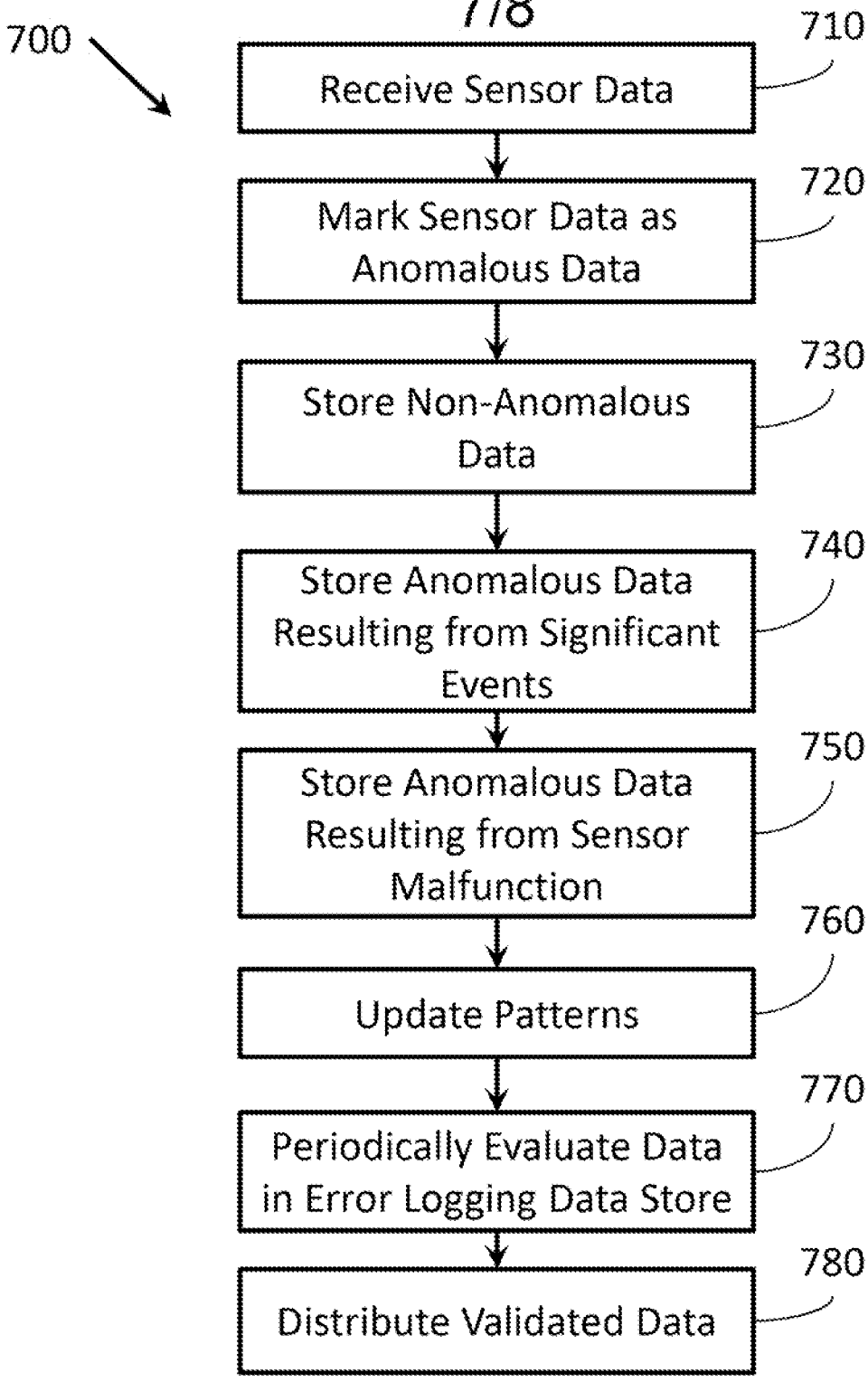
700
Receive Sensor Data
710
Mark Sensor Data as Anomalous Data
720
Store Non-Anomalous Data
730
Store Anomalous Data Resulting from Significant Events
740
Store Anomalous Data Resulting from Sensor Malfunction
750
Update Patterns
760
Periodically Evaluate Data in Error Logging Data Store
770
Distribute Validated Data
780

Figure 7

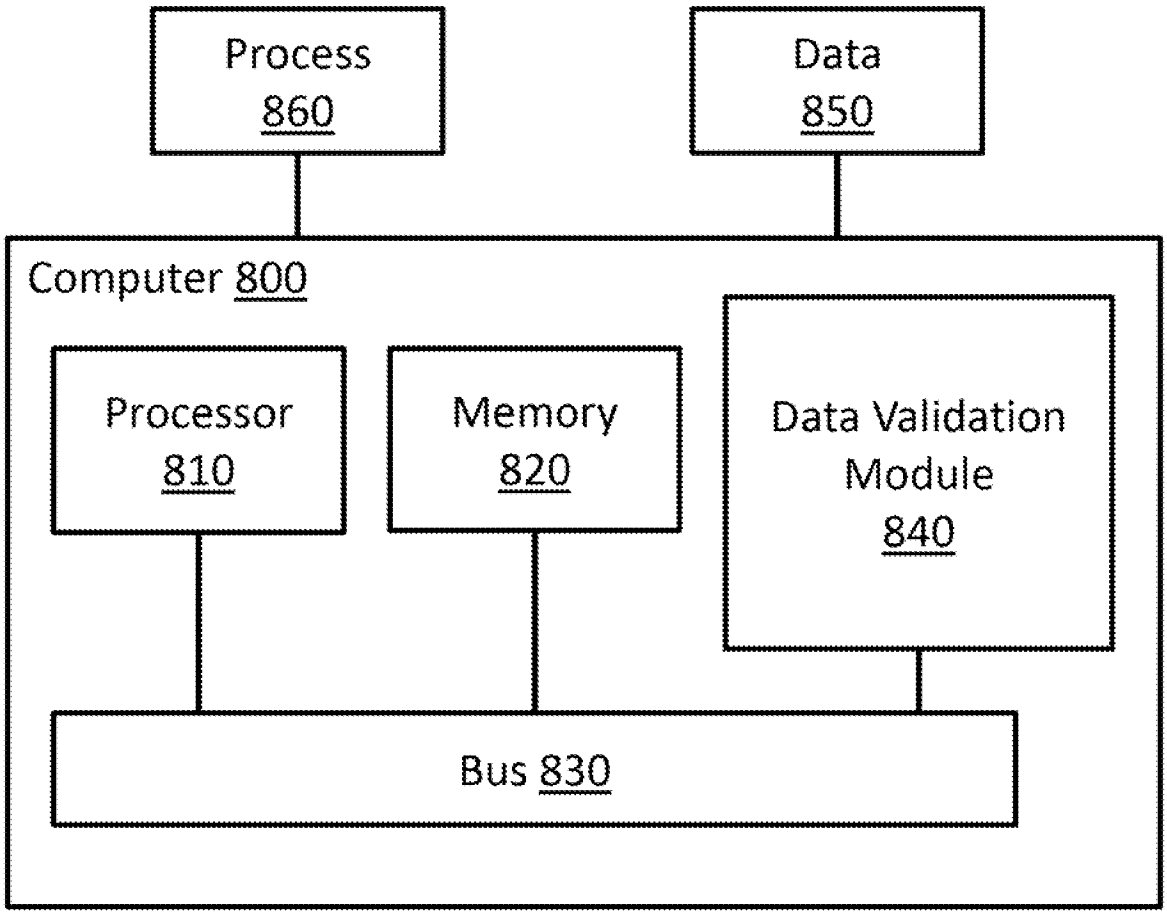
Process
860
Data
850
Computer 800
Processor
810
Memory
820
Data Validation
Module
840
Bus 830

Figure 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2015/032570

A. CLASSIFICATION OF SUBJECT MATTER

G06F 11/08(2006.01)i, G06N 3/08(2006.01)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)
G06F 11/08; G08B 19/00; G06F 19/00; G08B 1/08; G06F 17/00; G06F 17/30; G01D 3/02; G06N 5/02; G01P 21/00; G06N 3/08

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: data, validation, anomalous, sensor, pattern, compare, match, update, result, malfunction, normal, weight, level, and similar terms.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 7,230,528 B2 (LAWRENCE KATES) 12 June 2009 See column 8, lines 28-30; column 9, line 21 - column 10, line 3; column 11, line 49 - column 12, line 17; and figures 1 and 6.	1-15
A	US 7,603,222 B2 (MATTHEW WILLIAM WISEMAN et al.) 13 October 2009 See column 5, line 41 - column 6, line 7; claims 8 and 12; and figure 2.	1-15
A	US 8,341,106 B1 (HAGGAI SCOLNICOV et al.) 25 December 2012 See column 4, line 65 - column 9, line 49; and figure 1.	1-15
A	US 2012-0102002 A1 (VINAYA SATHYANARAYANA et al.) 26 April 2012 See paragraphs [0034]-[0039] and figure 1.	1-15
A	US 2007-0163324 A1 (LISA E. MCMAHAN et al.) 19 July 2007 See paragraphs [0022]-[0031] and figures 1-2.	1-15

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 March 2016 (30.03.2016)	**31 March 2016 (31.03.2016)**

Name and mailing address of the ISA/KR
International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea
Facsimile No. +82-42-481-8578

Authorized officer
HAN, JOONG SUB
Telephone No. +82-42-481-3578

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/US2015/032570

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 7230528 B2	12/06/2007	AU 2006-292764 A1	29/03/2007
		CA 2623003 A1	29/03/2007
		CN 101292273 A	22/10/2008
		EP 1927090 A1	04/06/2008
		JP 2009-509237 A	05/03/2009
		KR 10-2008-0045293 A	22/05/2008
		RU 2008114637 A	27/10/2009
		US 2007-0063833 A1	22/03/2007
		US 2007-0229237 A1	04/10/2007
		WO 2007-035219 A1	29/03/2007
US 7603222 B2	13/10/2009	CA 2568407 A1	18/05/2007
		CA 2568407 C	17/02/2015
		EP 1811133 A2	25/07/2007
		EP 1811133 A3	05/12/2012
		EP 1811133 B1	11/06/2014
		JP 2007-138937 A	07/06/2007
		JP 4948981 B2	06/06/2012
		US 2007-0118270 A1	24/05/2007
US 8341106 B1	25/12/2012	EP 2788724 A2	15/10/2014
		EP 2788724 A4	05/08/2015
		JP 2015-507245 A	05/03/2015
		JP 5666757 B1	19/12/2014
		KR 10-2014-0092385 A	23/07/2014
		US 2013-332090 A1	12/12/2013
		WO 2013-084068 A2	13/06/2013
		WO 2013-084068 A3	19/09/2013
US 2012-0102002 A1	26/04/2012	US 8468167 B2	18/06/2013
US 2007-0163324 A1	19/07/2007	EP 1728081 A1	06/12/2006
		JP 2007-530948 A	01/11/2007
		JP 2011-180154 A	15/09/2011
		US 2005-0210952 A1	29/09/2005
		US 7204123 B2	17/04/2007
		US 7661291 B2	16/02/2010
		WO 2005-098456 A1	20/10/2005