



- (51) International Patent Classification:
G06F 17/00 (2006.01)
- (21) International Application Number:
PCT/IB2015/051978
- (22) International Filing Date:
18 March 2015 (18.03.2015)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
14/220,921 20 March 2014 (20.03.2014) US
- (71) Applicant: INTERNATIONAL BUSINESS MACHINES CORPORATION [US/US]; New Orchard Road, Armonk, New York 10504 (US).
- (71) Applicants (for MG only): IBM UNITED KINGDOM LIMITED [GB/GB]; PO Box 41, North Harbour, Portsmouth Hampshire PO6 3AU (GB). IBM (CHINA) INVESTMENT COMPANY LIMITED [CN/CN]; 25/F,

Pangu Plaza, No.27, Central North 4th Ring Road, Chaoyang District, Beijing 100101 (CN).

- (72) Inventors: OLSON, John, Thomas; IBM Corporation, Mail Drop 9022-2 217, 9000 S Rita Road, Tucson, Arizona 85744-0002 (US). BALDWIN, Duane, Mark; IBM Corporation, Mail Drop 005-2, 3605 Highway 52 North, Rochester, Minnesota 55901-1407 (US). PUNADIKAR, Sachin, Chandrakant; IBM India Private Limited, Mail Drop Etb1 3f B161, Rajiv Gandhi Infotech Pk Phase 2, Plot No PI-3 Midc, Hinjewadi, Village Limit Of Marunji, Pune MH 411057 (IN). PATIL, Sandeep, Ramesh; IBM India Private Limited, Mail Drop Etb1 3f B109, Rajiv Gandhi Infotech Pk Phase 2, Plot No PI-3 Midc, Hinjewadi, Village Limit Of Marunji, Pune MH 411057 (IN).
- (74) Agent: SHAW, Anita; IBM United Kingdom Limited, Intellectual Property Law, Hursley Park, Winchester Hampshire SO21 2JN (GB).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,

[Continued on next page]

(54) Title: ACCELERATED ACCESS TO OBJECTS IN AN OBJECT STORE IMPLEMENTED UTILIZING A FILE STORAGE SYSTEM

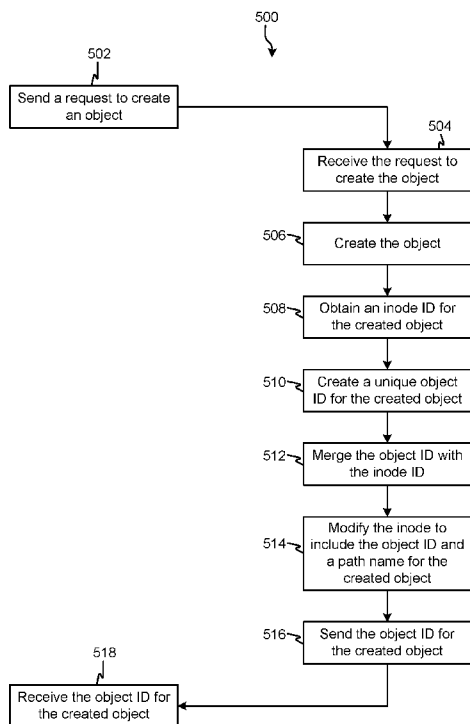


FIG. 5

(57) Abstract: In one embodiment, a method for creating an object includes creating an object in an object store making use of a file system, obtaining an index node (inode) identifier (ID) for an inode associated with the object, creating an object ID for the object, wherein the object ID is unique from any other object IDs in the object store, merging the object ID with the inode ID, and modifying the inode to include at least the object ID. In another embodiment, a system includes a processor and logic integrated with and/or executable by the processor, the logic being configured to: create an object in an object store making use of a file system, obtain an inode ID for an inode associated with the object, create an object ID for the object, the object ID including the inode ID, and modify the inode to include at least the object ID.

WO 2015/140728 A1

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))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

ACCELERATED ACCESS TO OBJECTS IN AN OBJECT STORE IMPLEMENTED UTILIZING A FILE STORAGE SYSTEM

FIELD OF THE INVENTION

[0001] The present invention relates to accessing objects in an object store, and more specifically, to accelerated access to objects that have been stored in an object store which utilizes a file storage system for its implementation.

BACKGROUND OF THE INVENTION

[0002] Object store is a storage technique that is useful in many different storage systems, including cloud storage, among others, where object store is the most prominent storage used. Object storage, is mainly intended to handle exponential growth of unstructured data. Unlike traditional storage of files in network-attached storage (NAS) or blocks in a storage area network (SAN), object store uses data objects. Each object is assigned a unique object identifier (ID) and each object contains its own metadata (part of which is user defined metadata, like tags associated with objects) along with actual data, thereby removing the need for centralized indexing. The user defined metadata (or tags) is a strong differentiator of object store-based storage techniques and is used extensively throughout object store. Thus, it enables massive scalability and geographic independence of storage locations, while maintaining reasonable costs.

[0003] The important elements of an object store are “objects” and “containers.” The objects and containers in the object store are identified with an object ID. The object ID is a universally unique ID (UUID) given to a particular object or container in a particular object store. The key usage and purpose of object ID is to allow for easy retrieval of objects and containers.

[0004] Depending on the particular implementation of object store, the object ID may be referred to as an object name, an object key, or by some other name, but will be referred to as object ID throughout. Accessing these objects/containers is made easy by just providing

the object ID while requesting the object/container. This hides the implementation details from the end user and management of the objects is simplified. Most implementations of object store use a file system to store the objects and containers. In a file system semantic, a full path is provided to a file and/or directory to access a particular file or directory. So in a file system semantic, depending on the location in the file system, the objects and/or containers will have different path names. This can be very easily observed when the object association is moved from one container to another, and then the same object gets a different path, thereby making it difficult to manage. Therefore, using an object ID is the preferred mechanism for accessing the objects/containers.

[0005] One of the biggest needs in cloud based object store is to improve overall performance. This calls for optimizing the object store system. One way to optimize the object store system is to optimize the object retrieval process which may lead to increased performance. Particularly, in an object store implementation, such as one based on cloud data management interface (CDMI), OpenStack Swift, etc., the objects are being stored on a file system, with each object mapping to a file and each container mapping to a directory. Inside the file system, there is no object and everything is being accessed/operated as a file. To the outside world, the objects are allowed and preferred to be referred to as object IDs. These object IDs are unique and both objects and containers may be accessed with an object ID.

[0006] When one tries to access the object/container using an object ID, then at the object store server (which is responding to the requests), there needs to be a mapping made available from the object ID to the file path name, and then the path name may be used for further processing. Considering a huge number of objects being stored in a cloud environment, mapping from object ID to file path name becomes time consuming and is one of the constraining factors for object store performance related to retrieval of data.

[0007] A generic method being followed to solve this issue is to maintain an in-memory copy of this mapping. When a very large number of such records is stored, this represents an impractical solution. Another mechanism would be to create a database which will keep this mapping information keeping the object ID as a primary key. The second method uses less

memory, but is more computationally intensive. The huge list of object IDs may need to be searched in order to find the one object in the request, and then to retrieve the associated file name.

[0008] Hence, smarter object ID generation and mapping would be beneficial to help optimize the system with regard to performance and memory.

DISCLOSURE OF THE INVENTION

[0009] In one embodiment, a method for creating an object includes creating an object in an object store making use of a file system, obtaining an index node (inode) identifier (ID) for an inode associated with the object, creating an object ID for the object, wherein the object ID is unique from any other object IDs in the object store, merging the object ID with the inode ID, and modifying the inode to include at least the object ID.

[0010] In another embodiment, a computer program product for accessing an object includes a computer readable storage medium having program code embodied therewith, the program code readable and/or executable by a processor to cause the processor to: extract, by the processor, an inode ID from an object ID associated with an object stored in an object store making use of a file system, access, by the processor, an inode associated with the inode ID and determine an object ID stored therein, determine, by the processor, whether the object ID from the inode matches the object ID associated with the object from which the inode ID was extracted, and perform, by the processor, an operation on the object.

[0011] In yet another embodiment, a system includes a processor and logic integrated with and/or executable by the processor, the logic being configured to: create an object in an object store making use of a file system, obtain an inode ID for an inode associated with the object, create an object ID for the object, the object ID including the inode ID, wherein the object ID is unique from any other object IDs in the object store, and modify the inode to include at least the object ID.

[0012] Other aspects and embodiments of the present invention will become apparent from the following detailed description, which, when taken in conjunction with the drawings, illustrate by way of example the principles of the invention.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0013] A preferred embodiment of the invention will now be described, by way of example only, and with reference to the following drawings:

FIG. 1 illustrates a network architecture, in accordance with one embodiment.

FIG. 2 shows a representative hardware environment that may be associated with the servers and/or clients of FIG. 1, in accordance with one embodiment.

FIG. 3 illustrates a tiered data storage system in accordance with one embodiment.

FIG. 4 shows an example of an index node (inode) structure, according to one embodiment.

FIG. 5 is a flowchart of a method for creating an object according to one embodiment.

FIG. 6 is a flowchart of a method for accessing an object according to one embodiment.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

[0014] The following description is made for the purpose of illustrating the general principles of the present invention and is not meant to limit the inventive concepts claimed herein. Further, particular features described herein can be used in combination with other described features in each of the various possible combinations and permutations.

[0015] Unless otherwise specifically defined herein, all terms are to be given their broadest possible interpretation including meanings implied from the specification as well as meanings understood by those skilled in the art and/or as defined in dictionaries, treatises, etc.

[0016] It must also be noted that, as used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless otherwise specified. It will be further understood that the terms "comprises" and/or "comprising," when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof.

[0017] The following description discloses several preferred embodiments of systems, methods and computer program products for enhancing a file system metadata layer to be cognizant of an object store residing over the file system and an object identifier (ID) associated with the object belonging to the object store. The technique helps improve on the performance of object look-up/retrieval for the object store.

[0018] In one general embodiment, a method for creating an object includes creating an object in an object store making use of a file system, obtaining an inode node (inode) identifier (ID) for an inode associated with the object, creating an object ID for the object, wherein the object ID is unique from any other object IDs in the object store, merging the object ID with the inode ID, and modifying the inode to include at least the object ID.

[0019] In another general embodiment, a computer program product for accessing an object includes a computer readable storage medium having program code embodied therewith, the program code readable and/or executable by a processor to cause the processor to: extract, by the processor, an inode ID from an object ID associated with an object stored in an object store making use of a file system, access, by the processor, an inode associated with the inode ID and determine an object ID stored therein, determine, by the processor, whether the object ID from the inode matches the object ID associated with the object from which the inode ID was extracted, and perform, by the processor, an operation on the object.

[0020] In yet another general embodiment, a system includes a processor and logic integrated with and/or executable by the processor, the logic being configured to: create an object in an object store making use of a file system, obtain an inode ID for an inode

associated with the object, create an object ID for the object, the object ID including the inode ID, wherein the object ID is unique from any other object IDs in the object store, and modify the inode to include at least the object ID.

[0021] FIG. 1 illustrates an architecture 100, in accordance with one embodiment. As shown in FIG. 1, a plurality of remote networks 102 are provided including a first remote network 104 and a second remote network 106. A gateway 101 may be coupled between the remote networks 102 and a proximate network 108. In the context of the present architecture 100, the networks 104, 106 may each take any form including, but not limited to a LAN, a WAN such as the Internet, public switched telephone network (PSTN), internal telephone network, etc.

[0022] In use, the gateway 101 serves as an entrance point from the remote networks 102 to the proximate network 108. As such, the gateway 101 may function as a router, which is capable of directing a given packet of data that arrives at the gateway 101, and a switch, which furnishes the actual path in and out of the gateway 101 for a given packet.

[0023] Further included is at least one data server 114 coupled to the proximate network 108, and which is accessible from the remote networks 102 via the gateway 101. It should be noted that the data server(s) 114 may include any type of computing device/groupware. Coupled to each data server 114 is a plurality of user devices 116. Such user devices 116 may include a desktop computer, lap-top computer, hand-held computer, printer or any other type of logic. It should be noted that a user device 111 may also be directly coupled to any of the networks, in one embodiment.

[0024] A peripheral 120 or series of peripherals 120, e.g., facsimile machines, printers, networked and/or local storage units or systems, etc., may be coupled to one or more of the networks 104, 106, 108. It should be noted that databases and/or additional components may be utilized with, or integrated into, any type of network element coupled to the networks 104, 106, 108. In the context of the present description, a network element may refer to any component of a network.

[0025] According to some approaches, methods and systems described herein may be implemented with and/or on virtual systems and/or systems which emulate one or more other systems, such as a UNIX (UNIX is a registered trademark of The Open Group in the United States and other countries) system which emulates an IBM z/OS (IBM and z/OS are registered trademarks of International Business Machined Corporation) environment, a UNIX system which virtually hosts a MICROSOFT WINDOWS (Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries, or both) environment, a MICROSOFT WINDOWS system which emulates an IBM z/OS environment, etc. This virtualization and/or emulation may be enhanced through the use of VMWARE software, in some embodiments.

[0026] In more approaches, one or more networks 104, 106, 108, may represent a cluster of systems commonly referred to as a "cloud." In cloud computing, shared resources, such as processing power, peripherals, software, data, servers, etc., are provided to any system in the cloud in an on-demand relationship, thereby allowing access and distribution of services across many computing systems. Cloud computing typically involves an Internet connection between the systems operating in the cloud, but other techniques of connecting the systems may also be used.

[0027] FIG. 2 shows a representative hardware environment associated with a user device 116 and/or server 114 of FIG. 1, in accordance with one embodiment. Such figure illustrates a typical hardware configuration of a workstation having a central processing unit 210, such as a microprocessor, and a number of other units interconnected via a system bus 212.

[0028] The workstation shown in FIG. 2 includes a Random Access Memory (RAM) 214, Read Only Memory (ROM) 216, an I/O adapter 218 for connecting peripheral devices such as disk storage units 220 to the bus 212, a user interface adapter 222 for connecting a keyboard 224, a mouse 226, a speaker 228, a microphone 232, and/or other user interface devices such as a touch screen and a digital camera (not shown) to the bus 212, communication adapter 234 for connecting the workstation to a communication network 235

(e.g., a data processing network) and a display adapter 236 for connecting the bus 212 to a display device 238.

[0029] The workstation may have resident thereon an operating system such as the Microsoft Windows® Operating System (OS), a MAC OS, a UNIX OS, etc. It will be appreciated that a preferred embodiment may also be implemented on platforms and operating systems other than those mentioned. A preferred embodiment may be written using JAVA (Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates), XML, C, and/or C++ language, or other programming languages, along with an object oriented programming methodology. Object oriented programming (OOP), which has become increasingly used to develop complex applications, may be used.

[0030] Now referring to FIG. 3, a storage system 300 is shown according to one embodiment. Note that some of the elements shown in FIG. 3 may be implemented as hardware and/or software, according to various embodiments. The storage system 300 may include a storage system manager 312 for communicating with a plurality of media on a higher storage tier 302 and a lower storage tier 306. The higher storage tier 302 preferably may include one or more random access and/or direct access media 304, such as hard disks in hard disk drives (HDDs), nonvolatile memory (NVM), solid state memory in solid state drives (SSDs), etc., and/or others noted herein. The lower storage tier 306 may preferably include one or more sequential access media 308, such as magnetic tape in tape drives, optical media, etc., and/or others noted herein. Additional storage tiers 316 may include any combination of storage memory media. The storage system manager 312 may communicate with the storage media 304, 308 on the higher and lower storage tiers 302, 306 through a network 310, such as a storage area network (SAN), as shown in FIG. 3. The storage system manager 312 may also communicate with one or more host systems (not shown) through a host interface 314, which may or may not be a part of the storage system manager 312. The storage system manager 312 and/or any other component of the storage system 300 may be implemented in hardware and/or software, and may make use of a processor (not shown) for executing commands of a type known in the art, such as a central processing unit (CPU), a field programmable gate array (FPGA), an application specific integrated circuit (ASIC), etc.

Of course, any arrangement of a storage system may be used, as will be apparent to those of skill in the art upon reading the present description.

[0031] In more embodiments, the storage system 300 may include any number of data storage tiers, and may include the same or different storage memory media within each storage tier. For example, each data storage tier may include the same type of storage memory media, such as HDDs, SSDs, sequential access media (tape in tape drives, optical disk in optical disk drives, etc.), direct access media (CD-ROM, DVD-ROM, etc.), or any combination of media storage types. In one such configuration, a higher storage tier 302, may include a majority of SSD storage media for storing data in a higher performing storage environment, and remaining storage tiers, including lower storage tier 306 and additional storage tiers 316 may include any combination of SSDs, HDDs, tape drives, etc., for storing data in a lower performing storage environment. In this way, more frequently accessed data, data having a higher priority, data needing to be accessed more quickly, etc., may be stored to the higher storage tier 302, while data not having one of these attributes may be stored to the additional storage tiers 316, including lower storage tier 306. Of course, one of skill in the art, upon reading the present descriptions, may devise many other combinations of storage media types to implement into different storage schemes, according to the embodiments presented herein.

[0032] According to some embodiments, the storage system (such as 300) may include logic adapted to receive a request to open a data set, logic adapted to determine if the requested data set is stored to a lower storage tier 306 of a tiered data storage system 300 in multiple associated portions, logic adapted to move each associated portion of the requested data set to a higher storage tier 302 of the tiered data storage system 300, and logic adapted to assemble the requested data set on the higher storage tier 302 of the tiered data storage system 300 from the associated portions.

[0033] Of course, this logic may be implemented as a method on any device and/or system or as a computer program product, according to various embodiments.

[0034] In order to implement an enhanced object ID in an object store environment, some structures and/or formats are described according to various embodiments. The structures and/or formats include the object ID format, the underlying file system inode tables that hold the various objects, and object creation processing. These structures and/or formats are integrated in an object store implementation on a file system that provides improved object retrieval performance.

[0035] To reduce the resource utilization in current methods of mapping an object ID to file, data set, directory, etc., a method is proposed as described below. First, the object ID is enhanced to include an additional inode identifier (such as a number, alphanumeric string, etc., that uniquely identifies the inode). The identifier may be any number, alphanumeric string, hash, etc., that uniquely identifies the inode. In place of an inode, which refers to an index node, depending on the file system being used, other file indexing structures may be utilized in the context of embodiments described herein. For example, UNIX-based systems may utilize an inode, but a MICROSOFT operating system may rely on a different indexing structure that acts similarly to an inode, as would be understood by one of skill in the art, and it is this different indexing structure which is utilized in the context of the embodiments described herein. However, for the sake of simplicity, all indexing structures will be referred to as an inode, with the assumption that any indexing structure may be used.

[0036] An inode represents metadata about an object in an object store (or a file in a file system semantic). Referring now to FIG. 4, the metadata may include some or all of the following information: size of the file, a device ID where the file is stored, an user ID associated with the file, a group ID associated with the file, file mode information and access privileges for the owner, the group, and possibly others, one or more file protection flags, a timestamp for file creation, a timestamp for file modification, a timestamp for file access, etc., a link counter to determine a number of hard links associated with the file, one or more pointers associated with the blocks storing the file's contents, extended attributes such as append only, immutability (no one can delete file including root user), etc., or any other information known in the art.

[0037] In one embodiment, the inode may be modified to include a unique ID of the object and a path name of the object within the extended attributes field, along with any other information desired to be stored in the extended attributes field.

[0038] Most object store implementations utilize object IDs that include a universally unique identifier (UUID) at least included therein. Thus, the object ID may be enhanced to store the inode number as well as any other information that is already stored therein and/or that may be useful for other identification purposes, depending on the file system being used, in one approach. In this way, implementations of object ID may be enhanced with the addition of the inode identifier (or some other identifier of the indexing structure depending on the file system being used) related to a particular object/container. The creation of any object/container will now cause an enhanced object ID to also be created which includes the inode identifier in the object ID.

[0039] Referring again to FIG. 4, an inode 402 of the particular file system being used may also be modified to store this enhanced object ID 426 (or at least the UUID included therein) and the explicit path name of the object/container 428.

[0040] Then, whenever an object is being accessed via an object ID, and as the server or system accessing the object extracts the inode identifier from that object ID, the server will then access the respective inode information present in an inode register (such as an inode table), which allows the inode to be identified, which may include information about the object/container.

[0041] The inode 402 may include the object ID 426 associated with the object/container, such as in the extended attributes field. Other information may also be included, such as type 404, owner 406, size 408, group 410, permissions 412, last access time 414, last modified time 416, created time 418, number of links 420, extended attributes field 422, pointers to data blocks 424, etc. In one embodiment, the object ID 426 and/or the explicit path name of the object/container 428 may be stored in the extended attributes field 422 of the inode 402, along with any other attributes 430 desired to be stored in the extended attributes field 422.

[0042] Next, the object ID may be passed/requested by the end user/application and the object ID stored in the inode will be matched thereto. When a match between object IDs is confirmed, it is assured that the inode being referred to belongs to the object being requested. Then, the file name will be read from the inode, and the path obtained will be used for further processing of retrieving the object/container. This direct low level inode relation with objects via mapping and integration of object IDs and inode results in faster retrieval of objects via the object ID associated therewith.

[0043] As accessing any inode is an $O(1)$ operation, it is guaranteed that this method of accessing an object/container is faster than any currently used methods. As there is no need to keep the mapping table/database either in-memory or on a HDD or other direct access storage device (DASD), less resources are consumed using the methods and structures described herein according to various embodiments.

[0044] For the following descriptions, it is assumed that in an object store implementation using a file system, an object is stored as a file, and a container is stored as a directory. In a file system semantic, file and directory are both stored as files in a LINUX/UNIX operating system. Information related to files/directories, referred to as metadata, may be stored in an inode.

[0045] To access any file/directory on a file system, the inode is accessed and relevant information is read therefrom. Any further tasks may also be carried out at this point. So in any case, accessing any file/directory/device in Linux/Unix occurs via an access of the inode.

[0046] In order to implement the methods described herein according to various embodiments, some changes may be made in the structure of the inode. In one embodiment, the object ID and the path to the object (such as a file or directory) may be stored in the extended attributes of the inode. The path to the object is used when the object store considers the permissions set on the container object.

[0047] In another embodiment, some changes may be made during an object creation process. As part of an object/container creation process, traditionally, the object ID is generated by the object store server. This object ID creation guarantees uniqueness across the object store.

[0048] However, according to one embodiment, the existing object ID is modified to include the inode identifier therein. The inode identifier may also be appended to the created object ID. This modified object ID and the path name (to the file holding the object ID on the file system) may be added to the extended attributes of that file.

[0049] Now referring to FIG. 5, a flowchart of a method 500 for creating an object is shown according to one embodiment. Method 500 may be executed in any desired environment, including those shown in FIGS. 1-4, among others. Furthermore, more or less operations than those specifically described in FIG. 5 may be included in method 500.

[0050] In one embodiment, operations 502 and 518 may be performed by a client application, device, system, etc., in an attempt to create an object on the object store. In this embodiment, operations 504-516 may be performed on the object store server, file system, or some other system, application, or device connected to the object store.

[0051] In operation 502, a request to create an object is sent from a device, system, application, etc., to the object store and/or file system implementing the object store. This request may be sent by the client, a system, an application, or any other system, device, application, etc., known in the art. The request may include any information that would be useful in creating the object and/or container as would be understood by one of skill in the art.

[0052] In operation 504, the request to create the object is received by the object store and/or the file system implementing the object store, and any useful information therein is identified.

[0053] In operation 506, the object is created on the file system. The object may be of any type known in the art, such as a file, data set, directory, container, etc.

[0054] In operation 508, an inode ID, such as an inode number, is obtained for the created object. Any process of obtaining the inode ID known in the art may be used, such as by requesting the inode ID via a user interface with the file system. In one such embodiment, a user may enter a command to retrieve an inode associated with a file/directory.

[0055] One exemplary command and response exchange is shown below:

```
>> # ls -i test
```

```
>> 21733378 test
```

In this exemplary exchange, 21733378 is the inode number which is returned in response to the command requesting the inode number for test.

[0056] In operation 510, a unique object ID is created for the object. The unique object ID may include a UUID, as would be understood by one of skill in the art upon reading the present descriptions. Furthermore, the object ID may take any form known in the art, such as a number, alphanumeric string, hash, etc.

[0057] In operation 512, the object ID is merged with the inode ID. In one embodiment, the inode ID may be appended to the created object ID (at a beginning thereof, an end thereof, or included in some other manner that is known to the system for extraction thereof), thereby ensuring that an association between the object and the inode may be recalled at any given time from the object ID.

[0058] In operation 514, the inode is modified to include the object ID and possibly the file path of the object as metadata to the inode. These values may be stored to the extended attributes of the inode, in one embodiment. The storage of these file storage search parameters allows a search to be performed for the object without needing to consult an index or other correlation structure to determine a location of the object outside of the inode. Furthermore, since the inode ID is provided in the object ID, it is known where to find the information to locate the object in the file system.

[0059] In operation 516, the object ID is sent (returned) to the client, system, application, or other device which sent the request.

[0060] In operation 518, the object ID is received. Once the client, system, application, or other device which requested the object to be created receives the object ID, any future access requests (read, write, etc.) to that object may be made via the object ID, thereby speeding up the process of accessing the object associated with the object ID.

[0061] Furthermore, in one approach, some changes may be made to object accessing methods. The client application, device, system, etc., may pass on the object ID (which may be provided by the object store server, at the time of creating the object) to the object store server for operations like read, modify, delete, etc. The server is configured to read the object ID. From that object ID, the server is configured to extract the inode ID (which was appended to and/or included in the object ID). Then, the server is configured to access the inode and any information stored therewith, such as in its extended attributes.

[0062] Next, the stored object ID is compared with the object ID that was received. When the two object IDs match, the intended operation is allowed, otherwise, a reply is generated indicating that the object is not found and/or invalid.

[0063] When the object store implementation mandates permissions on containers (directories in file system semantics), then the path name stored in the object ID (such as in the extended attributes) may be used to check the permissions. Unmodified object store has no concept of permissions on container objects.

[0064] In this way, when utilizing object store, a second metadata may be merged with each object ID to provide object storage search parameters to more efficiently access the object in the object store.

[0065] Now referring to FIG. 6, a flowchart of a method 600 for accessing an object is shown according to one embodiment. Method 600 may be executed in any desired

environment, including those shown in FIGS. 1-4, among others. Furthermore, more or less operations than those specifically described in FIG. 6 may be included in method 600.

[0066] In one embodiment, operations 602 and 618 may be performed by a client application, device, system, etc., in an attempt to create an object on the object store. In this embodiment, operations 604-616 may be performed on the object store server, file system, or some other system, application, or device connected to the object store.

[0067] In operation 602, a request to access an object is sent from a device, system, application, etc., to the object store and/or file system implementing the object store. This request may be sent by any client system, device, application, etc., known in the art. The request may include any information that would be useful in accessing the object and/or container as would be understood by one of skill in the art. In one embodiment, the request includes an object ID associated with the object to be accessed. In a further embodiment, an operation to be performed on the object is also included in the request (i.e., read/write/modify, etc.).

[0068] In operation 604, the request to access the object is received by the object store and/or the file system implementing the object store, and any useful information therein is identified, including the object ID for the object to be accessed.

[0069] In operation 606, an inode ID, such as an inode number, is extracted from the object ID.

[0070] In operation 608, the inode associated with the inode ID is accessed and an object ID stored therein is determined. This may be accomplished by reading the inode associated with the inode ID.

[0071] In operation 610, it is determined whether the object ID included in the request matches with an object ID included in the inode associated with the inode ID. When the object IDs match, method 600 continues to optional operation 612; otherwise, method 600 proceeds to operation 616.

[0072] In operation 612, a path name stored in the inode is determined, should the path name be required for further processing of the object. Operation 612 is not performed when the path name is not required for further processing of the object.

[0073] In operation 614, the operation included in the request is performed on the object, such as reading a portion or all of the object, writing over a portion or all of the object, appending information to the object, modifying the object, deleting the object, etc.

[0074] In operation 616, results of the access are sent, such as back to the client system, device, application, etc., which originally sent the request. When the object IDs do not match, the results indicate that the object does not exist or is not accessible, and/or the operation failed. For any of the other operations, results of the read/write/modify are indicated to the requesting entity.

[0075] In operation 618, results of the access are received, such as by the client, system, application, or other device which sent the request.

[0076] The present invention may be a system, a method, and/or a computer program product. The computer program product may include a computer readable storage medium (or media) having computer readable program instructions thereon for causing a processor to carry out aspects of the present invention.

[0077] For example, in one embodiment, a computer program product for accessing an object may comprise a computer readable storage medium having program code embodied therewith, the program code readable and/or executable by a processor to cause the processor to: extract, by the processor, an index node (inode) identifier (ID) from an object ID associated with an object stored in an object store making use of a file system, access, by the processor, an inode associated with the inode ID and determine an object ID stored therein, determine, by the processor, whether the object ID from the inode matches the object ID associated with the object from which the inode ID was extracted, and perform, by the processor, an operation on the object.

[0078] In various embodiments, the program code readable and/or executable by the processor may further causes the processor to: receive a request to access the object stored in the object store from a remote system and send results of accessing the object after performing the operation thereon to the remote system, wherein the request includes the object ID associated with the object to be accessed, and the operation to perform on the object. In another embodiment, the operation to perform on the object may include at least one of reading, writing, and modifying. According to another embodiment, the program code readable and/or executable by the processor may further cause the processor to determine a path name from the inode when the path name is required for further processing of the object.

[0079] In another example, a system may comprise a processor and logic integrated with and/or executable by the processor, the logic being configured to: create an object in an object store making use of a file system, obtain an inode ID for an inode associated with the object, create an object ID for the object, the object ID including the inode ID, wherein the object ID is unique from any other object IDs in the object store, and modify the inode to include at least the object ID.

[0080] In various embodiments, the logic may be further configured to: receive a request to create the object in the object store from a remote system prior to creating the object, and send the object ID for the object to the remote system after creating the object. In another embodiment, the logic may be further configured to modify the inode to include a path name for the object, wherein extended attributes of the inode may be modified to include the object ID and the path name for the object.

[0081] In another embodiment, the object ID may have the inode ID appended thereto, such as at a start or an end thereof. In further embodiments, the logic may be further configured to: extract the inode ID from the object ID associated with the object stored in the object store, access the inode associated with the inode ID and determine an object ID stored therein, determine whether the object ID from the inode matches the object ID associated with the object from which the inode ID was extracted, and perform an operation on the object.

[0082] In some approaches, the logic may be further configured to: receive a request to access the object stored in the object store from a remote system, and send results of accessing the object after performing the operation thereon to the remote system.

[0083] Furthermore, any of the following is possible: the request includes the object ID associated with the object to be accessed, and the operation to perform on the object, the operation to perform on the object includes at least one of reading, writing, and modifying, and the logic may be further configured to determine a path name from the inode when the path name is required for further processing of the object.

[0084] The computer readable storage medium can be a tangible device that can retain and store instructions for use by an instruction execution device. The computer readable storage medium may be, for example, but is not limited to, an electronic storage device, a magnetic storage device, an optical storage device, an electromagnetic storage device, a semiconductor storage device, or any suitable combination of the foregoing. A non-exhaustive list of more specific examples of the computer readable storage medium includes the following: a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), a static random access memory (SRAM), a portable compact disc read-only memory (CD-ROM), a digital versatile disk (DVD), a memory stick, a floppy disk, a mechanically encoded device such as punch-cards or raised structures in a groove having instructions recorded thereon, and any suitable combination of the foregoing. A computer readable storage medium, as used herein, is not to be construed as being transitory signals per se, such as radio waves or other freely propagating electromagnetic waves, electromagnetic waves propagating through a waveguide or other transmission media (e.g., light pulses passing through a fiber-optic cable), or electrical signals transmitted through a wire.

[0085] Computer readable program instructions described herein can be downloaded to respective computing/processing devices from a computer readable storage medium or to an external computer or external storage device via a network, for example, the Internet, a local area network, a wide area network and/or a wireless network. The network may comprise

copper transmission cables, optical transmission fibers, wireless transmission, routers, firewalls, switches, gateway computers and/or edge servers. A network adapter card or network interface in each computing/processing device receives computer readable program instructions from the network and forwards the computer readable program instructions for storage in a computer readable storage medium within the respective computing/processing device.

[0086] Computer readable program instructions for carrying out operations of the present invention may be assembler instructions, instruction-set-architecture (ISA) instructions, machine instructions, machine dependent instructions, microcode, firmware instructions, state-setting data, or either source code or object code written in any combination of one or more programming languages, including an object oriented programming language such as Java, Smalltalk, C++ or the like, and conventional procedural programming languages, such as the "C" programming language or similar programming languages. The computer readable program instructions may execute entirely on the user's computer, partly on the user's computer, as a stand-alone software package, partly on the user's computer and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through any type of network, including a local area network (LAN) or a wide area network (WAN), or the connection may be made to an external computer (for example, through the Internet using an Internet Service Provider). In some embodiments, electronic circuitry including, for example, programmable logic circuitry, field-programmable gate arrays (FPGA), or programmable logic arrays (PLA) may execute the computer readable program instructions by utilizing state information of the computer readable program instructions to personalize the electronic circuitry, in order to perform aspects of the present invention.

[0087] Aspects of the present invention are described herein with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems), and computer program products according to embodiments of the invention. It will be understood that each block of the flowchart illustrations and/or block diagrams, and combinations of blocks in the

flowchart illustrations and/or block diagrams, can be implemented by computer readable program instructions.

[0088] These computer readable program instructions may be provided to a processor of a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions, which execute via the processor of the computer or other programmable data processing apparatus, create means for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks. These computer readable program instructions may also be stored in a computer readable storage medium that can direct a computer, a programmable data processing apparatus, and/or other devices to function in a particular manner, such that the computer readable storage medium having instructions stored therein comprises an article of manufacture including instructions which implement aspects of the function/act specified in the flowchart and/or block diagram block or blocks.

[0089] The computer readable program instructions may also be loaded onto a computer, other programmable data processing apparatus, or other device to cause a series of operational steps to be performed on the computer, other programmable apparatus or other device to produce a computer implemented process, such that the instructions which execute on the computer, other programmable apparatus, or other device implement the functions/acts specified in the flowchart and/or block diagram block or blocks.

[0090] The flowchart and block diagrams in the Figures illustrate the architecture, functionality, and operation of possible implementations of systems, methods, and computer program products according to various embodiments of the present invention. In this regard, each block in the flowchart or block diagrams may represent a module, segment, or portion of instructions, which comprises one or more executable instructions for implementing the specified logical function(s). In some alternative implementations, the functions noted in the block may occur out of the order noted in the figures. For example, two blocks shown in succession may, in fact, be executed substantially concurrently, or the blocks may sometimes be executed in the reverse order, depending upon the functionality involved. It will also be noted that each block of the block diagrams and/or flowchart illustration, and combinations

of blocks in the block diagrams and/or flowchart illustration, can be implemented by special purpose hardware-based systems that perform the specified functions or acts or carry out combinations of special purpose hardware and computer instructions.

[0091] Moreover, a system according to various embodiments may include a processor and logic integrated with and/or executable by the processor, the logic being configured to perform one or more of the process steps recited herein. By integrated with, what is meant is that the processor has logic embedded therewith as hardware logic, such as an application specific integrated circuit (ASIC), a field programmable gate array (FPGA), etc. By executable by the processor, what is meant is that the logic is hardware logic, software logic such as firmware, operating system, etc., or some combination of hardware and software logic that is accessible by the processor and configured to cause the processor to perform some functionality upon execution by the processor. Software logic may be stored on local and/or remote memory of any memory type, as known in the art. Any processor known in the art may be used, such as a software processor module and/or a hardware processor such as an ASIC, a FPGA, a central processing unit (CPU), an integrated circuit (IC), etc.

[0092] It will be clear that the various features of the foregoing systems and/or methodologies may be combined in any way, creating a plurality of combinations from the descriptions presented above.

[0093] It will be further appreciated that embodiments of the present invention may be provided in the form of a service deployed on behalf of a customer to offer service on demand.

[0094] While various embodiments have been described above, it should be understood that they have been presented by way of example only, and not limitation. Thus, the breadth and scope of a preferred embodiment should not be limited by any of the above-described exemplary embodiments, but should be defined only in accordance with the following claims and their equivalents.

CLAIMS

1. A method for creating an object, the method comprising:
creating an object in an object store making use of a file system;
obtaining an index node (inode) identifier (ID) for an inode associated with the object;
creating an object ID for the object, wherein the object ID is unique from any other object IDs in the object store; and
modifying the inode to comprise at least the object ID.
2. The method as recited in claim 1, further comprising: merging the object ID with the inode ID.
3. The method as recited in claim 1 or claim 2, further comprising:
receiving a request to create the object in the object store from a remote system prior to creating the object; and
sending the object ID for the object to the remote system after creating the object.
4. The method as recited in any preceding claim, further comprising modifying the inode to comprise a path name for the object to provide search parameters for the object.
5. The method as recited in claim 4, wherein extended attributes of the inode are modified to comprise the object ID and the path name for the object.
6. The method as recited in any preceding claim, wherein the object ID has the inode ID appended thereto.
7. A method for accessing an object, the method comprising:
extracting an index node (inode) identifier (ID) from an object ID associated with an object stored in an object store making use of a file system;
accessing an inode associated with the inode ID and determine an object ID stored therein;

determining whether the object ID from the inode matches the object ID associated with the object from which the inode ID was extracted; and

performing an operation on the object when the object ID from the inode matches the object ID associated with the object from which the inode ID was extracted.

8. The method as recited in claim 7, further comprising:

receiving a request to access the object stored in the object store from a remote system; and

sending results of accessing the object after performing the operation thereon to the remote system.

9. The method as recited in claim 8, wherein the request comprises the object ID associated with the object to be accessed, and the operation to perform on the object.

10. The method as recited in any of claims 7 to 9, wherein the performing step further comprises at least one of reading, writing, and modifying.

11. The method as recited in any of claims 7 to 10, further comprising: determining a path name from the inode when the path name is required for further processing of the object.

12. A system, comprising a processor and logic integrated with and/or executable by the processor, the logic being configured to:

create an object in an object store making use of a file system;

obtain an index node (inode) identifier (ID) for an inode associated with the object;

create an object ID for the object, the object ID including the inode ID, wherein the object ID is unique from any other object IDs in the object store; and

modify the inode to comprise at least the object ID.

13. The system as recited in claim 12, wherein the logic is further configured to:

merge the object ID with the inode ID.

14. The system as recited in claim 12 or claim 13, wherein the logic is further configured to:
- receive a request to create the object in the object store from a remote system prior to creating the object; and
 - send the object ID for the object to the remote system after creating the object.
15. The system as recited in any of claims 12 to 14, wherein the logic is further configured to modify the inode to comprise a path name for the object.
16. The system as recited in claim 15, wherein extended attributes of the inode are modified to comprise the object ID and the path name for the object.
17. The system as recited in any of claims 12 to 16, wherein the object ID has the inode ID appended thereto.
18. A system, comprising a processor and logic integrated with and/or executable by the processor, the logic being configured to:
- extract the inode ID from the object ID associated with the object stored in the object store;
 - access the inode associated with the inode ID and determine an object ID stored therein;
 - determine whether the object ID from the inode matches the object ID associated with the object from which the inode ID was extracted; and
 - perform an operation on the object when the object ID from the inode matches the object ID associated with the object from which the inode ID was extracted.
19. The system as recited in claim 18, wherein the logic is further configured to:
- receive a request to access the object stored in the object store from a remote system; and
 - send results of accessing the object after performing the operation thereon to the remote system.

20. The system as recited in claim 19, wherein the request comprises the object ID associated with the object to be accessed, and the operation to perform on the object.

21. The system as recited in any of claims 18 to 20, wherein the operation to perform on the object comprises at least one of reading, writing, and modifying.

22. The system as recited in any of claims 18 to 21, wherein the logic is further configured to determine a path name from the inode when the path name is required for further processing of the object.

23. A computer program comprising computer program code stored on a computer readable medium to, when loaded into a computer system and executed thereon, cause said computer system to perform all the steps of a method according to any of claims 1 to 11.

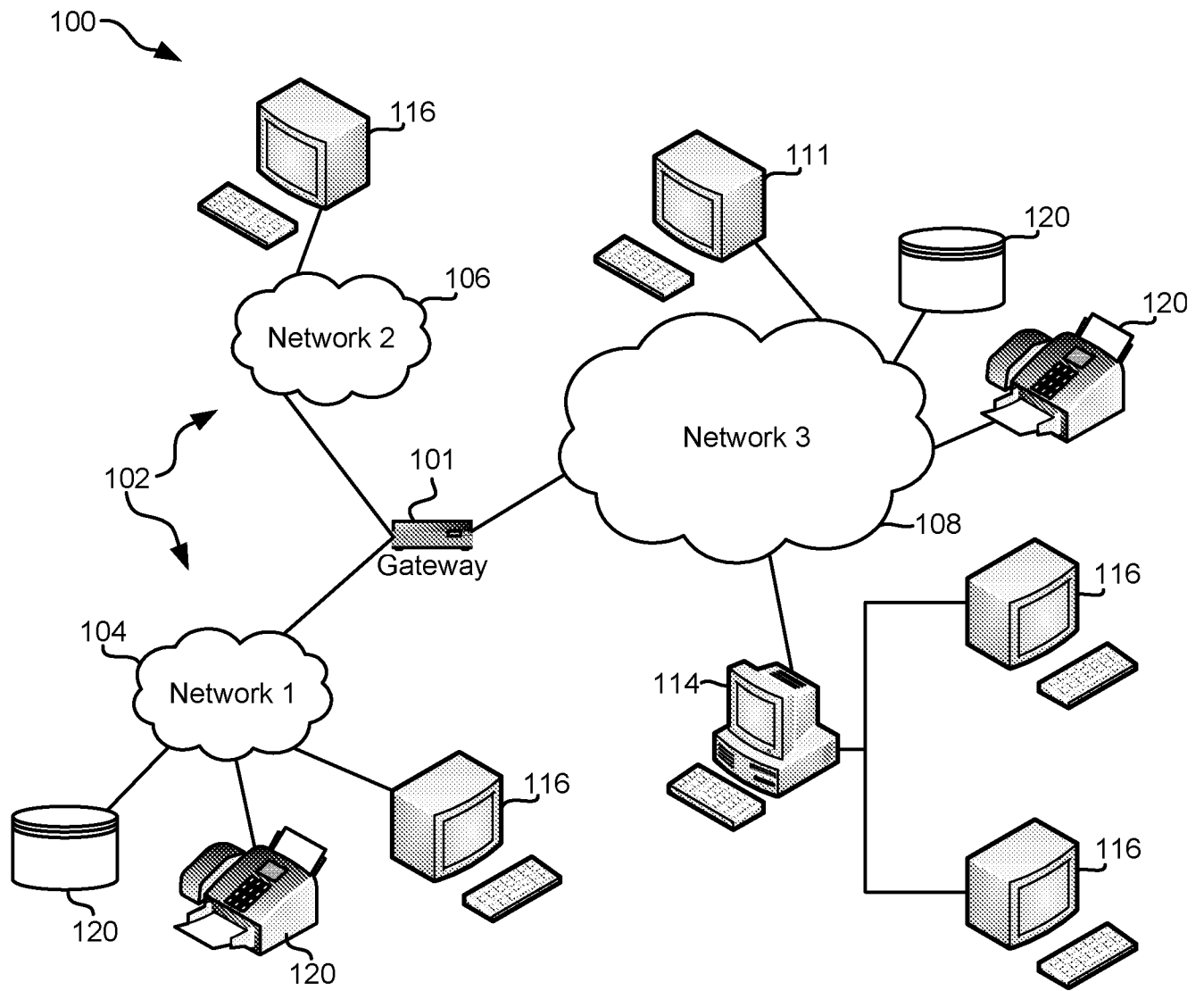


FIG. 1

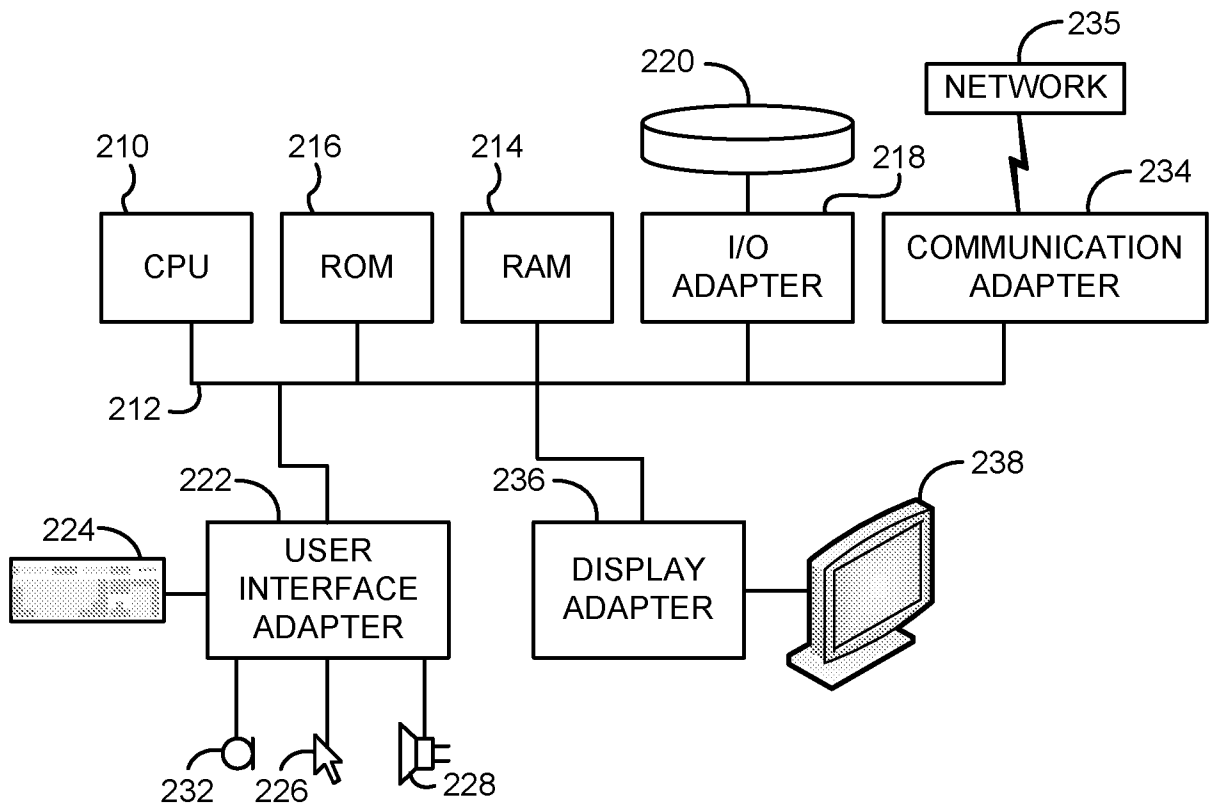


FIG. 2

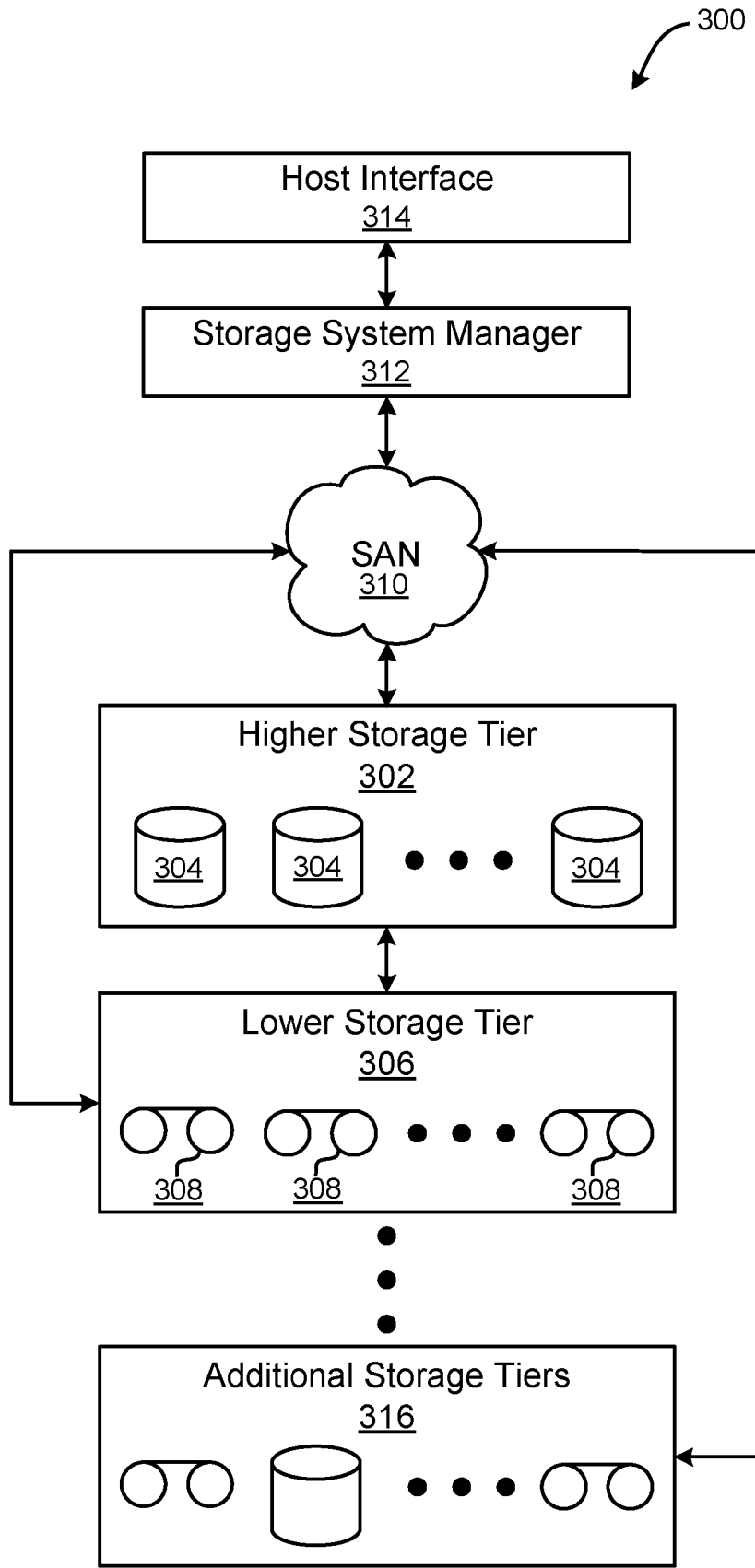


FIG. 3

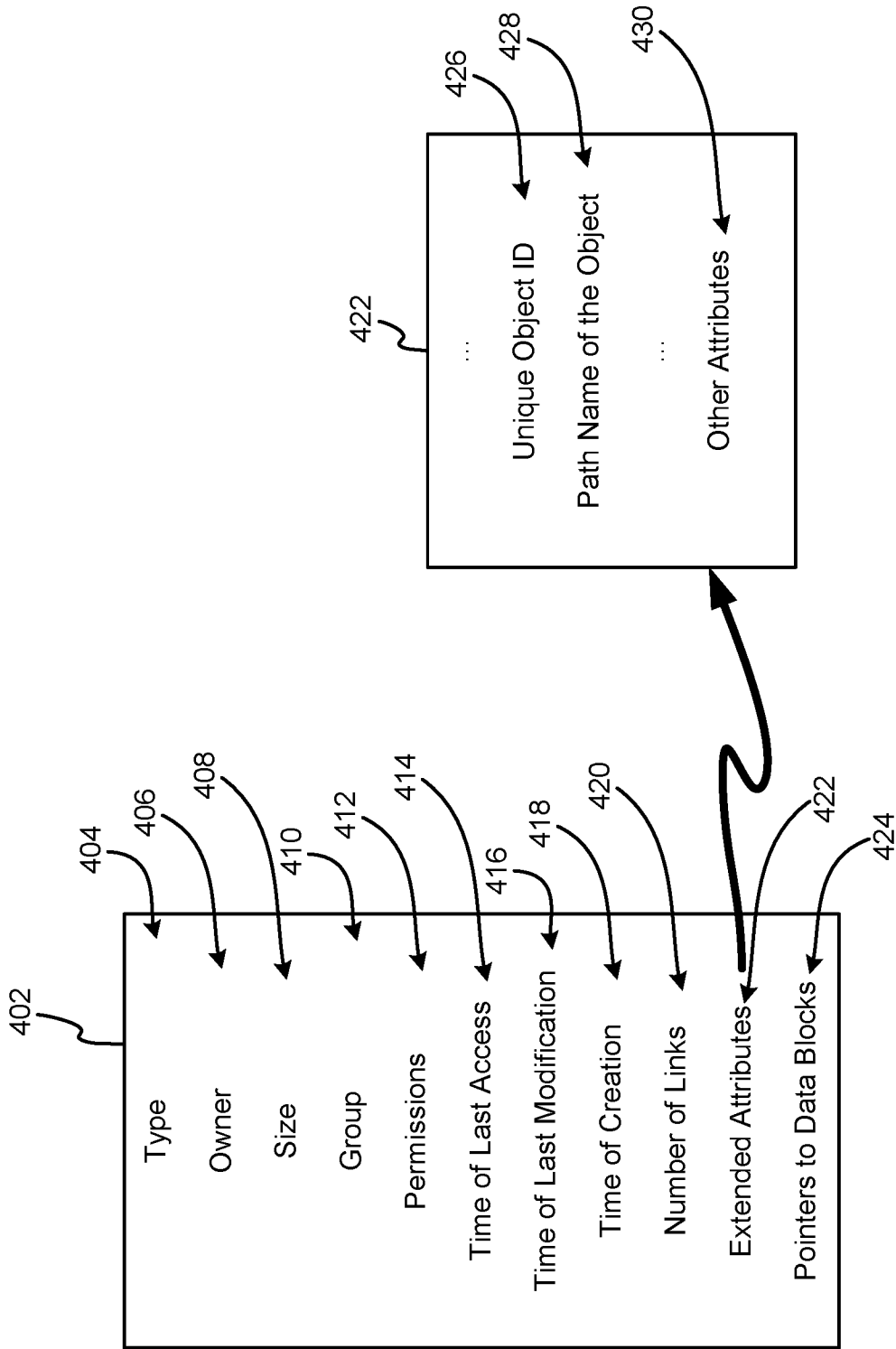


FIG. 4

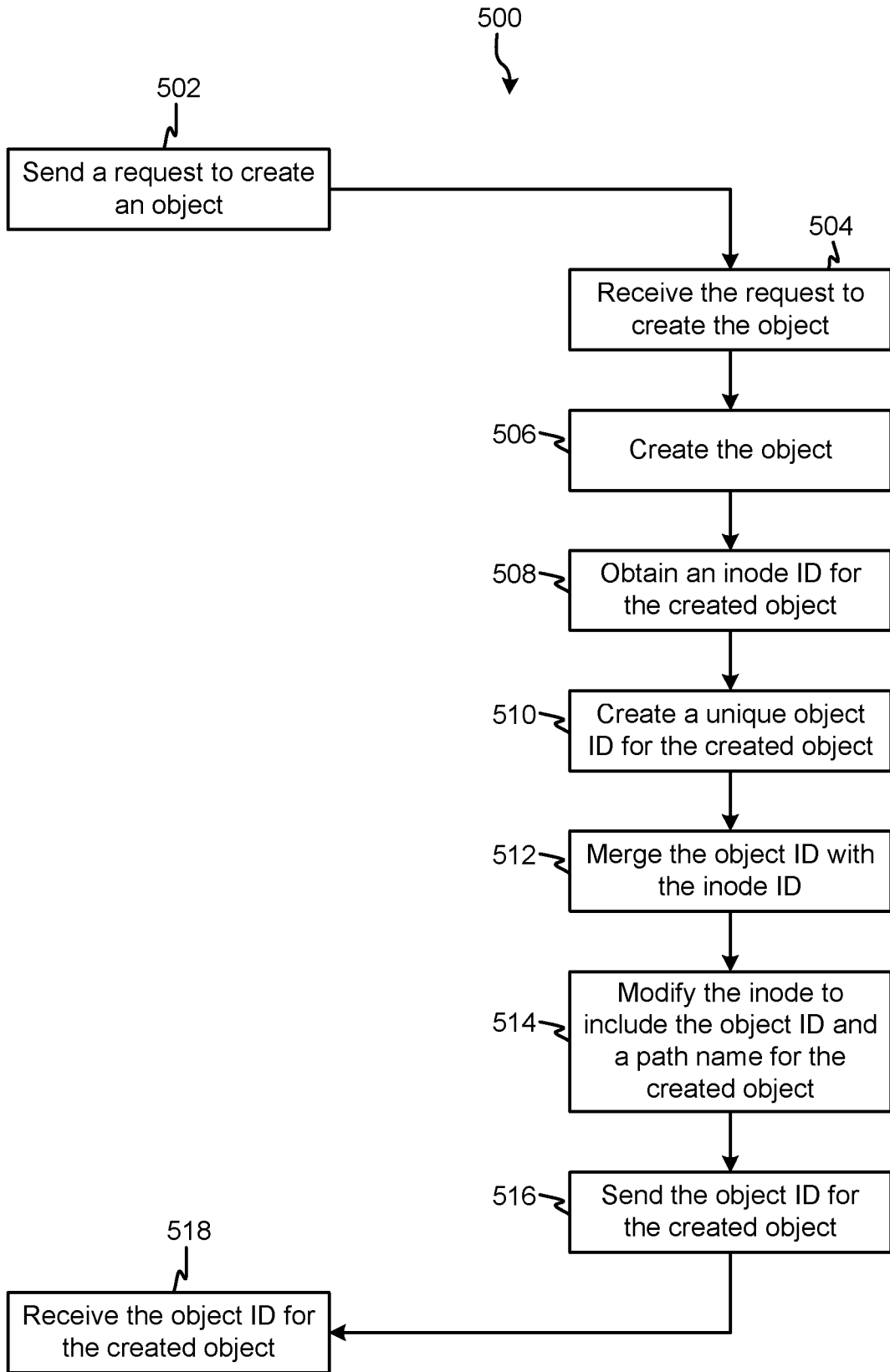


FIG. 5

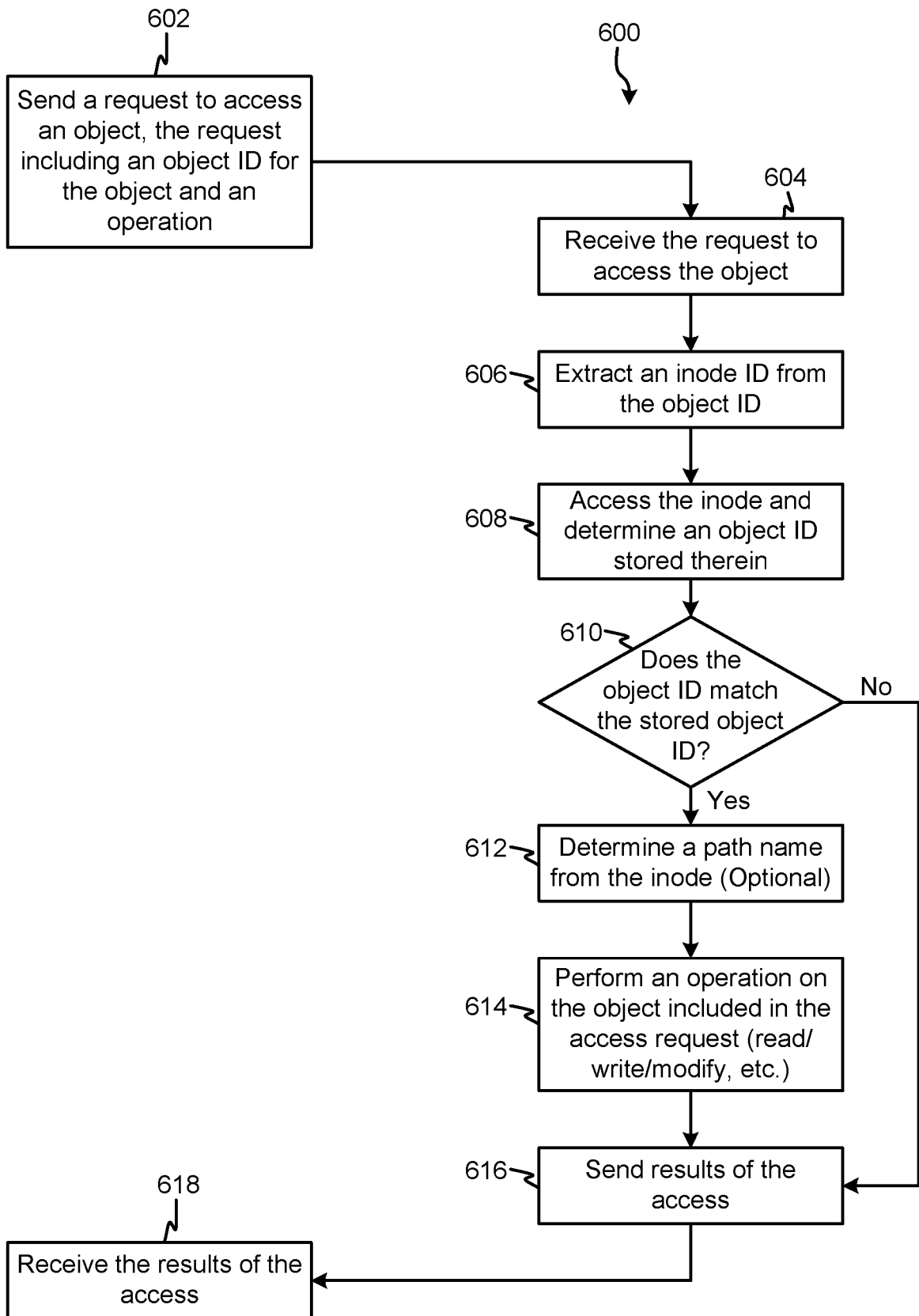


FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB2015/051978

A. CLASSIFICATION OF SUBJECT MATTER G06F 17/00(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) G06F12; G06F17 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) CNKI,CNABS, CNTXT, VEN:存储, 系统, 文件, 目录, 对象, 索引节点; store, system, file, object, entity, index node, directory		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2013110891 A1 (OGASAWARA HIROSHI ET AL.) 02 May 2013 (2013-05-02) figure 4	1-23
A	CN 103229173 A (HUAWEI TECHNOLOGIES CO., LTD.) 31 July 2013 (2013-07-31) the whole document	1-23
A	CN 102693286 A (UNIV HUAZHONG SCI & TECHNOLOGY) 26 September 2012 (2012-09-26) the whole document	1-23
A	US 2012290629 A1 (BEAVERSON ARTHUR JET AL.) 15 November 2012 (2012-11-15) the whole document	1-23
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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 15 July 2015		Date of mailing of the international search report 22 July 2015
Name and mailing address of the ISA/CN STATE INTELLECTUAL PROPERTY OFFICE OF THE P.R.CHINA 6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088, China		Authorized officer LIU, Xiaohua
Facsimile No. (86-10)62019451		Telephone No. (86-10)62089142

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/IB2015/051978

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
US	2013110891	A1	02 May 2013	EP	2728484	A1	07 May 2014
				WO	2013061463	A1	02 May 2013
				CN	103748564	A	23 April 2014
				EP	2728484	A4	03 June 2015
				US	2014237204	A1	21 August 2014
				US	8725695	B2	13 May 2014
CN	103229173	A	31 July 2013	WO	2014101000	A1	03 July 2014
CN	102693286	A	26 September 2012	CN	102693286	B	26 March 2014
US	2012290629	A1	15 November 2012	US	2015081720	A1	19 March 2015
				CN	103635902	A	12 March 2014
				EP	2712452	B1	22 April 2015
				US	8898195	B2	25 November 2014
				US	8346810	B2	01 January 2013
				JP	2014513849	A	05 June 2014
				AU	2012256137	A1	28 November 2013
				CA	2836026	A1	22 November 2012
				EP	2712452	A1	02 April 2014
				ES	2535943	T3	19 May 2015
				US	2013110894	A1	02 May 2013
WO	2012158501	A1	22 November 2012				