



(51) International Patent Classification:
H04L 12/26 (2006.01)

(21) International Application Number:
PCT/US2015/011108

(22) International Filing Date:
13 January 2015 (13.01.2015)

(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) Inventor: **GLIGOROV, Daniel**; 108 North Terrace, Adelaide, South Australia 5000 (AU).

(74) Agents: **BELL, Scott** et al.; 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, 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))

(54) Title: SERVER RESPONSE TIME MONITORING

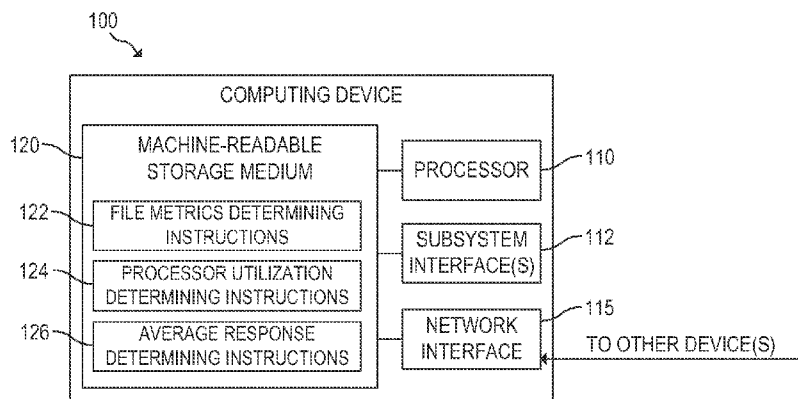


FIG. 1

(57) Abstract: Examples relate to monitoring server response times. In some examples, a metric interface is used to obtain performance metrics, where the metric interface provides access to a measuring module configured to monitor the performance metrics of a computing device. At this stage, response times are determined for the computing device based on the performance metrics.

SERVER RESPONSE TIME MONITORING

BACKGROUND

[0001] Response time is the amount of time that passes before a system is able to respond to a request. Server response times can be measured based on communication between a client device and a server. For example, timestamps of a request from the client and a response from the server can be compared to determine the response time for the request.

BRIEF DESCRIPTION OF THE DRAWINGS

[0002] The following detailed description references the drawings, wherein:

[0003] FIG. 1 is a block diagram of an example computing device for providing monitoring of server response times;

[0004] FIG. 2 is a block diagram of an example computing device in communication with a client computing device for providing monitoring of server response times;

[0005] FIG. 3 is a block diagram of an example computing device in communication with a transactional computing device for providing monitoring of server response times;

[0006] FIG. 4A is a flowchart of an example method for execution by a computing device for providing monitoring of server response times using a subsystem interface; and

[0007] FIG. 4B is a flowchart of an example method for execution by a computing device for providing monitoring of server response times using a communication interface.

DETAILED DESCRIPTION

[0008] As described above, server response times can be determined based on monitoring of network activity of a server. On the server-side, response times can also be monitored using customized code that is embedded in a server

application. For example, debugging code can be used during the execution of the server application to determine response times for server-side operations. However, such techniques are application specific and have limited access to system activities.

[0009] Examples herein describe a generic tool for capturing response times of single threaded servers, including response times per transaction type. Further, the generic tool is capable of being implemented in existing applications. In addition to capturing response times, the examples include a data storage system and presentation system for graphical and tabular presentation of data. Various server metrics such as system wide response times as well as various other measure entities such as structured query language (SQL), file, disk, central processing unit (CPU), device, and controller statistics.

[0010] In some examples, a metric interface is used to obtain performance metrics, where the metric interface provides access to a measuring module (e.g., external measuring module, internal measuring module, etc.) configured to monitor the performance metrics of a computing device. At this stage, response times are determined for the computing device based on the performance metrics

[0011] Referring now to the drawings, FIG. 1 is a block diagram of an example computing device 100 for providing monitoring of server response times. Computing device 100 may be any computing device (e.g., server, desktop computer, etc.) with access to other computing devices such as user devices and server devices. In the embodiment of FIG. 1, computing device 100 includes a processor 110, subsystem interface(s) 112, a network interface 115, and a machine-readable storage medium 120.

[0012] Processor 110 may be one or more central processing units (CPUs), microprocessors, and/or other hardware devices suitable for retrieval and execution of instructions stored in machine-readable storage medium 120. Processor 110 may fetch, decode, and execute instructions 122, 124, 126 to enable providing monitoring of server response times, as described below. As an alternative or in addition to retrieving and executing instructions, processor 110

may include one or more electronic circuits comprising a number of electronic components for performing the functionality of one or more of instructions 122, 124, 126.

[0013] Subsystem interface(s) 112 provides access to external measuring modules that are configured to measure performance metrics of subsystems of computing device 100. Examples of subsystems include a storage subsystem, a processor subsystem, an input/output subsystem, a networking subsystem, etc. Subsystem interface 112 provides an application programming interface (API) for accessing various metrics of its corresponding subsystem. For example, a subsystem interface 112 for a file system provides access to storage metrics such as access times, data transfer rates, etc. An external measuring module may be implemented in, for example, a controller (not shown) of a corresponding subsystem.

[0014] Network interface 115 may include a number of electronic components for communicating with cloud devices. For example, network interface 115 may be an Ethernet interface, a Universal Serial Bus (USB) interface, an IEEE 1394 (Firewire) interface, an external Serial Advanced Technology Attachment (eSATA) interface, or any other physical connection interface suitable for communication with other computing devices. Alternatively, network interface 115 may be a wireless interface, such as a wireless local area network (WLAN) interface or a near-field communication (NFC) interface. In operation, as detailed below, network interface 115 may be used to send and receive data to and from a corresponding interface of a client computing device.

[0015] Machine-readable storage medium 120 may be any electronic, magnetic, optical, or other physical storage device that stores executable instructions. Thus, machine-readable storage medium 120 may be, for example, Random Access Memory (RAM), an Electrically-Erasable Programmable Read-Only Memory (EEPROM), a storage drive, an optical disc, and the like. As described in detail below, machine-readable storage medium 120 may be encoded with executable instructions for providing monitoring of server response times.

[0016] File metrics determining instructions 122 determine file metrics of computing device 100 during a monitoring interval. The duration and/or timing of the monitoring interval may be preconfigured by an administrator of computing device 100. File metrics may describe file operations (e.g., create, modify attributes, open, read, write, close, etc.) and include characteristics such as timestamps, target files, storage address, etc. The file metrics can be logged by computing device 100 as the file operations are performed. In one example, the file metrics can be determined by accessing the log files. In another example, the file metrics can be determined in real-time using subsystem interface 112.

[0017] Processor utilization determining instructions 124 determine processor utilization metrics of computing device 100 during the monitoring interval. Processor utilization metrics describe the performance of processor 110 and can be determined for multiple cores, processors, etc. in a multi-processor system. Specifically, processor utilization describes the percentage of processing cycles of processor 110 that are being used to perform tasks. Further, processor utilization metrics can also describe processor utilization of particular processes being executed by computing device 100. Similar to file metrics, processor utilization metrics can be determined by accessing a log file and/or determining processing utilization in real-time using subsystem interface 112.

[0018] Average response determining instructions 126 determine average response times of computing device 100 for the monitoring interval based on the file metrics and the processor utilization metrics. Specifically, portions of the processor utilization metrics can be associated with a file operation and then used to determine a response time of the file operation. The associated portions excludes processor utilization that can be attributed to queuing and idle time; therefore, the associated portions are more relevant to the file operation (i.e., processor utilization related to handling a request for the file operation, waiting on input/output, management of computing device 100, etc.). Idle time includes processor cycles that are not being used by any program thereby allowing for optimizations such as reduced power consumption. Idle time is excluded from the average response time because the idle time is irrelevant to the processing of requests by computing device 100. In this example, response times for each of

the file operations can be determined as described above, where the individual response times can then be used to determine an average response time of computing device 100.

[0019] FIG. 2 is a block diagram of an example computing device 200 in communication via a network 245 with a client computing device 250. As illustrated in FIG. 2 and described below, computing device 200 may communicate with client computing device 250 to provide monitoring of server response times.

[0020] As illustrated, computing device 200 may include a number of modules 202-224, while client computing device 250 may include a number of modules 254-256. Each of the modules may include a series of instructions encoded on a machine-readable storage medium and executable by a processor of the respective device 200, 250. In addition or as an alternative, each module may include one or more hardware devices including electronic circuitry for implementing the functionality described below.

[0021] As with computing device 100 of FIG. 1, computing device 200 may be any computing device with access to client computing devices over a network 245 that is suitable for executing the functionality described below. As detailed below, computing device 200 may include a series of modules 202-224 for providing monitoring of server response times.

[0022] Network interface 202 may initiate and manage connections with client computing device 250. Specifically, the network interface 202 may receive client requests from the transactional computing device 250 for processing as described below.

[0023] Operating system (OS) 224 may manage hardware and software resources of computing device 200. OS 224 may also provide common services for programs/applications such as performance monitor 210 described below. For example, OS 224 may provide performance monitor 210 with access to a subsystem (not shown) of computing device 200. Specifically, OS 224 may provide functions that can be used to access subsystem interfaces 220, which in

turn us external measuring module 222 to access subsystem performance metrics of computing device 200.

[0024] Performance monitor 210 may monitor the performance of server computing devices. In this example, computing device 200 corresponds to the server computing device being monitored. In other examples, performance monitor 210 may be implemented into a separate computing device that is configured to monitor remote server computing devices. Although the components of performance monitor 210 are described in detail below, additional details regarding an example implementation of performance monitor 210 are provided above in connection with instructions 122, 124, and 126 of FIG. 1.

[0025] Subsystem module 212 is configured to interact with subsystem interface 220. Specifically, subsystem module 212 can interact with subsystem interface 220 to obtain performance metrics of computing device 200. Performance metrics may include, but are not limited to, file performance metrics and processor utilization metrics similar to as described above with respect to FIG. 1. In some cases, subsystem module 212 is configured to poll subsystem interface 220 to continuously monitor the performance metrics of computing device 200. For example, performance metrics can be obtained in a staggered fashion such that subsequent requests for metrics are submitted before previous results are received, allowing for more continuous coverage of the metrics. Subsystem module 212 may also be configured to obtain performance metrics using subsystem interface 220 in response to system events of computing device 200 (e.g., initiation of file operation, receipt of network request, etc.). In some cases, subsystem interfaces 220 correspond to OS function calls that are accessible to subsystem module 212. In this case, the OS function calls may access counters that are managed by external measuring module 222, which is described below.

[0026] Analysis module 216 is configured to analyze performance metrics to generate summary data. Analysis module 216 may generate reports that display the summary data. Examples of summary data include, but are not limited to, average response time of file operations and average response time for requests

made by a client computing device (e.g., client computing device 250). For example, analysis module 216 may analyze processor utilization data that is associated with file operations to determine a portion of the processor utilization data that is relevant to average response times of computing device 200. In this example, the relevant processor utilization data can be used along with the quantity of requests from client computing devices (e.g., client computing device 250) to determine an average response time for computing device 200. In some cases, the summary data can be grouped by volume in the case that computing device 200 includes a drive array (not shown) with multiple volumes.

[0027] Subsystem interface(s) 220 (i.e., metric interfaces) are configured to provide access to performance metrics of various subsystems of computing device 200. Specifically, subsystem interface(s) 220 may provide access to external measuring module 222, which in turn obtains the subsystem metrics from the subsystem as described below.

[0028] External measuring module 222 (i.e., measuring module) is configured to update counters in memory or other storage based on performance metrics that are obtained from the subsystem of computing device 200. External measuring module 222 may manage any number of counters for monitoring performance metrics such as processor utilization, file operation time stamps and statistics, etc. For example, a file open count counter can be managed to monitor the number of file open operations that have been initiated by computing device 200.

[0029] Computing device 200 includes storage devices (not shown), which may be any hardware storage device for maintaining data accessible to computing device 200. For example, storage devices may include one or more hard disk drives, solid state drives, tape drives, and/or any other storage devices. The storage devices may be located in computing device 200 and/or in another device in communication with computing device 200.

[0030] Client computing device 250 may be a server computing device, a notebook, desktop, tablet, workstation, mobile device, or any other device suitable for executing the functionality described below. As detailed below, client

computing device 250 may include a series of modules 254-256 for accessing services provided by computing device 200.

[0031] Client computing device 250 may use services of the computing device 200 to perform transactions. For example, client computing device 250 may be a banking device that processes money transfers or a commerce device that processes sales. Device interface 254 may initiate and manage connections with computing device 200. Specifically, the device interface 254 may send client requests to the computing device 200 for processing as described above.

[0032] Application module 256 is an application that performs transactions on behalf of users interacting with client computing device 250. For example, application module 256 may be banking web page that provides access to banking services via the Internet. In another example, application module 256 may manage online sales of merchandise via a commerce site.

[0033] Display device 270 may be a server computing device, a notebook, desktop, tablet, workstation, mobile device, or any other device suitable for executing the functionality described below. As detailed below, display device 270 may include a display module 272 for displaying summary data generated by analysis module 216. Specifically, display module 272 can collect summary data from multiple server computing devices so that reports including data from multiple systems (e.g., a distributed system including any number of server computing devices) can be shown. For example, a report showing the average response time for each server computing device can be generated based on data from multiple server computing devices.

[0034] FIG. 3 is a block diagram of an example computing device 300 in communication via a network 345 with a transactional computing device 350. As illustrated in FIG. 3 and described below, computing device 300 may communicate with transactional computing device 350 to provide monitoring of server response times.

[0035] As illustrated, computing device 300 may include a number of modules 302-324, while transactional computing device 350 may include a number of modules 354-356. Each of the modules may include a series of instructions

encoded on a machine-readable storage medium and executable by a processor of the respective device 300, 350. In addition or as an alternative, each module may include one or more hardware devices including electronic circuitry for implementing the functionality described below.

[0036] As with computing device 100 of FIG. 1 and computing device 200 of FIG. 2, computing device 300 may be any computing device with access to client/transactional computing devices over a network 345 that is suitable for executing the functionality described below. As detailed below, computing device 300 may include a series of modules 302-324 for providing monitoring of server response times.

[0037] Network interface 302, performance monitor 310, operating system 324, and analysis module 316 may be substantially similar to corresponding components described above with respect to FIG. 2.

[0038] Communication module 312 is configured to interact with communication interface 318 (i.e., metric interface). Specifically, communication module 312 can interact with communication interface 318 to obtain performance metrics of computing device 300. Performance metrics may include, but are not limited to, communication metrics related to messages processed by computing device 300 on behalf of transactional computing device 350. In some cases, communication module 312 is triggered to use communication interface 318 in response to messaging events (e.g., receiving a transaction request, initiating a transaction, completing a transaction, sending a transaction response, etc.). Communication module 312 may also be configured to poll communication interface 318 for communication metrics at regular time intervals.

[0039] Transactional module 314 is configured to track transactions (i.e., requests and subsequent responses) as they are processed by computing device 300. For example, multiple file operations and a portion of processor utilization can be attributed to a transaction as a request from transactional computing device 350 is processed. In this example, the response time for a request may be determined based on the relevant file operations and processor utilization. Because non-relevant delays in the transaction are ignored, transactional module

314 is able to more accurately determine the response times of transactions. In another example, transactional module 314 may track transactions that are independent of transactional computing device 350 (i.e., transactions that originate from internal components of computing device 300).

[0040] Transactional module 314 is also configured to use communication interface 318 to intercept transactions initiated by transactional computing device 350. The intercepted transactions can be analyzed when incoming and outgoing to determine when each transaction is received and then handled with a response. These transactional metrics can then be used by analysis module 316 to determine minimum and maximum response times for each transaction type.

[0041] Analysis module 316 is configured to analyze performance metrics and transactional data to generate summary data. Analysis module 316 may generate reports that display the summary data. Examples of summary data include, but are not limited to, average response time for transactions and average response times for transactions of particular transaction types (e.g., banking transactions, purchase transactions, data query transactions, etc.). In some cases, the summary data can be grouped by volume in the case that computing device 300 includes a drive array (not shown) with multiple volumes.

[0042] Communication interface(s) 318 are configured to provide access to communication metrics of various subsystems of computing device 300. Specifically, communication interface(s) 318 interrupt communication protocols as they occur on computing device to provide the communication metrics. For example, when a transaction is initiated, communication interface 318 can receive a start timestamp and transaction identifier for the transaction. In this example, when the transaction is completed, communication interface 318 can receive an end timestamp and the transaction identifier for the transaction. In this example, the start timestamp and the end timestamp are communication metrics that can be used by the analysis module to determine a response time for the transaction. Further, communication interface 318 also receives a transaction type for the transaction so that response time statistics can be grouped by transaction type.

[0043] In some cases, communication interface(s) 318 may be implemented as hooks of OS 324 that are triggered by messaging events. The OS hooks may interrupt protocol functions to provide communication metrics to performance monitor 310. In other cases, communication interface(s) 318 can be implemented as replacement functions of the protocol functions that include specialized code to capture the communication metrics.

[0044] Internal measuring module 322 (i.e., measuring module) is configured to extract the communication metrics from messages processed by operating system 324. For example, the various timestamps as described above can be extracted from messages as the messages are processed. Communication interface(s) 318 use internal measuring module 322 to access the communication metrics in real-time or as historical data.

[0045] Computing device 300 includes storage devices (not shown), which may be any hardware storage device for maintaining data accessible to computing device 300. For example, storage devices may include one or more hard disk drives, solid state drives, tape drives, and/or any other storage devices. The storage devices may be located in computing device 300 and/or in another device in communication with computing device 300.

[0046] Transactional computing device 350 may be a server computing device, a notebook, desktop, tablet, workstation, mobile device, or any other device suitable for executing the functionality described below. As detailed below, transactional computing device 350 may include a series of modules 354-356 for accessing services provided by computing device 300.

[0047] Transactional computing device 350 may use services of the computing device 300 to perform transactions. For example, transactional computing device 350 may be a banking device that processes money transfers or a commerce device that processes sales. Device interface 354 may initiate and manage connections with computing device 300. Specifically, the device interface 354 may send client requests to the computing device 300 for processing as described above.

[0048] Application module 356 is an application that performs transactions on behalf of users interacting with transactional computing device 350. For example, application module 356 may be banking web page that provides access to banking services via the Internet. In another example, application module 356 may manage online sales of merchandise via a commerce site.

[0049] Display device 370 may be a server computing device, a notebook, desktop, tablet, workstation, mobile device, or any other device suitable for executing the functionality described below. As detailed below, display device 370 may include a display module 372 for displaying summary data generated by analysis module 316. Specifically, display module 372 can collect summary data from multiple server computing devices so that reports including data from multiple systems (e.g., a distributed system including any number of server computing devices) can be shown. For example, a report showing the average response time for each transaction type can be generated based on data from one or more server computing devices.

[0050] FIG. 4A is a flowchart of an example method 400 for execution by a computing device 100 for providing monitoring of server response times using a subsystem interface. Although execution of method 400 is described below with reference to computing device 100 of FIG. 1, other suitable devices for execution of method 400 may be used, such as computing device 200 of FIG. 2. Method 400 may be implemented in the form of executable instructions stored on a machine-readable storage medium, such as storage medium 120, and/or in the form of electronic circuitry.

[0051] Method 400 may start in block 405 and continue to block 410, where computing device 100 uses a subsystem interface to determine file metrics of computing device 100 during a monitoring interval. The file metrics can be logged by computing device 100 as the file operations are performed. In block 415, computing device 100 uses the subsystem interface to determine processor utilization metrics of computing device 100 during the monitoring interval. Processor utilization metrics describe the performance of multiple cores or processors in a multi-processor system.

[0052] At block 420, computing device 100 determines average response times for the monitoring interval based on the file metrics and the processor utilization metrics. Specifically, portions of the processor utilization metrics can be associated with a file operation and then used to determine a response time of the file operation. For example, FILE and PROCESS performance matrixes can be used to calculate response times as shown below:

$$\text{Response time} = \text{SUM of (file-busy-time) of each file opened by the process} - \text{FILE-BUSY-TIME of the \$RECEIVE file} + \text{CPU-BUSY-TIME}$$

Where \$RECEIVE is a special file used by processing units to receive and respond to messages.

[0053] In another example, the FILE measure entity is used to calculate the response time as shown below:

$$\text{Response time} = \text{FILE-OPEN time of \$RECEIVE file} - \text{FILE-BUSY time of \$RECEIVE file}$$

[0054] The average response times can then be used to create reports for review by an administrator. Based on the reports, the administrator may reconfigure computing device 100 in order to minimize future average response times. Method 400 may then proceed to block 425, where method 400 may stop.

[0055] FIG. 4B is a flowchart of an example method 450 for execution by a computing device 300 for providing monitoring of server response times using a network interface. Although execution of method 450 is described below with reference to computing device 300 of FIG. 3, other suitable devices for execution of method 450 may be used. Method 450 may be implemented in the form of executable instructions stored on a machine-readable storage medium, such as storage medium 120, and/or in the form of electronic circuitry.

[0056] Method 450 may start in block 455 and continue to block 460, where computing device 300 uses a communication interface to intercept messages initiated by a transactional computing device. Specifically, requests from the transactional computing device may be intercepted so that the requests can be

examined as the requests are processed by computing device 300. Similarly, responses from computing device 300 to the transactional computing device may also be intercepted so that the responses can be examined before the responses are sent to the transactional computing device.

[0057] In block 465, minimum and maximum response times are determined for each message received from the transactional computing device. For example, if computing device 300 provides a web site, minimum and maximum response times can be determined for each web page. The minimum and maximum response times are determined based on the amount of time that elapses between the receipt of the request from the transactional computing device and the return of the response from computing device 300. For example, a response time of a message can be calculated as shown below:

$$\text{Response Time} = \text{TIMESTAMP of the RESPOND} - \text{TIMESTAMP of the RECEIVED message}$$

Time timestamps may be recorded in real time by intercepting the functions of the communication interface used by the process that is being measured to receive and respond to a message. Communication interface functions are typically part of the operating system, but a communication interface may also be supplied by a separate entity that did not develop the operating system. In the second case, the communication interface usually depends on similar operating system functions.

[0058] The related operating system (OS) functions can be intercepted in a variety of manners. For example, the OS functions can be replaced with replacement functions, which each has identical interface as the OS function but once called executes the OS function and records the time (either before or after the OS function call). In another example, the vendor of the OS function provides an interface and calls an external function to record the time and forward the time to memory accessible to a performance monitor. In yet another example, an intercept library is provided to intercept the OS function from within the object of the calling program. In this example, an intercept library is attached to each program to be monitored.

[0059] Because the response times are determined based on interactions with a communication interface, computing device 300 can perform the calculation regardless of the operating system that is installed on computing device 300. Method 450 may then proceed to block 475, where method 400 may stop.

[0060] In some cases, the minimum and maximum response times of each transaction type can be used to further determine average response times for each transaction type, as shown in block 470. The minimum and maximum response times along with the average times can then be displayed in a report for review by an administrator of computing device 300.

[0061] The foregoing disclosure describes a number of example embodiments for providing monitoring of server response times. In this manner, the embodiments disclosed herein improve determining response types by a subsystem or communication interface to obtain performance metrics independent of the implementation of the operating system of the computing device.

CLAIMSWe claim:

1. A computing device for monitoring server response times, the computing device comprising:
 - a measuring module to monitor performance metrics;
 - a metric interface to provide access to the measuring module;
 - a plurality of processors to perform a plurality of messages; and
 - a performance monitor for execution by at least one of the plurality of processors to:
 - use the metric interface to obtain the performance metrics; and
 - determine a plurality of response times for the computing device based on the performance metrics.
2. The computing device of claim 1, wherein the metric interface is a subsystem interface and the performance metrics comprise a plurality of file metrics for a plurality of file operations and processor utilization for the plurality of processors, and wherein the performance monitor is further to determine an average response time for the computing device based on the plurality of response times.
3. The computing device of claim 2, wherein the plurality of file operations are monitored in parallel such that the plurality of file metrics are staggered to provide continuous coverage.
4. The computing device of claim 2, wherein idle time of the plurality of processors is excluded from the determination of the plurality of response times.
5. The computing device of claim 2, wherein the subsystem interface comprises operating system (OS) function calls that can access a plurality of counters, wherein each of the plurality of counters is associated with one of the plurality of file metrics or the processor utilization.

6. The computing device of claim 1, wherein the metric interface is a communication interface and the performance metrics correspond to a plurality of communication metrics, and wherein the communication interface is used to interrupt the plurality of messages initiated by a plurality of transactional devices to obtain the plurality of communication metrics.

7. The computing device of claim 6, wherein the plurality of response times comprises a minimum response time and a maximum response time for each of the plurality of messages, and wherein the performance monitor is further to use the minimum response time and the maximum response time to determine the average response time for each transaction type of the plurality of messages.

8. The computing device of claim 1, wherein the performance monitor is further to send the plurality of response times to a display device that is configured to display a performance report based on the plurality of response times.

9. A method for monitoring server response times, the method comprising:
using a subsystem interface to determine a plurality of file metrics for a plurality of file operations of a computing device during a monitoring interval, wherein the subsystem interface provides access to an external measuring module configured to monitor performance metrics;

using the subsystem interface to determine processor utilization during the plurality of file operations to determine a plurality of processor metrics, wherein the plurality of processor metrics are associated with a plurality of processors of the computing device; and

determining an average response time for the computing device during the monitoring interval based on the plurality of file metrics and the plurality of processor metrics, wherein idle time of the plurality of processors is excluded from the determination of the average response time.

10. The method of claim 9, wherein the plurality of file operations are monitored in parallel such that the plurality of file metrics are staggered to provide continuous coverage.

11. The method of claim 9, wherein monitoring of each of the plurality of file metrics is initiated in response to a file open command.

12. The method of claim 9, wherein the subsystem interface uses operating system (OS) function calls to access a plurality of counters to determine the plurality of file metrics and the processor utilization, and wherein the plurality of counters are managed by the external measuring module.

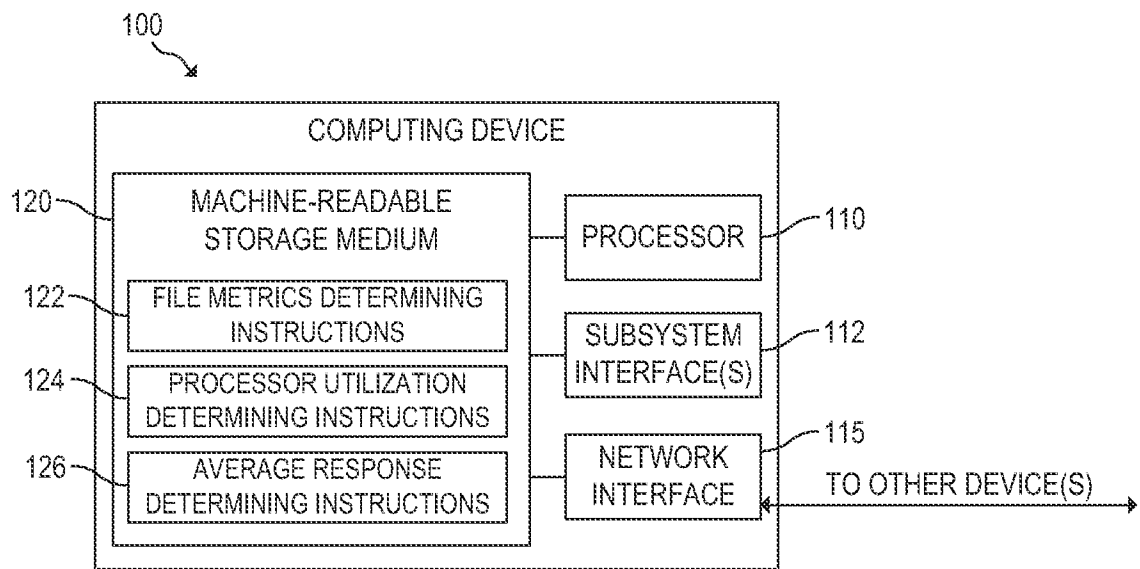
13. A non-transitory machine-readable storage medium encoded with instructions executable by a processor for monitoring server response times, the machine-readable storage medium comprising instructions to:

use a communication interface to determine a plurality of communication metrics for a plurality of messages of a computing device, wherein the communication interface provides access to an internal measuring module configured to monitor the plurality of messages; and

use the plurality of communication metrics to determine a minimum response time and a maximum response time for each of the plurality of messages.

14. The non-transitory machine-readable storage medium of claim 13, wherein the instructions are further to use the minimum response time and the maximum response time to determine an average response time for each transaction type of the plurality of messages.

15. The non-transitory machine-readable storage medium of claim 13, wherein the communication interface uses a plurality of hooks to intercept the plurality of messages during messaging events to determine the plurality of communication metrics.

**FIG. 1**

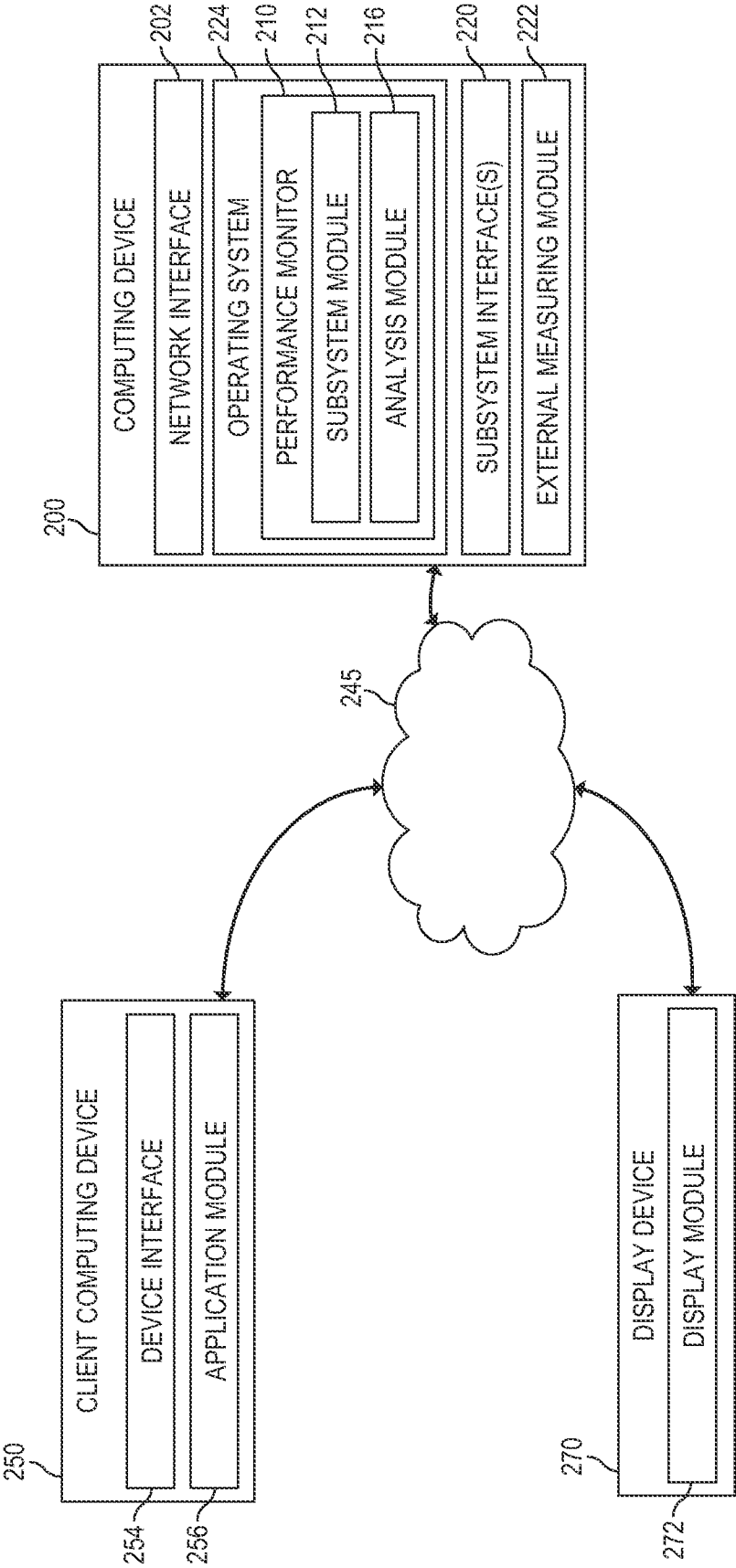


FIG. 2

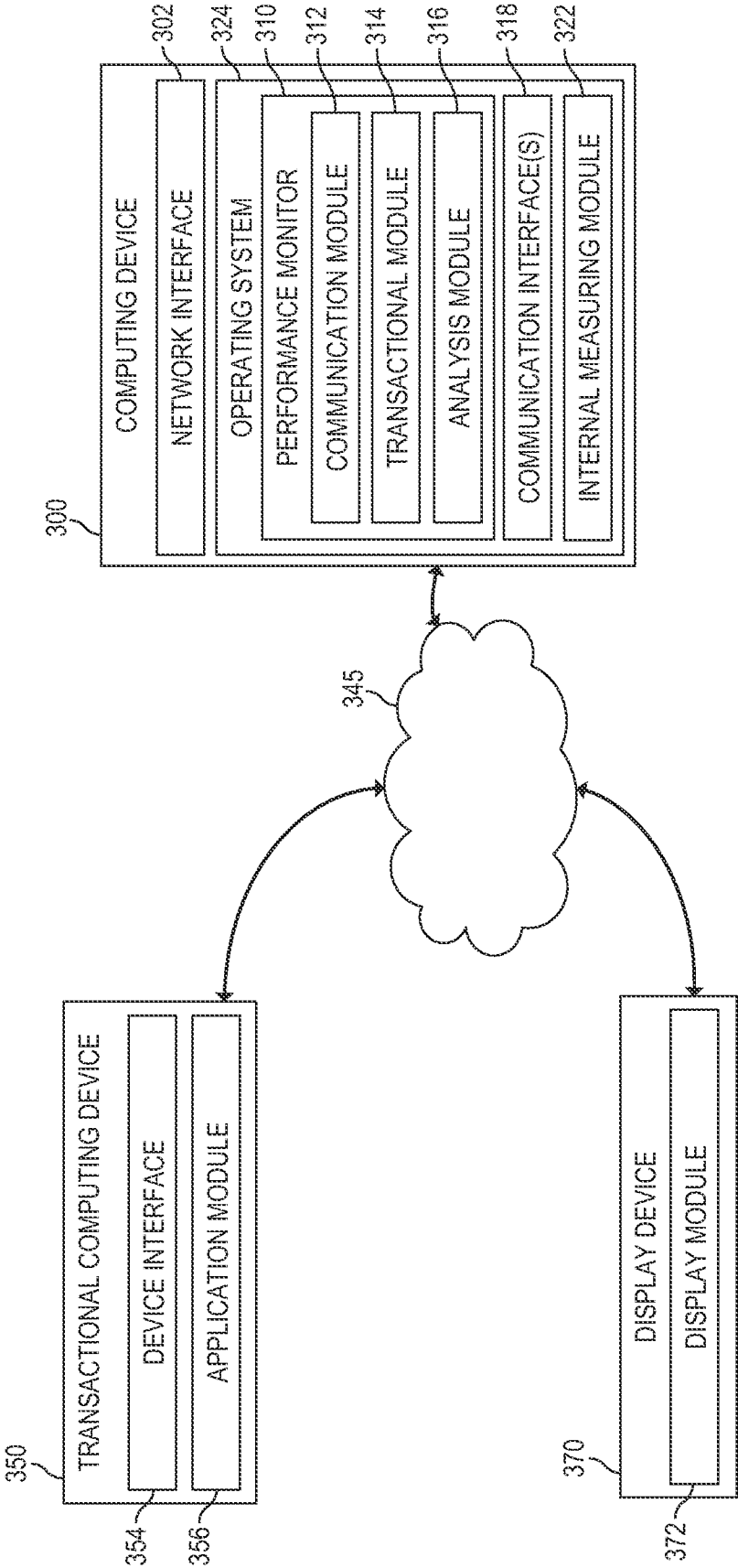
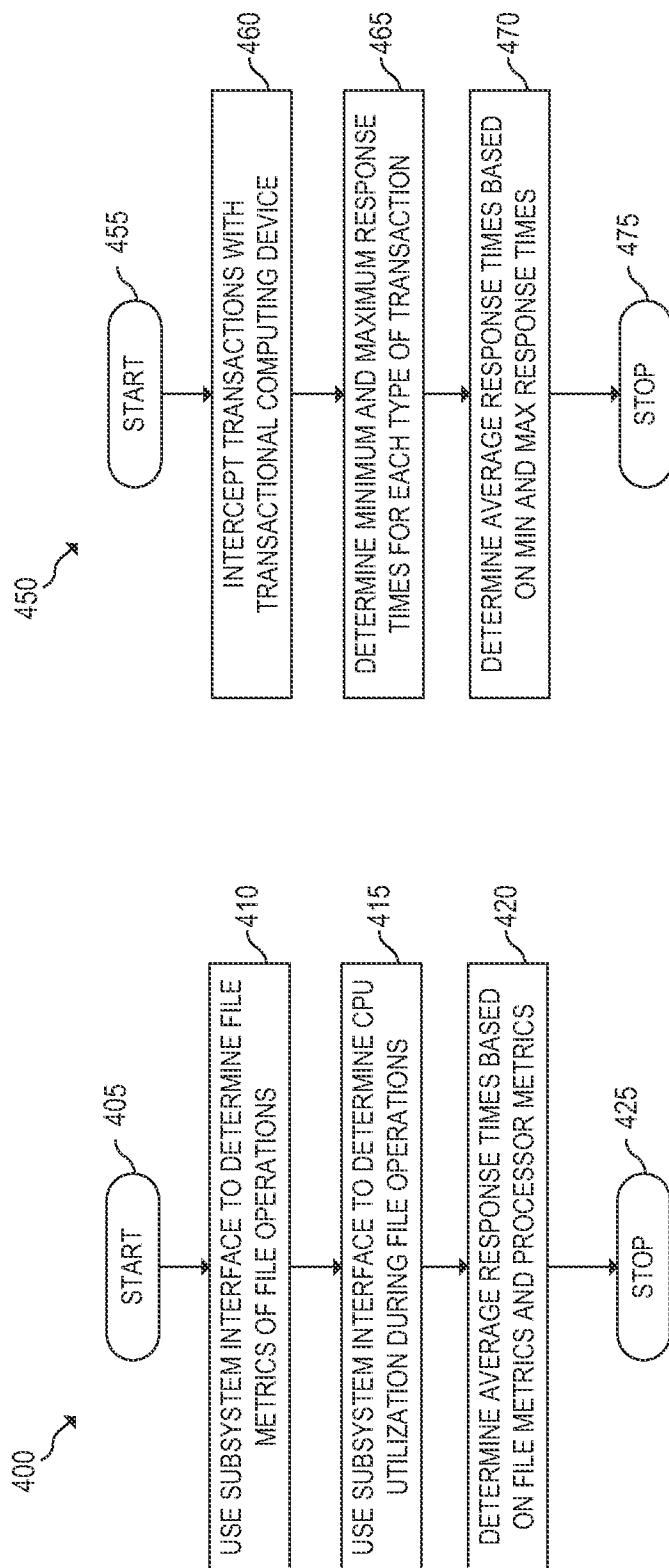


FIG. 3

**FIG. 4A****FIG. 4B**

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2015/011108**A. CLASSIFICATION OF SUBJECT MATTER****H04L 12/26(2006.01)i**

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

B. FIELDS SEARCHEDMinimum documentation searched (classification system followed by classification symbols)
H04L 12/26; G21C 17/00; G06F 15/16; G06F 15/173; G06F 15/00; G06F 15/18; G06F 9/45Documentation 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 modelsElectronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & Keywords: monitoring, server, response, time, performance, metric, file, operation, processor, utilization, 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.
Y	US 2004-0215768 A1 (YOSSI OULU et al.) 28 October 2004 See paragraphs [0007], [0009], [0027], [0031], [0036], [0061], [0092]; and figure 1A.	1-15
Y	US 2003-0110007 A1 (JOHN MCGEE et al.) 12 June 2003 See paragraphs [0054], [0057]; and figure 1.	1-15
A	US 2002-0198985 A1 (NOAM FRAENKEL et al.) 26 December 2002 See paragraphs [0134]-[0152]; and figures 17-20.	1-15
A	US 2008-0168177 A1 (SUBBU SUBRAMANIAM) 10 July 2008 See paragraphs [0026]-[0035]; and figures 3-4.	1-15
A	US 2010-0082320 A1 (TIMOTHY W. WOOD et al.) 01 April 2010 See paragraphs [0039]-[0059]; and figures 4-6.	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

07 October 2015 (07.10.2015)

Date of mailing of the international search report

08 October 2015 (08.10.2015)

Name and mailing address of the ISA/KR

International Application Division
Korean Intellectual Property Office
189 Cheongsu-ro, Seo-gu, Daejeon Metropolitan City, 302-701,
Republic of Korea

Facsimile No. +82-42-472-7140

Authorized officer

LEE, Dong Yun

Telephone No. +82-42-481-8734



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2015/011108

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2004-0215768 A1	28/10/2004	AU 2003-277267 A1 AU 2003-277267 A8 US 06792460 B2 US 07401143 B2 US 07822844 B2 US 2004-0068560 A1 US 2004-0215762 A1 WO 2004-031914 A2 WO 2004-031914 A3	23/04/2004 30/07/2009 14/09/2004 15/07/2008 26/10/2010 08/04/2004 28/10/2004 15/04/2004 18/06/2009
US 2003-0110007 A1	12/06/2003	AU 2002-316479 A1 US 06643613 B2 US 07076695 B2 US 07219034 B2 US 07444263 B2 US 2003-0079160 A1 US 2003-0088542 A1 US 2008-0016412 A1 WO 03005279 A1 WO 03005279 A9	21/01/2003 04/11/2003 11/07/2006 15/05/2007 28/10/2008 24/04/2003 08/05/2003 17/01/2008 16/01/2003 15/04/2004
US 2002-0198985 A1	26/12/2002	None	
US 2008-0168177 A1	10/07/2008	US 07779133 B2	17/08/2010
US 2010-0082320 A1	01/04/2010	US 08131519 B2	06/03/2012