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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
01 March 2018 (01.03.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/039424 A1

(51) International Patent Classification:
G06F 17/30 (2006.01)

(21) International Application Number:
PCT/US2017/048360

(22) International Filing Date:
24 August 2017 (24.08.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/378,728	24 August 2016 (24.08.2016)	US
15/684,325	23 August 2017 (23.08.2017)	US
15/684,273	23 August 2017 (23.08.2017)	US

(71) Applicant: **NEC LABORATORIES AMERICA, INC.** [US/US]; 4 Independence Way, Suite 200, Princeton, NJ 08540 (US).

(72) Inventors: **XIAO, Xusheng**; 10900 Euclid Avenue, Cleveland, OH 44106 (US). **LI, Zhichun**; 306 Sayer Drive, Princeton, NJ 08540 (US). **ZHANG, Mu**; 2809 Ravens Crest Drive, Plainsboro, NJ 08536 (US). **JIANG, Guofei**; 5 Danby Court, Princeton, NJ 08540 (US). **GUI, Jiaping**; 2700 Ellendale Place, Apt. 216, Los Angeles, CA 90007 (US).

(74) Agent: **BITETTO, James, J.**; Tutunjian & Bitetto, P.C., 401 Broadhollow Road, Suite 402, Melville, NY 11747 (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH,

(54) Title: PROGRESSIVE PROCESSING FOR QUERYING SYSTEM BEHAVIOR

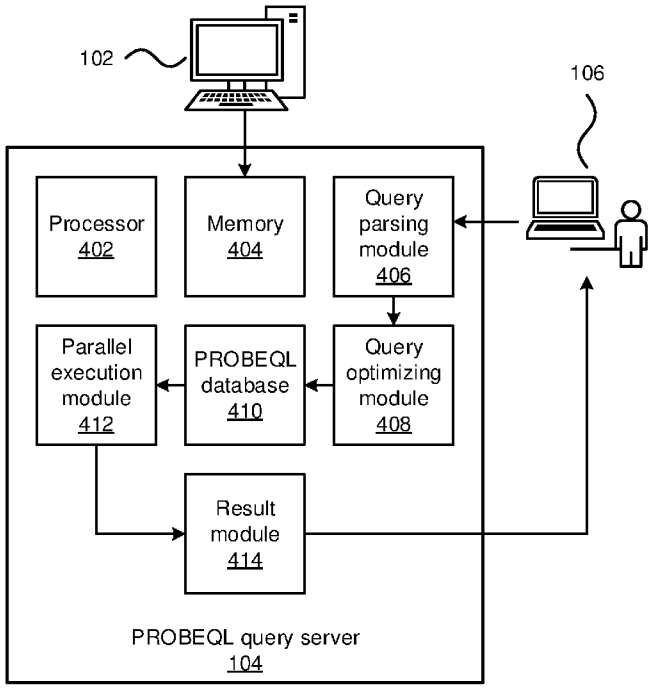

FIG. 4

(57) Abstract: Methods for querying a database and database systems include optimizing (304) a database query for parallel execution using spatial and temporal information relating to elements in the database, the optimized database query being split into sub-queries with sub-queries being divided spatially according to host and temporally according to time window. The sub-queries are executed (306) in parallel. The results of the database query are outputted (310) progressively.

WO 2018/039424 A1

[Continued on next page]

GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

PROGRESSIVE PROCESSING FOR QUERYING SYSTEM BEHAVIOR

RELATED APPLICATION INFORMATION

[0001] This application claims priority to U.S. Application Serial No. 62/378,728, filed on August 24, 2016, and to U.S. Application Serial Nos. 15/684,273 and 15/684,325, filed on August 23, 2017, incorporated herein by reference in its entirety.

BACKGROUND

Technical Field

[0002] The present invention relates to log management and, more particularly, to efficiently querying system monitoring data.

Description of the Related Art

[0003] Due to the complexity of modern computer systems and the many software programs that may be installed in a given system, system experts have limited visibility into the behaviors of the system. Application logs provide only partial information about the software's behaviors and therefore can be insufficient for monitoring the behaviors of those programs. Thus, relying solely on application logs can be overly limiting.

[0004] Furthermore, collecting data from system-level event audits produces such large amounts of information that searching the information can be difficult and time-consuming. Because multiple queries may be needed to gain an understanding of the state of the system in view of a particular need, searching through such a large amount of data imposes a heavy burden on the system operator.

SUMMARY

[0005] A method querying a database includes optimizing a database query for parallel execution using spatial and temporal information relating to elements in the database, the optimized database query being split into a plurality of sub-queries with sub-queries being divided spatially according to host and temporally according to time window. The sub-queries are executed in parallel. The results of the database query are outputted progressively.

[0006] A database system includes a query optimizing module that has a processor configured to optimize a database query for parallel execution using spatial and temporal information relating to elements in the database, the optimized database query being split into a plurality of sub-queries with sub-queries being divided spatially according to host and temporally according to time window. A parallel execution module is configured to execute the sub-queries in parallel. A results module is configured to output progressive results of the database query.

[0007] These and other features and advantages will become apparent from the following detailed description of illustrative embodiments thereof, which is to be read in connection with the accompanying drawings.

BRIEF DESCRIPTION OF DRAWINGS

[0008] The disclosure will provide details in the following description of preferred embodiments with reference to the following figures wherein:

[0009] FIG. 1 is a block diagram of a database query system that includes a progressive software behavioral query language (PROBEQL) query server in accordance with an embodiment of the present invention;

[0010] FIG. 2 is pseudo-code representing an online adaptive workload algorithm for splitting queries into sub-queries in accordance with an embodiment of the present invention;

[0011] FIG. 3 is a block/flow diagram of a method for optimizing and executing a PROBEQL query in accordance with an embodiment of the present invention;

[0012] FIG. 4 is a block diagram of a PROBEQL query server in accordance with an embodiment of the present invention;

[0013] FIG. 5 is a block diagram of a processing system in accordance with an embodiment of the present invention; and

[0014] FIG. 6 is a block diagram of an automated computer security system that identifies anomalous behavior and automatically performs a security control action in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

[0015] Embodiments of the present principles provide methods and systems for querying large amounts of log data using a domain-specific query language, referred to herein as Progressive Software Behavioral Query Language (PROBEQL). PROBEQL provides a query syntax that specifies event patterns for software behaviors. A query engine optimizes searches over the logs based on the domain-specific characteristics of the system monitoring data. PROBEQL further reports query results progressively, so that even lengthy searches can provide quick and continuous feedback.

[0016] The present embodiments have one application in computer security monitoring, where a PROBEQL system is used to efficiently execute queries dealing with security threats based on monitored system data, providing progressive results to

the user when the execution time is long, thereby permitting a user to rapidly adjust query terms in accordance with preliminary results.

[0017] Referring now in detail to the figures in which like numerals represent the same or similar elements and initially to FIG. 1, monitoring and analysis system 100 is shown. A set of monitored computer systems 102 may include any appropriate system, including mainframes, desktop workstations, laptops, embedded devices, single-purposes devices, network switches and routers, and any other device that may be part of a computer system or network that runs software and generates logged event information.

[0018] The system monitoring data generated by each monitored computer system 102 records the interactions between software programs and system resources as events. Each recorded interaction represents a system event that occurs at a particular host and includes information regarding the initiator of the interaction, the type of interaction, and the target of the interaction. Initiating processes originate from software programs (e.g., a web browser), while targets are system resources, such as files, other processes, and network connections. Each interaction is associated with a timestamp that indicates when the interaction occurred.

[0019] To monitor such event information, each monitored system 102 may implement a data collection agent to audit events about system calls. The monitored system calls are mapped to at least three categories of event, including process creation and termination, file accesses, and network accesses. The data may be stored in relational databases that come with mature indexing mechanisms and that are scalable to large amounts of data. The data is stored in partitions based on its temporal and spatial properties, with groups of hosts being separated into table partitions and with databases being dumped periodically (e.g., daily) for temporal grouping.

[0020] The logged event information is sent to a PROBEQL query server 104. It should be understood that the PROBEQL query server 104 may be implemented as a standalone device or may, in other embodiments, be implemented as a component within one or more of the monitored computer systems 102. A user 106 interfaces with the PROBEQL query server 104 to perform searches of the stored event information.

[0021] In particular, the user 106 creates a PROBEQL query and sends the PROBEQL query to the PROBEQL query server 104. It is specifically contemplated that PROBEQL has a syntax format of {*subject-operation-object*}, though it should be understood that other syntaxes may be used instead. The PROBEQL syntax specifies an event pattern for software behavior, where software programs are represented as subjects, system resources are represented as objects, and interactions are represented as operations that are initiated by a subject to target an object. An exemplary query is shown below:

[0022] host = 1 // host id

[0023] (from "02/01/2016" to "02/07/2016") // time window

[0024] proc p1 write file f1['/var/log/wtmp'||'/var/log/lastlog'] // event pattern

[0025] return distinct p1, f1 // result projection

[0026] update 5s // progressive update frequency

[0027] In this exemplary query, the PROBEQL syntax models software behaviors directly. The query looks for logs where a process p1 (the subject) write (the operation) to one of two files (the objects), returning each distinct pair of subject and object. The query specifies that it is to output its results progressively every five seconds. This allows the user to make changes to the query (e.g., filtering or stopping the query altogether) if the early results show that the query is not providing the information that is needed.

[0028] An exemplary partial PROBEQL grammar is set forth below. Although it is contemplated that the PROBEQL parsing system will include these features, it should be understood that the grammar described below may be altered, added to, or removed from as is appropriate to the specific needs at hand:

[0029] <probeql> ::= (<evt_cstr>)* <query>

[0030] <evt_cstr> ::= <cstr> | '(' <time_window> ')'

[0031] <query> ::= <evt_patt>+ <return> <progress>? <t_num>?

[0032] <evt_patt> ::= <entity> <op_exp> <entity> <evt>?

[0033] <entity> ::= <type> <entity_id>? ('[' <attr_cstr> ']')?

[0034] <return> ::= 'return' 'distinct'? <res> (',' <res>)* ', count'?

[0035] <res> ::= <entity_id>('.'<attr_name>)?

[0036] <progress> ::= 'update' <val> <timeunit>

[0037] <t_num> ::= 'using' <val> 'worker'

[0038] In this grammar, a PROBEQL query includes two major parts: event constraints <evt_cstr>, which specify constraints for hosts and time windows, and event patterns <evt_patt>, which specify subject, objects, operations, and optional event identifiers. The subject and object are specified as <entity> entries, including entity type <type>, an optional entity identifier <entity_id>, and optional attributes. The exemplary query above specifies a process entity p and a file entity f1. The operation <op_exp> specifies the operation initiated by the subject and targeting the object, such as reading a file. Logical operators (e.g., "," for AND, "||" for OR, "!" for NOT, etc.) can be used to combine multiple operations. The event return <return> specifies which attributes of the found events to return. Based on the events' attributes that are specified in the event return rule, the result tuples are output.

[0039] The PROBEQL query server 104 receives the PROBEQL query and executes it, optimizing the search based on domain-specific characteristics of the system monitoring data. Each logged system event is generated with a timestamp at a specific monitored system 102. These events exhibit strong spatial and temporal properties, where events in different monitored systems 102 are independent and events within a single monitored system 102 may be independent over time. Based on the spatial and temporal properties of the logged data, at least two types of parallel strategies are contemplated to partition a query into sub-queries, uniformly splitting a time window or workload. The PROBEQL query server executes the sub-queries in parallel.

[0040] Even with the sub-queries being executed in parallel and the increase in performance that results, the speed of the query is still limited by the speed of the hardware, and executing a search over a large amount of data will still take a long time. As noted above, progressive updating of query results helps the user 106 make rapid use of incomplete query outputs. Different update frequencies may be appropriate for different types of queries. For example, a query that looks for the existence of a behavior may need a shorter update frequency (e.g., every two seconds), while a query that looks for the exact files that a specific program writes may need only infrequent updates so that more results are collected for analysis.

[0041] In a PROBEQL query, the progressive processing clause <progress> specifies the update frequency, while thread number <t_num> specifies the number of worker threads used to execute subqueries. If the update frequency is not specified, the query execution will not report results until the end of the search, allowing the PROBEQL query server 104 to fully leverage the spatial and temporal properties of the system monitoring data to optimize the overall execution time, parallelizing the execution of

the PROBEQL query. If the thread number is not specified, the optimal number is inferred based on the number of events to be searched over.

[0042] The present embodiments provide at least four strategies for partitioning a workload into sub-queries for execution, each with smaller time windows and fewer involved hosts. The partitioning strategies described below include uniform time window, parallel workload, sequential workload, and sequential workload with initialization cost.

[0043] As events in system monitoring data are independent over time, one straightforward partitioning strategy is to uniformly split the time window to each sub-query. This strategy is referred to herein as uniform time window partitioning. However, uniform time window partitioning usually does not split the workload fairly for each sub-query. In practice, hosts often produce events at different rates for different times in the day.

[0044] Another set of strategies fall into a category of uniform workload partitioning. A first uniform workload strategy is parallel workload partitioning, where time windows are split into smaller time windows such that the number of events in each time window are the same. A second uniform workload strategy is sequential workload partitioning, which sequentially concatenates the events of all the hosts as an array of events and then divides the array uniformly for each sub-query.

[0045] While these two strategies uniformly partition the workload of a query, neither provides the same execution time among all sub-queries due to the initialization cost associated with accessing the data of a host for the first time. Subsequent accesses benefit from caching, reducing or eliminating the initialization cost going forward. Thus, because different sub-queries may have different initialization costs, the execution time may differ.

[0046] A third uniform workload strategy is thus sequential workload with initialization cost partitioning, which takes into account the initialization cost for the first access to each host's data. This strategy converts the initialization time to a virtual workload and inserts such virtual workloads into the workloads of each hosts. Sequential workload partitioning is then applied on the combined workloads to provide fair partitions.

[0047] In general, sequential workload with initialization cost partitioning performs better than sequential workload partitioning, which performs better than parallel workload partitioning, which in turn performs better than uniform time window partitioning. Parallelism improves query execution time, but there are diminishing returns when more worker threads are used than read/write channels are available in the PROBEQL query server 104.

[0048] In addition to partitioning to increase processing speed, to ensure quality in the progressive processing, the PROBEQL query server 104 partitions the query so that new results are reported in every update and so that the overhead is kept low. If no partitioning of the query is performed, a result can only be provided at the end. On the other hand, very frequent updates incur an unacceptably high overhead due to the cost of establishing connections to databases and parsing the sub-queries. To address this problem, some embodiments partition workloads to sub-queries such that each sub-query takes an amount of time equal to or just less than the requested update frequency. The present embodiments therefore employ an adaptive workload partition strategy, dynamically predicting the workloads for subsequent sub-queries based on the latest execution information of already finished sub-queries.

[0049] In one scenario, including an enterprise network of one hundred monitored systems 102, a set of commonly used queries for real-world suspicious behaviors were

issued over monitored system event data. The time window of the queries was set to cover five days of data (representing about 230GB of monitored system event data) with various update frequencies (e.g., 5 seconds, 10 seconds, 15 seconds, etc.). The results show that the adaptive workload partition strategy progressively provided the partial results and was 32.8% faster than a regular SQL query. Even compared to the parallel execution of the query without progressive processing and using four worker threads, the overhead of the progressive processing was only 1.2%.

[0050] Adaptive workload partitioning furthermore provides execution of sub-queries that have low deviations from the requested update frequencies relative to fixed time window and fixed workload partitioning strategies. Building off sequential workload partitioning with initialization cost partitioning for parallelization, the following partitioning strategies provide adaptive workload partitioning within the sequential workload partitioning with initialization cost framework to provide partitioning strategies that complete sub-queries with an execution time that is close to the requested progressive update frequency.

[0051] A first type of adaptive partitioning is fixed time window partitioning. For a query searching for events in n hosts $(h_1, \ldots, h_n)$, in the time window T, the query is based on a fixed time window T_i for events in the host h_i. T_i may be computed as $U_f ep_i / g_i$, where U_f is the update frequency, ep_i is the event processing rate of h_i, and g_i is the average data generating rate computed using the total events of h_i, divided by the time window T.

[0052] A second type of adaptive partitioning is fixed workload partitioning. For a query searching for events in n hosts $(h_1, \ldots, h_n)$, the query is partitioned based on a fixed workload Wd_i for events in each host h_i. Wd_i may be computed as $U_f ep_i$, where

U_f is the update frequency, ep_i is the event processing rate of h_i. The initialization costs are considered to be workloads unto themselves in each host.

[0053] Both fixed time window partitioning and fixed workload partitioning assume that event processing rates are constant. However, as database systems often employ caches, which greatly speed processing rates for events in the cache, the event processing rate can fluctuate significantly during runtime. The execution time of the sub-queries for both fixed time window partitioning and fixed workload partitioning can be far from the requested update frequency.

[0054] To address this, adaptive workload partitioning can dynamically adjust event processing rates during runtime. In particular, online adaptive workload prediction partitioning leverages a set of latest <workload, execution_time> pairs as feedback to learn new event processing rates and to predict workloads for subsequent sub-queries. Gradient descent may be used to guide the adaptive learning.

[0055] The goal of learning in online adaptive workload prediction is to adjust the event processing rate obtained in a non-cached environment to fit the latest execution results. A new data processing rate *ep* and initialization time *it* are computed that approximate the actual event processing rate in the current running environment. In other words, the local minimum ep', it' are found such that, for each new execution result <*x*,*y*>, where *x* is the execution time and *y* is the workload, the execution time x' of the next estimated workload y' is closest to the update frequency U_f. To compute the gradient *g* with respect to the new training data set **S**, a loss function is used where <*x*,*y*> are the execution results in **S** having a size *N*.

[0056] By taking derivatives of the equation:

$$loss(x, y) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - (epx_i + it)\right)^2$$

the gradient g is obtained as:

$$g_a = \frac{\partial loss(x,y)}{\partial ep} = \frac{2}{N}\sum_{i=1}^{N} -x_i(y_i - epx_i + it)$$

$$g_b = \frac{\partial loss(x,y)}{\partial it} = \frac{2}{N}\sum_{i=1}^{N} -(y_i - epx_i + it)$$

After computing the gradient, the event processing rate and initialization time are updated as $ep' = ep - \gamma g_a$ and $it' = it - \gamma g_b$, where the learning rate γ controls the weight of the latest execution results over the historical results.

[0057] To avoid over-fitting, restrictions are placed on the bounds of the newly learned event processing rate ep'. Over-fitting causes the prediction of a large workload, for which the sub-query has no way to return results within the update frequency. The present regularization uses the offline-measured event processing rates as the lower bound. For the upper bound, the instant event processing rate y/x is calculated for each latest execution and the largest observed instant event process rate is used as the upper bound.

[0058] Referring now to FIG. 2, pseudo-code for an online adaptive workload algorithm is shown. Online adaptive workload partitioning follows the structure of sequential workload partitioning with initialization cost by concatenating the events of all involved hosts as an array *Evt*, concatenating initialization costs as virtual events, and processing the events in the array sequentially. Inputs include the update frequency U_f, the current host index k, the index of the first unprocessed event S_i in *Evt*, the host measurement/prediction M, the latest execution information ΔE, the observed maximum instant event processing rate ep_{max}, and the learning rate γ. The host/prediction M stores the information of the involved hosts' offline measured (or predicted) event processing rates and initialization times.

[0059] The algorithm shown in FIG. 2 initializes the predicted workload D to zero and the remaining time within the update frequency U_f' to U_f. If S_i indicates that it is the first time the data of host h_k is accessed, the event processing rate ep and the initialization time it are retrieved from $M[k]$. In this case, if it is larger than U_f', such that even without any workload the execution of the sub-query will not finish within U_f', the default event size is assigned as the predicted workload D and D is returned. If this is not the first time that h_k is accessed, the initialization time is set to zero and adaptive learning is used to learn a new event processing rate. The maximum event processing rate ep_{max} is then updated with the instant event processing rate. Based on ep, the algorithm predicts the workload D'. To avoid over-fitting, regularization is applied to smooth D' if needed.

[0060] After regularization, it is determined whether D' exceeds the remaining size of the current host h_k. If so, the execution time of the remaining workload in the host is computed and deducted from U_f'. A new iteration is started to predict the workload for a new host h_{k+1}. Otherwise, the predicted workload D is updated with D' and is returned. The algorithm then updates S_i *based on* D.

[0061] The learning rate γ controls the fitting degree of the gradient descent. A learning rate that is selected to be too high or too low causes over- or under-fitting that inaccurately partitions the workload and, hence, makes the average sub-query execution time deviate from the requested update frequency. It has been empirically shown that learning rates of about $\gamma = .0005$ achieve the best performance in queries with more than twenty sub—queries. It should be understood that this value for the learning rate is purely exemplary and that particular embodiments may benefit from a lower or higher learning rate.

[0062] Providing progressive results using online adaptive workload partitioning as described above produces results close to the requested update frequency with a total execution time for queries that is significantly faster than unoptimized execution. Relative to sequential workload partitioning with initialization cost strategies that do not provide progressive results, online adaptive workload partitioning incurs a negligible overhead on total execution time.

[0063] Referring now to FIG. 3, a block/flow diagram of the execution of a PROBEQL query is shown. Block 302 receives and parses a PROBEQL query from a user 106. As noted above, the PROBEQL query may have a structure that specifies a subject, an operation, and an object, as well as a progressive update frequency. Block 304 optimizes the PROBEQL query for parallel execution, splitting the PROBEQL into multiple sub-queries, taking into account the independence of events that take place across time and across different hosts (e.g., their spatial and temporal properties).

[0064] Toward this end, the optimization splits the PROBEQL query according to, e.g., sequential workload partitioning with initialization cost. Block 306 then executes the sub-queries in parallel. If progressive updates are requested in the PROBEQL query, block 308 provides progressive updates of the execution results, with the partitioning of block 306 being directed in particular to an adaptive learning strategy that provides sub-query results on timescales close to the requested update frequency. Once the query execution has completed, with all sub-queries having been executed, block 310 provides the final query results.

[0065] Embodiments described herein may be entirely hardware, entirely software or including both hardware and software elements. In a preferred embodiment, the present invention is implemented in software, which includes but is not limited to firmware, resident software, microcode, etc.

[0066] Embodiments may include a computer program product accessible from a computer-usable or computer-readable medium providing program code for use by or in connection with a computer or any instruction execution system. A computer-usable or computer readable medium may include any apparatus that stores, communicates, propagates, or transports the program for use by or in connection with the instruction execution system, apparatus, or device. The medium can be magnetic, optical, electronic, electromagnetic, infrared, or semiconductor system (or apparatus or device) or a propagation medium. The medium may include a computer-readable storage medium such as a semiconductor or solid state memory, magnetic tape, a removable computer diskette, a random access memory (RAM), a read-only memory (ROM), a rigid magnetic disk and an optical disk, etc.

[0067] Each computer program may be tangibly stored in a machine-readable storage media or device (e.g., program memory or magnetic disk) readable by a general or special purpose programmable computer, for configuring and controlling operation of a computer when the storage media or device is read by the computer to perform the procedures described herein. The inventive system may also be considered to be embodied in a computer-readable storage medium, configured with a computer program, where the storage medium so configured causes a computer to operate in a specific and predefined manner to perform the functions described herein.

[0068] A data processing system suitable for storing and/or executing program code may include at least one processor coupled directly or indirectly to memory elements through a system bus. The memory elements can include local memory employed during actual execution of the program code, bulk storage, and cache memories which provide temporary storage of at least some program code to reduce the number of times code is retrieved from bulk storage during execution. Input/output or I/O devices

(including but not limited to keyboards, displays, pointing devices, etc.) may be coupled to the system either directly or through intervening I/O controllers.

[0069] Network adapters may also be coupled to the system to enable the data processing system to become coupled to other data processing systems or remote printers or storage devices through intervening private or public networks. Modems, cable modem and Ethernet cards are just a few of the currently available types of network adapters.

[0070] Referring now to FIG. 4, additional detail on the PROBEQL query server 104 is shown. The PROBEQL query server 104 includes a hardware processor 402 and a memory 404. The hardware processor 402 may be, for example, a single device having multiple processor cores, may be multiple discrete processor devices, or may be a single, unitary processor. The PROBEQL query server 104 further includes functional modules that may, in some embodiments, be implemented as software that is stored in the memory 404 and that is executed by processor 402. In other embodiments, the functional modules may be implemented as one or more discrete hardware components in the form of, e.g., application specific integrated chips or field programmable gate arrays.

[0071] A query parsing module 406 receives a PROBEQL query from a user 106 and parses the query according to the PROBEQL grammar, identifying the subject, operation, and object of the query as well as specifying, for example, a progressive update frequency. A query optimizing module 408 then breaks the PROBEQL query into sub-queries that are optimized for parallel execution.

[0072] A PROBEQL database 410 is stored in memory 404 as a relational database, storing the information collected from the monitored systems 102. A parallel execution module 412 executes the sub-queries on the PROBEQL database 410 in parallel using

processor 402 and/or multiple other processing elements. The execution of each sub-query produces results which may be optionally displayed to the user 106 by result module 414 in a progressive fashion. If progressive results are requested by the user 106, the query optimizing module creates sub-queries that have an expected execution time that is close to the requested progressive result update frequency, such that the result module 414 can display the results of each sub-query as requested.

[0073] Referring now to FIG. 5, an exemplary processing system 500 is shown which may represent the PROBEQL query server 104. The processing system 500 includes at least one processor (CPU) 504 operatively coupled to other components via a system bus 502. A cache 506, a Read Only Memory (ROM) 508, a Random Access Memory (RAM) 510, an input/output (I/O) adapter 520, a sound adapter 530, a network adapter 540, a user interface adapter 550, and a display adapter 560, are operatively coupled to the system bus 502.

[0074] A first storage device 522 and a second storage device 524 are operatively coupled to system bus 502 by the I/O adapter 520. The storage devices 522 and 524 can be any of a disk storage device (e.g., a magnetic or optical disk storage device), a solid state magnetic device, and so forth. The storage devices 522 and 524 can be the same type of storage device or different types of storage devices.

[0075] A speaker 532 is operatively coupled to system bus 502 by the sound adapter 530. A transceiver 542 is operatively coupled to system bus 502 by network adapter 540. A display device 562 is operatively coupled to system bus 502 by display adapter 560.

[0076] A first user input device 552, a second user input device 554, and a third user input device 556 are operatively coupled to system bus 502 by user interface adapter 550. The user input devices 552, 554, and 556 can be any of a keyboard, a

mouse, a keypad, an image capture device, a motion sensing device, a microphone, a device incorporating the functionality of at least two of the preceding devices, and so forth. Of course, other types of input devices can also be used, while maintaining the spirit of the present principles. The user input devices 552, 554, and 556 can be the same type of user input device or different types of user input devices. The user input devices 552, 554, and 556 are used to input and output information to and from system 500.

[0077] Of course, the processing system 500 may also include other elements (not shown), as readily contemplated by one of skill in the art, as well as omit certain elements. For example, various other input devices and/or output devices can be included in processing system 500, depending upon the particular implementation of the same, as readily understood by one of ordinary skill in the art. For example, various types of wireless and/or wired input and/or output devices can be used. Moreover, additional processors, controllers, memories, and so forth, in various configurations can also be utilized as readily appreciated by one of ordinary skill in the art. These and other variations of the processing system 500 are readily contemplated by one of ordinary skill in the art given the teachings of the present principles provided herein.

[0078] Referring now to FIG. 6, an automated computer security system is shown. A set of monitored systems 12, which may for example include desktop computers, servers, individual applications, mobile devices, laptops, etc., are each monitored by one or more system monitors 13 that may represent, for example, software that probe system states, logging systems, physical sensors that measure physical states of the monitored systems 12, etc.

[0079] The system monitors 13 output the information they collect to a PROBEQL database server 14. The PROBEQL database server 14 receives PROBEQL queries from a user 16, optimizing the queries for execution across the large volume of collected data and providing progressive results. In another embodiment, an automated security control system 18 can provide queries to the PROBEQL database server 14 that assess particular risks, for example automatically searching for signs of intrusion. These automated queries may be executed periodically or may be triggered by meeting some triggering condition (e.g., the receipt of a security alert or a change in security policies).

[0080] The security control system 18 can then use the progressive results provided by the PROBEQL database server 14 to execute a security control action at the monitored systems 12. Examples of security control actions include, for example, changing a security level, powering devices on and off, managing network security policies (e.g., by restricting certain kinds of traffic or access to certain ports), and issuing alerts. The user 16 may also interact with the security control system 18 to manually trigger such security control actions based on the results of the user's PROBEQL queries. By using progressive query results, security control actions can be issued rapidly to respond to ongoing security threats.

[0081] The foregoing is to be understood as being in every respect illustrative and exemplary, but not restrictive, and the scope of the invention disclosed herein is not to be determined from the Detailed Description, but rather from the claims as interpreted according to the full breadth permitted by the patent laws. It is to be understood that the embodiments shown and described herein are only illustrative of the principles of the present invention and that those skilled in the art may implement various modifications without departing from the scope and spirit of the invention. Those skilled in the art could implement various other feature combinations without departing

from the scope and spirit of the invention. Having thus described aspects of the invention, with the details and particularity required by the patent laws, what is claimed and desired protected by Letters Patent is set forth in the appended claims.

WHAT IS CLAIMED IS:

1. A method for querying a database, comprising:

optimizing (304) a database query for parallel execution using spatial and temporal information relating to elements in the database, the optimized database query being split into a plurality of sub-queries with sub-queries being divided spatially according to host and temporally according to time window;

executing (306) the sub-queries in parallel; and

outputting (310) progressive results of the database query.

2. The method of claim 1, wherein the database query comprises a subject, an operation, and an object that the subject operates on.

3. The method of claim 1, wherein the database query comprises an update frequency.

4. The method of claim 3, wherein optimizing the database query comprises splitting the database query into sub-queries that have an expected execution time based on the update frequency.

5. The method of claim 4, wherein optimizing the database query comprises adapting the expected execution time based on previous sub-query execution times.

6. The method of claim 3, wherein outputting progressive results of the database query comprises outputting results from executed sub-queries in in accordance with the update frequency.

7. The method of claim 1, wherein optimizing the database query comprises splitting the database query into sub-queries in accordance with a sequential workload partitioning with initialization cost strategy.

8. The method of claim 7, wherein the sequential workload partitioning with initialization cost strategy is online adaptive workload prediction partitioning.

9. The method of claim 7, wherein optimizing the database query comprises compute initialization costs as separate workloads for the purpose of partitioning.

10. The method of claim 1, further comprising:

monitoring system state information from a plurality of monitored systems;

storing the monitored system state information in the database; and

performing a security control action in accordance with the progressive results.

11. A database system, comprising:

a query optimizing module (408) comprising a processor configured to optimize a database query for parallel execution using spatial and temporal information relating to elements in the database, the optimized database query being

split into a plurality of sub-queries with sub-queries being divided spatially according to host and temporally according to time window;

a parallel execution module (412) configured to execute the sub-queries in parallel; and

a results module (414) configured to output progressive results of the database query.

12. The system of claim 11, wherein the database query comprises a subject, an operation, and an object that the subject operates on.

13. The system of claim 11, wherein the database query comprises an update frequency.

14. The system of claim 13, wherein the query optimizing module is further configured to split the database query into sub-queries that have an expected execution time based on the update frequency.

15. The system of claim 14, wherein the query optimizing module is further configured to adapt the expected execution time based on previous sub-query execution times.

16. The system of claim 13, wherein the results module is further configured to output progressive results from executed sub-queries in in accordance with the update frequency.

17. The system of claim 11, wherein the query optimizing module is further configured to split the database query into sub-queries in accordance with a sequential workload partitioning with initialization cost strategy.

18. The system of claim 17, wherein the sequential workload partitioning with initialization cost strategy is online adaptive workload prediction partitioning.

19. The system of claim 17, wherein the query optimizing module is further configured to compute initialization costs as separate workloads for the purpose of partitioning.

20. The system of claim 11, further comprising:

a plurality of monitored systems, each having one or more corresponding monitors configured to record system state information;

a database configured to store the system state information from the plurality of monitored systems, wherein the database query is a query to the PROBEQL database; and

a security control system configured to perform a security control action in accordance with the progressive results.

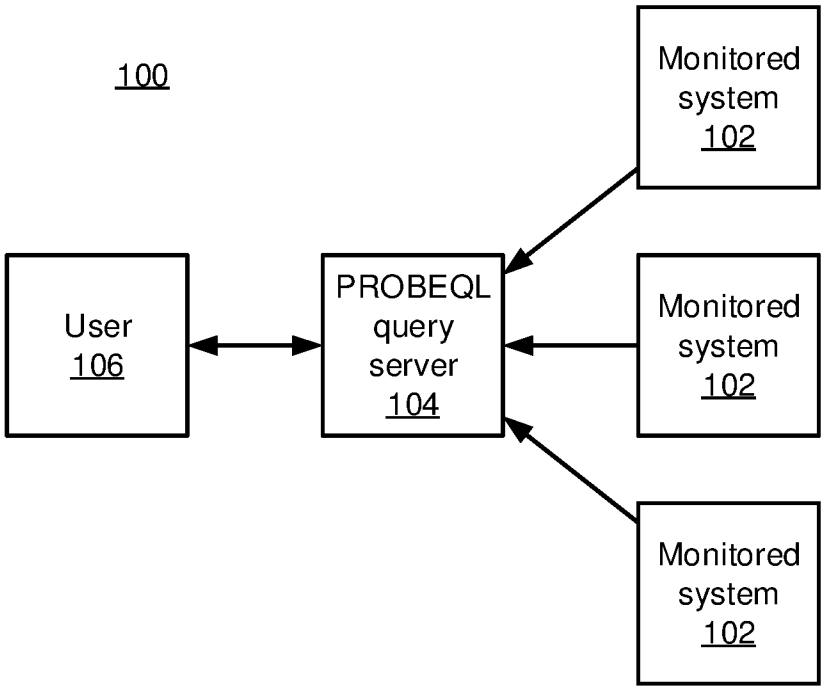
100
Monitored system 102
User 106
PROBEQL query server 104
Monitored system 102
Monitored system 102

FIG. 1

Algorithm 1: Online Adaptive Workload Prediction

Input: U_f: update frequency from user input
k: current host
S_i: index of first unprocessed event
$M[1, ..., n]$: host measurement/prediction
ΔE: latest execution information $\langle x, y \rangle$
ep_{max}: observed maximum event processing rate
γ: online learning rate

Output: D: workload for the subsequent sub-query

Initialize $D \leftarrow 0$ and $U'_f \leftarrow U_f$;
while more workload needed within U'_f **do**
  $D' \leftarrow 0$;
  $isNew \leftarrow$ **false**;
  if $Evt[S_i]$ is in a new host h_k **then**
    $ep, it \leftarrow getEPandIT(M, k)$;
    $ep_{max} \leftarrow ep$;
    $isNew \leftarrow$ **true**;
    if $it \geq U'_f$ **then**
      $D \leftarrow D + 1 * ep$;
      $S_i \leftarrow S_i + D$;
      return D;
  else
    $ep \leftarrow adaptiveLearning(ep, \Delta E, \gamma)$;
    if $ep_{max} \leq \frac{y}{x}$ **then**
      $ep_{max} \leftarrow \frac{y}{x}$;
  if $isNew ==$ **true then**
    $D' \leftarrow (U'_f - it) * ep$;
  else
    $D' \leftarrow U'_f * ep$;
  // regularization
  $maxD \leftarrow U'_f * ep_{max}$;
  if $D' > maxD$ **then**
    $D' \leftarrow maxD$;
  $D_r \leftarrow remainingEvent(M, k)$;
  if $D' > D_r$ **then**
    $U'_f \leftarrow U'_f - \frac{D_r}{ep}$;
    $D \leftarrow D + D_r$;
    $k \leftarrow k + 1$
  else
    $D \leftarrow D + D'$;
    break;
$S_i \leftarrow S_i + D$;
return D;

FIG. 2

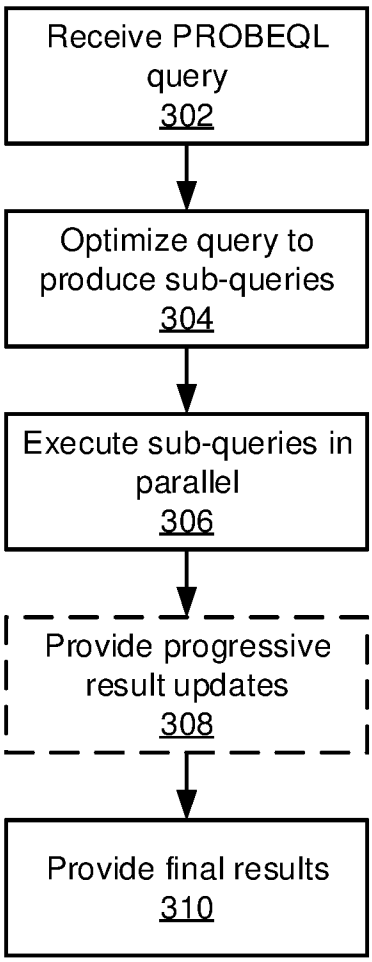
Receive PROBEQL query
302
Optimize query to produce sub-queries
304
Execute sub-queries in parallel
306
Provide progressive result updates
308
Provide final results
310

FIG. 3

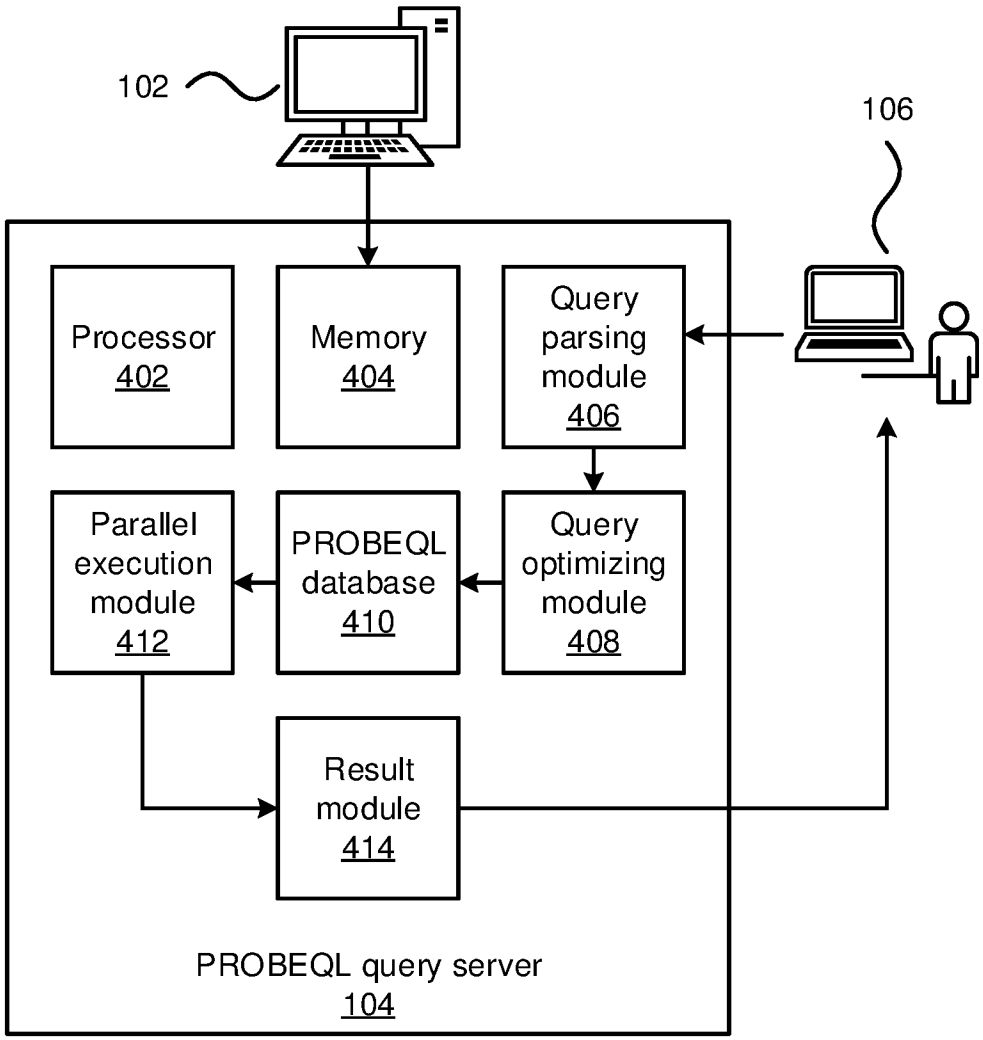
102
106
Processor
402
Memory
404
Query
parsing
module
406
Parallel
execution
module
412
PROBEQL
database
410
Query
optimizing
module
408
Result
module
414
PROBEQL query server
104

FIG. 4

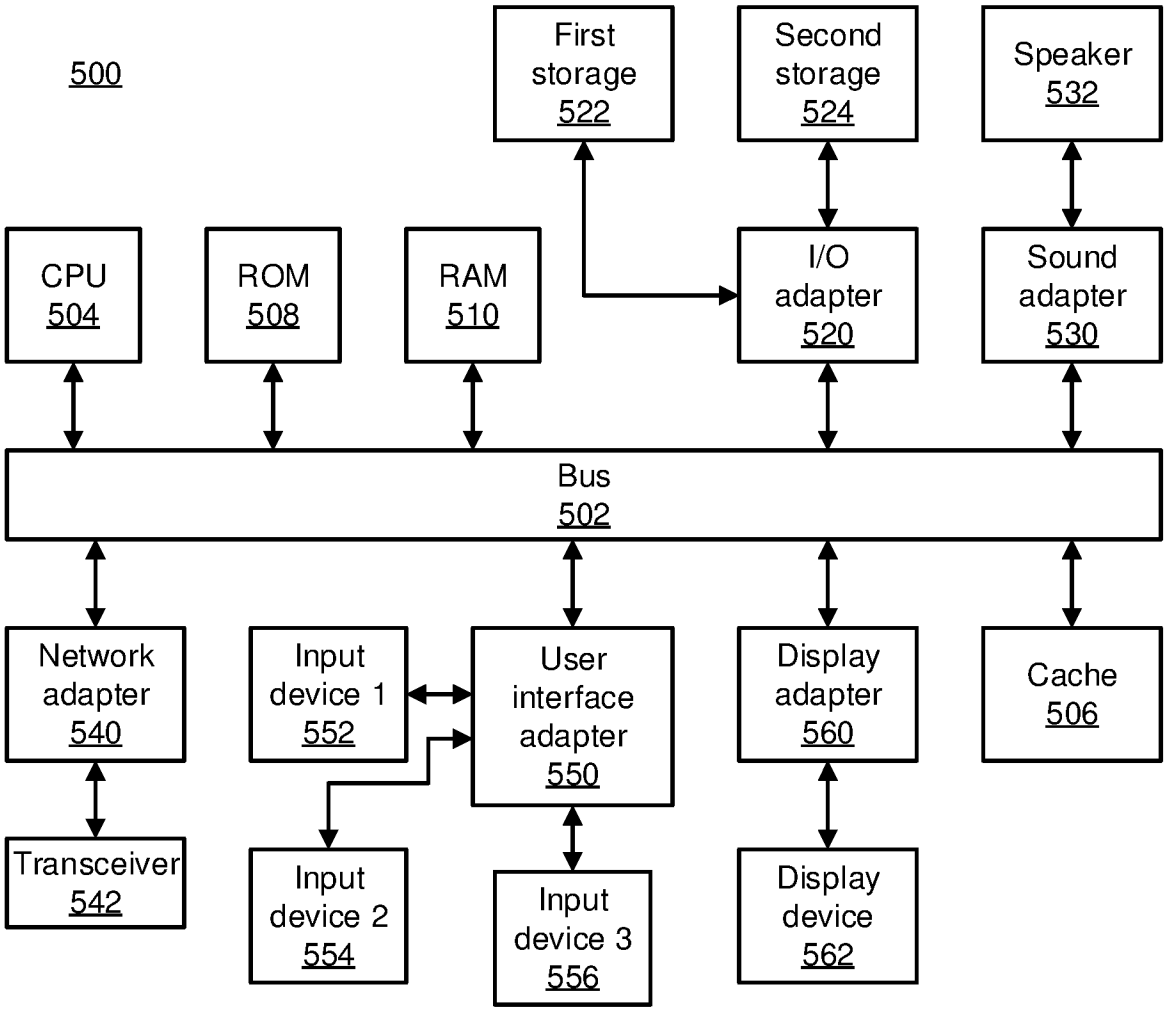
500
First storage 522
Second storage 524
Speaker 532
CPU 504
ROM 508
RAM 510
I/O adapter 520
Sound adapter 530
Bus 502
Network adapter 540
Input device 1 552
User interface adapter 550
Display adapter 560
Cache 506
Transceiver 542
Input device 2 554
Input device 3 556
Display device 562

FIG. 5

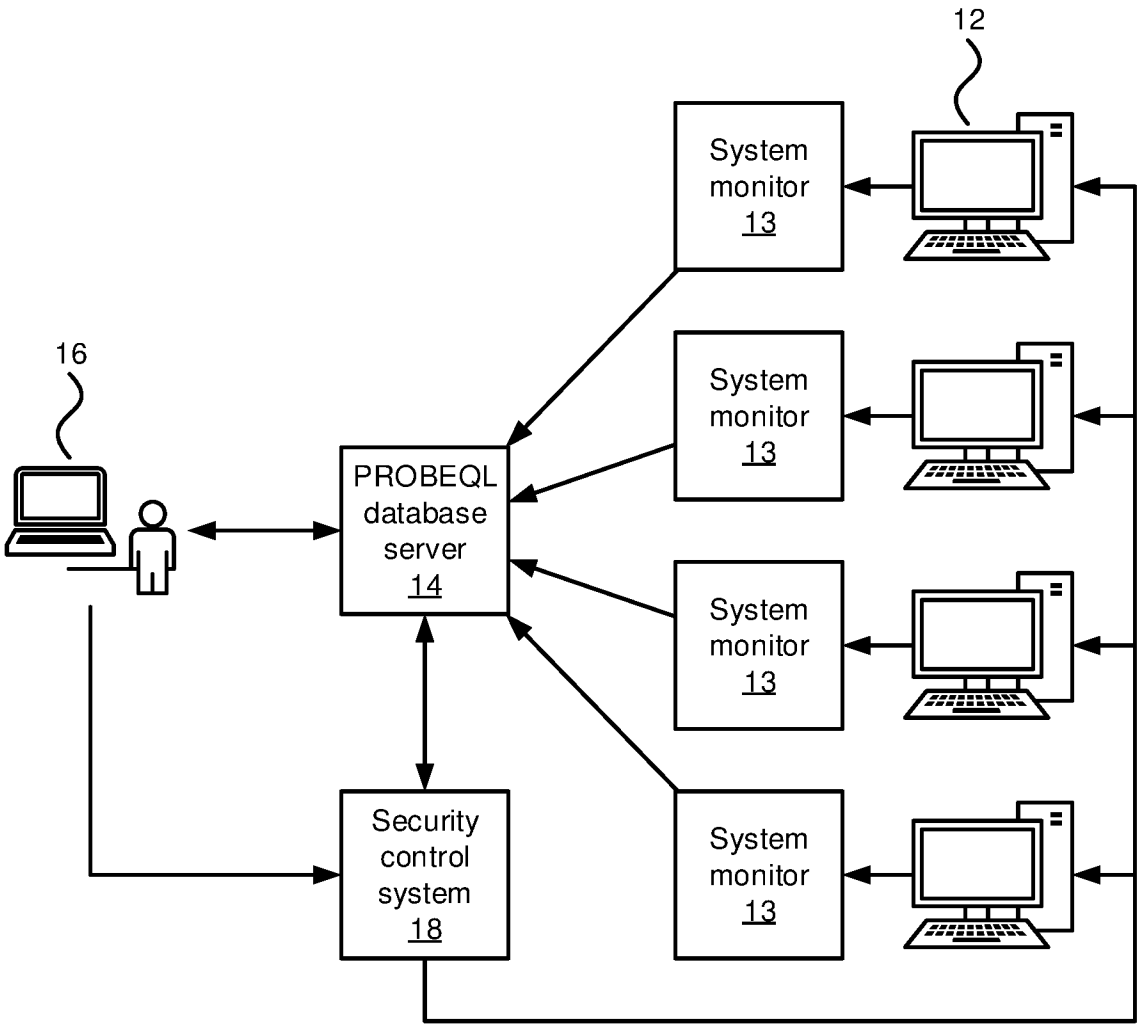
12
System monitor 13
System monitor 13
System monitor 13
System monitor 13
16
PROBEQL database server 14
Security control system 18

FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2017/048360

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30(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 17/30; G06F 15/173

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: query, optimize, split, sub-queries, spatial, host, temporal, time window

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2006-0080285 A1 (SUDIPTO R. CHOWDHURI) 13 April 2006 See paragraphs [0013]-[0015] and claims 1, 3 and 6.	1-20
A	US 5884299 A (BHASHYAM RAMESH et al.) 16 March 1999 See column 2, lines 15-33 and figure 2.	1-20
A	US 2014-0095473 A1 (ORACLE INTERNATIONAL CORP.) 03 April 2014 See paragraphs [0005]-[0008]; and claim 1; and figures 1-2.	1-20
A	US 2012-0166620 A1 (BARE SAID et al.) 28 June 2012 See paragraphs [0002]-[0004] and [0039].	1-20
A	US 2016-0034530 A1 (METANAUTIX, INC.) 04 February 2016 See paragraphs [0004]-[0024] and claim 1.	1-20

☐ 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
13 December 2017 (13.12.2017)	**13 December 2017 (13.12.2017)**

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
LEE, Chang Ho
Telephone No. +82-42-481-8288

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/US2017/048360

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2006-0080285 A1	13/04/2006	US 7574424 B2	11/08/2009
US 5884299 A	16/03/1999	None	
US 2014-0095473 A1	03/04/2014	CN 104756111 A	01/07/2015
		CN 104756112 A	01/07/2015
		CN 104885077 A	02/09/2015
		EP 2901319 A2	05/08/2015
		EP 2901320 A2	05/08/2015
		EP 2901321 A2	05/08/2015
		JP 2015-536001 A	17/12/2015
		JP 2016-500167 A	07/01/2016
		JP 2016-500168 A	07/01/2016
		US 2014-0095444 A1	03/04/2014
		US 2014-0095445 A1	03/04/2014
		US 2014-0095446 A1	03/04/2014
		US 2014-0095447 A1	03/04/2014
		US 2014-0095462 A1	03/04/2014
		US 2014-0095471 A1	03/04/2014
		US 2014-0095483 A1	03/04/2014
		US 2014-0095525 A1	03/04/2014
		US 2014-0095529 A1	03/04/2014
		US 2014-0095533 A1	03/04/2014
		US 2014-0095535 A1	03/04/2014
		US 2014-0095537 A1	03/04/2014
		US 2014-0095540 A1	03/04/2014
		US 2014-0095541 A1	03/04/2014
		US 2014-0095543 A1	03/04/2014
		US 2016-0103882 A1	14/04/2016
		US 2016-0140180 A1	19/05/2016
		US 2016-0154855 A1	02/06/2016
		US 9256646 B2	09/02/2016
		US 9262479 B2	16/02/2016
		US 9286352 B2	15/03/2016
		US 9292574 B2	22/03/2016
		US 9361308 B2	07/06/2016
		US 9563663 B2	07/02/2017
		US 9703836 B2	11/07/2017
		US 9715529 B2	25/07/2017
		WO 2014-052675 A2	03/04/2014
		WO 2014-052675 A3	04/09/2014
		WO 2014-052677 A2	03/04/2014
		WO 2014-052677 A3	21/08/2014
		WO 2014-052679 A2	03/04/2014
		WO 2014-052679 A3	21/08/2014
US 2012-0166620 A1	28/06/2012	US 9508048 B2	29/11/2016
US 2016-0034530 A1	04/02/2016	WO 2016-018947 A1	04/02/2016