



(12) **DEMANDE DE BREVET CANADIEN**
CANADIAN PATENT APPLICATION

(13) **A1**

(86) Date de dépôt PCT/PCT Filing Date: 2018/12/10
(87) Date publication PCT/PCT Publication Date: 2019/07/04
(85) Entrée phase nationale/National Entry: 2020/06/16
(86) N° demande PCT/PCT Application No.: US 2018/064659
(87) N° publication PCT/PCT Publication No.: 2019/133228
(30) Priorités/Priorities: 2017/12/28 (US62/611,473);
2017/12/29 (US15/858,357); 2018/01/12 (US15/870,221)

(51) Cl.Int./Int.Cl. *G06F 16/178* (2019.01)
(71) Demandeur/Applicant:
DROPBOX, INC., US
(72) Inventeurs/Inventors:
GOLDBERG, ISAAC, US;
LAI, JOHN, US;
JAYAKAR, SUJAY, US
(74) Agent: GOWLING WLG (CANADA) LLP

(54) Titre : ATTRIBUTION ET REATTRIBUTION D'IDENTIFIANTS UNIQUES POUR LA SYNCHRONISATION
D'ELEMENTS DE CONTENU
(54) Title: ALLOCATION AND REASSIGNMENT OF UNIQUE IDENTIFIERS FOR SYNCHRONIZATION OF CONTENT
ITEMS

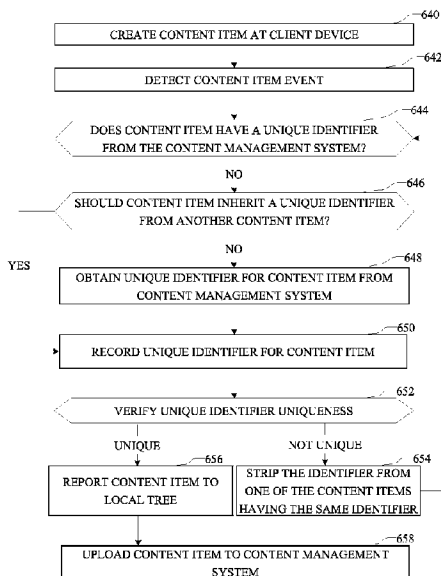


FIG. 6C

(57) **Abrégé/Abstract:**

A client can allocate and reassociate unique identifiers to local content items associated with an account at a content management system, and use the unique identifiers to commit operations for the content items on the content management system. For example, a client can create a content item and determine the content item does not have an identifier from the content management system. The client obtains an identifier for the content item and asks the content management system to verify a uniqueness of the identifier. When the identifier is unique, the client adds a node corresponding to the content item to a local tree representing a state at the client of content items associated with the account, and uploads the content item with the identifier to the content management system. When the identifier is not unique, the client obtains a new identifier for the content item.

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
04 July 2019 (04.07.2019)



(10) International Publication Number
WO 2019/133228 A1

(51) International Patent Classification:

G06F 16/178 (2019.01)

(21) International Application Number:

PCT/US2018/064659

(22) International Filing Date:

10 December 2018 (10.12.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/611,473 28 December 2017 (28.12.2017) US
15/858,357 29 December 2017 (29.12.2017) US
15/870,221 12 January 2018 (12.01.2018) US

(71) Applicant: **DROPBOX, INC.** [US/US]; 333 Brannan Street, San Francisco, CA 94107 (US).(72) Inventors: **GOLDBERG, Isaac**; Dropbox, Inc., 333 Brannan Street, San Francisco, CA 94107 (US). **LAI, John**; Dropbox, Inc., 333 Brannan Street, San Francisco, CA 94107 (US). **JAYAKAR, Sujay**; Dropbox, Inc., 333 Brannan Street, San Francisco, CA 94107 (US).(74) Agent: **MCKNIGHT, Brian**; Polsinelli PC, 1401 I Street, Suite 800, Washington, DC 20005 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, 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: ALLOCATION AND REASSIGNMENT OF UNIQUE IDENTIFIERS FOR SYNCHRONIZATION OF CONTENT ITEMS

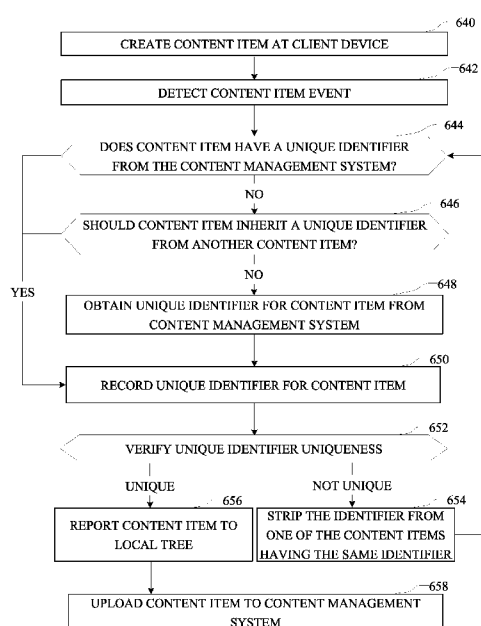


FIG. 6C

(57) Abstract: A client can allocate and reassociate unique identifiers to local content items associated with an account at a content management system, and use the unique identifiers to commit operations for the content items on the content management system. For example, a client can create a content item and determine the content item does not have an identifier from the content management system. The client obtains an identifier for the content item and asks the content management system to verify a uniqueness of the identifier. When the identifier is unique, the client adds a node corresponding to the content item to a local tree representing a state at the client of content items associated with the account, and uploads the content item with the identifier to the content management system. When the identifier is not unique, the client obtains a new identifier for the content item.

WO 2019/133228 A1

ALLOCATION AND REASSIGNMENT OF UNIQUE IDENTIFIERS FOR SYNCHRONIZATION OF CONTENT ITEMS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. non-provisional application number 15/870,221, filed on January 12, 2018, U.S. non-provisional application number 15/858,357, filed on December 29, 2017, and U.S. provisional application number 62/611,473, filed on December 28, 2017, all of which are expressly incorporated by reference herein in their entirety.

TECHNICAL FIELD

[0002] The present technology pertains to distributed storage, collaboration and synchronization systems.

BACKGROUND

[0003] Cloud storage systems allow users to store and access data on the cloud. Some cloud storage systems allow users to share data with other users and access the data in a collaborative fashion. In some cases, users may also store and access local copies of the data on their client devices. The local copies of the data may provide users with faster access to the data. Additionally, the local copies can allow the user to access the data when the user is offline. Cloud storage systems may also allow users to synchronize their local copies of the data with the data on the cloud to ensure consistency. Cloud storage systems may attempt to synchronize copies of data across a number of client devices and servers so each copy of data is identical. However, synchronization of data across multiple devices can be an extremely difficult task, often resulting in undesirable loss of data and inconsistencies.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] The above-recited and other advantages and features of the present technology will become apparent by reference to specific implementations illustrated in the appended drawings. A person of ordinary skill in the art will understand that these drawings only show some examples of the present technology and would not limit the scope of the present technology to these examples. Furthermore, the skilled artisan

will appreciate the principles of the present technology as described and explained with additional specificity and detail through the use of the accompanying drawings in which:

[0005] FIG. 1A shows an example of a content management system and client devices;

[0006] FIG. 1B shows an example of a client synchronization service in accordance with some aspects;

[0007] FIG. 2A shows a schematic diagram of an example architecture for synchronizing content between the content management system and client devices shown in FIG. 1A;

[0008] FIG. 2B shows an example configuration for storing and tracking blocks of content items in the example architecture for synchronizing content between the content management system and client devices shown in FIG. 2A;

[0009] FIG. 3A shows a diagram of example communications processed by a file journal interface between a client device and a server file journal on a content management system;

[0010] FIG. 3B shows a diagram of an example process for translating communications between a client device and a server file journal on a content management system;

[0011] FIG. 4A shows a diagram of an example translation and linearization process for translating server file journal data to linearized operations;

[0012] FIG. 4B shows a diagram of an example translation and linearization process for translating operations from a client device to revisions for a server file journal;

[0013] FIG. 4C shows an example linearization of cross-namespace operations;

[0014] FIG. 5A shows an example method of creating a namespace view in accordance with some aspects of the present technology;

[0015] FIG. 5B shows an example constructed namespace directory in accordance with some aspects of the present technology;

[0016] FIG. 6A shows an example of tree data structures in accordance with various aspects;

[0017] FIG. 6B shows an example of an update to the tree data structures shown in FIG. 6A;

[0018] FIG. 6C shows an example method for allocating and reassigning unique identifiers for content items created at a client device and synchronized between the client device and a content management system;

[0019] FIG. 6D shows an example method for downloading a content item to a client device from a content management system and recording the downloaded content item and unique identifier at the client device;

[0020] FIG. 7A shows an example method for synchronizing a server state and a file system state using tree data structures;

[0021] FIG. 7B shows an example method for resolving conflicts when synchronizing a server state and a file system state using tree data structures;

[0022] FIG. 8A shows a diagram of operations showing dependencies causalities between the operations;

[0023] FIG. 8B shows a diagram of events across namespaces ordered according to lamport clocks calculated for the events;

[0024] FIG. 9A shows an example mount state violation generated by a series of mount operations executed for a user;

[0025] FIG. 9B shows an example method for calculating lamport clocks for mount and unmount operations in a cross-namespace context;

[0026] FIG. 9C shows an example lamport clock configuration calculated for mount operations based on the example method shown in FIG. 9B;

[0027] FIG. 9D shows an example listing of operations serialized according to lamport clocks;

[0028] FIG. 10A shows an example process for updating lamport clocks based on a mount operation;

[0029] FIG. 10B shows an example process for updating lamport clocks based on an unmount operation;

[0030] FIG. 11A shows example tables in a server file journal for tracking move operations;

[0031] FIG. 11B shows a diagram of an example sequence for processing cross-namespace moves with lamport clocks;

[0032] FIG. 11C shows an example state machine defining an example flow of operations at various states of a move operation;

[0033] FIG. 11D shows a diagram of example move operations across locations based on a unique identifier of a content item and linearized based on causal relationships;

[0034] FIG. 12 shows an example update to the tree data structures shown in FIG. 6A, reflecting an intent to modify a content item based on a content item operation;

[0035] FIG. 13A shows an example method for committing a content item operation to a server file journal;

[0036] FIG. 13B shows an example method for processing a request to commit a content item operation from a client device;

[0037] FIG. 14 shows a diagram of an example symbolic link;

[0038] FIG. 15A shows a table of an example scenario for synchronizing symbolic links when the target and the symbolic link are contained in the same namespace;

[0039] FIG. 15B shows a table of an example scenario for synchronizing symbolic links when the target and symbolic link are contained on different namespaces;

[0040] FIG. 15C shows a table of an example scenario for synchronizing symbolic links when the target of the symbolic link is an internal target;

[0041] FIG. 15D shows a table of an example scenario for synchronizing symbolic links when the target of the symbolic link is an external target; and

[0042] FIG. 16 shows an example of a system for implementing various aspects of the present technology.

DETAILED DESCRIPTION

[0043] Various examples of the present technology are discussed in detail below. While specific implementations are discussed, it should be understood that this is done for illustration purposes only. A person skilled in the relevant art will recognize that other components and configurations may be used without parting from the spirit and scope of the present technology.

[0044] Cloud storage systems allow users to store and access content items across multiple devices. The content items may include, but are not limited to, files, documents, messages (e.g., email messages or text messages), media files (e.g., photos, videos, and audio files), folders, or any other unit of content. Content items may be shared with multiple users, edited, deleted, added, renamed, or moved. However, synchronizing content items shared or stored across several devices and user accounts has remained flawed and rife with technical obstacles.

[0045] To illustrate, a first machine (e.g., a client device or server) may send communications to a second machine that provides information about how a user's modification of content items on a cloud storage system. These communications may

be used by the second machine to synchronize the content items on the second machine such that actions performed on content items on the first machine are reflected in content items on the second machine, and the content items on the first machine are substantially identical to the content items on the second machine.

[0046] However, in many cases, there may be several communications sent between the various machines, which may be difficult to manage. Moreover, some of the communications may be received out of order as a result of various issues, such as client or network problems. This often results in conflicts and errors between content items at the various machines. The user's activity may also generate a large number of revisions which can further complicate synchronization efforts and exacerbate inconsistencies. For example, a user may perform a large number of modifications to various content items, undo modifications in a short period of time, or quickly perform additional modifications to a previously modified content item. This increases the likelihood that changes and revisions from users are received out of order, causing outdated modifications and conflicting content items. As a result, some operations may not be compatible with the current state of the content items. Moreover, it can be extremely difficult to detect whether operations are in conflict.

[0047] There is also an inherent latency with synchronization actions. For example, actions taken on the first machine are first detected by the first machine, and a communication is then generated and transmitted through a network. The communication is received by the second machine which may still be processing previous communications, and actions detailed in the communications may be taken at the second machine. In this illustrative scenario, there are several possible points of latency, including the first machine, the second machine, and the network. As latency increases, the likelihood of conflicts between content items also increases. Processing such conflicted communications and resolving conflicts are extremely difficult and computationally expensive tasks.

[0048] Further complexity is introduced when the same or different user on the second machine or other machines with access to the content items make modifications to the content items. Moreover, it is difficult to uniquely identify content items stored across multiple systems, such as a cloud storage system and client devices. When a content item is generated at a client device, the client device cannot guarantee that an identifier assigned by the client device to the content item is unique at other systems, even if the identifier is randomly generated and otherwise

unique at the client device. In addition, as content items are created, modified, moved, and deleted across devices, the content items can be extremely difficult to track across the various locations, often resulting in duplicate identifiers and metadata at one or more locations. Duplicate identifiers and metadata can create inconsistencies between content items and limit the ability to process operations and synchronize changes across systems.

[0049] Additional technical issues arise when content items are modified locally and remotely in a large collaboration environment, and race conditions are created by various operations generated across systems. As illustrated here, these issues can quickly multiply and grow in complexity, creating a wide array of problems and inconsistencies in the content items.

[0050] Content Management System

[0051] In some embodiments the disclosed technology is deployed in the context of a content management system having content item synchronization capabilities and collaboration features, among others. An example system configuration 100 is shown in FIG. 1A, which depicts content management system 110 interacting with client device 150.

[0052] Accounts

[0053] Content management system 110 can store content items in association with accounts, as well as perform a variety of content item management tasks, such as retrieve, modify, browse, and/or share the content item(s). Furthermore, content management system 110 can enable an account to access content item(s) from multiple client devices.

[0054] Content management system 110 supports a plurality of accounts. An entity (user, group of users, team, company, etc.) can create an account with content management system, and account details can be stored in account database 140. Account database 140 can store profile information for registered entities. In some cases, profile information for registered entities includes a username and/or email address. Account database 140 can include account management information, such as account type (e.g. various tiers of free or paid accounts), storage space allocated, storage space used, client devices 150 having a registered content management client application 152 resident thereon, security settings, personal configuration settings, etc.

[0055] Account database 140 can store groups of accounts associated with an entity. Groups can have permissions based on group policies and/or access control lists, and

members of the groups can inherit the permissions. For example, a marketing group can have access to one set of content items while an engineering group can have access to another set of content items. An administrator group can modify groups, modify user accounts, etc.

[0056] Content Item Storage

[0057] A feature of content management system 110 is the storage of content items, which can be stored in content storage 142. Content items can be any digital data such as documents, collaboration content items, text files, audio files, image files, video files, webpages, executable files, binary files, etc. A content item can also include collections or other mechanisms for grouping content items together with different behaviors, such as folders, zip files, playlists, albums, etc. A collection can refer to a folder, or a plurality of content items that are related or grouped by a common attribute. In some embodiments, content storage 142 is combined with other types of storage or databases to handle specific functions. Content storage 142 can store content items, while metadata regarding the content items can be stored in metadata database 146. Likewise, data regarding where a content item is stored in content storage 142 can be stored in content directory 144. Additionally, data regarding changes, access, etc. can be stored in server file journal 148. Each of the various storages/databases such as content storage 142, content directory 144, server file journal 148, and metadata database 146 can be comprised of more than one such storage or database and can be distributed over many devices and locations. Other configurations are also possible. For example, data from content storage 142, content directory 144, server file journal 148, and/or metadata database 146 may be combined into one or more content storages or databases or further segmented into additional content storages or databases. Thus, content management system 110 may include more or less storages and/or databases than shown in FIG. 1A.

[0058] In some embodiments, content storage 142 is associated with at least one content storage service 116, which includes software or other processor executable instructions for managing the storage of content items including, but not limited to, receiving content items for storage, preparing content items for storage, selecting a storage location for the content item, retrieving content items from storage, etc. In some embodiments, content storage service 116 can divide a content item into smaller chunks for storage at content storage 142. The location of each chunk making up a content item can be recorded in content directory 144. Content directory 144 can

include a content entry for each content item stored in content storage 142. The content entry can be associated with a unique ID, which identifies a content item.

[0059] In some embodiments, the unique ID, which identifies a content item in content directory 144, can be derived from a deterministic hash function. This method of deriving a unique ID for a content item can ensure that content item duplicates are recognized as such since the deterministic hash function will output the same identifier for every copy of the same content item, but will output a different identifier for a different content item. Using this methodology, content storage service 116 can output a unique ID for each content item.

[0060] Content storage service 116 can also designate or record a content path for a content item in metadata database 146. The content path can include the name of the content item and/or folder hierarchy associated with the content item. For example, the content path can include a folder or path of folders in which the content item is stored in a local file system on a client device. While content items are stored in content storage 142 in blocks and may not be stored under a tree like directory structure, such directory structure is a comfortable navigation structure for users. Content storage service 116 can define or record a content path for a content item wherein the “root” node of a directory structure can be a namespace for each account. Within the namespace can be a directory structure defined by a user of an account and/or content storage service 116. Metadata database 146 can store the content path for each content item as part of a content entry.

[0061] In some embodiments the namespace can include additional namespaces nested in the directory structure as if they are stored within the root node. This can occur when an account has access to a shared collection. Shared collections can be assigned their own namespace within content management system 110. While some shared collections are actually a root node for the shared collection, they are located subordinate to the account namespace in the directory structure, and can appear as a folder within a folder for the account. As addressed above, the directory structure is merely a comfortable navigation structure for users, but does not correlate to storage locations of content items in content storage 142.

[0062] While the directory structure in which an account views content items does not correlate to storage locations at content management system 110, the directory structure can correlate to storage locations on client device 150 depending on the file system used by client device 150.

[0063] As addressed above, a content entry in content directory 144 can also include the location of each chunk making up a content item. More specifically, the content entry can include content pointers that identify the location in content storage 142 of the chunks that make up the content item.

[0064] In addition to a content path and content pointer, a content entry in content directory 144 can also include a user account identifier that identifies the user account that has access to the content item and/or a group identifier that identifies a group with access to the content item and/or a namespace to which the content entry belongs.

[0065] Content storage service 116 can decrease the amount of storage space required by identifying duplicate content items or duplicate blocks that make up a content item or versions of a content item. Instead of storing multiple copies, content storage 142 can store a single copy of the content item or block of the content item and content directory 144 can include a pointer or other mechanism to link the duplicates to the single copy.

[0066] Content storage service 116 can also store metadata describing content items, content item types, folders, file path, and/or the relationship of content items to various accounts, collections, or groups in metadata database 146, in association with the unique ID of the content item.

[0067] Content storage service 116 can also store a log of data regarding changes, access, etc. in server file journal 148. Server file journal 148 can include the unique ID of the content item and a description of the change or access action along with a time stamp or version number and any other relevant data. Server file journal 148 can also include pointers to blocks affected by the change or content item access. Content storage service can provide the ability to undo operations, by using a content item version control that tracks changes to content items, different versions of content items (including diverging version trees), and a change history that can be acquired from the server file journal 148.

[0068] Content Item Synchronization

[0069] Another feature of content management system 110 is synchronization of content items with at least one client device 150. Client device(s) can take different forms and have different capabilities. For example, client device 150₁ is a computing device having a local file system accessible by multiple applications resident thereon. Client device 1452 is a computing device wherein content items are only accessible to

a specific application or by permission given by the specific application, and the content items are typically stored either in an application specific space or in the cloud. Client device 150₃ is any client device accessing content management system 110 via a web browser and accessing content items via a web interface. While example client devices 150₁, 1452, and 150₃ are depicted in form factors such as a laptop, mobile device, or web browser, it should be understood that the descriptions thereof are not limited to devices of these example form factors. For example a mobile device such as client 1452 might have a local file system accessible by multiple applications resident thereon, or client 1452 might access content management system 110 via a web browser. As such, the form factor should not be considered limiting when considering client 150's capabilities. One or more functions described herein with respect to client device 150 may or may not be available on every client device depending on the specific capabilities of the device – the file access model being one such capability.

[0070] In many embodiments, client devices are associated with an account of content management system 110, but in some embodiments client devices can access content using shared links and do not require an account.

[0071] As noted above, some client devices can access content management system 110 using a web browser. However, client devices can also access content management system 110 using client application 152 stored and running on client device 150. Client application 152 can include a client synchronization service 156.

[0072] Client synchronization service 156 can be in communication with server synchronization service 112 to synchronize changes to content items between client device 150 and content management system 110.

[0073] Client device 150 can synchronize content with content management system 110 via client synchronization service 156. The synchronization can be platform agnostic. That is, content can be synchronized across multiple client devices of varying type, capabilities, operating systems, etc. Client synchronization service 156 can synchronize any changes (new, deleted, modified, copied, or moved content items) to content items in a designated location of a file system of client device 150.

[0074] Content items can be synchronized from client device 150 to content management system 110, and vice versa. In embodiments wherein synchronization is from client device 150 to content management system 110, a user can manipulate content items directly from the file system of client device 150, while client

synchronization service 156 can monitor directory on client device 150 for changes to files within the monitored folders.

[0075] When client synchronization service 156 detects a write, move, copy, or delete of content in a directory that it monitors, client synchronization service 156 can synchronize the changes to content management system service 116. In some embodiments, client synchronization service 156 can perform some functions of content management system service 116 including functions addressed above such as dividing the content item into blocks, hashing the content item to generate a unique identifier, etc. Client synchronization service 156 can index content within client storage index 164 and save the result in storage index 164. Indexing can include storing paths plus a unique server identifier, and a unique client identifier for each content item. In some embodiments, client synchronization service 156 learns the unique server identifier from server synchronization service 112, and learns the unique client identifier from the operating system of client device 150. ,

[0076] Client synchronization service 156 can use storage index 164 to facilitate the synchronization of at least a portion of the content within client storage with content associated with a user account on content management system 110. For example, client synchronization service 156 can compare storage index 164 with content management system 110 and detect differences between content on client storage and content associated with a user account on content management system 110. Client synchronization service 156 can then attempt to reconcile differences by uploading, downloading, modifying, and deleting content on client storage as appropriate. Content storage service 116 can store the changed or new block for the content item and update server file journal 148, metadata database 146, content directory 144, content storage 142, account database 140, etc. as appropriate.

[0077] When synchronizing from content management system 110 to client device 150, a mount, modification, addition, deletion, move of a content item recorded in server file journal 148 can trigger a notification to be sent to client device 150 using notification service 117. When client device 150 is informed of the change a request changes listed in server file journal 148 since the last synchronization point known to the client device. When client device 150 determines that it is out of synchronization with content management system 110, client synchronization service 156 requests content item blocks including the changes, and updates its local copy of the changed content items.

[0078] In some embodiments, storage index 164 stores tree data structures wherein one tree reflects the latest representation of a directory according to server synchronization service 112, while another tree reflects the latest representation of the directory according to client synchronization service 156. Client synchronization service can work to ensure that the tree structures match by requesting data from server synchronization service 112 or committing changes on client device 150 to content management system 110.

[0079] Sometimes client device 150 might not have a network connection available. In this scenario, client synchronization service 156 can monitor the linked collection for content item changes and queue those changes for later synchronization to content management system 110 when a network connection is available. Similarly, a user can manually start, stop, pause, or resume synchronization with content management system 110.

[0080] Client synchronization service 156 can synchronize all content associated with a particular user account on content management system 110. Alternatively, client synchronization service 156 can selectively synchronize a portion of the content of the total content associated with the particular user account on content management system 110. Selectively synchronizing only a portion of the content can preserve space on client device 150 and save bandwidth.

[0081] In some embodiments, client synchronization service 156 selectively stores a portion of the content associated with the particular user account and stores placeholder content items in client storage for the remainder portion of the content. For example, client synchronization service 156 can store a placeholder content item that has the same filename, path, extension, metadata, of its respective complete content item on content management system 110, but lacking the data of the complete content item. The placeholder content item can be a few bytes or less in size while the respective complete content item might be significantly larger. After client device 150 attempts to access the content item, client synchronization service 156 can retrieve the data of the content item from content management system 110 and provide the complete content item to accessing client device 150. This approach can provide significant space and bandwidth savings while still providing full access to a user's content on content management system 110.

[0082] Collaboration features

[0083] Another feature of content management system 110 is to facilitate collaboration between users. Collaboration features include content item sharing, commenting on content items, co-working on content items, instant messaging, providing presence and seen state information regarding content items, etc.

[0084] Sharing

[0085] Content management system 110 can manage sharing content via sharing service 128. Sharing content by providing a link to the content can include making the content item accessible from any computing device in network communication with content management system 110. However, in some embodiments a link can be associated with access restrictions enforced by content management system 110 and access control list 145. Sharing content can also include linking content using sharing service 128 to share content within content management system 110 with at least one additional user account (in addition to the original user account associated with the content item) so that each user account has access to the content item. The additional user account can gain access to the content by accepting the content, which will then be accessible through either web interface service 124 or directly from within the directory structure associated with their account on client device 150. The sharing can be performed in a platform agnostic manner. That is, the content can be shared across multiple client devices 150 of varying type, capabilities, operating systems, etc. The content can also be shared across varying types of user accounts.

[0086] To share a content item within content management system 110 sharing service 128 can add a user account identifier or multiple user account identifiers to a content entry in access control list database 145 associated with the content item, thus granting the added user account access to the content item. Sharing service 128 can also remove user account identifiers from a content entry to restrict a user account's access to the content item. Sharing service 128 can record content item identifiers, user account identifiers given access to a content item, and access levels in access control list database 145. For example, in some embodiments, user account identifiers associated with a single content entry can specify different permissions for respective user account identifiers with respect to the associated content item.

[0087] To share content items outside of content management system 110, sharing service 128 can generate a custom network address, such as a uniform resource locator (URL), which allows any web browser to access the content item or collection in content management system 110 without any authentication. To accomplish this,

sharing service 128 can include content identification data in the generated URL, which can later be used to properly identify and return the requested content item. For example, sharing service 128 can include the account identifier and the content path or a content item identifying code in the generated URL. Upon selection of the URL, the content identification data included in the URL can be transmitted to content management system 110, which can use the received content identification data to identify the appropriate content item and return the content item.

[0088] In addition to generating the URL, sharing service 128 can also be configured to record in access control list database 145 that a URL to the content item has been created. In some embodiments, the content entry associated with a content item can include a URL flag indicating whether a URL to the content item has been created. For example, the URL flag can be a Boolean value initially set to 0 or false to indicate that a URL to the content item has not been created. Sharing service 128 can change the value of the flag to 1 or true after generating a URL to the content item.

[0089] In some embodiments, sharing service 128 can associate a set of permissions to a URL for a content item. For example, if a user attempts to access the content item via the URL, sharing service 128 can provide a limited set of permissions for the content item. Examples of limited permissions include restrictions that the user cannot download the content item, save the content item, copy the content item, modify the content item, etc. In some embodiments, limited permissions include restrictions that only permit a content item to be accessed from with a specified domain, i.e., from within a corporate network domain, or by accounts associated with a specified domain, e.g., accounts associated with a company account (e.g., @acme.com).

[0090] In some embodiments, sharing service 128 can also be configured to deactivate a generated URL. For example, each content entry can also include a URL active flag indicating whether the content should be returned in response to a request from the generated URL. For example, sharing service 128 can only return a content item requested by a generated link if the URL active flag is set to 1 or true. Thus, access to a content item for which a URL has been generated can be easily restricted by changing the value of the URL active flag. This allows a user to restrict access to the shared content item without having to move the content item or delete the generated URL. Likewise, sharing service 128 can reactivate the URL by again changing the value of the URL active flag to 1 or true. A user can thus easily restore access to the content item without the need to generate a new URL.

[0091] In some embodiments, content management system 110 can designate a URL for uploading a content item. For example, a first user with a user account can request such a URL, provide the URL to a contributing user and the contributing user can upload a content item to the first user's user account using the URL.

[0092] Team Service

[0093] In some embodiments content management system 110 includes team service 130. Team service 130 can provide functionality for creating and managing defined teams of user accounts. Teams can be created for a company, with sub-teams (e.g., business units, or project teams, etc.), and user accounts assigned to teams and sub-teams, or teams can be created for any defined group of user accounts. Teams service 130 can provide a common shared space for the team, private user account folders, and access limited shared folders. Teams service can also provide a management interface for an administrator to manage collections and content items within team, and can manage user accounts that are associated with the team.

[0094] Authorization Service

[0095] In some embodiments, content management system 110 includes authorization service 132. Authorization service 132 ensures that a user account attempting to access a namespace has appropriate rights to access the namespace. Authorization service 132 can receive a token from client application 152 that follows a request to access a namespace and can return the capabilities permitted to the user account. For user accounts with multiple levels of access (e.g. a user account with user rights and administrator rights) authorization service 132 can also require explicit privilege escalation to avoid unintentional actions by administrators.

[0096] Presence and Seen State

[0097] In some embodiments, content management system can provide information about how users with which a content item is shared are interacting or have interacted with the content item. In some embodiments, content management system 110 can report that a user with which a content item is shared is currently viewing the content item. For example, client collaboration service 160 can notify notifications service 117 when client device 150 is accessing the content item. Notifications service 117 can then notify all client devices of other users having access to the same content item of the presence of the user of client device 150 with respect to the content item.

[0098] In some embodiments, content management system 110 can report a history of user interaction with a shared content item. Collaboration service 126 can query data

sources such as metadata database 146 and server file journal 148 to determine that a user has saved the content item, that a user has yet to view the content item, etc., and disseminate this status information using notification service 117 to other users so that they can know who currently is or has viewed or modified the content item.

[0099] Collaboration service 126 can facilitate comments associated with content, even if a content item does not natively support commenting functionality. Such comments can be stored in metadata database 146.

[0100] Collaboration service 126 can originate and transmit notifications for users. For example, a user can mention another user in a comment and collaboration service 126 can send a notification to that user that he has been mentioned in the comment. Various other content item events can trigger notifications, including deleting a content item, sharing a content item, etc.

[0101] Collaboration service 126 can provide a messaging platform whereby users can send and receive instant messages, voice calls, emails, etc.

[0102] Collaboration Content Items

[0103] In some embodiments content management service can also include Collaborative document service 134 which can provide an interactive content item collaboration platform whereby users can simultaneously create collaboration content items, comment in the collaboration content items, and manage tasks within the collaboration content items. Collaboration content items can be files that users can create and edit using a collaboration content item editor, and can contain collaboration content item elements. Collaboration content item elements may include a collaboration content item identifier, one or more author identifiers, collaboration content item text, collaboration content item attributes, interaction information, comments, sharing users, etc. Collaboration content item elements can be stored as database entities, which allows for searching and retrieving the collaboration content items. Multiple users may access, view, edit, and collaborate on collaboration content items at the same time or at different times. In some embodiments this can be managed by requiring two users access a content item through a web interface and there they can work on the same copy of the content item at the same time.

[0104] Collaboration Companion Interface

[0105] In some embodiments client collaboration service 160 can provide a native application companion interface for the purpose of displaying information relevant to a content item being presented on client device 150. In embodiments wherein a

content item is accessed by a native application stored and executed on client device 150, where the content item is in a designated location of the file system of client device 150 such that the content item is managed by content application 152, the native application may not provide any native way to display the above addressed collaboration data. In such embodiments, client collaboration service 160 can detect that a user has opened a content item, and can provide an overlay with additional information for the content item, such as collaboration data. For example, the additional information can include comments for the content item, status of the content item, activity of other users previously or currently viewing the content item. Such an overlay can warn a user that changes might be lost because another user is currently editing the content item.

[0106] In some embodiments, one or more of the services or storages/databases discussed above can be accessed using public or private application programming interfaces.

[0107] Certain software applications can access content storage 142 via an API on behalf of a user. For example, a software package such as an application running on client device 150, can programmatically make API calls directly to content management system 110 when a user provides authentication credentials, to read, write, create, delete, share, or otherwise manipulate content.

[0108] A user can view or manipulate content stored in a user account via a web interface generated and served by web interface service 124. For example, the user can navigate in a web browser to a web address provided by content management system 110. Changes or updates to content in the content storage 142 made through the web interface, such as uploading a new version of a content item, can be propagated back to other client devices associated with the user's account. For example, multiple client devices, each with their own client software, can be associated with a single account and content items in the account can be synchronized between each of the multiple client devices.

[0109] Client device 150 can connect to content management system 110 on behalf of a user. A user can directly interact with client device 150, for example when client device 150 is a desktop or laptop computer, phone, television, internet-of-things device, etc. Alternatively or additionally, client device 150 can act on behalf of the user without the user having physical access to client device 150, for example when client device 150 is a server.

[0110] Some features of client device 150 are enabled by an application installed on client device 150. In some embodiments, the application can include a content management system specific component. For example, the content management system specific component can be a stand-alone application 152, one or more application plug-ins, and/or a browser extension. However, the user can also interact with content management system 110 via a third-party application, such as a web browser, that resides on client device 150 and is configured to communicate with content management system 110. In various implementations, the client-side application 152 can present a user interface (UI) for a user to interact with content management system 110. For example, the user can interact with the content management system 110 via a file system explorer integrated with the file system or via a webpage displayed using a web browser application.

[0111] In some embodiments, client application 152 can be configured to manage and synchronize content for more than one account of content management system 110. In such embodiments client application 152 can remain logged into multiple accounts and provide normal services for the multiple accounts. In some embodiments, each account can appear as folder in a file system, and all content items within that folder can be synchronized with content management system 110. In some embodiments, client application 152 can include a selector to choose one of the multiple accounts to be the primary account or default account.

[0112] While content management system 110 is presented with specific components, it should be understood by one skilled in the art, that the architectural configuration of system 100 is simply one possible configuration and that other configurations with more or fewer components are possible. Further, a service can have more or less functionality, even including functionality described as being with another service. Moreover, features described herein with respect to an embodiment can be combined with features described with respect to another embodiment.

[0113] While system 100 is presented with specific components, it should be understood by one skilled in the art, that the architectural configuration of system 100 is simply one possible configuration and that other configurations with more or fewer components are possible.

[0114] Client Synchronization Service

[0115] FIG. 1B shows an example of a client synchronization service 156, in accordance with some embodiments. According to some embodiments, client

synchronization service 156 may be implemented in client device 150 shown in FIG. 1A. However, in other embodiments, client synchronization service 156 may be implemented on another computing device. Client synchronization service 156 is configured to synchronize changes to content items between a content management system and the client device on which client synchronization service 156 runs.

[0116] Client synchronization service 156 may include file system interface 170, server interface 172, tree storage 174, planner 176, and scheduler 178. Additional or alternative components may also be included. High level descriptions of client synchronization service 156 and its components are discussed below with respect to FIG. 1B. However, further details and embodiments of client synchronization service 156 and its components are discussed throughout.

[0117] File system interface 170 is configured to process changes to content items on the local filesystem of the client device and update the local tree. For example, file system interface 170 can be in communication with client synchronization service 156 to detect changes to content items on the local filesystem of the client device. Changes may also be made and detected via client application 152 of FIG. 1A. File system interface 170 may make updates to the local tree may be made based on the changes (new, deleted, modified, copied, renamed, or moved content items) to content items on the client device.

[0118] Server interface 172 is configured to aid in the processing of remote changes to content items at a remote storage of the content management system and updating of the remote tree. For example, server interface 172 can be in communication with server synchronization service 112 of FIG. 1A to synchronize changes to content items between client device 150 and content management system 110. Changes (new, deleted, modified, copied, renamed, or moved content items) to content items at content management system 110 may be detected and updates may be made to the remote tree to reflect the changes at content management system 110.

[0119] Tree storage 174 is configured to store and maintain the tree data structures used by client synchronization service 156. For example, tree storage 174 may store the local tree, the sync tree, and the remote tree. According to some embodiments, tree storage 200 may store the tree data structures in persistent memory (e.g., a hard disk or other secondary storage device) as well as in main memory (e.g., RAM or other primary storage device) in order to reduce latency and response time. For example, on start-up of the client device or client synchronization service 156, the tree

data structures may be retrieved from persistent memory and loaded into main memory. Tree storage 174 may access and update the tree data structures on main memory and, before the client device or client synchronization service 156 is shut down, tree storage 174 may store the updated tree data structures on persistent memory. Because main memory is expensive in cost and often limited in size on most client devices, additional technological improvements are implemented to decrease the footprint of the tree data structures on main memory. These technological solutions are described further below.

[0120] Planner 176 is configured to detect differences between the server state associated with the content management system and the file system state associated with the client device based on the state of the tree data structures. For example, planner 176 may determine if there is a difference between the remote tree and the sync tree. A difference between the remote tree and the sync tree indicates that an action performed remotely on one or more content items stored at the content management system has caused the server state and the file system state to become out of sync. Similarly, planner 176 may also determine if there is a difference between the local tree and the sync tree. A difference between the local tree and the sync tree indicates that an action performed locally on one or more content items stored on the client device has caused the server state and the file system state to become out of sync. If a difference is detected, planner 176 generates a sequence of operations that synchronize the tree data structures.

[0121] In some scenarios, a sequence of operations generated based on a difference between the remote tree and the sync tree and a sequence of operations generated based on a difference between the local tree and the sync tree may conflict. Planner 176 is may also be configured to merge the two sequences of operations into a single merged plan of operations.

[0122] Scheduler 178 is configured to take the generated sequence of operations and manage the execution of those operations. According to some embodiments, scheduler 178 converts each operation in the sequence of operations into a series of one or more tasks that need to be executed in order to perform the operation. In some scenarios, some tasks may become out dated or no longer relevant. Scheduler 178 is configured to identify those tasks and cancel them.

File Journal and Storage Systems (Tomahawk and Waterfall)

[0123] FIG. 2A illustrates a schematic diagram of an example architecture for synchronizing content between content management system 110 and client device 150 in system configuration 100. In this example, client device 150 interacts with content storage 142 and server file journal 148 respectively via content storage interface 206 and file journal interface 202. Content storage interface 206 can be provided or managed by content storage service 116, and file journal interface 202 can be provided or managed by server synchronization service 112. For example, content storage interface 206 can be a subcomponent or subservice of content storage service 116, and file journal interface 202 can be a subcomponent or subservice of server synchronization service 112.

[0124] Content storage interface 206 can manage communications, such as content requests or interactions, between client device 150 and content storage 142. Content storage interface 206 can process requests from client device 150 to upload and download content to and from content storage 142. Content storage interface 206 can receive content requests (e.g., downloads, uploads, etc.) from client device 150, verify permissions in access control list 145, communicate with authorization service 132 to determine if client device 150 (and/or the request from client device 150) is authorized to upload or download the content to or from content storage 142, and interact with content storage 142 to download or upload the content in content storage 142 to client device 150. If the request from client device 150 is a request to download a content item, content storage interface 206 can retrieve the content item from content storage 142 and provide the content item to client device 150. If the request from client device 150 is a request to upload a content item, content storage interface 206 can obtain the content item from client device 150 and upload the content item to content storage 142 for storage.

[0125] When processing content requests from client device 150, content storage interface 206 can communicate with storage index 210 to check the availability and/or storage location of the requested content in content storage 142, and track content items in content storage 142. Storage index 210 can maintain an index of content items on content storage 142 which identifies the content items on content storage 142 and can also identify a respective location of the content items within content storage 142. Thus, storage index 210 can track content items on content storage 142 as well

as storage locations of the content items. Storage index 210 can track entire content items, such as files, and/or portions of the content items, such as blocks or chunks. In some cases, content items can be split into blocks or chunks which can be stored at content storage 142 and tracked in storage index 210. For example, content storage 142 can store a content item as blocks or chunks of data which include respective data portions of the content item. Storage index 210 can track the blocks or chunks of the content item stored in content storage 142. FIG. 2B described below illustrates an example configuration for storing and tracking blocks of content items.

[0126] File journal interface 202 can manage communications, such as metadata requests and content synchronizations and operations, between client device 150 and server file journal 148. For example, file journal interface 202 can translate, validate, authenticate, and/or process operations, configurations, and state information between client device 150 and server file journal 148. File journal interface 202 can verify permissions from an FSAuth token in a cursor or through authorization service 132 to authorize, or verify authorization of, requests sent by client device 150 to server file journal 148. When processing requests or operations from client device 150, file journal interface 202 can access namespace membership store 208 to determine or verify namespace ownership information for any namespaces associated with the requests or operations from client device 150, and retrieve permissions information from access control list 145 to verify permissions of content associated with the requests or operations from client device 150.

[0127] Translation service 204 in file journal interface 202 can perform linearization and translation operations for communications between client device 150 and server file journal 148. For example, translation service 204 can translate communications from client device 150 to a different format consistent with the structure and format of data in server file journal 148, and vice versa. To illustrate, in some cases, client device 150 can process content item information (e.g., state, changes, versions, etc.) at client device 150 as operations, while server file journal 148 can process the same information as content item revisions reflected by rows in a data structure such as a database table. To enable synchronization of content item information between client device 150 and server file journal 148, translation service 204 can translate operations from client device 150 into revisions suitable for server file journal 148, and can translate revisions reflected in rows of data on server file journal 148 to operations suitable for client device 150.

[0128] In some cases, authorization service 132 can generate a token that verifies or indicates that client device 150 is authorized to access, update, download, or upload a requested content item. The token can include a device identifier associated with client device 150, an account identifier associated with a user account authenticated or authorized at client device 150, a session identifier associated with an authorized session at client device 150, a view context, and access permissions to identified collections. The token can be included in a cryptographically signed data object called a cursor, which will be described in greater detail below. Content management system 110 and/or authorization service 132 can send the token(s) to client device 150, and client device 150 can provide the token to content management system 110 when requesting content item revisions and/or updates to server file journal 148 as further described below. Client device 150 can also provide the token to content storage interface 206 to validate any content requests (e.g., downloads, uploads, etc.). Content storage interface 206 can use the token to authorize queries to storage index 210 and upload or download content items to or from content storage 142.

[0129] For example, client device 150 can send to content storage interface 206 a request to upload a content item to content storage 142. The request can include the token and the content item to be uploaded. Content storage interface 206 can use the token to authorize a query to storage index 210 to check if the content item already exists on content storage 142, and authorize the upload of the content item to content storage 142. Client device 150 can also provide the token to file journal interface 202 to authorize a request to store metadata on server file journal 148 to track the upload and revision of the content item.

[0130] FIG. 2B illustrates an example block storage and synchronization configuration. In this example, content storage 142 can store blocks of data, which can be opaque chunks of content items (e.g., files) up to a particular size (e.g., 4MB). Content items can be split into blocks and the blocks can be stored at content storage 142 for access. Storage index 210 can track blocks stored at content storage 142, as well as the respective locations of the blocks stored at content storage 142. File journal interface 202 can interact with server file journal 148 to track revisions to the content items and/or blocks stored at content storage 142.

[0131] For example, content item 220 (e.g., MyFile.abc) can be split into blocks 220A, 220B, 220C, 220N. Content storage interface 206 can receive blocks 220A, 220B, 220C, 220N and send block data 222B to content storage 142 for storage at

content storage 142. Block data 222B can include blocks 220A, 220B, 220C, 220N associated with content item 220.

[0132] Blocks 220A, 220B, 220C, 220N can be stored on one or more storage devices or volumes at content storage 142 and/or aggregated within one or more logical storage containers (e.g., buckets) or data clusters. In some cases, blocks 220A, 220B, 220C, 220N can be stored together on a same location (e.g., storage device, volume, container, and/or cluster). In other cases, some or all of blocks 220A, 220B, 220C, 220N can be stored on two or more different locations (e.g., two or more different storage devices, volumes, containers, and/or clusters).

[0133] Content storage interface 206 can also store block metadata 222A at storage index 210. Block metadata 222A can identify blocks 220A, 220B, 220C, 220N, and allows storage index 210 to track blocks 220A, 220B, 220C, 220N at content storage 142. Block metadata 222A can include an identifier for each block 220A, 220B, 220C, 220N. The identifier for a block can be a name or key, such as a hash of the block, which identifies the block.

[0134] Block metadata 222A can also include location information for blocks 220A, 220B, 220C, 220N, which indicates the respective storage location of blocks 220A, 220B, 220C, 220N. The location information of a block can identify the storage device or volume where the block is stored and/or a logical storage container or data cluster where the block is contained. The location information can be used to access or retrieve the associated block.

[0135] Content storage interface 206 can store block metadata 222A at storage index 210 before or after storing blocks 220A, 220B, 220C, 220N at content storage 142. For example, content storage interface 206 can store blocks 220A, 220B, 220C, 220N at content storage 142 and subsequently store block metadata 222A at storage index 210 to indicate that blocks 220A, 220B, 220C, 220N have been stored at content storage 142.

[0136] In some cases, content storage interface 206 can query storage index 210 prior to storing blocks 220A, 220B, 220C, 220N at content storage 142, to determine if (or where) blocks 220A, 220B, 220C, 220N are stored at content storage 142. For example, content storage interface 206 can query storage index 210 based on block metadata 222A to check if blocks 220A, 220B, 220C, 220N are stored at content storage 142. Storage index 210 can compare block identifiers in block metadata 222A with block identifiers at storage index 210 to check for any matches. A match

between block identifiers indicates that an associated block is stored at content storage 142.

[0137] As previously mentioned, server file journal 148 tracks content item revisions, including content item adds, edits, moves or renames, deletes, etc. Accordingly, file journal interface 202 can store revision 222C at server file journal 148 to indicate that content item 220 and/or blocks 220A, 220B, 220C, 220N were added to content storage 142. Revision 222C can represent a revision of content item 220 within a journal of content item revisions at server file journal 148.

[0138] Revision 222C can identify content item 220 and an operation associated with content item 220, such as an add operation (e.g., upload), edit operation, move or rename operation, delete operation, etc. Revision 222C can also identify a namespace in content management system 110 where content item 220 is stored, and a row in a journal of content item revisions at server file journal 148 for storing revision 222C. The row within the journal of content item revisions can represent a revision number associated with revision 222C for content item 220.

[0139] File Journal Interface (Tomahawk)

[0140] FIG. 3A illustrates a diagram of communications processed by file journal interface 202 between client device 150 and server file journal 148. Server file journal 148 tracks content item state and changes (e.g., revisions) as values in rows and fields in server file journal 148. For example, server file journal 148 can maintain one or more journals of revisions to content items in content storage 142. The one or more journals can track revisions of each content item on each namespace. A row of values in a journal on server file journal 148 can identify a content item in a namespace and reflects a state of the content item in the namespace. A subsequent row in the journal corresponding to the same content item in the namespace can reflect a subsequent revision to the content item in the namespace. Thus, rows in server file journal 148 associated with a content item can identify the current state of the content item and any revisions to the content item from creation to the current state.

[0141] To synchronize content item information (e.g., state, changes or revisions, etc.) with client device 150, server file journal 148 can send or receive revisions data 304 to or from file journal interface 202, which represent revisions tracked or stored in server file journal 148 for one or more content items. Revisions data 304 can include, for example, a log of content item revisions corresponding to rows in server

file journal 148. Server file journal 148 can send revisions data 304 to file journal interface 204, which can translate revisions data 304 into operations data 302 for client device 150, as further described below.

[0142] Client device 150 can perform content operations to update or modify content items at client device 150. To synchronize content item information with server file journal 148, client device 150 can send or receive operations data 302 to or from file journal interface 202. Client device 150 can send operations data 302 to file journal interface 202 to report changes at client device 150 to content items, and receive operations data 302 from file journal interface 202 to obtain the latest state of content items from server file journal 148 (e.g., revisions data 304).

[0143] For example, client device 150 can edit content item A at client device 150 and report to file journal interface 202 an edit operation indicating the edit to content item A. The edit operation can be included in operations data 302 communicated with file journal interface 202 to indicate the revision to content item A. File journal interface 202 can receive operations data 302 including the edit operation and generate a revision for storage at server file journal 148, tracking the edit to content item A. File journal interface 202 can include the revision associated with the edit operation in revisions data 304 to server file journal 148, in order to update server file journal 148 to store the revision representing the edited state of content item A.

[0144] As further described below, operations data 302 can include a cursor which identifies the latest state or revision obtained by client device 150 for each namespace associated with client device 150. For example, the cursor can identify the latest revision in server file journal 148 obtained by client device 150 for each namespace associated with client device 150. The information in the cursor allows file journal interface 202 to determine whether an operation in operations data 302 from client device 150 reflects the latest state or revisions in server file journal 148 for the namespace(s) associated with the operation. This can help file journal interface 202 ensure that operations in operations data 302 from client device 150 that correspond to older revisions in server file journal 148 are not written to server file journal 148, which can create a conflict between existing revisions in server file journal 148 and revisions translated from operations data 302.

[0145] To enable synchronization of content item information between client device 150 and server file journal 148, file journal interface 202 can translate (e.g., via translation service 204) operations data 302 to revisions data 304, and vice versa.

When receiving operations data 302 from client device 150, file journal interface 202 can convert operations data 302 to revisions data 304, which includes content item revisions interpreted from operations in operations data 302. When receiving revisions data 304 from server file journal 148, file journal interface 202 can convert revisions data 304 to operations data 302, which include operations for implementing revisions in revisions data 304 at client device 150. Revisions data 304 includes data in server file journal 148 describing what happened to one or more content items (i.e., revisions to the one or more content items), and operations data 302 includes operations that have been executed or should be executed at client device 150 to modify the one or more content items. Thus, file journal interface 202 can translate data describing revisions to one or more content items from server file journal 148 (e.g., operations data 304) to operations that have or should be executed at client device 150 to modify the one or more content items at client device 150.

[0146] As previously noted, in addition to translating operations data 302 from client device 150 to revisions data 304 for server file journal 148, file journal interface 202 can convert revisions data 304 from server file journal 148 to operations data 302 for client device 150. File journal interface 202 can obtain revisions data 304 from server file journal 148 and translate revisions in revisions data 304 to operations for execution at client device 150 to revise one or more content items at client device 150 according to such revisions. The operations generated from the revisions in revisions data 304 are included in operations data 302 provided by file journal interface 202 to client device 150. This translation between operations data 302 and revisions data 304 allows client device 150 and server file journal 148 to synchronize content item information with each other as necessary.

[0147] Prior to writing to server file journal 148 any revision data 304 generated from operations data 302 provided by client device 150, file journal interface 202 can check a cursor in operations data 302 and/or query server file journal 148 to ensure any revisions in revisions data 304 do not create a conflict in server file journal 148. For example, file journal interface 202 can query server file journal 148 to check whether the version of a content item associated with a revision in revisions data 304 is the same the version of the content item at server file journal 148, or whether the version of the content item at server file journal 148 is an updated or different version as the content item to which the revision in revisions data 304 pertains. If server file journal

148 shows that the latest version of the content item is a different version than the version to which revision data 304 pertains, the two versions are in conflict.

[0148] File journal interface 202 can update server file journal 148 to store new revisions included in revisions data 304 derived from operations data 302. When querying and/or updating revisions in server file journal 148, file journal interface 202 can query namespace membership store 208 to retrieve namespace ownership information associated with any namespaces affected by the revisions in revisions data 304. The namespace ownership information can indicate which user account(s) own or are members of a particular namespace, and thus are able to access the particular namespace. Thus, file journal interface 202 can analyze the namespace ownership information to ensure server file journal 148 is not updated to include a revision to a namespace from a user account that is not a member of the namespace.

[0149] With reference to FIG. 3B, server file journal 148 can store journals 310, 312 to track and identify content item revisions and state. In this example, journal 310 includes records containing a namespace identifier (NS_ID), server journal identifier (SJ_ID), path, block, previous revision (Prev_Rev), and target namespace (Target_NS). NS_ID can include one or more values for uniquely identifying a namespace in server file journal 148. SJ_ID include monotonically increasing values which map to a row in a given namespace and provides an ordering of operations or revisions within that namespace. The path can be a namespace-relative path that identifies an associated content item. Prev_Rev identifies the SJ_ID of the row which corresponds to the previous state of the content item associated with the path. Target_NS identifies the NS_ID of the target namespace for a mount point of a mounted namespace. The Target_NS field is not set for rows (e.g., revisions) which do not correspond to mount points.

[0150] Journal 312 includes records containing an NS_ID, SJ_ID, clock (e.g., timestamp), file identifier (File_ID), extended attribute(s) (xattr), etc. The xattr can store metadata associated with content items or operations.

[0151] In some cases, journal 310 can include other fields such as a size field which represents the size of an associated content item, a directory field (e.g., Is_Dir) which can be set to indicate when a content item is a directory, a file identifier that uniquely identifies the associated file, a clock or timestamp field, etc.

[0152] File journal interface 202 can perform translation 320 based on operations data 302 and revisions data 304 as previously mentioned. When performing translation

320, translation service 204 can transform operations data 302 into revisions 322, which include linearized revisions for storage at server file journal 148. Translation service 204 can also transform revisions data 304 into linearized operations 324A, included in operations data 302 sent to client device 150, which can be applied by client device 150 to update content item information (e.g., state, changes, etc.) at client device 150. Translation service 204 can also generate or update cursor 324B and provide cursor 324B in operations data 302 to client device 150. Cursor 324B identifies a respective revision or row in server file journal 148 corresponding to each namespace and/or content item associated with linearized operations 324B.

[0153] For example, cursor 324B can identify a namespace (e.g., NS_ID) and row in server file journal 148 for that namespace (e.g., SJ_ID), which indicate the latest revision in server file journal 148 for that namespace. The namespace and row in cursor 324B can be associated with an operation in linearized operations 324A. Cursor 324B can identify a specific position on a log of revisions in server file journal 148 for the particular namespace, indicating the revision or state of the namespace in server file journal 148 after and/or before linearized operations 324A are applied at client device 150. Thus, cursor 324B can indicate the state of a namespace and/or content item in server file journal 148 before or after linearized operations 324A, which can help avoid revision conflicts and track the order of revisions before and after linearized operations 324A are applied.

[0154] FIG. 4A illustrates a diagram of an example translation and linearization process for translating server file journal data to linearized operations. In this example, journal 310 in server file journal 148 includes rows 402 with revisions 322 tracked by server file journal 148. Revisions 322 in journal 310 are associated with namespaces 100 and 101 (i.e., NS_IDs 100 and 101). In some cases, server file journal 148 can store namespace-specific journals that track revisions specific to respective namespaces. The rows (e.g., 402) in a namespace-specific journal include data specific to that namespace, and each row reflects a revision specific to that namespace.

[0155] Each row (402) in journal 310 includes a namespace identifier field (NS_ID) for uniquely identifying a namespace associated with that row, a server journal identifier field (SJ_ID) that includes monotonically increasing values which map to a row in a given namespace and provides an ordering of operations or revisions within that namespace. Journal 310 also includes a path field (Path) for identifying a

namespace-relative path of a content item, a block field (Block) for identifying a block or blocklist associated with the content item, a previous revision field (Prev_Rev) for identifying the row (i.e., SJ_ID) in journal 310 that represents the previous state or revision of the content item, and a target namespace field (Target_NS) for identifying a target namespace for a mount point of a mounted namespace (if the row corresponds to a mount). There is no data for the Target_NS field for rows (e.g., revisions) which do not correspond to mount points.

[0156] The first of rows 402 in journal 310 identifies the first revision (SJ_ID 1) for “File1” (Path field value File1) in namespace “100” (NS_ID 100), which corresponds to block “h1” and has no previous revisions (Prev_Rev) or target namespaces (Target_NS). Since the row does not include a previous revision or a target namespace, the revision represented by the row corresponds to an addition at namespace “100” of “File1” associated with block “h1”. The row in journal 310 containing SJ_ID “4” represents the last revision in journal 310 for “File1” on namespace “100”, since this row is the last row or SJ_ID in journal 310 corresponding to “File1” on namespace “100”. This row containing SJ_ID “4” indicates that “File1” on namespace “100” was edited after being added in SJ_ID “1”, and the edit corresponds to block “h4”.

[0157] Modifications 404 depict an example of modifications representing revisions 322. In this example, each of modifications 404 illustrates a content revision from a corresponding row (402) in journal 310. Each modification corresponds to an SJID and NSID in journal 310, and a file associated with the corresponding SJID and NSID in journal 310. In this example, the content associated with modifications 404 represents example content values of the blocks (e.g., “h1”, “h2”, “h3”, “h4”) in journal 310. The content values in modifications 404 are provided for illustration purposes to depict example modifications to content associated with each revision.

[0158] For example, the first modification in modifications 404 represents SJID “1” and NSID “100” in journal 310, and depicts “File1” in namespace “100” being added. Content “aaa” represents a value of “h1” for “File1” at SJID “1” of NSID “100”. Modifications 404 also depict an edit of “File1” in namespace “100” representing SJID “4” and NSID “100” in journal 310, which illustrates the content “aaa” (e.g., “h1”) associated with “File1” in namespace “100” being modified to “aa2” (e.g., “h4”).

[0159] In translation 320, revisions 322 from rows 402 in journal 310 are converted to linearized operations 324A. Linearized operations 324A are generated from revisions 322 in journal 310 and represent modifications 404 after linearization. As illustrated by linearized operations 324A, an operation in linearized operations 324A can be based on multiple revisions (322) and/or modifications (404), or a single revision (322) and/or modification (404).

[0160] For example, modifications 404 depict a revision adding “File1” to namespace “100”, which corresponds to SJID “1” and NSID “100” in journal 310, and a revision editing “File1” in namespace “100”, which corresponds to SJID “4” and NSID “100” in journal 310. The add revision can be inferred from the content value “aaa” (e.g., “h1”) associated with “File1” and NSID “100” and the lack of any previous revisions for “File1” and NSID “100”. In other words, the content “aaa” indicates that content (e.g., “h1”) was either added or edited, and the lack of a previous revision for “File1” and NSID “100” suggests that the content “aaa” represents content (e.g., “h1”) being added as opposed to edited. The edit revision can be inferred from the content value “aa2” (e.g., “h4”) associated with “File1” and NSID “100” and the previous revision (SJID “1” and NSID “100”) associated with “File1” and NSID “100”. In other words, the change from content “aaa” to “aa2” associated with “File1” and NSID “100” suggests that the content “aa2” represents an edit.

[0161] In linearized operations 324A, the add and edit modifications (404) corresponding to SJID “1” and SJID “4” for NSID “100” can be converted into a single linearized operation (Edit operation) which edits the content value associated with “File1” from “aaa” (e.g., “h1”) to “aa2” (e.g., “h4”). The single linearized operation editing content (e.g., “h1”) of “File1” to “aa2” (e.g., “h4”) reflects the modification adding “File1” associated with content “aaa” (e.g., “h1”) to namespace “100”, as well as the modification editing content “aaa” (e.g., “h1”) associated with “File1” in namespace “100” to “aa2” (e.g., “h4”). Accordingly, this linearized operation is based on two modifications 404 and two corresponding revisions in revisions 322.

[0162] The modification in modifications 404 corresponding to SJID “2” and NSID “100” in journal 310 represents a revision adding “File2” associated with content “bbb” (e.g., “h2”) to namespace “100”. This modification represents the only revision 322 from journal 310 corresponding to “File2” on namespace “100”. Accordingly, linearized operations 324A include a single operation for “File2” on namespace

“100”, which adds “File2” associated with content “bbb” (e.g., “h2”) to namespace “100” and is based on a single modification 404 (add of “File2” on namespace “100”) and revision 322.

[0163] Modifications 404 in this example also include for a modification adding “File3” associated with content “ccc” (e.g., “h3”) to namespace “100”, which corresponds to SJID “3” and NSID “100” in journal 310, and a delete (represented as “-1”) of “File3” from namespace “100”, which corresponds to SJID “5” and NSID “100” in journal 310. Thus, revisions 322 include two modifications 404 associated with “File3” on namespace “100”. Since the last revision in journal 310 associated with “File3” and namespace “100” corresponds to the delete modification representing SJID “5” and NSID “100” in journal 310, the add and delete modifications 404 associated with “File3” and namespace “100” from revisions 322 can be linearized to a single operation deleting “File3” from namespace “100”. Accordingly, linearized operations 324A include a single operation for “File3” and namespace “100”, which is the single operation deleting “File3” from namespace “100”.

[0164] SJIDs “6” and “7” for NSID “100” and SJID “1” for NSID “101” in journal 310 represent “Dir” being added to namespace “100” and later moved from namespace “100” to namespace “101”. For example, SJID “6” and NSID “100” identifies “Dir” and namespace “100” and does not include a previous revision, which indicates “Dir” was added to namespace “100” at SJID “6”. SJID “7” identifies “Dir” being moved from namespace “100” to namespace “101”, as reflected by the block field (“-“), the previous revision field (SJID “6”), and the target namespace field (“101”). SJID “1” for NSID “101” then identifies “Dir” being added to namespace “101”, as indicated by the lack of prior rows or revisions for “Dir” and namespace “101”. The add and move revisions in SJIDs “6” and “7” in NSID “100” and SJID “1” in NSID “8” are depicted by three modifications 404: an add of “Dir” to namespace “100” which corresponds to SJID “6” and NSID “100”, a delete of “Dir” from namespace “100” which corresponds to SJID “7” and NSID “100”, and an add of “Dir” to namespace “101” which corresponds to SJID “1” and NSID “101”.

[0165] The add and delete modifications 404 of “Dir” and namespace “100”, which respectively correspond to SJIDs “6” and “7” of NSID “100” in journal 310, are linearized to a single operation deleting “Dir” from namespace “100”, since the last revision in journal 310 corresponding to “Dir” and namespace “100” is a delete of

“Dir” from namespace “100” at SJID “7” and NSID “100”. The add of “Dir” to namespace “101”, which corresponds to SJID “1” and NSID “101” in journal 310, is the only modification 404 and revision 322 corresponding to “Dir” and namespace “101”. Accordingly, the add is provided in linearized operations 324A as a single mount operation for “Dir” and namespace “101”. Therefore, the three modifications 404 from revisions 322 corresponding to SJIDs “6” and “7” in NSID “100” and SJID “1” in NSID “101” (i.e., the add and delete of “Dir” on namespace “100”, and the add of “Dir” on namespace “101”), are linearized to two operations in linearized operations 324A: a delete operation for “Dir” in namespace “100” and a mount operation for “Dir” in namespace “101”.

[0166] As illustrated above, linearized operations 324A include an edit operation for “File1” and namespace “100”, an add operation for “File2” and namespace “100”, a delete operation of “File3” in namespace “100”, a delete operation for “Dir” in namespace “100”, and a mount operation for adding “Dir” to namespace “101”. These operations in linearized operations 324A are generated from revisions 322 and reflect the latest state of each content item in journal 310. File journal interface 202 can generate linearized operations 324A and send linearized operations 324A to client device 150 to ensure client device 150 contains the latest state from revisions 322 in journal 310.

[0167] When providing linearized operations 324A to client device 150, file journal interface 202 can include cursor 324B along with linearized operations 324A to client device 150. Cursor 324B can identify the last revision (SJID) for each namespace (NSID) in journal 310. In some embodiments, cursor 324B can also include an FSAuth token including the user ID, and the last observed access permissions to the NS_ID provided in the cursor. The last revision for each namespace can indicate a position in journal 310 corresponding to the latest revisions sent to client device 150 for each namespace.

[0168] In some cases, cursor 324B can also map each operation in linearized operations 324A to a namespace (NSID) and row (SJID) in journal 310. The namespace and row associated with an operation can indicate the position in journal 310 corresponding to the operation. In other words, the namespace and row associated with an operation can indicate the revision number in journal 310 represented by that operation. The namespaces and rows in cursor 324B correspond to the latest state in journal 310 for each namespace and content item associated with

linearized operations 324A. Cursor 324B can be provided to client device 150 as a tool for client device 150 to identify to file journal interface 202 the latest state or revisions obtained by client device 150 for one or more namespaces and/or content items when attempting to apply changes (e.g., via operations data 302) from client device 150 to the one or more namespaces and/or content items. When file journal interface 202 receives cursor 324B from client device 150, it can use cursor 324B to identify the position of client device 150 at journal 310 (e.g., the latest revisions from journal 310 obtained by client device 150) and detect or avoid conflicts caused by operations from client device 150.

[0169] For example, if file journal interface 202 receives an operation from client device 150 modifying “File1” in namespace “100”, file journal interface 202 can use cursor 324B, which it receives from client device 150 along with the operation, to check whether journal 310 has any newer revisions for “File1” in namespace “100” than the revision identified in cursor 324B from client device 150. If the revision in cursor 324B is the most current revision in journal 310, file journal interface 202 can commit the edit operation as a new revision in journal 310 (e.g., SJID “8” in NSID “100”) for “File1” in namespace “100”.

[0170] Alternatively, if the revision in cursor 324B is not the most current revision in journal 310 for “File1” in namespace “100”, file journal interface 202 can determine that the edit operation from client device 150 is not based on the most current version in journal 310 for “File1” in namespace “100”. For example, if cursor 324B identifies SJID “4” and NSID “100” in journal 310 and file journal interface 202 determines that journal 310 includes a revision at SJID “12” and NSID “100” for “File1” in namespace “100”, file journal interface 202 can determine that the edit operation from client device 150 pertains to an older version of “File1” on namespace “100” (e.g., SJID “4” and NSID “100”), and the edit operation can create a conflict as it edits a file that has since been modified. File journal interface 202 can detect this conflict created by the edit operation and reject the edit operation, attempt to reconcile the conflict, or provide the latest revisions to client device 150 and allow client device 150 to reconcile the conflict.

[0171] Each time file journal interface 202 sends linearized operations to client device 150, it can include a cursor as described here which identifies a respective position in journal 310 for each namespace and/or content item. Similarly, any time client device 150 sends an operation to file journal interface 202, it can include its

latest cursor which file journal interface 202 can use to map the state at client device 150 with the state at journal 310.

[0172] Journal 310 in this example depicts a journal with multiple namespaces. As previously noted, in some examples, server file journal 148 can maintain namespace-specific journals. Cursor 324B may include an SJID and NSID for each namespace, to indicate the latest revision for each namespace. Based on cursor 324B, file journal interface 200 can query multiple journals, in embodiments where multiple journals are maintained, and/or retrieve revisions from multiple journals, as further explained herein.

[0173] FIG. 4B illustrates a diagram of an example process for linearization 410 to convert operations data 302 from client device 150 to revisions 322 for journal 310 at server file journal 148. Client device 150 can provide operations data 302 to file journal interface 202. Operations data 302 in this example includes operations 412 at client device 150, such as content item edit, add, rename, move, mount, or delete operations. In some cases, operations 412 can include multiple operations to a same content item. For example, operations 412 can include an operation editing “File4” on namespace “100” and an operation deleting “File4” from namespace “100”.

[0174] Operations data 302 also includes cursor 324B previously received by client device 150 from file journal interface 202. Cursor 324B can identify the state (e.g., NSID and SJID) or latest revisions in journal 310 for one or more namespaces and/or content items. Client device 150 can provide cursor 324B to file journal interface 202 as a reference point for operations 412. In this example, cursor 324B provides the latest state for namespace “100”, which is represented by SJID “9”.

[0175] In some cases, the cursor is cryptographically signed by content management system 110, which allows file journal interface 202 to determine that the cursor has not been tampered with. Further, since client device 150 commit revisions to server file journal 148 when it has received the most recent revisions from server file journal 148 for the namespace, file journal interface 202 can accept that the last observed access permissions to the NS_ID are still valid, and therefore client device 150 has access to the namespace.

[0176] File journal interface 202 can receive operations 412 and cursor 324B and perform linearization 410, to linearize and transform operations 412 from client device 150 to revisions 322 for journal 310. Based on operations 412, file journal interface 202 can generate log 414 of operations. Log 414 can include a list of

operations from operations 412 mapped to respective namespace(s) in journal 310. In some cases, log 414 can include linearized operations (324A) generated from operations 412 as previously explained.

[0177] File journal interface 202 can use cursor 324B to verify that operations 412 reflect the latest state or revisions in journal 310 before updating journal 310 to reflect the operations in log 414. If file journal interface 202 confirms that cursor 324B reflects the latest state or revisions in journal 310 for the namespaces and/or content items associated with log 414, file journal interface 202 can add revisions 322 to journal 310 based on log 414. Revisions 322 can include the latest state or revision of each content item and/or namespace associated with the operations in log 414.

[0178] The operations in log 414 include an add and edit operation for "File5". Accordingly, revisions 322 include the edit of "File5", which file journal interface 202 can write to journal 310 as the latest state of "File5" (i.e., the state after the add and edit operations are applied to "File5" in a linearized fashion). The operations in log 414 also include an add operation for "Dir2" as well as edit and delete operations for "File4" on namespace "100". Revisions 322 can thus include an operation adding "Dir2" to namespace "100" and an operation deleting "File4" from namespace "100" as the latest state of "Dir2" and "File4" respectively.

[0179] In FIG. 4B, the revisions (322) depicted in journal 310 reflect the latest state of each content item ("File4", "File5", "Dir2") associated with operations 412. However, it should be noted that, in some cases, file journal interface 202 can write every revision represented by log 414 to journal 310 in order to reflect not only the latest state revision of each namespace and/or content item resulting from log 414, but also any previous states or revisions leading up to the latest state or revision. For example, file journal interface 202 can write a revision in journal 310 for the edit of "File4" and a subsequent revision for the delete of "File4", as opposed to only writing the edit of "File4" reflecting the latest state from operations 412, to indicate in journal 310 the full sequence of revisions of "File4" from operations 412.

[0180] File journal interface 202 can transform operations in log 414 to revisions 322 and update journal 310 to include revisions 322. File journal interface 202 can write revisions 322 to journal 310 at respective rows in journal 310. File journal interface 202 can add revisions 322 to the next available rows (e.g., SJIDs) in journal 310. In some cases, file journal interface 202 can add revisions 322 based on a relative order

which can be determined based on linearization 410 and/or respective timestamps or clocks.

[0181] As shown in FIG. 4B, the delete operation of “File4” in namespace “100” is included in row “11” or SJID “11” for namespace “100”. The revision in SJID “11” of journal 310 indicates that “File4” in namespace “100” has been deleted, as reflected by the minus symbol in the block field, and identifies SJID “9” as the previous revision in journal 310 for “File4” in namespace “100”. The addition of “Dir2” and edit of “File5” are included respectively in rows or SJIDs 12 and 14.

[0182] Journal 310 in FIG. 4B has been updated to include revisions 322 based on log 414 and cursor 324B, to reflect the state of each content item modified in log 414. The path field at each row in journal 310 identifies a content item within the associated namespace (e.g., namespace “100”). The path field of a row is based on the file and namespace from a corresponding operation in log 414. The block field in journal 310 represents the content item. In some cases, the block field can include a hash of a respective content item or data block. The block field can be empty if the content item has been deleted and/or is a directory, folder, mount, etc.

[0183] When updating journal 310 to include revisions 322 based on log 414 and cursor 324B, translation service 204 can identify the path of each content item to include in the path field of journal 310. In some cases, translation service 204 can translate an identifier of a content item (e.g., File ID) to a path of the content item (e.g., /directory/filename). For example, client device 150 can use identifiers to identify content items (e.g., content items in operations data 302) without having to track or calculate respective paths for the content items. Journal 310 may instead use a content item’s path to identify the content item. Translation service 204 can use the identifiers of content items from client device 150 to calculate the paths of the content items for journal 310, and update journal 310 using the paths calculated for the content items. Translation service 204 can also perform a reverse translation to obtain a content item’s identifier based on the content item’s path, and use the content item’s identifier when referencing the content item in communications with client device 150.

[0184] For example, translation service 204 can use the path in journal 310, NSID in journal 310, and/or a directory field in journal 310 (or elsewhere in server file journal 148) to identify a content item and obtain an identifier (e.g., File ID) of that content item. If file journal interface 202 sends an update or information to client device 150

pertaining to that content item, file journal interface 202 can provide the identifier of the content item to client device 150, which client device 150 can use to identify the content item with or without the path of the content item.

[0185] As previously mentioned, before writing revisions 322 to journal 310 from operations 412, file journal interface 202 can check if cursor 324B reflects the latest state or revision in journal 310 for each namespace and/or content item associated with operations 412. In some cases, after confirming that cursor 324B reflects the latest state or revisions in journal 310, file journal interface 202 can also perform a second check to ensure that a revision generated from operations 412 will not conflict with an existing revision in journal 310. For example, if SJID “5” in namespace “100” at journal 310 represents a delete operation of “File5”, the edit revision 322 of “File5” depicted in SJID “14” emitted from operations 412 received by file journal interface 202 from client device 150 would create a conflict by attempting to edit “File5” even though “File5” was deleted at SJID “5”. Thus, file journal interface 202 can reject the edit operation and revision in this example, and communicate to client device 150 that the edit operation is invalid. File journal interface 202 can update cursor 324B and provide the updated cursor to client device 150 to inform client device 150 of the latest state or revision in journal 310 for “File5” (and any other content item) as necessary.

[0186] FIG. 4C illustrates a diagram of an example linearization of cross-namespace operations. Cross-namespace linearization and cross-shard or cross-namespace listing can be performed via clock ordering. Tables 452A, 452B (collectively “452”) illustrate a batch of cross-namespace operations for linearization. Tables 452A, 452B respectively include columns 456A, 458A, which are namespace (NSID) fields for identifying a namespace for the records in tables 452A, 452B, columns 456B, 458B are SJID fields for identifying rows or SJIDs in tables 452A, 452B for respective namespaces in columns 456A, 458A, columns 456C, 458C are operations fields for identifying operations associated with each SJID, and columns 456D, 458D are clock fields for identifying a timestamp associated with the operations in columns 456C, 458C.

[0187] In this example, table 452A depicts SJIDs “100” and “101” for NSID “1”. SJID “100” is associated with an operation adding “foo.txt” to namespace “1” at timestamp “1000”, and SJID “101” is associated with an operation mounting namespace “2” at timestamp “1001”. Table 452B depicts SJIDs “1” and “2” for

NSID “2”. SJID “1” is associated with an operation adding “bar.txt” to namespace “2” at timestamp “500”, and SJID “2” is associated with an operation editing “bar.txt” at timestamp “1002”.

[0188] A linearizer (e.g., translation service 204) can obtain the batch of operations in tables 452 and emit a single stream of operations (462) with a cursor (464). The linearizer can identify all namespaces having at least one operation in tables 452 and linearize the operations for all namespaces based on the respective timestamps, NSIDs, SJIDs. In this example, the batch of operations in tables 452 linearize to the stream of operations shown in table 454.

[0189] Table 454 includes NSID column 460 which includes NSID fields for identifying the namespace of each operation, operations column 462 which includes operation fields for identifying the operations in table 454, and cursor column 464 which includes cursor fields for identifying a cursor state for each operation. Row 454A in table 454 includes the add operation from SJID “100” of namespace “1” in table 452A. The cursor state in cursor column 464 for row 454A is namespace “1” and SJID “100”, which indicates the add operation corresponds to SJID “100” in namespace “1” shown in table 452A. Row 454B in table 454 does not include a value in NSID column 460 or operations column 462, but updates the cursor state in cursor column 464 to include a cross-namespace cursor state, which in this example adds SJID “0” for namespace “2”.

[0190] Row 454C in table 454 includes the add operation from SJID “1” in namespace “2” shown in table 452A. The cursor state in cursor column 464 for row 454C includes the respective SJIDs “100” and “1” for namespaces “1” and “2” associated with the add operation in row 454C. As shown, the cursor state indicates the cursor is at SJID “100” in namespace “1” and SJID “1” in namespace “2”. In other words, the row or SJID in namespace “1” has not increased as the add operation does not affect the state of namespace “1”, but the row or SJID in namespace “2” has increased by one as the add operation represents a revision in namespace “2” and affects the state of namespace “2”. Thus, the cursor state in row 454C tracks the respective SJIDs for namespace “1” and namespace “2” after the add operation at SJID “1” in namespace “2”.

[0191] Row 454D in table 454 includes the mount operation at SJID “101” and namespace “1” at table 452A. The mount operation mounts namespace “2” at namespace “1”. The mount operation increases the SJID in namespace “1” from

“100” to “101”, but does not increase the SJID in namespace “2”. Accordingly, the cursor state in cursor column 464 for row 454D includes SJID “101” for namespace “1” and remains SJID “1” for namespace “2”. This cursor state reflects the state and/or order at namespaces “1” and “2”.

[0192] Row 454E in table 454 includes the edit operation at SJID “2” and namespace “2” in table 452A, which according to the respective timestamps of the mount and edit operations, is after the mount operation at SJID “101” in namespace “1”. The cursor state in cursor column 464 of row 454E maintains the cursor state for namespace “1” at SJID “101” but increases the cursor state for namespace “2” to SJID “2”.

[0193] As illustrated in table 454, operations 462 are listed as a stream of operations linearized based on causality and timestamps across namespaces “1” and “2”. Once operations 462 are linearized in table 454 to reflect cross-namespace causality and sequencing, operations 462 can be converted to revisions in server file journal 148 (e.g., revisions 322 in journal 310) and written to server file journal 148.

[0194] For example, a journal for namespace “1” in server file journal 148 can be updated to include a revision at SJID “100” representing the add operation adding “foo.txt” to namespace “1”, and a revision at SJID “101” representing the mount operation mounting namespace “2” on namespace “1”. Moreover, a journal for namespace “2” in server file journal 148 can be updated to include a revision at SJID “1” representing the add operation adding “bar.txt” to namespace “2”, and a revision at SJID “2” representing the edit operation editing “bar.txt” on namespace “2”.

[0195] Namespace Views and Snapshots

[0196] FIG. 5A illustrates an example process for constructing an up-to-date namespace view by content management system 110. First content management system 110 can determine (502) a most efficient mechanism to create namespace view.

[0197] A first method is to choose an arbitrary point in server file journal from which to read for the namespace, and to identify all paths in the namespace at that point in time. As addressed above, server file journal 148 is a collection of rows describing revisions to content items. The rows are organized by namespace identifiers and server file journal identifiers. The combination of a namespace identifier and a server file journal identifier (ns_id, sj_id) corresponds to a specific row in server file journal 148 and a specific revision within the namespace that is identified. Content

management system 110 can read (504) from server file journal 148 for a specific namespace ID at or before a recent row in server file journal 148. The revisions in server file journal 148 can be filtered by a latest tag, which identifies rows in server file journal 148 as being the last revision to any specific path. Accordingly, by reading all rows of server file journal 148 for a specified namespace having the latest tag applied will yield a list of revisions that identify every file path for the specified namespace. These file paths can be extracted (508) from the server file journal 148 entries and used to construct (510) a namespace view directory structure.

[0198] The second method is to read from a snapshot service provided by content storage service 116, which stores condensed snapshots of a namespace in content storage 142 at a recent point in time. The snapshot service is described in more detail below, however, by way of introduction, the snapshot service stores a condensed view of a namespace - including all of the content items in it - for a recent point in time. Large namespaces may need to be sharded over multiple discs, and shard names can include a hash that identifies each file path stored on that shard. A benefit of storing path information in the shard names is that the shards don't need to be opened to extract file paths, and you can use file path information to only open the shard(s) you are looking for when using the snapshot service. Each snapshot is current for a particular namespace at a specific server file journal ID (ns_is, sj_id).

[0199] Returning to FIG. 5A, file paths can be extracted (506) from the snapshot service for a namespace, and from these file paths are used to construct (510) a namespace view directory structure.

[0200] Content management system 110 can heuristically determine (502) which mechanism is likely to be the most efficient mechanism. If the namespace is not very active, and therefore there are not too many rows of revisions in server file journal 148 for any given content item, content management system may decide that the server file journal method is preferred. If however the namespace is very active, or some other factor causes content management system 110 to determine that it is more efficient to read from snapshot service, that method can be used.

[0201] Both mechanisms result in using extracted file paths to construct (510) a namespace view directory structure for a particular point in time in the past. In the server file journal method, content management system reads entries from server file journal 148 at or before a given server file journal ID, and in the snapshot method, the paths were extracted from a snapshot that was current as of a given server file

journal ID. Accordingly content management system 110 can bring the namespace view into a more current state by processing revisions (512) stored in rows in server file journal 148 that are after the server file journal ID that was used when paths were extracted from server file journal entries or snapshot service as described above.

[0202] Now that the server has a current view of a particular namespace that view needs to be filtered for the user account requesting the namespace view since, as described above, a given user account may not be able to see every content item in a namespace. Content management system can determine (514) access for the user account in the namespace using principles describe with respect to figure 4 above, and can filter (516) the constructed namespace view according to access permissions for the user account. Content management system 110 can send (518) the filtered namespace view two client device 1502 construct a remote tree.

[0203] Client device 150 can update (520) its local tree to match the remote tree. When updating its local tree, creating new directories is a relatively easy task however; storing the content items can require downloading a large amount of data. Fortunately, any content items already stored on client device do not need to be downloaded. Client device 150 can recreate the content items using from a block cache on client device 150 that contains blocks for deleted content items and comments for the content items. Using those already downloaded blocks is more efficient than downloading the blocks. However, any blocks not found in the block cache need to be downloaded (524) from the server.

[0204] The process illustrated in FIG. 5A can be repeated for each namespace mounted in organization directory 532.

[0205] FIG. 5B illustrates example information used to construct a user's (User 1) view (530) of shared directory 532. For any namespace, the namespace view builder process described with respect to FIG. 5A can result in extracted paths for the namespace that are filtered according to user 1's access rights. View 530 illustrates paths in the namespace for user account folder 534 of user 1. This can be used to determine a directory tree for the namespace, and the process can be repeated for each namespace. Each namespace can be located within organization directory based on information from the mount table.

[0206] View 530 for user 1 provides a view of namespaces and shared folders on content management system 110 from the perspective of user 1. As illustrated, view

530 includes user account folder 534 for user 1, and shared folders 540, which are visible to other user accounts on content management system 110.

[0207] Content management system 110 can provide a variety of access levels including: no access (user account can't access or see a folder); view name only (allows a user account to view the existence of a folder), traversal (allows a user account to view the existence of the folder, and open the folder, but does not allow the user account to see any folder(s) or content item(s) with the folder other than a specified folder(s) or content item(s)); read access (can view and open, but cannot change any content item or write to the folder); write access (can modify the folder or content item), etc. In some cases, these access levels can be used by content management system 110 to determine that a folder (such as those of user 1) should not be synchronized with client device 150. In such cases, client device 150 for user 1 may not be aware of the existence of other user account folders.

[0208] The access levels can thus determine what content items, including folders, user 1 can see, access, navigate, modify, etc. For example, in view 530, shared folders 540 include confidential shared folder 542. Confidential shared folder 542 is a confidential folder, which may be seen by members of directory 532 that do not have access to confidential shared folder 542, but such members of directory 532 which do not have access to confidential shared folder 542 may not be able to view content items in confidential shared folder 542.

[0209] In some instances, it can be useful to have a complete namespace saved together in one place. For example, in instances when an entire namespace needs to be constructed, such as when a namespace is shared with a new account, when a new client device is added to user account, or when transiting to an organization directory, it can be more efficient to download an entire namespace from one place instead of reading each revision from server file journal 148 and using content storage service 116 to extract blocks making up a content item from content storage 142.

[0210] Accordingly, content storage service 116 can include a snapshot service that stores namespace views at a point in time, or more specifically, as of a server journal identifier (SJID) for that namespace. A namespace view is a compressed snapshot of the contents of a namespace at that namespace (NSID). The snapshot includes the contents of a namespace, and may not be filtered for a user account's permissions, or hidden files, etc. Namespace snapshots may be large and need to be divided into

pages. Each page can include representations of paths within the namespace (e.g., a subset of paths that make up the namespace).

[0211] Since each page is compressed, it may be preferable to avoid opening a page on content management system 110, where it could potentially utilize a large amount of memory. Accordingly, in some cases, each path and prefix of the path can be hashed, and the hashes can be stored in the name of the snapshot page. Therefore, it is possible to search for the hash of a path without having to open pages of the snapshot to find that path. This optimization also works in reverse where the path names can be extracted from the snapshot pages to reconstruct a directory tree for a namespace.

[0212] To reconstruct a directory tree, the paths can be topologically sorted (parent_prefix_hash $\rightarrow$ path_hash). The initial set of paths with no dependencies are the children of the root of the namespace, and the full path (e.g., "/foo/bar/baz") can be reconstructed from the set of paths (e.g., "/foo," "/foo" $\rightarrow$ "/foo/bar", "/foo/bar" $\rightarrow$ "/foo/bar/baz"). Each path state along this set contains the last component of its path, e.g. "foo", "bar", "baz".

[0213] As addressed above, the namespace snapshots are not filtered for user account access permissions. Instead, after constructing the directory tree from the namespace, the paths in the directory tree can be filtered according to user account access permissions. User account access permissions can be obtained from content management system 110. Each path to which the user account has access to can be designated as allowed, e.g., represent the set of filters as follows:

- ALLOW_ONLY (hash1, hash2, hash3, ...)
- ALLOW_MOUNT (hash4, hash5, ...)

[0214] Each filter can contain full path hashes. If a filter is specified, content storage service can filter the paths out at runtime without doing a full topological sort. This allows only looking at data in a single page of a snapshot to apply the filters to that page.

[0215] Tree Data Structures

[0216] FIG. 6A shows an example of tree data structures, in accordance with various embodiments. The tree data structures may be stored at the client device and managed by a client synchronization service such as client synchronization service 156. In FIG. 6A, the tree data structures are shown including remote tree 610, sync tree 620, and local tree 630.

[0217] Remote tree 610 represents a server state or the state of content items stored remotely from the client device (e.g., on a server of the content management system). Local tree 630 represents a file system state or the state of the corresponding content items stored locally on the client device. Sync tree 620 represents a merge base for the local tree and the remote tree. The merge base may be thought of as a common ancestor of the local tree and the remote tree or a last known synced state between the local tree and the remote tree.

[0218] Each tree data structure (e.g., remote tree 610, sync tree 620, or local tree 630) may include one or more nodes. Each node may have one or more child nodes and the parent-child relationship is represented by an edge. For example, remote tree 610 includes nodes 602 and 604. Node 602 is a parent of node 604 and node 604 is a child of node 602. This parent-child relationship is represented by edge 606. A root node, such as root node 602, does not have a parent node. A leaf node, such as node 604, does not have a child node.

[0219] Each node in a tree data structure may represent a content item (e.g., a file, document, folder, etc.). For example, root node 602 may represent the root folder associated with the content management system and node 604 may represent a file (e.g., a text file named “Foo.txt”) located in that root folder. Each node in a tree data structure may contain data such as, for example, a directory file identifier (“DirFileID”) specifying the file identifier of a parent node of the content item, a file name for the content item, a file identifier for the content item, and metadata for the content item.

[0220] As described above, a client synchronization service may determine that the server state and the file system state of the client device are synchronized when all 3 trees (e.g., remote tree 610, sync tree 620, and local tree 630) are identical. In other words, the trees are in sync when their tree structures and the relationships that they express are identical and the data contained in their nodes are identical as well. Conversely, the trees are not in sync if the 3 trees are not identical. In the example scenario illustrated in FIG. 3, remote tree 610, sync tree 620, and local tree 630 are shown as being identical and in sync and, as a result, the server state and the file system state are synchronized.

[0221] Tracking Changes Using Tree Data Structures

[0222] FIG. 6B shows an example of tree data structures, in accordance with various embodiments. This example shows a scenario after a previously synchronized state,

such as the scenario illustrated in FIG. 6A, additional actions are performed on the content items represented in the trees to modify the content items such that the trees are no longer in sync. Sync tree 620 maintains a representation of the previously known synchronized state and may be used by the client synchronization service to identify the differences between the server state and the file system state as well as generate operations for the content management system and/or the client device to perform to converge so that the server state and the file system state are synchronized.

[0223] For example, a user (the same user as the user associated with the client device or a different user with access to the content item) may make modifications to the “foo.txt” content item stored by the content management system. This content item is represented by node 604 in remote tree 610. The modification shown in the remote tree 610 is a removal (e.g., a removal of the content item from a space managed by the content management system) or delete of the foo.txt content item. These modifications may be performed, for example, on another client device and the modifications were synced to the content item stored by the content management system or content item stored by the content management system via a web browser.

[0224] When the change is made on the content management system, the content management system generates modification data specifying the change made and transmits the modification data to the client synchronization service on the client device. The client synchronization service updates the remote tree representing the server state for the content items stored by the content management system based on the modification data. For example, in remote tree 610, node 604 representing the foo.txt content item is shown as deleted.

[0225] The client synchronization service may identify a difference between remote tree 610 and sync tree 620 and, as a result, determine that a modification of the content items at the content management system has caused the server state and the file system state to no longer be in sync. The client synchronization service may further generate and execute a sequence of operations for the content items stored on the client device that are configured to converge the server state and the file system state so that they will be in sync.

[0226] Additionally or alternatively, a user (the same user as the user associated with modifications at the content management system or a different user with access to the content item) may make modifications to the content items stored locally on the client device that are associated with the content management system. For example, the

user may add a folder “/bar” to the “/root” folder and add a “Hi.doc” document to the “/bar” folder.

[0227] When the change is made on the client device, the client device (e.g., client synchronization service 156 or client application 152) generates modification data specifying the change made and passes the modification data to the client synchronization service on the client device. The client synchronization service updates the local tree representing the file system state for the content items stored on the client device based on the modification data. For example, in local tree 630, node 612 and node 614 are shown as added. Node 612 and node 614 represent the “/bar” folder and the “Hi.doc” document respectively.

[0228] The client synchronization service may identify a difference between local tree 630 and sync tree 620 and, as a result, determine that a modification of the content items at the client device has caused the server state and the file system state to no longer be in sync. The client synchronization service may further generate a sequence of operations for the content items stored by the content management system that are configured to converge the server state and the file system state so that they will be in sync. This sequence of operations may be transmitted to the content management system for execution.

[0229] As seen in FIG. 6B, modifications to content items stored on the client device and content items stored by the content management system may occur at substantially the same time or within a particular time period. These modifications can be reflected in the tree data structures and used by the client synchronization service to generate sequences of operations for the client device and for the content management system in parallel. In other scenarios, however, modifications may not necessarily occur within the same time period and sequences of operations may be generated in an as-needed manner. Furthermore, although FIG. 6B illustrates scenarios for adding content items and deleting content items, other types of modifications such as, editing, renaming, copying, or moving content items are also supported.

[0230] According to various embodiments, identifying a difference between two tree data structures and generating operations may involve checking each node in both tree data structures and determining whether an action has been performed on the node. The actions may include, for example, the addition of the node, the deletion of the node, the editing of the node, or the moving of the node. These actions may then be

used to generate the sequence of operations configured to converge the server state and the file system state.

[0231] For example, if the two tree data structures are a sync tree and a remote tree, the client synchronization service may identify each node in the sync tree by, for example, requesting the file identifiers of all nodes in the sync tree. For each node or file identifier for the node in the sync tree, the client synchronization service may determine if the node or file identifier is also in the remote tree. A node or file identifier in the sync tree that is not found in the remote tree may indicate that the node has been deleted from the server state that is represented by the remote tree. Accordingly, the client synchronization service may determine that a delete action has occurred on the remote tree. If the node or file identifier for the node is found in the remote tree, the client synchronization service may check whether the node in the remote tree has been edited or moved.

[0232] To determine whether the node in the remote tree has been edited with respect to the node in the sync tree, the client synchronization service may compare the metadata for the node in the sync tree with the metadata for the corresponding node (e.g., the node with the same file identifier) in the remote tree. The metadata may include information that may be used to determine whether the content item represented by the node has been edited. For example, the metadata may include one or more hash values that are generated based on the data in the content item or a portion thereof. The metadata may additionally or alternatively include a size value, a last modified value, or other value for the content item. The metadata for the node in the client synchronization service may be compared with the metadata for the node in the remote tree. If the metadata do not match, an edit of the content item may have been edited in the server state represented by the remote tree. Accordingly, the client synchronization service may determine that an edit action has occurred for the node on the remote tree. If the metadata matches, no edit may have occurred.

[0233] To determine whether the node in the remote tree has been moved, the client synchronization service may compare the location for the node in the sync tree with the location for the corresponding node (e.g., the node with the same file identifier) in the remote tree. The location may include, for example, a path where the node is located, a file name, and/or a directory file identifier (“DirFileID”) specifying the file identifier of the node’s parent. If the locations match, no move may have occurred. On the other hand, if the locations do not match, a move of the content item may have

occurred in the server state represented by the remote tree. Accordingly, the client synchronization service may determine that a move action has occurred for the node on the remote tree.

[0234] To determine whether a node has been added to the remote tree, the client synchronization service may identify any nodes or file identifiers in the remote tree that are not found in the sync tree. If a node or file identifier is found in the remote tree and not found in the sync tree, the client synchronization service may determine that an add action of this node has occurred on the remote tree representing the server state.

[0235] Although the example above is described with respect to the sync tree and the remote tree, in other embodiments, a similar process may occur with the sync tree and a local tree in order to identify a difference between the sync tree and the local tree and determine which actions have occurred on the local tree representing the file system state.

[0236] File ID Allocation and Reassignment

[0237] FIG. 6C shows an example method for allocating and reassigning unique identifiers for content items created at client device 150 and synchronized between client device 150 and content management system 110. Although the methods and processes described herein may be shown with certain steps and operations in a particular order, additional, fewer, or alternative steps and operations performed in similar or alternative orders, or in parallel, are within the scope of various embodiments unless otherwise stated. The method may be implemented by a system such as, for example, client synchronization service 156 on client device 150.

[0238] At step 640, client device 150 creates a content item. At step 642, client application 152 (e.g., via client synchronization service 156) detects a content item event based on the content item created at step 640.

[0239] At step 644, client application 152 determines whether the content item has a unique identifier (e.g., File ID) from content management system 110. The unique identifier can refer to a globally unique identifier assigned and/or stored at content management system 110. Content management system 110 may assign unique identifiers for content items stored on content management system 110 to uniquely identify each content item at content management system 110 and any respective synchronized copies on client devices (e.g., client device 150). In some cases, a unique identifier can be associated with a particular namespace(s) on content

management system 110 corresponding to the content item assigned to the unique identifier. However, in other cases, the unique identifiers may be agnostic or independent of namespaces.

[0240] Moreover, the unique identifiers assigned by content management system 110 can be created by content management system 110 and differ from any local content identifiers generated by client devices for specific content items stored and/or generated at the client devices, such as a local filesystem identifier or a local identifier generated by the operating system at a client device.

[0241] If client application 152 determines that the content item already has a unique identifier from content management system 110, client application 152 proceeds to step 650 described below. If client application 152 determines that the content item does not have a unique identifier from content management system 110, client application 152 proceeds to step 646, where it determines if the content item should inherit a unique identifier from another content item. In some cases, this process for determining whether the content item should inherit a unique identifier refers to “path parenting”, and involves comparing the path of the content item with the path of another content item.

[0242] For example, if content item “A” is created at step 640 for path “/Root/Directory1/” and client application 152 determines that a previous content item with a unique identifier of “123” was previously or recently stored at “/Root/Directory1/”, client application 152 may determine that the new content item, content item “A”, is a version or copy of the previous content item with the unique identifier of “123” or that the previous content item was moved out of “/Root/Directory1/” and content item “A” represents the content item being re-added or moved back to “Root/Directory1”, and therefore should inherit the unique identifier “123” of the previous content item.

[0243] In some cases, client application 152 determines that the content item should inherit a unique identifier from another content item that client application 152 deems to be a previous version or copy of the content item. For example, when a content item stored in a specific location (e.g., path) on a client device is edited or modified, some applications or operating systems may move the content item to a temporary location on the client device and then add the edited or modified content item back to the specific location of the content item. Thus, when client application 152 detects that a content item was created and does not have a unique identifier, and also

identifies another content item that previously had the same path and/or name as the content item created, client application 152 may determine that the other content item was moved and is related or represents the same content item as the new content item. Client application 152 may therefore determine that the new content item should inherit the unique identifier of the other content item.

[0244] In some examples, client application 152 can compare metadata of the content item with metadata of other content items on client device 150 to determine if the content item should inherit the unique identifier of another content item – either because the content items are deemed to be related, equivalent, versions of each other, etc. The metadata used to compare the content items can include, for example, respective content item paths, filenames, directory identifiers, directory names, namespace identifiers (NSIDs), journal identifiers (SJIDs), revision identifiers, etc. A match of some or all of the metadata of two content items can trigger a determination that the newly-created content item should be assigned the unique identifier of the other content item. For example, if content item “A” created at step 640 has the same path, filename, directory identifier, NSID, and/or SJID as content item “B” at client device 150, client application 152 may assign a current, unique identifier of content item “B” to content item “A”.

[0245] If client application 152 determines at step 646 that the content item should inherit the unique identifier of another content item, client application 152 proceeds to step 650 further described below. If client application 152 determines at step 646 that the content item should not inherit a unique identifier from another content item on client device 150 (e.g., because it fails to find a content item associated with the same path, filename, and/or any other portion of metadata), client application 152 proceeds to step 648 where it obtains a unique identifier for the content item from content management system 110.

[0246] At step 648, client application 152 can request a unique identifier for the content item from content management system 110 and receive a unique identifier assigned to the content item by content management system 110. The unique identifier can be unique across content management system 110 as previously explained. Thus, the unique identifier can be used to uniquely identify the content item at content management system 110 and client device 150.

[0247] At step 650, when the content item has a unique identifier, client application 152 records the unique identifier for the content item. For example, client application

152 can write the unique identifier as metadata for the content item. In some cases, client application 152 appends the unique identifier to the content item or assigns the unique identifier to content item as an attribute, such as an extended attributed (xattr).

[0248] At step 652, client application 152 verifies that the unique identifier recorded for the content item is unique at content management system 110. As previously mentioned, content management system 110 can require that identifiers assigned to content items be unique at content management system 110 in order to allow the identifier to uniquely identify each content item at content management system 110. Content management system 110 can use the unique identifier to track, manage, synchronize, modify, and recognize content items. This uniqueness attribute can be verified by client application 152 at step 652 to prevent inadvertently duplicating identifiers or creating conflicts or errors as a result of a same identifier being assigned to more than one content item.

[0249] To verify the uniqueness of the unique identifier, client application 152 can notify content management system 110 of the unique identifier associated with the content item and request verification or acknowledgment from content management system 110 that the unique identifier is indeed unique at content management system 110. The request for verification or acknowledgement from client application 152 to content management system 110 can identify the unique identifier and may include any other information that may help content management system 110 verify uniqueness. For example, in some cases, the request for verification or acknowledgement from client application 152 to content management system 110 can also include a block or blocklist associated with the content item, a hash of at least a portion of the content item, and/or metadata associated with the content item, such as a path, a directory identifier, a filename, etc.

[0250] Content management system 110 can receive the request from client application 152 and search for the unique identifier in content management system 110. For example, content management system 110 can query server file journal 148 for any records containing the unique identifier. If content management system 110 identifies a matching identifier assigned to another content item and/or contained in a record on server file journal 148, content management system 110 can determine that the unique identifier is not unique – at least as it pertains to the content item and any other content items on content management system 110. Content management system 110 can notify client application 152 that the unique identifier is not unique or failed a

uniqueness test. In some cases, content management system 110 can also identify which other content item(s) the identifier is assigned to, as well as any information or metadata about those content items.

[0251] If content management system 110 does not identify a matching identifier assigned to another content item or contained in a record on server file journal 148, content management system 110 can acknowledge or verify that the unique identifier is indeed unique. In some cases, content management 110 can perform an additional search or verification before acknowledging to client application 152 the uniqueness of the unique identifier. For example, content management system 110 can search records in server file journal 148 for metadata associated with the content item, such as a path, filename, etc.

[0252] At step 654, if client application 152 determines that the unique identifier is not unique across content management system 110 (i.e., uniqueness verification fails), client application 152 can strip the identifier from one of the content items having the identifier and causing a duplicate identifier. For example, client application 152 can remove or unassign the identifier from the content item or from the other content item determined to have the same identifier. In some cases, client application 152 can delete an attribute (e.g., xattr), metadata, or tag containing the identifier from one of the content items having the same identifier in order to strip the identifier from that content item.

[0253] Once client application 152 has stripped the identifier from one of the content items having the same identifier, client application 152 returns to step 644 to process the content item that was stripped of the identifier. At step 644, client application 152 can process the content item without an identifier and proceed through steps 644, 646, 648, 650, 652 to obtain a unique identifier for such content item. The logic and heuristics from steps 644, 646, 648, 650, 652 allow client application 152 to obtain and verify a unique identifier for the content item and avoid duplicate identifiers.

[0254] In some cases, steps 644, 646, 648, 650, 652 can also be used to resurrect a content item that has been moved or deleted from a location. For example, if a content item needs to be re-added or moved at client device 150, client application 152 can perform steps 644, 646, 648, 650, 652 to obtain a unique identifier for the content item and store the content item at a particular path for the content item. As previously explained, in some examples, the path and/or content item identifier can be

determined by analyzing content items previously associated with the path or filename of the content item.

[0255] At step 656, if client application 152 determines that the unique identifier is unique across content management system 110 (i.e., uniqueness verification succeeds), client application 152 can report the content item to local tree 630. Here, client application 152 can add a node to local tree 630 representing the content item. The node in local tree 630 for the content item can identify the content item based on the unique identifier and may include other metadata associated with the content item, such as a path, a directory identifier, a filename, a timestamp, a size, a hash, etc.

[0256] At step 658, client application 152 can upload the content item to content management system 110 with the unique identifier for the content item. When uploading the content item, client application 152 can provide the unique identifier in order to identify the content item to content management system 110. Client application 152 can provide to content management system 110 the actual data or blocks (e.g., 220A, 220B, 220C, 220N) of the content item, the unique identifier, and any other metadata such as a path, directory identifier, filename, namespace identifier, timestamp, etc.

[0257] In some cases, client application 152 can upload the content item (e.g., 220) and/or the blocks (e.g., 220A, 220B, 220C, 220N) of the content item to content storage interface 206 for storage on content storage 142, and send operations data (e.g., 302) to file journal interface 202 including a cursor (e.g., 324B) indicating the position of client device 150 at server file journal 148 and one or more operations (e.g., 412) identifying the add operation adding the content item at client device 150. File journal interface 202 can use the cursor and the one or more operations from client application 152 as previously explained to add a revision to server file journal 148 (e.g., revision 322 to journal 310) indicating the content item was added at client device 150. As previously explained, file journal interface 202 can verify that the cursor from client device 150 is “at head”, meaning it reflects the latest revisions or state at server file journal 148, and convert the one or more operations to revisions for server file journal 148. Moreover, content storage interface 206 can receive the content item (e.g., 220) and/or the blocks (e.g., 220A, 220B, 220C, 220N), upload the content item (e.g., 220) and/or the blocks (e.g., 220A, 220B, 220C, 220N) to content storage 142, and record or index the content item (e.g., 220) and/or the blocks (e.g., 220A, 220B, 220C, 220N) in storage index 210.

[0258] When uploading the content item, client application 152 commit the addition of the content item to content management system 110 as further described below with reference to FIG. 13A. Committing the operation (e.g., the add of the content item at client device 150) allows client application 152 to report or record a revision to content management system 110 (e.g., server file journal 148) tracking the add of the content item. While uploading the content item and/or committing the add operation to content management system 110, client application 152 can implement crash safety procedures to prevent errors and enable recovery in the event that the upload and/or commit is not successful or fails at any point. In some cases, client application 152 can durably record the intent to commit the operation (e.g., the add and/or upload of the content item) on client device 150 to keep track of the commit until the commit either succeeds (e.g., is committed or recorded at content management system 110) or fails. Once the commit succeeds or fails, client application 152 can clear the intent to commit recorded to avoid confusion or errors, as further described in FIG. 13A.

[0259] If client application 152 succeeds in updating local tree 630 to include a node for the content item added and uploading the content item to content management storage 110, client application 152 can synchronize tree data structures 610, 620, 630 (i.e., remote tree, sync tree, and local tree) to ensure the remote and local metadata for content items at client device 150 and content management system 110 is synchronized, consistent, and error free. The synchronization can involve an incremental synchronization and/or a conflict handling procedure as described herein to synchronize remote tree 610, sync tree 620, and local tree 630.

[0260] FIG. 6D shows an example method for adding at client device 150 a content item downloaded from content management system 110 and recording the downloaded content item and unique identifier. At step 670, client application 152 downloads a content item from content management system 110. The content item can be a content item added to content management system 110 by another user account, client device, or through a web interface (e.g., 124), and synchronized to client device 150.

[0261] The content item downloaded can include a unique identifier assigned to the content item by content management system 110. In some cases, client application 152 can receive, for the download, an operation for adding the content item to client device 150. For example, client application 152 can receive an add operation

identifying the filename, path, unique identifier, etc. To illustrate, client application 152 can receive the operation “Add(/Directory/file.jpg : 321)”, where “Add” identifies the operation, “file.jpg” identifies the filename, “/Directory/” identifies the directory where the file resides, “/Directory/file.jpg” identifies the full path and/or filename, and “321” identifies the unique identifier of the file.

[0262] At step 672, client application 152 records the unique identifier of the content item for use at client device 150 to identify the content item. For example, client application 152 can write the unique identifier as an attribute of the content item (e.g., xattr), append the unique identifier to the content item, tag the content item with the unique identifier, and/or write the unique identifier to a metadata record or file for the content item (e.g., com.application.fileID).

[0263] At step 674, client application 152 can store the content item at a destination for the content item on client device 150. The destination can be the path defined for the content item. For example, the destination can include a specific directory within a namespace on client device 150.

[0264] Client application 152 can store data and/or various type of information on client device 150 for the content item. For example, client application 152 can store the contents or data of the content item, a modification date, application specific metadata, etc.

[0265] At step 676, client application 152 can report the content item to local tree 630, sync tree 620, and/or remote tree 610. For example, client application 152 can report the content item to remote tree 610, which adds a node on remote tree 610 for the content item. Client application 152 can then synchronize local tree 630, sync tree 620, and remote tree 610 based on the node added to remote tree 610 for the content item.

[0266] Synchronization Using Tree Data Structures

[0267] FIG. 7A shows an example method for synchronizing a server state and a file system state using tree data structures, in accordance with various embodiments of the subject technology. Although the methods and processes described herein may be shown with certain steps and operations in a particular order, additional, fewer, or alternative steps and operations performed in similar or alternative orders, or in parallel, are within the scope of various embodiments unless otherwise stated. The method may be implemented by a system such as, for example, client synchronization service 156 on client device 150.

[0268] The system is configured to identify a difference between a remote tree representing a server state for content items stored by the content management system, a local tree representing the file system state for the corresponding content items stored on the client device, and a sync tree representing a known synced state between the server state and the file system state. Based on these differences, a sequence of operations may be generated that, if executed, are configured to converge the server state and the file system state towards a synchronized state where the three tree data structures would be identical.

[0269] For example, at operation 702, the system may receive modification data for content items stored by a content management system or on a client device. The modification data may be used to update a remote tree or a local tree at operation 704.

[0270] The modification data specifies what changes are done to one or more content items associated with a content management service. Accordingly, the modification data may be received from the content management system or from the client device (e.g., from client application 152). Modification data received from the content management system may be referred to as server modification data. Server modification data specifies what changes are done to one or more content items by the content management system and may be used to update the remote tree at operation 704. Modification data received from the client device may be referred to as client modification data. Client modification data specifies what changes are done to one or more content items on the client device and may be used to update the local tree at operation 704.

[0271] At operation 706, the system may determine whether a server state for content items stored by the content management system and a file system state for the content items stored on the client device are in sync. Because the local tree and the remote tree are representative of the file system state and the server state and are continually being updated to track changes that occur at the content management system and the client device, determining whether the server state and the file system state are in sync may be done by comparing the local tree and/or the remote tree to the sync tree to find differences between the trees. This process of finding differences between the trees is sometimes referred to as “diffing” the trees.

[0272] According to some embodiments and scenarios, determining whether the server state and the file system state are in sync may include one or more of identifying differences between the remote tree and the sync tree and/or identifying

differences between the local tree and the sync tree. Differences between the remote tree and sync tree may indicate the occurrence of changes to content items stored by the content management system that may not be reflected at the client device. Similarly, differences between the local tree and sync tree may indicate the occurrence of changes to content items stored at the client device that may not be reflected at the content management system.

[0273] If there are no differences between the trees, the server state and the file system state are in sync and no synchronization actions are needed. Accordingly, the method may return to operation 702 and await new modification data. On the other hand, if differences are detected, the system may generate a sequence of operations configured to converge the server state and the file system state at operation 708.

[0274] The sequence of operations generated depends on the one or more differences that are detected. For example, if the difference between two trees is an added content item, the generated sequence of operations may include retrieving the added content item and adding it. If the difference between two trees is a deletion of a content item, the generated sequence of operations may include deleting the content item. The sequence of operations may also include a number of checks to ensure tree constraints are maintained. As will be described further below, the sequence of operations may conflict with the current state of the server state, the file system state, or other operations that are pending execution. Accordingly, the system may also resolve these conflicts before proceeding.

[0275] As noted above, if there are differences between the remote tree and sync tree, changes to content items stored by the content management system may have occurred that may not be reflected at the client device. Accordingly, in this scenario, the system may generate a client sequence of operations configured to operate on the content items stored on the client device to converge the server state and the file system state and this client sequence of operations may be provided to the client device for execution at operation 710.

[0276] On the other hand, if there are differences between the local tree and sync tree, changes to content items stored at the client device may have occurred that may not be reflected at the content management system. Accordingly, in this scenario, the system may generate a server sequence of operations configured to operate on the content items stored by the content management system to converge the server state and the file system state and this server sequence of operations may be provided to the

content management system for execution at operation 710. In some cases, both cases may be true and a client sequence of operations and a server sequence of operations may be generated and provided to their intended recipients at operation 710.

[0277] Once the sequence(s) of operations are provided to the intended recipient(s), the method may return to operation 702 and await new modification data. The sequence(s) of operations may provide one or more steps towards the convergence of the server state and the file system state or provide all steps needed to sync the server state and the file system state. For example, the content management system may receive the server sequence of operations and execute the server sequence of operations on content items stored by the content management system. This execution of the server sequence of operations causes changes to the content items stored by the content management system, which are detected and specified in server modification data, which is transmitted back to the system. The system may then update the remote tree and determine whether the server state and the file system state are in sync.

[0278] The client device may receive the client sequence of operations and execute the client sequence of operations on content items stored on the client device. This execution of the client sequence of operations causes changes to the content items stored on the client device, which are detected and specified in client modification data, which is passed to the system. The system may then update the local tree and determine whether the server state and the file system state are in sync. These operations of method 700 may continue until the server state and the file system state are in sync.

[0279] The operations of method 700 are described with respect to a client side and a server side (e.g., a local tree and a remote tree, a file system state and a server state, a client sequence of operations and a server sequence of operations, client modification data and server modification data). In various embodiments the operations associated with the two sides may occur in parallel, in sequence, in isolation of the other side, or a combination.

[0280] Conflict Handling

[0281] As described above with respect to FIG. 7A, differences between a sync tree and a remote tree are identified and used to generate a client sequence of operations configured to converge the server state and the file system state. However, in some cases, the client sequence of operations may conflict with the current state of a local

tree. Similarly, differences between the sync tree and the local tree are identified and used to generate a server sequence of operations configured to converge the server state and the file system state. However, the server sequence of operations may conflict with the current state of the remote tree. Additionally or alternatively, the client sequence of operations and the server sequence of operations may conflict with one another. Accordingly, various embodiments of the subject technology provide additional technical improvements by resolving these conflicts.

[0282] For example, the client synchronization service may identify an operation in a sequence of operations (e.g., the client sequence of operations or the server sequence of operations) that conflicts with a rule. Each rule used to identify a conflict may also be associated with a resolution for the conflict. The client synchronization service may update the sequence of operations based on the resolution for the conflict or perform resolve the conflict by performing operations associated with the resolutions for the conflict before providing the sequence of operations for execution.

[0283] FIG. 7B shows an example method for resolving conflicts when synchronizing a server state and a file system state using tree data structures, in accordance with various embodiments of the subject technology. Although the methods and processes described herein may be shown with certain steps and operations in a particular order, additional, fewer, or alternative steps and operations performed in similar or alternative orders, or in parallel, are within the scope of various embodiments unless otherwise stated. The method 600 may be implemented by a system such as, for example, client synchronization service 156, running on a client device.

[0284] The system may receive a sequence of operations configured to converge a server state and a file system state at operation 720. The sequence of operations may be, for example, the client sequence of operations or the server sequence of operations generated and described with respect to the method 700 of FIG. 7A.

[0285] At operation 720, the system identifies one or more violations in the sequence of operations based on a set of rules. The set of rules may be stored by client synchronization service 156 and specify a number of constraints, invariants, or conflicts that need to be resolved. The set of rules are applied to the tree data structures and help control sync behavior. Each rule in the set of rules may also be associated or otherwise linked to a resolution to a violation of that rule. For example, the resolution may include an alteration of one or more operations in the sequence of operations, a removal off one or more operations, an addition of one or more

operations, one or more additional actions to the server state or the file state, or a combination of actions.

[0286] For each operation in a sequence of operations, the system may determine whether any rule in the set of rules is violated. If a rule is violated, the system identifies a resolution of the violation and, at operation 722, performs the resolution. The resolution may include actions such as modifying one or more operations in the sequence of operations, a removing or adding one or more operations, or additional actions on the server state or the file state.

[0287] Once the resolution actions are performed, the system may generate a resolved or rebased sequence of operation based on the resolution and the sequence of operations at operation 724 and, at operation 728, provide the resolved sequence of operations to the appropriate entity for execution. For example, if the sequence of operations is a client sequence of operations, the resolved sequence of operations may be provided to the client device. If the sequence of operations is a server sequence of operations, the resolved sequence of operations may be provided to the content management service. Additionally, the method may be performed on client sequence of operations and server sequence of operations in sequence, in parallel, or in various different orders.

[0288] According to some embodiments, each type of operation may be associated with the same or a different set of rules. For example, operation types may include, for example, adding a content item, deleting a content item, editing a content item, moving a content item, renaming a content item, etc. The sequence of operations may consist of operations each belonging to one of the operation types above. Each operation type may be associated with a specific set of rules.

[0289] For illustrative purposes, a set of rules for an “Add” operation type may include rules such as file identifiers for content items must be unique in a tree (e.g., no two nodes in a tree may have the same file identifier), a directory file identifier (“DirFileID”) specifying the file identifier of a parent node of the content item must exist in the opposite tree data structure, and a DirFileID and file name combination for a content item are not used in the opposite tree.

[0290] Opposite tree, as used here, refers to the tree data structure that represents the state of the opposing entity. For example, a client sequence of operations configured to operate on the client device and the resulting changes to the file system on the client device will be reflected in the local tree. Accordingly, the opposite tree for the

client sequence of operations is the remote tree. Similarly, a server sequence of operations is configured to be transmitted to the content management system to be executed and the resulting changes to the server state will be reflected in the remote tree. Accordingly, the opposite tree for the server sequence of operations is the local tree.

[0291] Lamport Clocks

[0292] FIG. 8A illustrates a diagram of operations 802 showing dependencies and causality between the various operations (802). Operations 802 include an add operation, delete operation, edit operation, move operation, mount operation and unmount operation. Each of the operations has an intent or function. The add operation has an add intent (802A), the delete operation has a delete intent (802B), the edit operation has an edit intent (802C), the move operation has a move intent (802D), the mount operation has a mount intent (802E), and the unmount operation has an unmount intent (802F). Operations 802 can include dependencies or causal relationships, and may involve one or more atomic operations for achieving their respective intents (802A-F).

[0293] For example, the add operation is an atomic operation and simply involves add 804A for performing add intent 802A. The delete operation involves delete 804B which depends on add 804A, meaning the delete operation is executed for a content item after an add operation for that content item. The edit operation involves edit 804C which similarly depends on add 804A, meaning the edit operation is executed for a content item after the add operation for that content item.

[0294] The move operation involves delete 804B and add 804A, which are atomic operations that are executed to achieve move intent 802D for the move operation. The mount operation can involve mount 804E for executing mount intent 802E. However, in some cases, the mount operation may involve unmount 804F and mount 804E for executing mount intent 802E. For example, in a cross-namespace context, a mount operation in a namespace may involve an unmount (804F) and a mount (804E). In some cases, content management system 110 may include rules that prohibit two mounts from being executed if the two mounts create two paths to a same namespace, as further described below with reference to FIG. 8B. In this case, the second mount of the same namespace may trigger an unmount of that namespace to ensure the second mount does not lead to multiple paths to the same namespace for that user. Thus, in this example, at least one of the two mount operations would involve

unmount 804F and mount 804E. Unmount 804F would be a dependency of the mount operation.

[0295] The unmount operation involves mount 804E and unmount 804F for executing unmount intent 802F. In other words, to execute unmount intent 802F for a content item, the content item would have first been mounted through a mount (804E). The content item can then be unmounted through an unmount (804F).

[0296] The intents (802A-F) and atomic operations, dependencies, causalities, etc., for operations 802 can be used when calculating lamport clocks for operations 802. Lamport clocks can be used to determine an ordering and causality of events at one or more namespaces and ensure the state of events does not violate one or more rules. Lamports clocks can be used to ensure operations 802 are properly synchronized, linearized, serialized, etc., to avoid conflicts at one or more devices.

[0297] FIG. 8B illustrates a diagram of events across namespaces with lamport clocks calculated for the events. In this example, various operations have been executed across namespaces NSID 1, NSID 2, and NSID 3. Each namespace maintains an SJID for every operation at that namespace in order to determine the ordering of operations within the namespace. However, the SJID of a namespace does not identify ordering and causality of operations across namespaces. Accordingly, lamport clocks are calculated for the operations in the namespaces NSID 1, 2, 3 to determine causality and obtain a cross-namespace ordering of operations.

[0298] At NSID 1, operation 810 has SJID 1 and clock 1. At NSID 2, operation 816 has SJID 1 and clock 1. At NSID, operation 820 has SJID 1 and clock 1. Operations 810, 816, 820 span multiple namespaces and do not have causal relationships. Accordingly, operations 810, 816, 820 do not affect each other's clocks.

[0299] Ordering of operations within the namespace can be determined based on the SJID at the namespace. Clocks for operations within the same namespace can simply be incremented by 1. Thus, at SJID 2 in NSID 1, the clock for operation 812 is incremented to 2.

[0300] Operation 812 in NSID 1 is a move of File1 to NSID 2. Accordingly, operation 812 triggers operation 818 at NSID 2, which is the add of File1 at NSID 2. Since operation 818 at NSID 2 is causally dependent on another operation from a different namespace, namely operation 812 from NSID 1, the clock for operation 818 is calculated based on the clock at NSID 1 and the clock at NSID 2. The algorithm can be expressed as: $\text{TargetNS_clock}_{t1} = \max(\text{Source_NS}_{\text{clock}}, \text{TargetNS_clock}_{t0}) + 1$.

Thus, in this example, the clock for operation 818 at NSID 2 is 3 (e.g., $\max(2, 1) + 1$). Accordingly, operation 818 at NSID 2 has SJID 2 and clock 3.

[0301] Similarly, operation 816 at NSID is a move of File2 from NSID 2 to NSID 1. Operation 816 thus triggers operation 822 at NSID 1, for adding File2 at NSID 1. The clock for operation 822 is calculated based on the clock algorithm, which equals 3. Thus, operation 822 has SJID 3 at NSID 1 and clock 3.

[0302] Operation 824 at NSID 3 is causally dependent on an operation in the same namespace, namely operation 820 at NSID 3. Thus, the clock for operation 824 can be calculated by incrementing the clock of operation 820 at NSID 3. In this example, the clock for operation 824 is therefore 2. Operation 824 at NSID 3 has SJID 2 and clock 2. Since operation 824 is a move operation for moving Dir to NSID 1, operation 824 triggers operation 826 at NSID 1, adding Dir to NSID 1.

[0303] Since operation 826 is triggered by operation 824 in a different namespace (NSID 3), the clock for operation 826 is calculated based on the clock at NSID 1 and the clock for operation 824. Accordingly, the clock for operation 826 is set to 4 (e.g., $\max(2, 3) + 1$). Operation 826 thus has SJID 4 at NSID 1 and clock 4.

[0304] Operation 828 at NSID 1 adds File3 to NSID 1, and is not a cross-namespace operation. Accordingly, the clock for operation 828 is calculated by incrementing the clock at NSID 1. The clock for operation 828 is thus set to 8.

[0305] Operation 830 is causally dependent on operation 828 also within NSID 1. The clock for operation 830 is thus set to 6 by incrementing the clock of operation 828 at NSID 1. Operation 830 has SJID 6 at NSID 1 and clock 6.

[0306] Operation 830 is a move operation which moves File3 to NSID 3. Operation 830 thus triggers operation 832 at NSID 3. Since operation 832 is based on an operation from a different namespace, its clock is calculated using the clock algorithm based on the clock at NSID 3 and the clock of operation 830. In this case, the clock for operation 832 is set to 7. Operation 832 thus has SJID 3 at NSID 3 and clock 7.

[0307] Operations 834, 836 are not cross-namespace operations and are causally related to operation 830 at NSID 3. Thus, the clock for operations 834, 836 can be calculated by incrementing the clock of operation 832. In this example, the clocks for operations 834, 836 are set to 8 and 9 respectively.

[0308] FIG. 9A illustrates an example mount state violation 900 generated by a series of mount operations executed for a user. Mount state violation is based on an

invariant defined for mount operations which prohibits User 1 from mounting multiple paths (e.g., /a/c/ and /b/) to the same namespace (e.g., NSID 2).

[0309] In this example, mount operations 908, 910, 912 are executed at steps 902, 904, 906. At step 902, User 1 performs mount operation 908 which mounts namespace 1 (NSID 1) at path /a. Thus, at this point, User 1 has a mount to NSID 1 through path /a.

[0310] At step 904, User 1 then performs mount operation 910 which mounts namespace 2 (NSID 2) at path /c. Thus, at this point, User 1 has a mount to NSID 1 through path /a and a mount to NSID 2 through path /a/c.

[0311] At step 906, User 1 then performs mount operation 912 which mounts namespace 2 (NSID 2) at path /b. At this point, User 1 has a mount to NSID 1 through path /a, a mount to NSID 2 through path /a/c, and a mount to NSID 2 through path /b. Thus, User 1 has two paths to NSID 2, namely /a/c and /b. This violates an invariant which prevents multiple paths from pointing to the same namespace, and thus results in mount state violation 900. To avoid violating this invariant, operation 912 which mounts NSID 2 at path /b and establishes a second path to NSID 2 can depend on an unmount operation that breaks one of the paths to NSID 2. For example, operation 912 can depend on an unmount operation for unmounting NSID 2 at path /a/c. The unmount operation should be executed before operation 912 to avoid mount state violation 900. This would resolve mount state violation 900 created by operation 912 at step 906, and leave User 1 with one path to NSID 2, namely path /b to NSID 2.

[0312] Lamport clocks and dependencies can be leveraged to ensure the unmount operation is executed before operation 912. By properly setting lamport clocks on operations 902, 904, 906 as well as any unmounts from which operations 902, 904, 906 depend on, mounts and unmounts operations can be serialized and executed in a way that prevents mount state violation 900. On the other hand, improperly set lamport clocks can result in mount state violation 900, as well as other violations.

[0313] As previously shown in FIG. 8B, lamport clocks for events in the same namespace can be calculated by incrementing the clock of each new operation in the namespace. However, in the context of multiple namespaces, the calculation of lamport clocks can become more difficult, as operations, including mounts and unmounts, are generated across multiple namespaces. Moreover, when considering the invariant for mount operations that prohibits two paths pointing to the same

namespace, additional rules for calculating lamport clocks of mount and unmount operations may be necessary to ensure compliance with such invariant for mount and unmount operations. The lamport clocks, if properly calculated, can be used to enforce such invariants and prevent mount state violation 900. If not properly calculated, the lamport clocks can otherwise lead to violations such as mount state violation 900.

[0314] FIG. 9B illustrates a method for calculating lamport clocks for mount and unmount operations in a cross-namespace context. The method calculates lamport clocks in a way that allows the lamport clocks to be used to prevent violations of the invariant which provides that a mount operation should not lead to multiple paths to a same namespace (e.g., mount state violation 900). The lamport clocks are calculated such that mount and unmount operations can be linearized, serialized, and/or executed in a way that ensures that a consistent cut or state of the namespaces in content management system 110 will not lead client device 150 to have multiple paths to the same namespace. For example, the method can calculate the lamport clocks for operations 908, 910, 912 shown in FIG. 9A, to avoid mount state violation 900 by preventing the two paths to namespace 2 from being created as a result of operation 912 at step 906 mounting NSID 2 at path /b despite NSID 2 being already mounted at path /a/c/.

[0315] According to the example method, at step 920, content management system 110 detects a mount operation (e.g., operation 912) mounting a namespace (e.g., NSID 2) at a particular path (e.g., /b). At step 922, content management system 110 determines that the mount operation creates two paths to the same namespace (e.g., NSID 2). This condition is illustrated in step 906 of FIG. 9A. As previously mentioned, this condition violates an invariant prohibiting multiple paths to the same namespace, resulting in mount state violation 900. Accordingly, steps 924, 926, 928 of this example method, as described below, can be implemented to prevent this condition (e.g., mount state violation 900) from occurring.

[0316] At step 924, content management system 110 identifies dependencies for the mount operation. For example, content management system 110 can perform a mount dependency computation that identifies dependencies for the mount operation. The dependencies involve operations that have to occur prior to the mount operation for the mount operation to occur.

[0317] For example, if the mount operation violates the invariant prohibiting two paths to the same namespace, the mount operation should depend on an unmount operation that unmounts one of the two paths that points to the same namespace in order to prevent the invariant from being violated by the mount operation. To illustrate based on the example from step 906 in FIG. 9A, mount operation 912 mounting NSID 2 at path /b may depend on an unmount operation that unmounts NSID 2 at path /a/c/. As another example, a mount operation can involve multiple atomic operations, such as an unmount of a source and a mount of a destination, which can generate dependencies for the mount.

[0318] At step 926, content management system 110 identifies any unmount operations in the dependencies of the mount operation. Such unmount operations would be unmount dependencies for the mount operation. For example, content management system 110 may identify an unmount operation in the dependencies for mount operation 912. The unmount operation can be an operation that unmounts NSID 2 at path /a/c/. This unmount operation would be a dependency of mount operation 912.

[0319] At step 928, content management system 110 assigns the mount operation a higher lamport clock value than all unmount operations in the dependencies. Content management system 110 can either increase the clock value of the mount operation to ensure it exceeds the clock value of all unmount operations in its dependencies, or decrease the clock value of one or more of the unmount operations in its dependencies as necessary to yield a higher clock value for the mount operation. Content management system 110 can also ensure that each unmount operation has a higher clock value than a respective mount operation from which such unmount operation depends.

[0320] With reference to the example from FIG. 9B, content management system 110 can assign mount operation 912 a higher clock value than an unmount operation for unmounting NSID 2 from path /a/c/. Content management system 110 can also set the clock value of mount operation 910, which mounts NSID 2 at path /a/c/, to be lower than the clock value of the unmount operation for unmounting NSID 2 from path /a/c/. This calculation and configuration of lamport clocks can ensure that an unmount operation for NSID 2 is executed/processed between mount operation 910 mounting NSID 2 at path /a/c/ and mount operation 912 mounting NSID 2 at path /b/. As a result, the state of operations 908, 910, 912 will never lead to multiple paths

pointing to the same namespace (NSID 2) and mount state violation 900 will be prevented.

[0321] FIG. 9C illustrates an example lamport clock configuration 930 calculated for mount operations 908, 910, 912 based on the method shown in FIG. 9B. In this example, lamport clock configuration 930 depicts mount operation 908 mounting NSID 1 at path /a for User 1, mount operation 910 mounting NSID 2 at /a/c/ for User 1, and mount operation 912 mounting NSID 2 at /b/ for User 1. Lamport clock configuration 930 also includes unmount operation 914, which unmounts NSID 2 from /a/c/. Consistent with the invariant prohibiting multiple paths to the same namespace, mount operation 912 depends from unmount operation 914. Accordingly, unmount operation 914 is a dependency of mount operation 912. Content management system 110 can thus identify unmount operation 914 as an unmount operation in the dependencies of mount operation 912, as previously described in step 926 of FIG. 9B.

[0322] Lamport clock configuration 930 includes lamport clocks 932, 934, 936 calculated and assigned to operations 908, 910, 912, 914. Lamport clocks 932, 934, 936 are calculated to prevent mount state violation 900 by ensuring that unmount operation 914 is listed prior to mount operation 912. For example, because mount operation 912 depends on unmount operation 914, mount operation 912 is assigned a higher clock value than unmount operation 914. In this example, unmount operation 914 has lamport clock 934 with a clock value of “1”, and mount operation 912 has lamport clock 936 with a clock value of “2”. Mount operation 912 is thus ordered after unmount operation 914. Moreover, since unmount operation 914 depends on mount operation 910, lamport clock 932 for mount operation 910 is set lower than lamport clock 934 for unmount operation 914. As illustrated, lamport clock 932 has a value of “0”.

[0323] In some cases, content management system 110 can set lamport clocks 932, 934, 936 by decreasing the clock value for unmount operation 914 below the clock value for mount operation 912, and decreasing the clock value for mount operation 910 below the clock value for unmount operation 914. In other examples, content management system 110 can set lamport clocks 932, 934, 936 by incrementing the clock value of mount operation 912 to exceed the clock values of mount operation 910 and unmount operation 914.

[0324] Lamport clocks 932, 934, 936 can ensure that operations 908, 910, 912, 914 are properly serialized to prevent mount state violation 900 and any other mount/unmount invariant. FIG. 9D illustrates an example listing 940 of operations 908, 910, 912, 914 serialized according to lamport clocks 932, 934, 936.

[0325] Listing 940 includes operations 908, 910, 912, 914 and cursor states 942, 944, 946, 948. Listing 940 first includes mount operation 908 at NSID 1, which correlates to cursor state 942 identifying NSID 1 and SJID 1 for mount operation 908. Next, listing 940 includes mount operation 910 which has a clock value of “0” in lamport clock 932. Cursor state 944 corresponding to mount operation 910 includes SJID 1 at NSID 1 and SJID 0 at NSID 2.

[0326] Listing 940 next includes unmount operation 914 which has a clock value of “1” in lamport clock 934 and thus is ordered after mount operation 910. Cursor state 946 corresponding to unmount operation 914 includes SJID 1 at NSID 1 and SJID 1 at NSID 2. Cursor state 946 has thus incremented the SJID in NSID 2 after cursor state 944.

[0327] Listing 940 finally includes mount operation 912 which has a clock value of “2” in lamport clock 936 and thus is ordered after unmount operation 914 and mount operation 910. Cursor state 948 corresponding to mount operation 912 includes SJID 1 at NSID 1 and SJID 2 at NSID 2. Cursor state 948 has thus incremented the SJID in NSID 2 after cursor state 946.

[0328] Based on listing 940, mount operations 908, 910 would be executed first to mount NSID 1 at /a and NSID 2 at /a/c/. Before executing mount operation 912 for mounting NSID 2 at /b, unmount operation 914 would be executed to unmount NSID 2 from /a/c. Once NSID 2 has been unmounted, mount operation 912 would be executed to mount NSID 2 at /b. The resulting state would include a single path to NSID 2 and a single path to NSID 1, namely /b and /a respectively.

[0329] Mounts or namespaces can have numerous content items with different paths and timestamps. As a result, when a mount operation occurs, the mount operation can affect numerous content items in the mount, including other mounts within a mount. Improper updates to content items or lamport clocks in a mount can cause significant conflicts and inconsistencies. Accordingly, lamport clocks can be updated based on mount operations to ensure the mount has a proper lamport clock as well as any contents within the mount. Individual calculations of lamport clocks for each subitem within a mount can be computationally expensive, particularly as the number of

mounts and subitems increases. To increase performance of lamport clock computations when performing mounts and reduce the computational burdens, lamport clock calculations can be propagated based on relationships between mounts and/or content items.

[0330] FIG. 10A illustrates an example process for updating lamport clocks based on a mount operation. In this example, User 1, User 2, and User 3 have mounted namespaces 1002 which are depicted in tree structure 1000 based on the interrelationships between mounted namespaces 1002. For example, NSID 3 is mounted within NSID 2, and NSID 2 is mounted within NSID 1. Accordingly, NSID 1, 2, and 3 are depicted as a branch of tree structure 1000. Similarly, NSID 4 is mounted within NSID 1 and is thus depicted as another branch within tree structure 1000.

[0331] Moreover, mounted namespaces 1002 have lamport clocks 1004 calculated for each corresponding mount. In this example, NSID 3 has a clock value of “3”, NSID 2 has a clock value of “2”, NSID 1 has a clock value of “1”, and NSID 4 has a clock value of “4”.

[0332] Mount operation 1010 represents a mount operation for mounting namespace 1006 (NSID 5) to NSID 4. When NSID 5 is mounted according to mount operation 1010, lamport clock 1008 is calculated for NSID 5. The new clock value for lamport clock 1008 is calculated by determining the max clock value of the new mount (1006) and every mount 1002 up the tree (1000) from the new mount (1006), and incrementing the max clock value determined. In this example, the clock value for lamport clock 1008 is calculated by incrementing the max of the clock values of NSID 5, 4, and 1. Based on tree 1000, mount operation 1010 for NSID 5 yields a clock value of “5”. NSID 4 has a clock value of “4”, and NSID 1 has a clock value of “1”. Therefore, the max clock value of NSID 5, 4, and 1 is “5”, corresponding to NSID 5. The max clock value of “5” is then incremented to “6” and the clock value “6” is assigned as the clock value of lamport clock 1008.

[0333] The clock value of lamport clock 1008 as previously calculated is then propagated to all lamport clocks 1004 up the tree (1000) from the new mount (1006). Thus, updates 1012 are generated to set the clock values of NSID 4 and NSID 1, which are parents of NSID 5 in tree 1000, to “6” according to lamport clock 1008. As a result, the lamport clock values stored for NSID 1, NSID 4, and NSID 5 will be “6” based on lamport clock 1008 calculated in response to mount operation 1010. Tree

1000 thus allows lampport clocks 1004 to quickly be updated as necessary when a new mount is created within tree 1000.

[0334] Tree 1000 can also be used to calculate clock values for unmount operations by incrementing the lampport clock of an unmount and propagated the clock value to the lampport clocks up the tree from the unmount. For example, with reference to FIG. 10B, when unmount operation 1022 is generated for unmounting NSID 5 (1006), lampport clock 1024 is calculated based on the max clock value of lampport clocks 1008, 1004 for NSID 5, NSID 4, and NSID 1. The max clock value calculated is incremented and set as the clock value of the lampport clocks for each mount up the tree (1000) from NSID 5, including NSID 5, NSID 4, and NSID 1. Updates 1026 can propagate the clock value calculated for lampport clock 1024 to lampport clocks 1004 associated with NSID 4 and NSID 1.

[0335] FIG. 11A illustrates example tables in server file journal 148 for tracking move operations. Table 1102 includes journal records for operations. In some examples, table 1102 can store operations, clocks (e.g., timestamps) for the operations, namespaces (NSIDs) associated with the operations, journal identifiers (SJIDs) associated with the namespaces, etc. Example operations can include add operations, delete operations mount operations, unmount operations, move operations, etc. The operations can also include control operations. For example, a move can be associated with various move control operations which define an intent at each stage of the move. Example control operations include, without limitation, an outgoing move operation, an incoming move operation, a finish operation, etc. In some cases, table 1102 can also include an operation identifier. For example, table 1102 can include a move identifier (Move_ID) which identifies a particular move operation.

[0336] Tables 1104A, 1104B can host records defining intents for in-progress moves. For example, table 1104A can store incoming move records which reflect intents for in-progress moves on a destination namespace, and table 1104B can store outgoing move records which reflect intents for in-progress moves on a source namespace.

[0337] In table 1104A, the incoming move records can include, for example, a move identifier (Move_ID) for the move operation, an indication of the operation, a source namespace identifier (Source NSID), a destination namespace identifier (Destination NSID), a hash of the destination path, the destination path, a state of the move operation, a start clock (e.g., start timestamp), an end clock (e.g., end timestamp), etc. The state can indicate whether the move is in progress, aborted, or finished. Thus, the

state can identify whether the move is being processed or has been completed or aborted.

[0338] In table 1104B, the outgoing move records can include, for example, a move identifier (Move_ID) for the move operation, an indication of the operation, a source namespace identifier (Source NSID), a destination namespace identifier (Destination NSID), a hash of the source path, the source path, a state of the move operation, a clock (e.g., timestamp), a cursor (e.g., move identifier and state of operation), etc. As previously mentioned, the state can indicate whether the move is in progress, aborted, or finished.

[0339] As operations are detected, file journal interface 202 can write entries or records to tables 1102, 1104A, 1104B to track operations and state, serialize the operations, or synchronize the operations to client device 150.

[0340] FIG. 11B illustrates a diagram of an example sequence for processing cross-namespace moves with lamport clocks. In this example, the process depicts a cross-namespace move from NSID 1 (namespace 1) to NSID 2 (namespace 2). The various operations for the move are processed and serialized for NSID 1 and NSID 2 until the move is complete at both namespaces and can be emitted to client device 150.

[0341] Content management system 110 first records add operations 1106 for NSID 1 at server file journal 148. Add operations 1106 add “/A/Dir/”, “/A/Dir/foo.txt”, and “/A/Dir/bar.txt” to NSID 1. Here, “Dir” is a subdirectory within directory “A”, and files “foo.txt” and “bar.txt” are added within subdirectory “Dir”.

[0342] Content management system 110 then increments lamport clock 1120 for add operations 1106 and records the incremented lamport clock for add operations 1106.

[0343] After add operations 1106, content management system 110 detects a move of directory “/A/” from NSID 1 to NSID 2. Content management system 110 then records outgoing move operation 1108 for NSID 1 at table 1102. In this case, NSID 1 is the source namespace for the move operation of directory “/A/” at NSID 1. When processing outgoing move operation 1108, content management system 110 assigns move identifier 1112 to the move of “/A/” from NSID 1 to NSID 2. Moreover, content management system 110 can record information associated with the outgoing move at table 1104B containing outgoing moves. For example, content management system 110 can record move identifier 1112, identify NSID 1 as the source namespace, indicate that NSID 2 is the destination namespace, identify the source path at NSID 1, indicate a state of the move (e.g., in progress), define a cursor based

on outgoing move operation 1108, and store lampport clock 1120 for outgoing move operation 1108.

[0344] When recording lampport clock 1120 for outgoing move operation 1108, content management system 110 increments the value of lampport clock 1120 based on outgoing move operation 1108 and records the incremented lampport clock value.

[0345] Next, content management system 110 records incoming move operation 1110 for NSID 2 at table 1102. Incoming move operation 1110 corresponds to the incoming move of directory “/A/” at NSID 2. Content management system 110 can also store an incoming move record for incoming move operation 1110 at table 1104A, which contains incoming moves. Incoming move operation 1110 can be associated or recorded with move identifier 1112 to correlate incoming move operation 1110 at NSID 2 to the move associated with outgoing move operation 1108 at NSID 1. Content management system 110 can record at table 1104A move identifier 1112, the source namespace (NSID 1), the destination namespace (NSID 2), the source path at NSID 1, a state of the move (e.g., in progress), a start clock, and an end clock. The start clock can be the current value of lampport clock 1120 and the end clock can be the incremented value of lampport clock 1120.

[0346] Content management system 110 then processes delete operations 1114 at NSID 1, which delete “/A/Dir/”, “/A/Dir/foo.txt”, and “/A/Dir/bar.txt” from NSID 1. Content management system 110 associates delete operations 1114 with move identifier 1112 to indicate that delete operations 1114 correspond to the move of directory “/A/” from NSID 1 to NSID 2 associated with outgoing move operation 1108, and distinguish delete operations 1114 from any other delete operations unrelated to the move. In some cases, content management system 110 tags deletes (e.g., “/A/Dir/”, “/A/Dir/foo.txt”, and “/A/Dir/bar.txt”) with an attribute correlating move identifier 1112 to the deletes, such as an extended attribute (xattr).

[0347] Content management system 110 then increments lampport clock 1120 for delete operations 1114 at NSID 1 and records the incremented clock value.

[0348] Content management system 110 subsequently processes add operations 1116 at NSID 2, which add “/A/Dir/”, “/A/Dir/foo.txt”, and “/A/Dir/bar.txt” to NSID 2. Content management system 110 correlates add operations 1116 to move identifier 1112 to indicate that add operations 1116 correspond with the move associated with incoming move operation 1110, and distinguish the adds from any other adds

unrelated to the move. As previously explained, content management system 110 can similarly tag the adds with an attribute (e.g., xattr) identifying move identifier 1112.

[0349] Content management system 110 then increments lamport clock 1120, and records (e.g., table 1102) the incremented clock for add operations 1116 at NSID 2.

[0350] Content management system 110 then records finish operation 1118A for NSID 1 at table 1102. Content management system 110 records move identifier 1112 with finish operation 1118A to identify finish operation 1118A at NSID 1 and correlate the finish operation with the move from outgoing move operation 1108. Content management system 110 also increments lamport clock 1120 and records the incremented clock for finish operation 1118A at NSID 1. After finish operation 1118A, the move associated with outgoing move operation 1108 has completed at NSID 1 and can be cleared. Content management system 110 can write the finished state in table 1104B containing the records for outgoing moves.

[0351] Content management system 110 then records finish operation 1118B for NSID 2 at table 1102 to complete the move at NSID 2. Content management system 110 associates finish operation 1118B with move identifier 1112 to identify finish operation 1118B and correlate it to the move associated with incoming move operation 1110 at NSID 2. Content management system 110 also increments lamport clock 1120 and records the incremented clock value for finish operation 1118B at NSID 2. At this point, the move operation has completed at NSID 2 and content management system 110 can write a finished state in table 1104A containing incoming move records.

[0352] As illustrated in this example, when processing the move, content management system 110 can process delete operations 1114 at NSID 1 prior to add operations 1116 at NSID 2. This ordering or sequence allows client device 150 to list and process the delete and add operations (1114, 1116) according to their lamport clock order and avoid a scenario where the move results in client device 150 having two instances of the same content item or content item identifier (e.g., File ID). For example, this ordering or sequence prevents a state where client device 150 processes add operations 1116 and adds “/A” to NSID 2 while “/A” is also at NSID 1 and has not been removed from NSID 1. Such a sequence could result in two instances of “/A” at client device 150, both instances corresponding to the same content item path and ID.

[0353] To prevent such scenario, the delete operations (1114) in a move can be serialized before the add operations for the move based on lamport dependencies as described herein. Since the lamport clock (1120) is incremented at each process, the adds will causally depend on the deletes and will be listed and processed after the deletes.

[0354] In some cases, incoming move operations are also serialized before outgoing move operations. This way, incoming moves are recorded and processed before their corresponding outgoing moves. Thus, if client device 150 detects the incoming move (1110) before the outgoing move (1108), it can continue processing the move as described here with a guarantee that it will not detect adds for the move until the deletes have been processed.

[0355] Moreover, when client device 150 receives from content management system 110 an outgoing move without a prior incoming move, client device 150 can process the outgoing move as a delete. For example, if client device 150 has access to NSID 1 but does not have access to NSID 2, it may list outgoing move operation 1108 without seeing incoming move operation 1110. Since client device 150 does not have access to NSID 2, it can process delete operations 1114 and complete after the deletes. To illustrate, if client device 150 (or the user at client device 150) does not have access to NSID 2, client device 150 may detect the outgoing move of directory “/A/” at NSID 1, but may not have access to view or process operations 1110, 1116, 1118B associated with NSID 2 for the move. From the perspective of client device 150, the move may appear as a delete.

[0356] The serialization and lamport clock calculations here can therefore ensure proper ordering and processing of operations in a move and prevent errors and conflicts in the various scenarios that may be encountered by client device 150, such as moves associated with a source namespace (e.g., NSID 1) or a destination namespace (e.g., NSID 2) that client device 150 does not have access to. Proper serialization and lamport clock dependencies also prevent issues and conflicts resulting from changes generated during a move. Non-limiting example scenarios that can be encountered during a move include a mount or unmount of a source and/or destination namespace during the move, a remount of a source and/or destination namespace during the move, an add or delete operation during the move, a crash during the move, etc.

[0357] To address the various scenarios that may be encountered in a move, a state machine can define a flow of tasks and operations for the different scenarios at each move state. FIG. 11C illustrates an example move state machine (1130) which defines the flow of operations at various states of a move. Move state machine 1130 can be processed for every move identifier in tables 1102, 1104A, 1104B. Move state machine 1130 addresses various scenarios or conditions that may be encountered in a move, including mounts or unmounts detected during a cross-namespace move as well as any other operations or access conditions during the move.

[0358] As illustrated, move state machine 1130 enters start state 1150 when outgoing move operation 1108 is detected. Outgoing move operation 1108 thus triggers move state machine 1130 to enter start state 1150 and begin processing the move. When processing a stream of operations, if move state machine 1130 detects an operation other than an outgoing move (1108), move state machine 1130 will abort 1140 and will not transition to start state 1150. For example, if move state machine 1130 detects an add operation (e.g., 1116), a delete operation (e.g., 1114), a finish operation (e.g., 1118A, 1118B), an unmount source operation, or an unmount destination operation, move state machine 1130 aborts 1140 without entering start state 1150 to initiate a move.

[0359] Once outgoing move operation 1108 is detected, move state machine 1130 transitions to start state 1150 to process the move. At start state 1150, when incoming move operation 1110 is detected, the move is emitted to client device 150 and move state machine 1130 transitions to emitted state 1160.

[0360] Incoming move operation 1110 can be recorded at table 1102 and associated at table 1104A with a lamport clock value of $x-1$. Outgoing move operation 1108 can be recorded at table 1102 and associated at table 1104B with a lamport clock value of x . This way, incoming move operation 1110 is serialized before outgoing move operation 1108 as previously described.

[0361] After the move is emitted (e.g., emitted state 1160), the move is set to finished or unmounted 1134 and move state machine 1130 reaches end state 1190. A finished state or operation can be processed or recorded for the move, and the finished state or operation can be correlated with the move identifier (e.g., 1112) associated with the move.

[0362] If the move is within the same namespace (e.g., asynchronous), the source in the namespace can be set to finished 1134 or unmounted. If the move is across

namespaces (e.g., NSID 1 to NSID 2), the move can be set to finished 1134 at the source namespace (e.g., NSID 1) and/or the destination namespace (e.g., NSID 2). In some cases, a finished operation is correlated to the move identifier (e.g., 1112) of the move and recorded for the source namespace and/or the destination namespace.

[0363] If add operations 1114 or delete operations 1116 are detected after the move has been emitted (e.g., emitted state 1160), move state machine 1130 ignores the adds and deletes (1114, 1116) and proceeds to end state 1190. Thus, move state machine 1130 skips add operations 1114 and delete operations 1116 detected between emitted state 1160 and end state 1190.

[0364] In a cross-namespace move (e.g., NSID 1 to NSID 2), if unmount source operation 1136A is detected after the move is emitted (e.g., emitted state 1160) but before the move reaches end state 1190, move state machine 1130 transitions to unmount source state 1170. At unmount source state 1170, move state machine 1130 ignores any add operations 1116 detected. Move state machine 1130 sets the destination namespace (e.g., NSID 2) to finished or unmounted 1118B and proceeds to end state 1190. Setting the destination namespace to finished or unmounted 1118B can involve processing a finished or unmounted operation with the move identifier and/or recording a finished or unmounted state for the move identifier.

[0365] At unmount source state 1170, if remount scenario 1132 is detected, move state machine 1130 processes the outgoing move (e.g., 1108) for the source namespace (e.g., NSID 1), processes or emits delete operations (e.g., 1114) for the source namespace (e.g., NSID 1), and ignores any add operations 1116 detected for the destination namespace (e.g., NSID 2). From unmount source state 1170, move state machine 1170 proceeds when the source namespace is finished or unmounted, and sets the destination namespace to finished or unmounted 1118B and reaches end state 1190.

[0366] Remount scenario 1132 after unmount source operation 1136A can include remounting the source namespace (NSID 1) at any point before end state 1190. In some cases, unmount source operation 1136A and remount scenario 1132 can occur more than once prior to end state 1190. For example, a user can unmount the source namespace and remount the source namespace multiple times during the move. Each time the source namespace is unmounted and remounted, move state machine 1130 processes remount scenario 1132 at unmount source state 1170, until the source namespace is finished or unmounted and move state machine 1130 proceeds to end

state 1190. Thus, in some cases, move state machine 1130 can loop through remount scenario 1132 at unmount source state 1170 until the source and destination namespaces are finished and end state 1190 is reached.

[0367] If unmount destination operation 1136B is detected after the move is emitted (e.g., emitted state 1160), move state machine 1130 transitions to unmount destination state 1180. At unmount destination state 1180, move state machine 1130 determines the source namespace (e.g., NSID 1) is finished or unmounted 1118A and proceeds to end state 1190. The source namespace can be set to a finished or unmounted state and a finished or unmounted record or operation can be correlated to the move identifier of the move.

[0368] If remount scenario 1132 is detected at unmount destination state 1180, move state machine 1130 processes the incoming move (e.g., 1110) on the destination namespace (e.g., NSID 2), processes add operations 1116 at the destination namespace (e.g., NSID 2), and ignores delete operations 1114 detected. Move state machine 1130 proceeds from unmount destination state 1180 when the destination namespace finishes or unmounts. After unmount destination state 1180, move state machine 1130 proceeds to end state 1190 when the source namespace is finished or unmounted 1118A.

[0369] Remount scenario 1132 after unmount destination operation 1136B can include remounting the destination namespace (NSID 2) at any point after unmount destination operation 1136B and prior to end state 1190. In some cases, unmount destination operation 1136B and remount scenario 1132 can occur more than once prior to end state 1190, causing move state machine 1130 to loop through remount scenario 1132 at unmount destination state 1180 until proceeding to end state 1190.

[0370] FIG. 11D shows a diagram of example move operations across locations based on a unique identifier of a content item and linearized based on causal relationships. In this example, move operations 1197A-B are performed across location 1, location 2, and location 3 based on operations 1195A, 1195B. Locations 1, 2, 3 can represent different namespaces or shared folders, for example. Revision numbers 1192 represent the ordering of operations 1197A-B within locations 1, 2, 3.

[0371] At add operation 1195A in location 1, file “Foo.txt” with unique identifier “123” is added to location 1. Add operation 1195A is assigned revision number 1 in location 1 and lamport clock 1. When move operation 1197A is generated to move file “Foo.txt” with unique identifier “123” from location 1 to location 2, delete

operation 1195B is first processed at location 1 and designated revision number 2 in location 1 with lamport clock 2. Add operation 1195A is then processed at location 2 to add file “Foo.txt” with unique identifier “123” to location 2. Add operation 1195A at location 2 obtains revision number 1 at location 2 and lamport clock 3. As illustrated, the file (“Foo.txt”) retains its unique identifier “123” when moved from location 1 to location 2. Thus, the unique identifier “123” uniquely identifies the file at any location.

[0372] When processing move operation 1197A, delete operation 1195B at location 1 and add operation 1195A at location 2 are serialized such that the delete of the file from the source (i.e., location 1) occurs before the add of the file to the destination (i.e., location 2). This serialization can be guaranteed by causally relating the delete and add operations and assigning the add operation on the destination (i.e., add operation 1195A at location 2) a higher lamport clock (clock 3) than the delete operation on the source (i.e., delete operation 1195B at location 1 with clock 2).

[0373] Move operation 1197B triggers a move of file “Foo.txt” with unique identifier “123” from location 2 to location 3. Move operation 1197B involves delete operation 1195B at location 2, which deletes file “Foo.txt” with unique identifier “123” from location 2, and add operation 1195A at location 3, which adds the file “Foo.txt” with unique identifier “123” to location 3. Again, the file “Foo.txt” retains its unique identifier “123” before and after the move. Delete operation 1195B at location 2 is assigned revision number 2 at location 2 and lamport clock 4, and thus is causally after add operation 1195A at location 2, which has lamport clock 3. Add operation 1195A at location 3 is assigned revision number 1 at location 3 and lamport clock 5.

[0374] As illustrated, move operations 1197A-B are processed using the unique identifier of the content item being moved (i.e., file “Foo.txt”), and the operations for implementing the moves are serialized according to respective lamport clocks and ordered by revision numbers 1192 within their respective locations. When serializing the add and delete operations for the moves, delete operations at the source are listed before add operations at the destination and designated lower lamport clocks than the add operations at the destination.

Commit Protocol

[0375] FIG. 12 illustrates an example state of tree data structures 610, 620, 630 reflecting an operation at client device 150. Intent 1206 represents the intended result of the operation at client device 150. In this example, intent 1206 shows an add

operation for adding “Foo.txt” at client device 150. Local tree 630 has been modified to add node 1204 based on intent 1206. Node 1204 modifies local tree 630 to depict an add of “Foo.txt” within root node 1202. As previously explained, when intent 1206 is synchronized with content management system 110, client device 150 can update remote tree 610 and sync tree 620 to include node 1204 and thus synchronize tree data structures 610, 620, 630 at client device 150. When synchronized, tree data structures 610, 620, 630 reflect a synchronized state at client device 150.

[0376] To synchronize intent 1206 with content management system 110, client device 150 can commit intent 1206 to content management system 110. In this example, client device 150 commits the add of “Foo.txt” to content management system 110 in order to synchronize intent 1206 with content management system 110.

[0377] FIG. 13A illustrates an example method for committing intent 1206 to content management system 110. At step 1302, client synchronization service 156 on client device 150 records an intent to commit operation (e.g., operation from intent 1206) at client device 150. Client synchronization service 156 can record the intent to commit the operation durably on disk or memory at client device 150 to track the pending commit. Client device 150 can store dirty commit records and track modifications until an event triggers removal of the pending commit(s), such as a failure or success.

[0378] At step 1304, client synchronization service 156 commits the operation to content management system 110 (e.g., file journal interface 202). Client synchronization service 156 can send a message to content management system 110 requesting to commit the operation. The message can include a cursor, as previously explained, which content management system 110 can use to determine the state of content items at client device 150 and the position of client device 150 at server file journal 148. The cursor can include, for example, a server journal ID (SJID) for each namespace associated with client device 150. The SJID indicates the position of client device 150 at server file journal 148 for each namespace, and thus provides the state of each namespace at client device 150.

[0379] At step 1306, client synchronization service 156 determines whether the commit to content management system 110 from step 1304 succeeded. In some cases, content management system 110 can automatically send an error or acknowledgement to client synchronization service 156 in response to the commit, indicating whether the commit succeeded or failed. In some cases, client synchronization service 156 can contact content management system 110 to request

an acknowledgement or error notification. For example, client synchronization service 156 can send a message to content management system 110 requesting acknowledgment that the commit succeeded.

[0380] If the commit succeeded, content management system 110 can respond to client synchronization service 156 and notify client device 150 that the commit succeeded. If the commit did not succeed, content management system 110 may respond with a failure response or may not respond to the request from client synchronization service 156. If client synchronization service 156 determines the commit failed, at step 1310 client synchronization service 156 clears the intent to commit operation from client device 150. If the state at client device 150 is outdated relative to the state at server file journal 148, client synchronization service 156 can receive or request any necessary updates to synchronize the state at client device 150 with the state at server file journal 148.

[0381] If client synchronization service 156 determines the commit succeeded, at step 1308, client synchronization service 156 updates remote tree 610 and sync tree 620 to include node 804 as shown in FIG. 8. At step 1310, client synchronization service 156 then clears the intent to commit the operation from client device 150.

[0382] By determining whether the commit to content management system 110 at step 1306 succeeded, client synchronization service 156 can prevent conflicts resulting from scenarios where client synchronization service 156 commits an operation and updates its content and tree data structures (610, 620, 630) but the commit does not succeed at content management system 110. This would create a conflict in state between client device 150 and content management system 110.

[0383] For example, after client synchronization service 156 commits the operation at step 1304, various scenarios can prevent the commit from actually being successfully processed and applied at content management system 110, such as a crash at client device 150 or content management system 110, a network condition (e.g., latency or congestion, network failure, etc.), processing conditions (e.g., long queue at content management system 110, memory errors at content management system 110, etc.), and so forth. At step 1306, client synchronization service 156 can thus verify whether the commit succeeded or failed before deciding whether to apply the operation or clear the stored or cached intent to commit at client device 150.

[0384] Checking if the commit succeeded and clearing the intent to commit if the commit failed also allows client device 150 to distinguish self-authored operations

and avoid conflicting itself through self-authored operations. For example, client device 150 can record the intent to commit for any operations authored by client device 150 as well as an indication that the operations are self-authored. Client device 150 then checks with content management system 110 to verify whether the commit succeeded before applying the operation (e.g., updating the tree data structures). As described further below with respect to FIG. 13B, content management system 110 can guarantee that if it reports to client device 150 that an intent to commit has failed or was not received, such intent to commit will not subsequently succeed if subsequently received by content management system 110. Thus, client device 150 can similarly guarantee that a commit will not succeed after client device 150 indicates the intent to commit failed and clears the intent to commit after receiving a fail or success acknowledgment from content management system 110.

[0385] This can prevent a scenario where, for example, client device 150 believes an attempt to commit an operation failed and later receives from content management system 110 an update or indication of a revision based on the operation being subsequently approved by content management system 110. For example, client device 150 may send a commit to content management system 110. If the commit is temporarily lost or delayed, content management system 110 and client device 150 may believe that the commit failed. If the commit later resurfaces and content management system 110 processes and approves the commit after client device 150 has cleared the intent to commit or otherwise marked the intent to commit as failed, the operation associated with the failed commit could inadvertently be propagated to client device 150 even though the operation should have never been approved. Client device 150 may receive the operation and not realize the operation is its own operation from the previous failed commit, and apply the operation believing the operation was generated and synchronized from another device. Unable to distinguish the operation as the self-authored operation associated with the previous failed commit, client device 150 may apply the operation and create a conflict in the state of the associated content item(s) or namespace(s). The client device's own operation may thus create a conflict at the client device caused by the client device failing its own operation and subsequently applying the same operation under the belief the operation is a new revision generated by another device and propagated to the client device.

[0386] FIG. 13B illustrates an example method for processing commits at content management system 110. At step 1320, file journal interface 202 monitors commit requests from client devices (e.g., client device 150). At step 1322, file journal interface 202 receives a request to commit an operation from client device 150 (e.g., client synchronization service 156). The request can identify the operation associated with the commit and a cursor associated with client device 150. For example, the request can be a request from client synchronization service 156 to commit intent 1206 as shown in FIG. 13A. Thus, the request can identify the add operation (e.g., Add “Foo.txt”) corresponding to intent 1206 shown in FIG. 8, and the cursor at client device 150. The cursor at client device 150 can be the last or most current cursor received by client device 150 from file journal interface 202.

[0387] As previously explained, the cursor can identify the position of client device 150 at server file journal 148, which reflects the latest revisions or state of each namespace at client device 150 relative to the revisions or state at server file journal 148 for each of the namespaces. For example, the cursor can identify a server file journal identifier (SJID) for each namespace at client device 150. The SJID of a namespace indicates the latest SJID (e.g., row, revision, or state) for that namespace obtained by client device 150. The cursor can thus indicate whether client device 150 has received the latest revision in server file journal 148 for each namespace at client device 150.

[0388] In response to the request to commit the operation, at step 1324, file journal interface 202 checks the cursor associated with the commit from client device 150. At step 1326, file journal interface 202 determines if the cursor is at head. Here, file journal interface 202 determines if the revision (SJID) identified by the cursor for each namespace in the cursor is the latest revision for that namespace on server file journal 148.

[0389] For example, if the cursor identifies SJID 50 for namespace 2, file journal interface 202 determines if SJID 50 is the latest SJID (e.g., the latest row or revision) for namespace 2 in server file journal 148 (e.g., journal 310). If so, then the cursor from client device 150 is at head, meaning it is at the end of server file journal 148 for each namespace at client device 150, which indicates that client device 150 has obtained the latest revisions or state in server file journal 148 for each namespace at client device 150. If not, the cursor from client device 150 is not at head, meaning it is not at the end of server file journal 148 for each namespace at client device 150,

which indicates that client device 150 has not obtained the latest revisions or state in server file journal 148 for each namespace at client device 150. In other words, the content items at client device 150 associated with one or more namespaces are outdated.

[0390] If the cursor is not at head, at step 1328, file journal interface 202 rejects the commit from client device 150. For example, based on the cursor not being at head, file journal interface 202 can determine that the operation associated with the commit from client device 150 modifies a content item(s) and/or namespace(s) that is outdated (e.g., does not reflect the latest revisions at server file journal 148) and may create a conflict with the revisions at server file journal 148 for that content item(s) and/or namespace(s). File journal interface 202 can thus reject the commit to prevent a conflict being created by the operation associated with the commit. File journal interface 202 can send a response to client device 150 indicating that the commit has been rejected. File journal interface 202 can also send the latest revisions or state in server file journal 148 to client device 150, or prompt client device 150 to perform an update.

[0391] If the cursor is at head, at step 1330, file journal interface 202 can accept the commit. In some cases, file journal interface 202 can perform another check to verify the operation will not create a conflict prior to accepting the commit. For example, in addition to determining that the cursor is at head, file journal interface 202 can also compare the revision reflected by the operation associated with the commit (e.g., intent 1206) with the revisions at server file journal 148 to verify that the operation will not create a conflict even if the cursor is at head. To illustrate, assume the operation is a delete operation for a file and file journal interface 202 determines that the cursor is at head. Before accepting the commit, file journal interface 202 can verify that server file journal 148 includes a previous revision adding the file to a specific namespace. If file journal interface 202 determines that server file journal 148 does not include a revision for adding the file to the namespace, file journal interface 202 can identify a conflict between the delete operation and the lack of an add revision. File journal interface 202 can then reject the commit, reconcile the conflict, and/or ask client device 150 to reconcile the conflict.

[0392] After accepting the commit, at step 1332, file journal interface 202 updates server file journal 148 based on the commit. For example, file journal interface 202 can write a revision to server file journal 148 reflecting the operation associated with

the commit. File journal interface 202 can translate the operation to a revision, as previously described, and write the revision for the namespace associated with the operation.

[0393] At step 1334, file journal interface 202 can determine if it received an acknowledgment request from client device 150. For example, client device 150 can send a request to file journal interface 202 for acknowledgment of the commit in order to determine if the commit failed or succeeded. At step 1336, file journal interface 202 can respond to the acknowledgment request from client device 150. File journal interface 202 can notify client device 150 of the status of the commit, including whether the commit succeeded or failed.

[0394] At step 1338, file journal interface 202 can bump the cursor associated with client device 150. For example, assume the cursor from client device 150 identifies SJID 6 for namespace 1 as the latest revision for namespace 1 at client device 150, and the operation associated with the commit corresponds to namespace 1. File journal interface 202 can bump the cursor associated with namespace 1 from SJID 6 to SJID 7. File journal interface 202 can provide the updated cursor to client device 150, and client device 150 can use the updated cursor for future commits.

[0395] File journal interface 202 can bump the cursor after updating server file journal 148 or after receiving an acknowledgment request from client device 150. In some cases, file journal interface 202 can bump the cursor after receiving an acknowledgment request from client device 150 even if the commit associated with the acknowledgement request was not accepted by file journal interface 202. For example, if file journal interface 202 receives an acknowledgment request from client device 150, it can bump the cursor based on the acknowledgment request even if file journal interface 202 never received or approved the associated request to commit. By bumping the cursor, file journal interface 202 can guarantee that once it indicates to client device 150 that the commit has not been accepted or recorded, the commit will not be subsequently accepted or recorded.

[0396] For example, when client device 150 sends a request to commit to file journal interface 202, in some cases the request can be temporarily lost or delayed for a variety of reasons, such as network or computing issues (e.g., latency, congestion, crashes, client-side task cancellation, etc.). As a result, file journal interface 202 may not receive the request to commit until later than expected. For example, file journal interface 202 may receive an acknowledgment request from client device 150,

respond to the acknowledgment request, only to later receive the original request to commit from client device 150. In this example, if file journal interface 202 responds to client device 150 with an indication that the commit was not recorded, received, or approved, and subsequently receives and approves/records the commit, such commit can create a conflict between the state at content management system 110 and client device 150.

[0397] Thus, to guarantee that when file journal interface 202 tells client device 150 that a commit was not recorded or approved, such commit will not be later committed even if subsequently received by file journal interface 202, file journal interface 202 can bump the cursor to ensure a lost or delayed commit will be rejected if it resurfaces. For example, when file journal interface 202 receives a request to commit after responding to client device 150 that the commit was not recorded or approved, file journal interface 202 checks if the cursor is at head as explained in step 1326. Since file journal interface 202 has bumped the cursor since the request to commit was initially sent by client device 150, file journal interface 202 will determine that the cursor associated with the request to commit is not at head and thereafter reject the request to commit as explained in step 1328. Accordingly, file journal interface 202 can prevent conflicts or errors resulting from commits delayed or lost during processing or transmission. In other words, if a commit has not been approved or received when the cursor is bumped, bumping the cursor will ensure that such commit is never approved or recorded.

[0398] This also prevents conflicts at client device 150 from self-authored operations, as previously explained. For example, if client device 150 fails an intent to commit at client device 150, bumping the cursor will ensure a failed commit is not subsequently recorded by file journal interface 202 and propagated to client device 150.

[0399] Bumping the cursor can also prevent various modification or commit races that could be potentially created with other approaches, such as a “commit again” approach where client device 150 and/or file journal interface 202 can retry a failed commit. Below is a non-limiting example of such a race or condition prevented by bumping the cursor.

[0400] Client device 150 records intent 1206 to add “Foo.txt” with blocklist A. The commit request from client device 150 is dispatched to file journal interface 202 but delayed at the network. In addition, client device 150 crashes after sending the commit. Client device 150 then comes back online and in recovery mode attempts to

re-commit the add. File journal interface 202 rejects the commit because “Foo.txt” is already latest elsewhere. Client device 150 interprets the rejected commit as proof that its own commit which is in flight has failed and consequently clears the intent to commit cache/log at client device 150. Client device 150 then edits “Foo.txt” on client device 150 to blocklist B. “Foo.txt” is deleted remotely by another user or device. The original commit request for intent 1206 from client device 150 then arrives at file journal interface 202 and is accepted by file journal interface 202. Client device 150 then pulls the add of “Foo.txt” with blocklist A from file journal interface 202 at the same path as the edited “Foo.txt” with blocklist B. Client device 150 now has a conflict between the edited “Foo.txt” with blocklist B and its own previous add of “Foo.txt” with blocklist A. Accordingly, client device 150 has inadvertently conflicted itself.

[0401] This example scenario would be prevented by bumping the cursor. For example, when file journal interface 202 later receives the add commit for “Foo.txt” with blocklist A, file journal interface 202 checks the cursor and determines that the cursor is not at head (e.g., step 1326), as the cursor has been bumped since the commit was sent by client device 150. File journal interface 202 will then reject the commit and prevent client device 150 from pulling the add of “Foo.txt” with blocklist A from file journal interface 202. Thus, once client device 150 determines a commit has failed and clears the intent to commit cache/log, the commit is guaranteed to never succeed. Accordingly, file journal interface 202 and client device 150 can guarantee that client device 150 will not pull the add of “Foo.txt” with blocklist A after the edit of “Foo.txt” in the previous example, and create a conflict between “Foo.txt” with blocklist A and “Foo.txt” with blocklist B at the same path.

Symbolic Links

[0402] FIG. 14 shows an example symbolic link. A symbolic link (“sym link”) is a content item (e.g., file) that contains a reference or pointer to another content item (e.g., a file or directory). The reference or pointer is a path to the target content item, such as a relative or absolute path. The symbolic link may appear as a file, folder, or shortcut on client device 150, but when selected, the symbolic link redirects to the target content item. Thus, the symbolic link can be used to access the target content item from the symbolic link.

[0403] In FIG. 14, symbolic link 1410 depicts a symbolic link with path 1415 to target 1420. When a user clicks on symbolic link 1410, the user will be redirected to

target 1420 via path 1415. If path 1415 is valid and the user does not encounter any problems accessing target 1420, the user will gain access to the content of target 1420 through symbolic link 1410.

[0404] As illustrated here, path 1415 can be an absolute path (1430) or a relative path (1440). Absolute path 1430 contains the full path of target 1420 on a destination filesystem or location regardless of the relative location of symbolic link 1410, and can be accessed via symbolic link 1410 irrespective of any relative location between symbolic link 1410 and target 1420.

[0405] Relative path 1440 contains a path relative to symbolic link 1410. For example, symbolic link 1410 is located in this example at “/Root/Dir” and target 1420 is located at “/Root/Dir/A/b.txt”. In absolute path 1430, symbolic link 1410 points to the full path of target 1420, namely “/Root/Dir/A/b.txt”. In relative path 1440, symbolic link 1410 instead points to “../A/b.txt”, which is the path to target 1420 relative to “/Root/Dir”, where symbolic link 1410 is located.

[0406] In the synchronization context, symbolic links present various challenges. First, the synchronization behavior of symbolic links may vary on different operating systems. Moreover, if the system synchronizes the symbolic link but does not also store and synchronize the content of the target content item, other users or devices receiving a copy of the symbolic link may not be able to access the target. Thus, the symbolic link may not properly synchronize to other user accounts or devices and may become invalid for other user accounts or devices. The link created by the symbolic link to the target may thus be broken when synchronized. This can further create problems for users. For example, certain content items, such as application bundles, may often include files that are symbolic links pointing to other files or folders. If those symbolic links are broken through synchronization, the content items (e.g., application bundles) will become corrupt, invalid, etc.

[0407] On the other hand, if the system synchronizing a symbolic link follows the target and synchronizes the target’s content, this can also create multiple problems for other users or hosts. For example, in some instances, the target may not be accessible to other users or hosts and the synchronization may thus create inconsistencies. Different users and devices could end up with different synchronization results or behavior for the symbolic links based on various factors, such as access, link validity, etc. Moreover, the contents of the target could be added to a user’s account or device when the user did not intend for such content to be included, and may even fill up the

storage available for the user account or device. Such synchronization can also create security concerns by copying target content from a user's account or device which the user did not intend to share with other users or devices or may allow other users or devices to gain unauthorized access to the user's content and/or device.

[0408] Depending on whether the path in the symbolic link is absolute or relative, other inconsistencies can also result based on invalid links resulting at other hosts or accounts. For example, an absolute path used in a symbolic link may point to a valid location on a local filesystem of a client device, which will not be accessible from another device. Thus, when the symbolic link with the absolute path is synchronized to other devices, those devices will not have access to the target and the symbolic link will become invalid at those devices.

[0409] Various other problems can result when synchronizing symbolic links. FIGs. 15A-D illustrate various scenarios for synchronizing symbolic links, and describe different options for synchronization as well as potential benefits and disadvantages.

[0410] FIG. 15A shows a table of an example scenario (1500) for synchronizing symbolic links when the target and the symbolic link are contained in the same namespace (e.g., NSID). The table includes options 1505 for handling the synchronization of the symbolic link (e.g., 1410) depending on whether the path (e.g., 1415) in the symbolic link is a relative path (1440) or an absolute path (1430). Options 1505 depict various advantages and disadvantages for situations when the path is a relative path (1440) and when the path is an absolute path (1430), which are further described below.

[0411] Option 1 includes synchronizing the symbolic link as a symbolic link. This refers to treating the symbolic link as a symbolic link and synchronizing the symbolic link itself without following the path (1440 or 1430) and synchronizing the target, even if synchronization of the symbolic link may render the symbolic link invalid for some users or devices.

[0412] At option 1, synchronization of the symbolic link may ensure data integrity (e.g., data synchronized, including symbolic links, reflects the data as stored by the user) when path 1415 is relative path 1440 or absolute path 1430. Option 1 also ensures consistency across platforms (e.g., operating systems) when the path (1415) is either relative (1440) or absolute (1430).

[0413] Consistency across platforms refers to the synchronization behavior of client application 152 (e.g., via client synchronization service 156) across different

platforms (e.g., operating systems). For example, the ability or timing for detecting content operations or events and trigger synchronization can vary when client application 152 is hosted on different operating systems. This variation can depend on where the target resides, and may be affected by other factors as well. For example, as further explained below with respect to option 2 as well as FIGs. 15C and 15D, the timing or ability to detect changes and trigger synchronization when a change occurs may vary between operating systems depending on whether the change occurred on a target content item that is an “internal target” or an “external target”.

[0414] An internal target can refer to a target contained within an area or location (e.g., a directory, a folder, a volume, a partition, a disk, etc.) on client device 150 that is set or enabled for synchronization with content management system 110 and/or available to content management system 110 and/or client application 152 for management and/or synchronization, and an external target can refer to a target contained outside of such area or location. For example, a directory, namespace or folder on client device 150 that is tracked and synchronized by content management system 110 and contains any or all of the content items maintained by client device 150 for a user account at content management system 110 is an internal target, while anything outside of such directory, namespace or folder is an external target.

[0415] To illustrate, client device 150 may have a directory “/Home/” (not shown) which serves as the top directory of content items for user account “Lisa” (not shown) on content management system 110. Content items stored inside of “/Home/” can be synchronized with content management system 110, and content items residing outside of “/Home/” may not be synchronized with content management system 110. Thus, the directory “/Home/” and anything within “/Home/” can be considered an internal target by content management system 110. By contrast, anything outside of the directory “/Home/” can be considered an external target by content management system 110. Accordingly, in this example, the folder “/Home/Folder A/” (not shown) would be considered an internal target and any content items (e.g., folders, files, etc.) residing at, say, “/” or “/bin/” (not shown) on client device 150 may be considered an external target.

[0416] Referring back to the notion of consistency across platforms described in option 1, when a target is an internal target, client application 152 will generally be able to detect a change to the internal target and trigger synchronization when the change occurs on clients with different operating systems. Thus, this synchronization

behavior (i.e., the ability to detect and synchronize a change when it occurs) is consistent across platforms (e.g., operation systems). However, as further described below, synchronization behavior may vary across platforms when the change occurs at an external target.

[0417] As illustrated in scenario 1500 shown in FIG. 15A, option 1 can, however, result in inconsistencies across hosts that have synchronization disabled for the symbolic link or target. This applies when the path (1415) is relative (1440) or absolute (1430). For example, if a symbolic link points to “Folder A” (not shown) and a user has synchronization of “Folder A” to client device 150 disabled (e.g., by selectively disabling synchronization of “Folder A” to exclude “Folder A” from synchronizations operations), client device 150 will not store a local copy of “Folder A”, as “Folder A” will be removed from client device 150 for the user and may only be stored remotely at content management system 110 and/or at other client devices. Accordingly, when the symbolic link pointing to “Folder A” is synchronized at client device 150, the symbolic link will be invalid as it will point to a target (i.e., “Folder A”) that does not exist on client device 150.

[0418] If the path is an absolute path (1430), option 1 can also result in inconsistencies across hosts for any hosts where the absolute path is invalid. For example, as shown in FIG. 14, absolute path 1430 of symbolic link 1410 points to target 1420 at “/Root/Dir/A/b.txt”. If symbolic link 1410 is synchronized to a client device that does not contain one or more of the folders or directories in absolute path 1430 (e.g., “/Root/”, “/Dir/” and/or “/A/”), then symbolic link 1410 at that client device will be invalid.

[0419] Option 2 involves following the symbolic link to the target and synchronizing the target. Thus, option 2 synchronizes the actual target and content of the target. In the example of symbolic link 1410, option 2 would involve synchronizing target 1420 instead of symbolic link 1410.

[0420] In option 2, if the path of the symbolic link is a relative path (1440) or an absolute path (1430), synchronization option 2 can corrupt content items which rely on the symbolic link. For example, application bundles often include symbolic links pointing to other items and rely on those symbolic links being valid and their paths properly pointing to their targets. Thus, in option 2, when a target is synchronized instead of the symbolic link to that target, the application bundle relying on that symbolic link can become corrupt and/or may generate errors. To illustrate, many

different operating systems and applications use bundles to encapsulate code and resources for the operating systems and applications. Bundles can use directories and files for organization and may contain symbolic links. When option 2 synchronizes the target of a symbolic link instead of the symbolic link, the symbolic link is replaced with a copy of the target which generates an error when the operating system or application expects a symbolic link and attempts to read the symbolic link. This can in turn render the code, resources, and/or content implemented via the symbolic link unreadable or inaccessible by the operating system or application.

[0421] Synchronizing the target in option 2 can also lead to large storage utilization, as previously mentioned, when the path is either relative or absolute. For example, the symbolic link may lead to the target being synchronized and stored on a client device even though the user may not want or intend for that content to be stored on the client device. The synchronization of the target can also lead to higher storage utilization for a user account at content management system 110, which can result in higher costs to the user account or limited availability for other content items. In some cases, synchronization of the target can lead to a significant increase in storage utilization and may also lead to duplicate content items. For example, some media applications, such as a photos application, may use symbolic links in their libraries to modify a library folder structure or transform an older library folder structure to a new compatible structure. If the libraries reside in a synchronized directory and are enabled for synchronization with content management system 110, option 2 may cause the targets for every symbolic link in the libraries to be synchronized and copied across devices. This can result in duplication of data. For example, the content of the target may be stored and synchronized at their actual location as well as at the location of the target's symbolic link. If the volume of target content synchronized via option 2 is large or the size of the target content items is large – as is often the case with photos or videos for example – the duplicated data can result in a significant increase in storage utilization across client devices as well as on user account(s) at content management system 110.

[0422] For both relative and absolute paths, option 2 can also result in synchronization asymmetry where the access, views, and/or behavior differs between user accounts and/or devices. For example, assume a first client adds symbolic link 1410 in “/Root/” pointing to target 1420 at “/Root/Dir/A/b.txt”, and “/Root/” is later synchronized to a second client. If symbolic link 1410 is synchronized according to

option 2, namely by following target 1420 and synchronizing target 1420 (i.e., “/Root/Dir/A/b.txt”), the second client will receive a copy of “b.txt” from target 1420 at “/Root/Dir/”, while the first client will have symbolic link 1410 at “/Root/Dir/”. Such asymmetric behavior can thus cause synchronization inconsistencies where the first client and the second client may have different content after synchronization is complete. Moreover, such synchronization asymmetry can occur when the path of the symbolic link is a relative path (1440) or an absolute path (1430).

[0423] Further, at option 2, the synchronization of the target can create potential security problems. Potential security problems can result when content of the target is synchronized to other user accounts or devices, when the target may not have been intended for sharing or may expose sensitive data. Consider the following example. Assume a symbolic link in “Folder A” (not shown) points to “/etc/password” (not shown), and “Folder A” is later uploaded to content management system 110 and shared and/or synchronized with other client devices. When “Folder A” is shared and/or synchronized with a user account or client device, the content in “Folder A” will include the content from the target “/etc/password” of the symbolic link in “Folder A”. However, content item “/etc/password” may be a sensitive or security item. Thus, the inadvertent sharing or synchronization of “/etc/password” as part of the content of “Folder A” can create a security hole, and could be leveraged to extract data from a user’s computer or gain unauthorized access to a user’s data.

[0424] At option 3, the synchronization simply ignores symbolic links and doesn’t synchronize the symbolic links or the target content. In both cases of relative and absolute paths, this option can create a consistent behavior and view across platforms and hosts, but may corrupt content items that rely on the symbolic link or need access to the target content.

[0425] FIG. 15B shows a table of an example scenario (1510) for synchronizing symbolic links when the target and symbolic link are contained on different namespaces. Options 1, 2, and 3 in scenario 1510 have similar advantages and disadvantages for relative paths and absolute paths as scenario 1500. However, in addition, scenario 1510 introduces additional, potential problems caused by the content spanning multiple namespaces. For example, if a symbolic link points to a target on a different namespace, the symbolic link may become invalid if synchronized to users or devices that do not have access to the different namespace,

and the target content may not be accessible at option 2 if it resides on a namespace which a user or device cannot access.

[0426] FIG. 15C shows a table of an example scenario (1520) for synchronizing symbolic links when the target of the symbolic link is an internal target. As previously described with reference to FIG. 15A, an internal target can refer to a target contained within an area or location (e.g., a directory, a folder, a volume, a partition, a disk, etc.) on client device 150 that is set or enabled for synchronization with content management system 110 and/or available to content management system 110 and/or client application 152 for management and/or synchronization.

[0427] At option 1, if the path is a relative path (1440) or an absolute path (1430), synchronization of the symbolic link can ensure data integrity and consistency across platforms, but may create inconsistencies across hosts for targets that have synchronization disabled at those hosts and may result in inconsistencies for symbolic links spanning multiple namespaces. If the path is an absolute path, option 1 can also result in inconsistencies across hosts if the path is invalid at those hosts.

[0428] At option 2, if the path is a relative path (1440) or an absolute path (1430), synchronization of the target can lead to consistency across platforms, but may result in synchronization asymmetry, potential inconsistencies for symbolic links spanning multiple namespaces, and corruption of content items that rely on symbolic links being valid and functional, such as bundles. In addition, if the path is an absolute path, option 2 can create potential security problems as previously explained.

[0429] At option 3, when the path is either relative or absolute, ignoring the symbolic link can result in consistency across platforms and hosts, but may corrupt content items that rely on symbolic links being valid and functional.

[0430] FIG. 15D shows a table of an example scenario (1530) for synchronizing symbolic links when the target of the symbolic link is an external target. As previously described with reference to FIG. 15A, an external target can refer to a target that is not contained within an area or location (e.g., a directory, a folder, a volume, a partition, a disk, etc.) on client device 150 that is set or enabled for synchronization with content management system 110 and/or available to content management system 110 and/or client application 152 for management and/or synchronization.

[0431] At option 1, when the path is either relative or absolute, synchronizing the symbolic link can ensure data integrity, but may result in invalid links for user

accounts or hosts that do not have access to the target, and may result in inconsistencies across platforms.

[0432] Inconsistencies across platforms can result from client application 152 being implemented on different operating systems. For example, in UNIX-based operating systems such as LINUX and MACINTOSH operating systems developed by APPLE INC. (e.g., macOS or also referred as MAC OS X), client synchronization service 156 in client application 152 may detect changes occurring at an external target, and may trigger synchronization of the contents of the external target when the external target is updated. This means that web clients and other clients will have the latest version of the external target when the target is updated. However, in WINDOWS-based operating systems developed by MICROSOFT CORPORATION, client synchronization service 156 in client application 152 may not detect changes to external targets when the external targets are modified. Instead, client synchronization service 156 may detect such changes when client application 152 or client synchronization service 156 is restarted at the host. Thus, synchronization of changes to the external targets may not occur when the external targets are modified but later when client application 152 or client synchronization service 156 is restarted at the host. At various times, this can result in inconsistencies in the content on some user accounts or devices, and may lead some users to believe that the symbolic link synchronization is broken or not supported.

[0433] Moreover, the synchronized symbolic links are guaranteed to be valid only on the hosts where the symbolic links are created. In other words, the symbolic links can become invalid in some hosts, as the target may not exist in those hosts.

[0434] At option 2, when the path is either relative or absolute, synchronization of the target corrupts content items that rely on symbolic links being valid and functional, creates synchronization asymmetry, and results in inconsistencies across platforms. In the case of symbolic links that use an absolute path to the target, synchronization of the target can also lead to security risks as previously explained.

[0435] At option 3, ignoring the symbolic links can yield consistency across platforms and hosts in either absolute or relative path scenarios. However, in both absolute and relative paths, ignoring the symbolic links corrupts any content items that rely on symbolic links.

[0436] As illustrated in scenarios 1500, 1510, 1520, 1530, there are various options (1505) for synchronizing symbolic links, and the various options can have different

advantages and disadvantages for the various scenarios. In each case, the advantages and disadvantages are affected by a number of factors. Accordingly, rules and/or requirements can be configured for handling synchronization of symbolic links in the different scenarios. In some cases, rules can be set which require symbolic links to always be synchronized as symbolic links regardless of whether they contain an absolute or relative path or which scenario applies to the particular synchronization scenario. This requirement can ensure consistency across hosts, platforms, scenarios, etc., and provide predictable results and behavior for symbolic links. At this same time, as illustrated in FIGs. 15A-D, this approach can result in inconsistencies and errors in certain scenarios.

[0437] In some cases, rules can be created to define different synchronization behaviors for symbolic links based on options 1505, scenarios 1500, 1510, 1520, 1530, the type of path in a symbolic link (e.g., absolute path 1430 or relative path 1440), and/or the various advantages and disadvantages described in scenarios 1500, 1510, 1520, 1530. For example, a rule can be defined that requires the actual symbolic links to be synchronized in every instance, but also provides that if synchronization of the symbolic link breaks/invalidates the symbolic link, content management system 110 can generate a prompt for one or more user accounts affected which requests user input, and allows the user to accept the synchronization, reject the synchronization, request the symbolic link be modified to preserve its functionality or validity post synchronization, or request that the target be also shared and synchronized with a user who does not have access to the target and thus would not be able to use the symbolic link.

[0438] For example, if a symbolic link is being synchronized with a set of user accounts or devices that also have access to the target, content management system 110 can determine that synchronization of the symbolic link will not break or invalidate the symbolic link, and simply synchronize the symbolic link to those user accounts. If the symbolic link has an absolute path which may become invalid when synchronized with other user accounts or devices, but the other user accounts or devices otherwise have access to the target, content management system 110 can notify the user account that created the symbolic link that the symbolic link will not function for the other user accounts or devices and ask the user account if the path in the symbolic link should be converted from absolute to relative in order to preserve the functionality or validity of the link for the other user accounts or devices. The

user can allow modification of the symbolic link to preserve its validity after synchronization, and content management system 110 can change the absolute path to a relative path that is valid for all user accounts or devices receiving a synchronized copy of the symbolic link. Since the other user accounts or devices in this example already have access to the target, the symbolic link will work properly for each user account or device and allow the target to be accessed from the symbolic link.

[0439] If a user account or device has disabled synchronization of the target, the symbolic link will be invalid for that user account or device even if the path is modified to a relative path. In this case, content management system 110 can determine that the symbolic link will not work for that user account or device because synchronization for the target has been selectively disabled for that user account or device, and generate a notification for the user account or device. The notification can indicate that the symbolic link will not function while synchronization is selectively disabled for the target, and provide an option for the user account or device to modify the synchronization settings for the target and selectively enable synchronization of the target. If the user account or device accepts enabling of synchronization of the target, the target can be synchronized with the user account or device and the symbolic link can then function properly.

[0440] If a user account or device does not have access to the target, content management system 110 can synchronize the symbolic link with that user account or device but also generate an alert to the user account or device that created the symbolic link (or any other user account with access to both the symbolic link and the target), indicating that the symbolic link will be invalid for the particular user account or device that does not have access to the target. The user account or device may not have access to the target for a number of reasons, as described in scenarios 1500, 1510, 1520, 1530, such as the target being located in a location (e.g., a namespace, an external location, etc.) that is not accessible by the user account or device. In some cases, the alert may also provide an option to share the target with the user account or device that does not have access to the target. For example, the alert may indicate that the symbolic link will be invalid for a particular user account or device receiving a copy of the symbolic link because the particular user account or device does not have access to the target, and provide an option to share the target with the particular user account or device. If a user who receives the alert and has permissions that allow that user to share the target with the particular user account or device accepts sharing the

target with the particular user account or device, content management system 110 can in turn synchronize both the symbolic link and the target with the particular user account or device. This in turn may ensure that the symbolic link is valid for that particular user account or device.

[0441] If content management system 110 determines that, after synchronizing the target to that particular user account or device, the symbolic link will still be invalid for that particular user account or device because the symbolic link includes an absolute path that will not be valid for the particular user account or device, then content management system 110 can additionally include an option in the alert for modifying the path in the symbolic link from an absolute path to a relative path that works within the structure of the synchronized content items.

[0442] FIG. 16 shows an example of computing system 1600, which can be for example any computing device making up client device 150, content management system 110 or any component thereof in which the components of the system are in communication with each other using connection 1605. Connection 1605 can be a physical connection via a bus, or a direct connection into processor 1610, such as in a chipset architecture. Connection 1605 can also be a virtual connection, networked connection, or logical connection.

[0443] In some embodiments computing system 1600 is a distributed system in which the functions described in this disclosure can be distributed within a datacenter, multiple datacenters, a peer network, etc. In some embodiments, one or more of the described system components represents many such components each performing some or all of the function for which the component is described. In some embodiments, the components can be physical or virtual devices.

[0444] Example system 1600 includes at least one processing unit (CPU or processor) 1610 and connection 1605 that couples various system components including system memory 1615, such as read only memory (ROM) 1620 and random access memory (RAM) 1625 to processor 1610. Computing system 1600 can include a cache of high-speed memory 1612 connected directly with, in close proximity to, or integrated as part of processor 1610.

[0445] Processor 1610 can include any general purpose processor and a hardware service or software service, such as services 1632, 1634, and 1636 stored in storage device 1630, configured to control processor 1610 as well as a special-purpose processor where software instructions are incorporated into the actual processor

design. Processor 1610 may essentially be a completely self-contained computing system, containing multiple cores or processors, a bus, memory controller, cache, etc. A multi-core processor may be symmetric or asymmetric.

[0446] To enable user interaction, computing system 1600 includes an input device 1645, which can represent any number of input mechanisms, such as a microphone for speech, a touch-sensitive screen for gesture or graphical input, keyboard, mouse, motion input, speech, etc. Computing system 1600 can also include output device 1635, which can be one or more of a number of output mechanisms known to those of skill in the art. In some instances, multimodal systems can enable a user to provide multiple types of input/output to communicate with computing system 1600. Computing system 1600 can include communications interface 1640, which can generally govern and manage the user input and system output. There is no restriction on operating on any particular hardware arrangement and therefore the basic features here may easily be substituted for improved hardware or firmware arrangements as they are developed.

[0447] Storage device 1630 can be a non-volatile memory device and can be a hard disk or other types of computer readable media which can store data that are accessible by a computer, such as magnetic cassettes, flash memory cards, solid state memory devices, digital versatile disks, cartridges, random access memories (RAMs), read only memory (ROM), and/or some combination of these devices.

[0448] The storage device 1630 can include software services, servers, services, etc., that when the code that defines such software is executed by the processor 1610, it causes the system to perform a function. In some embodiments, a hardware service that performs a particular function can include the software component stored in a computer-readable medium in connection with the necessary hardware components, such as processor 1610, connection 1605, output device 1635, etc., to carry out the function.

[0449] For clarity of explanation, in some instances the present technology may be presented as including individual functional blocks including functional blocks comprising devices, device components, steps or routines in a method embodied in software, or combinations of hardware and software.

[0450] Any of the steps, operations, functions, or processes described herein may be performed or implemented by a combination of hardware and software services or services, alone or in combination with other devices. In some embodiments, a service

can be software that resides in memory of a client device and/or one or more servers of a content management system and perform one or more functions when a processor executes the software associated with the service. In some embodiments, a service is a program, or a collection of programs that carry out a specific function. In some embodiments, a service can be considered a server. The memory can be a non-transitory computer-readable medium.

[0451] In some embodiments the computer-readable storage devices, mediums, and memories can include a cable or wireless signal containing a bit stream and the like. However, when mentioned, non-transitory computer-readable storage media expressly exclude media such as energy, carrier signals, electromagnetic waves, and signals per se.

[0452] Methods according to the above-described examples can be implemented using computer-executable instructions that are stored or otherwise available from computer readable media. Such instructions can comprise, for example, instructions and data which cause or otherwise configure a general purpose computer, special purpose computer, or special purpose processing device to perform a certain function or group of functions. Portions of computer resources used can be accessible over a network. The computer executable instructions may be, for example, binaries, intermediate format instructions such as assembly language, firmware, or source code. Examples of computer-readable media that may be used to store instructions, information used, and/or information created during methods according to described examples include magnetic or optical disks, solid state memory devices, flash memory, USB devices provided with non-volatile memory, networked storage devices, and so on.

[0453] Devices implementing methods according to these disclosures can comprise hardware, firmware and/or software, and can take any of a variety of form factors. Typical examples of such form factors include servers, laptops, smart phones, small form factor personal computers, personal digital assistants, and so on. Functionality described herein also can be embodied in peripherals or add-in cards. Such functionality can also be implemented on a circuit board among different chips or different processes executing in a single device, by way of further example.

[0454] The instructions, media for conveying such instructions, computing resources for executing them, and other structures for supporting such computing resources are means for providing the functions described in these disclosures.

[0455] Although a variety of examples and other information was used to explain aspects within the scope of the appended claims, no limitation of the claims should be implied based on particular features or arrangements in such examples, as one of ordinary skill would be able to use these examples to derive a wide variety of implementations. Further and although some subject matter may have been described in language specific to examples of structural features and/or method steps, it is to be understood that the subject matter defined in the appended claims is not necessarily limited to these described features or acts. For example, such functionality can be distributed differently or performed in components other than those identified herein. Rather, the described features and steps are disclosed as examples of components of systems and methods within the scope of the appended claims.

CLAIMS

What is claimed is:

1. A method comprising:
 - adding a content item at a client device, the content item being associated with a user account at a content management system;
 - determining, via a client application at the client device, that the content item does not have a unique identifier from the content management system, the client application being associated with the content management system;
 - obtaining the unique identifier for the content item, the unique identifier being defined by the content management system;
 - sending, by the client application to the content management system, a request to verify a uniqueness of the unique identifier at the content management system;
 - based on the request, determining, by the client application, whether the unique identifier is verifiably unique at the content management system;
 - when the unique identifier is verifiably unique at the content management system:
 - adding a node representing the content item to a local tree at the client device, the node comprising the unique identifier and the local tree representing a filesystem state at the client device for content items associated with the user account;
 - and
 - uploading the content item with the unique identifier to the content management system;
 - when the unique identifier is not verifiably unique at the content management system, obtaining a new unique identifier for the content item.

2. The method of claim 1, further comprising:
when the unique identifier is not verifiably unique at the content management system, removing the unique identifier from the content item; and
after removing the unique identifier from the content item, obtaining the new unique identifier for the content item.
3. The method of claim 1, wherein obtaining the unique identifier comprises determining whether the content item should inherit the unique identifier from a second content item previously associated with at least one of a path corresponding to the content item or a name corresponding to the content item.
4. The method of claim 3, further comprising:
determining that the second content item was previously located at the path corresponding to the content item; and
in response to determining that the second content item was previously located at the path corresponding to the content item, determining that the content item should inherit the unique identifier from the second content item.
5. The method of claim 4, further comprising:
in response to determining that the content item should inherit the unique identifier from the second content item, reassigning the unique identifier from the second content item to the content item.
6. The method of claim 5, wherein reassigning the unique identifier from the second content item to the content item comprises:

determining that the unique identifier is assigned to the second content item as a respective unique identifier of the second content item;

extracting the unique identifier from the second content item; and

associating the unique identifier with the content item.

7. The method of claim 6, further comprising appending the unique identifier to the content item as an extended attribute.

8. The method of claim 1, wherein obtaining the unique identifier comprises:
requesting, via the client application, the unique identifier for the content item from the content management system;

receiving, via the client application, the unique identifier for the content item from the content management system; and

associating, via the client application, the unique identifier with the content item.

9. The method of claim 8, wherein associating the unique identifier with the content item comprises writing the unique identifier to the content item as an extended attribute of the content item.

10. The method of claim 1, wherein obtaining the unique identifier comprises:
identifying a path associated with the content item;
determining a previous path associated with a second content item matches the path associated with the content item;

in response to determining the previous path associated with the second content item matches the path associated with the content item, determining that the content item comprises a version of the second content item being resurrected or moved to the path associated with the content item;

based on determining that the content item comprises the version of the second content item, determining that the content item should inherit the unique identifier from the second content item; and

assigning the unique identifier from the second content item to the content item.

11. A non-transitory computer readable medium comprising instructions, the instructions, when executed by one or more processors, cause a client device to:

store a content item associated with a user account at a content management system;

determine, via a client application, that the content item does not have a unique identifier from the content management system, the client application being associated with the content management system;

obtain the unique identifier for the content item, the unique identifier being defined by the content management system;

send, by the client application to the content management system, a request to verify a uniqueness of the unique identifier at the content management system;

based on the request, determine, by the client application, whether the unique identifier is verifiably unique at the content management system;

when the unique identifier is verifiably unique at the content management system:

add a node representing the content item to a local tree at the client device, the node comprising the unique identifier and the local tree representing a filesystem state at the client device for content items associated with the user account; and

upload the content item with the unique identifier to the content management system;

when the unique identifier is not verifiably unique at the content management system, obtain a new unique identifier for the content item.

12. The non-transitory computer readable medium of claim 11, storing instructions which, when executed by the one or more processors, cause the client device to:

when the unique identifier is not verifiably unique at the content management system, remove the unique identifier from the content item; and

after removing the unique identifier from the content item, obtain the new unique identifier for the content item.

13. The non-transitory computer readable medium of claim 11, wherein obtaining the unique identifier comprises determining whether the content item should inherit the unique identifier from a second content item previously associated with at least one of a path corresponding to the content item or a name corresponding to the content item.

14. The non-transitory computer readable medium of claim 11, storing instructions which, when executed by the one or more processors, cause the client device to:

determine that the second content item was previously located at the path corresponding to the content item; and

in response to determining that the second content item was previously located at the path corresponding to the content item, determine that the content item should inherit the unique identifier from the second content item; and

in response to determining that the content item should inherit the unique identifier from the second content item, reassign the unique identifier from the second content item to the content item.

15. The non-transitory computer readable medium of claim 11, wherein obtaining the unique identifier comprises:

identifying a path associated with the content item;

determining a previous path associated with a second content item matches the path associated with the content item;

in response to determining the previous path associated with the second content item matches the path associated with the content item, determining that the content item comprises a version of the second content item being resurrected or moved to the path associated with the content item;

based on determining that the content item comprises the version of the second content item, determining that the content item should inherit the unique identifier from the second content item; and

reassigning the unique identifier from the second content item to the content item, the reassigning comprising associating the unique identifier with the content item as an extended attribute of the content item.

16. A computing device comprising:

one or more processors; and

at least one non-transitory computer readable medium having stored therein instructions which, when executed by the one or more processors, cause the computing device to:

store a content item associated with a user account at a content management system;

determine, via a client application, that the content item does not have a unique identifier from the content management system, the client application being associated with the content management system;

obtain the unique identifier for the content item, the unique identifier being defined by the content management system;

send, by the client application to the content management system, a request to verify a uniqueness of the unique identifier at the content management system;

based on the request, determine, by the client application, whether the unique identifier is verifiably unique at the content management system;

when the unique identifier is verifiably unique at the content management system:

add a node representing the content item to a local tree at the client device, the node comprising the unique identifier and the local tree representing

a filesystem state at the computing device for content items associated with the user account; and

upload the content item with the unique identifier to the content management system;

when the unique identifier is not verifiably unique at the content management system, obtain a new unique identifier for the content item.

17. The computing device of claim 16, storing instructions which, when executed by the one or more processors, cause the computing device to:

when the unique identifier is not verifiably unique at the content management system, strip the unique identifier from the content item; and

after stripping the unique identifier from the content item, obtain the new unique identifier for the content item.

18. The computing device of claim 16, wherein obtaining the unique identifier comprises:

determining that a second content item was previously located at a path associated with the content item on the computing device;

in response to determining that the second content item was previously located at the path associated with the content item, determining that the content item should inherit a respective identifier assigned to the second content item, the respective identifier comprising the unique identifier; and

in response to determining that the content item should inherit the respective identifier associated with the second content item:

obtaining the unique identifier from the second content item; and

associating the unique identifier with the content item as an attribute of the content item.

19. The computing device of claim 16, wherein obtaining the unique identifier comprises:

requesting, via the client application, the unique identifier for the content item from the content management system;

receiving, via the client application, the unique identifier for the content item from the content management system; and

appending, via the client application, the unique identifier to the content item as an extended attribute of the content item.

20. The computing device of claim 16, wherein obtaining the unique identifier comprises:

identifying a path associated with the content item;

determining the path associated with the content item comprises a previous path of a second content item;

in response to determining the path associated with the content item comprises the previous path of the second content item, determining that the content item comprises a version of the second content item being resurrected or moved to the path on the computing device;

determining, via the client application, that the second content item is associated with the unique identifier;

based on determining that the content item comprises the version of the second content item, determining that the content item should inherit the unique identifier from the second content item; and

assigning the unique identifier from the second content item to the content item.

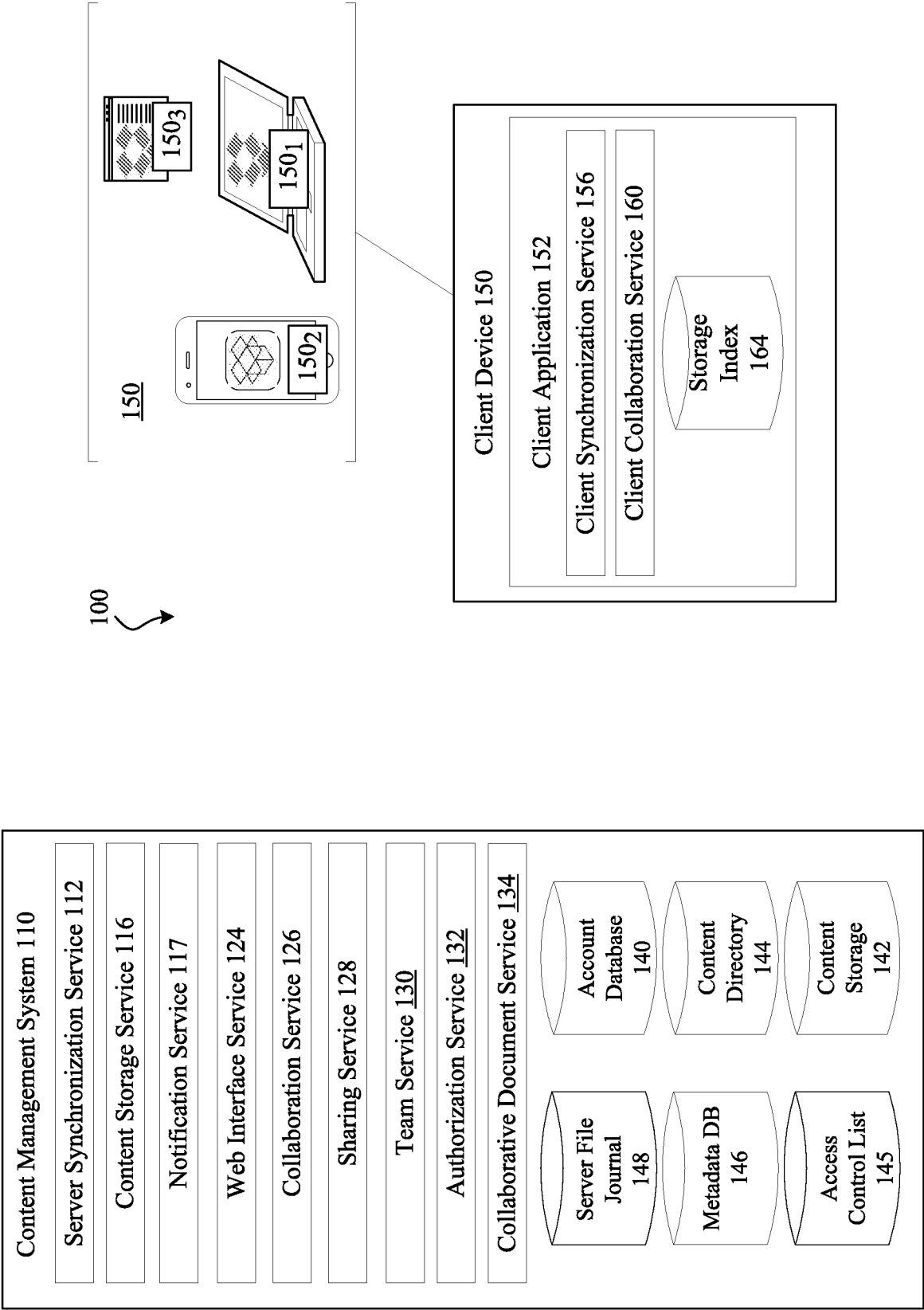


FIG. 1A

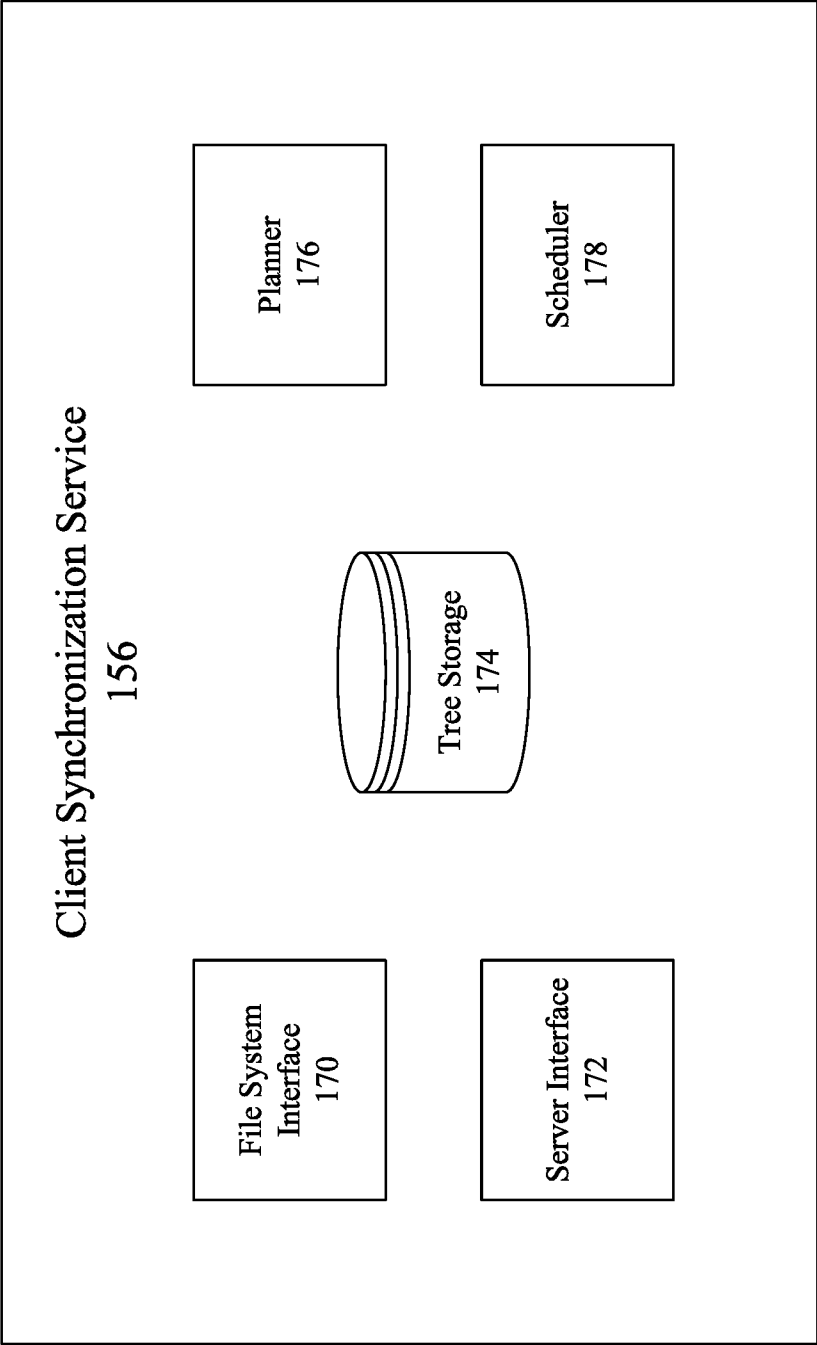


FIG. 1B

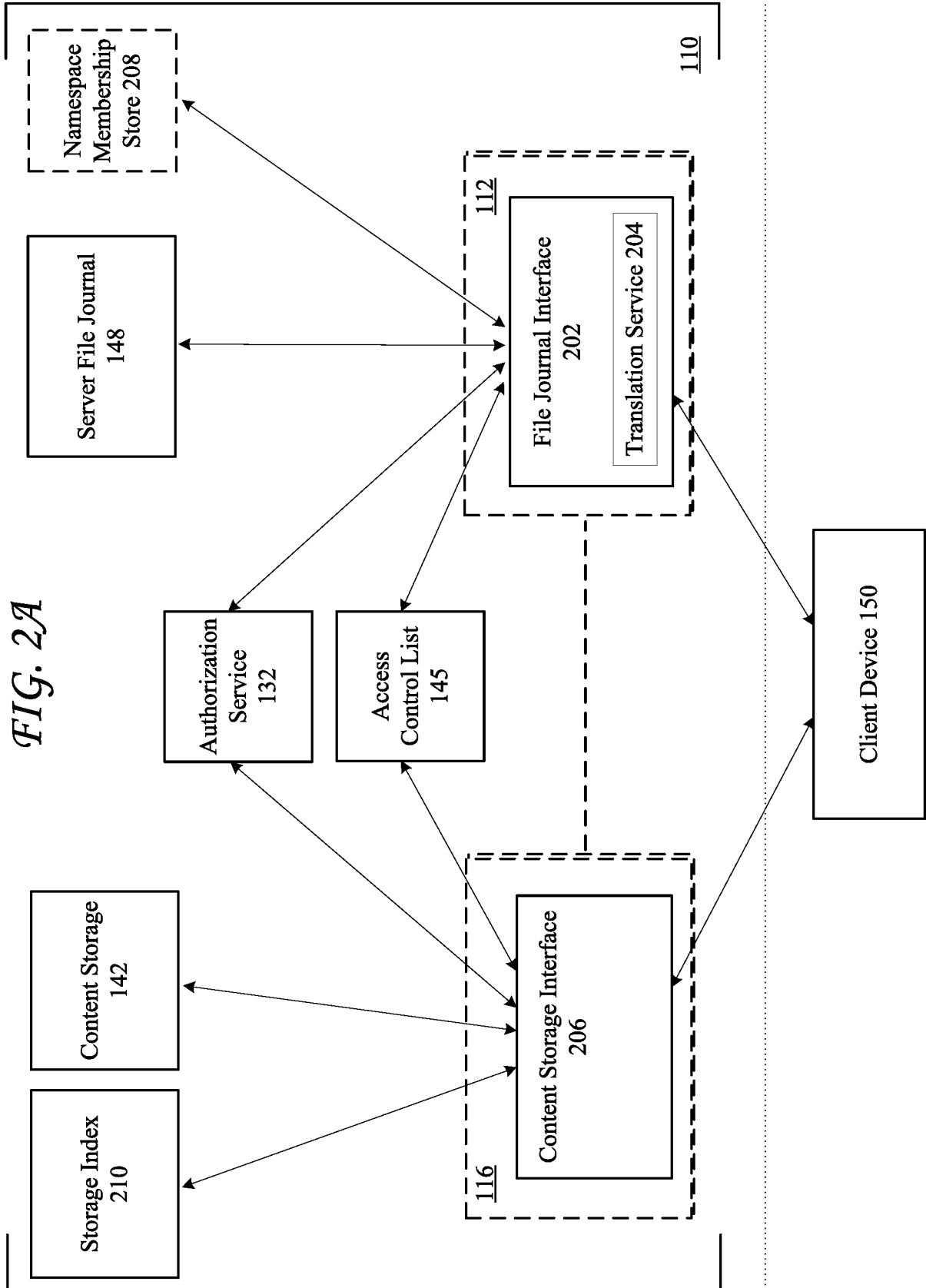
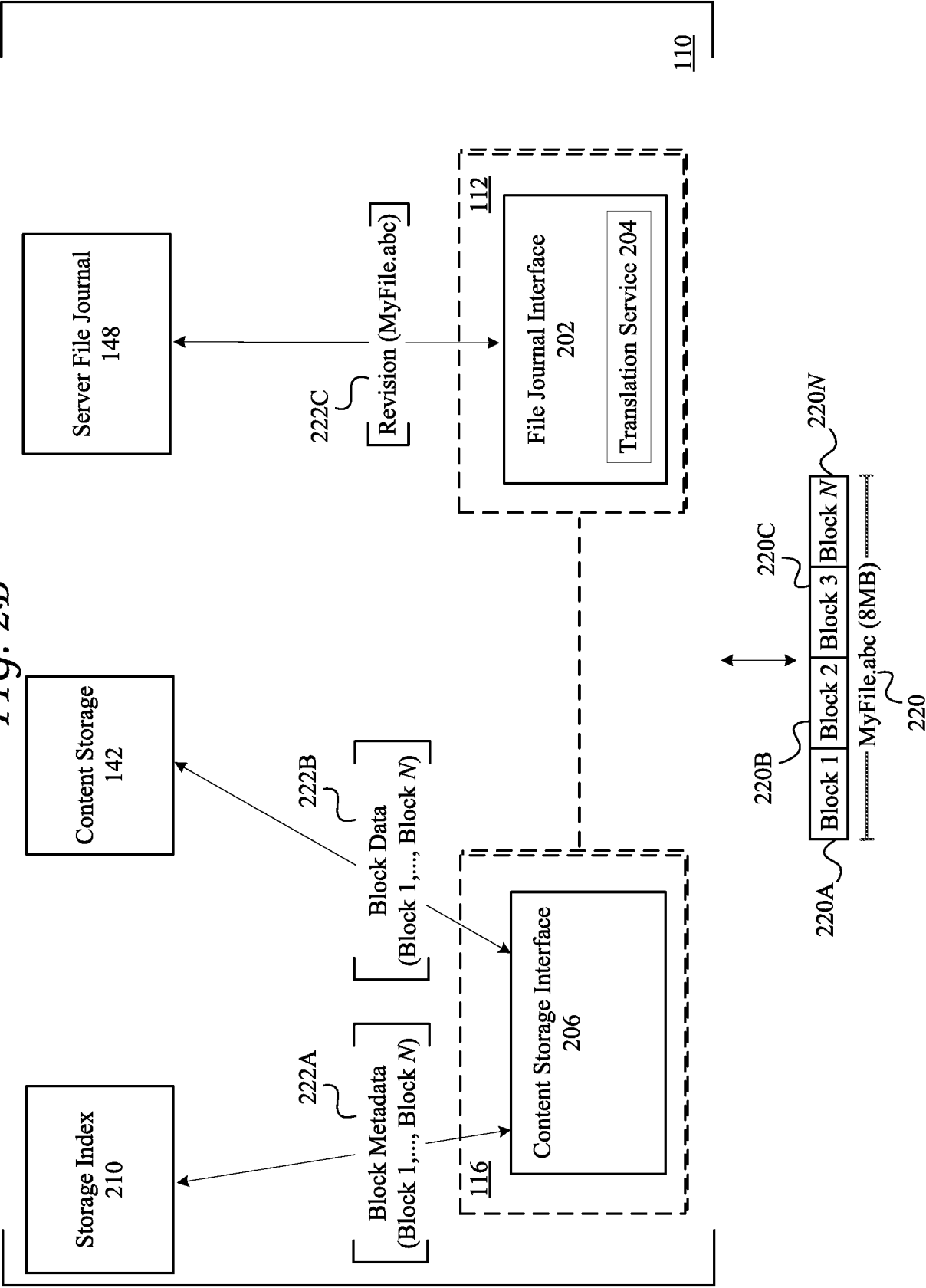


FIG. 2B



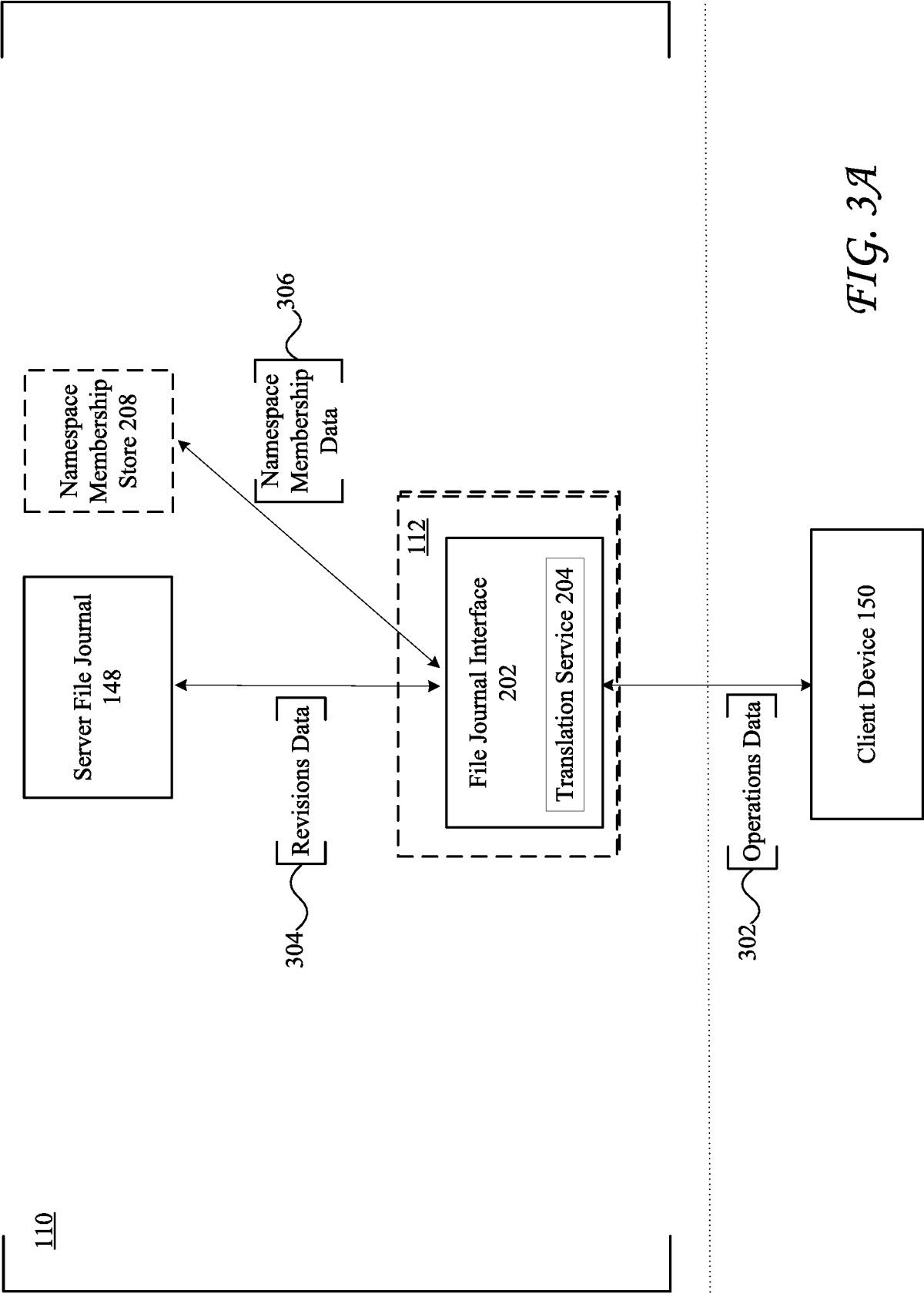


FIG. 3B

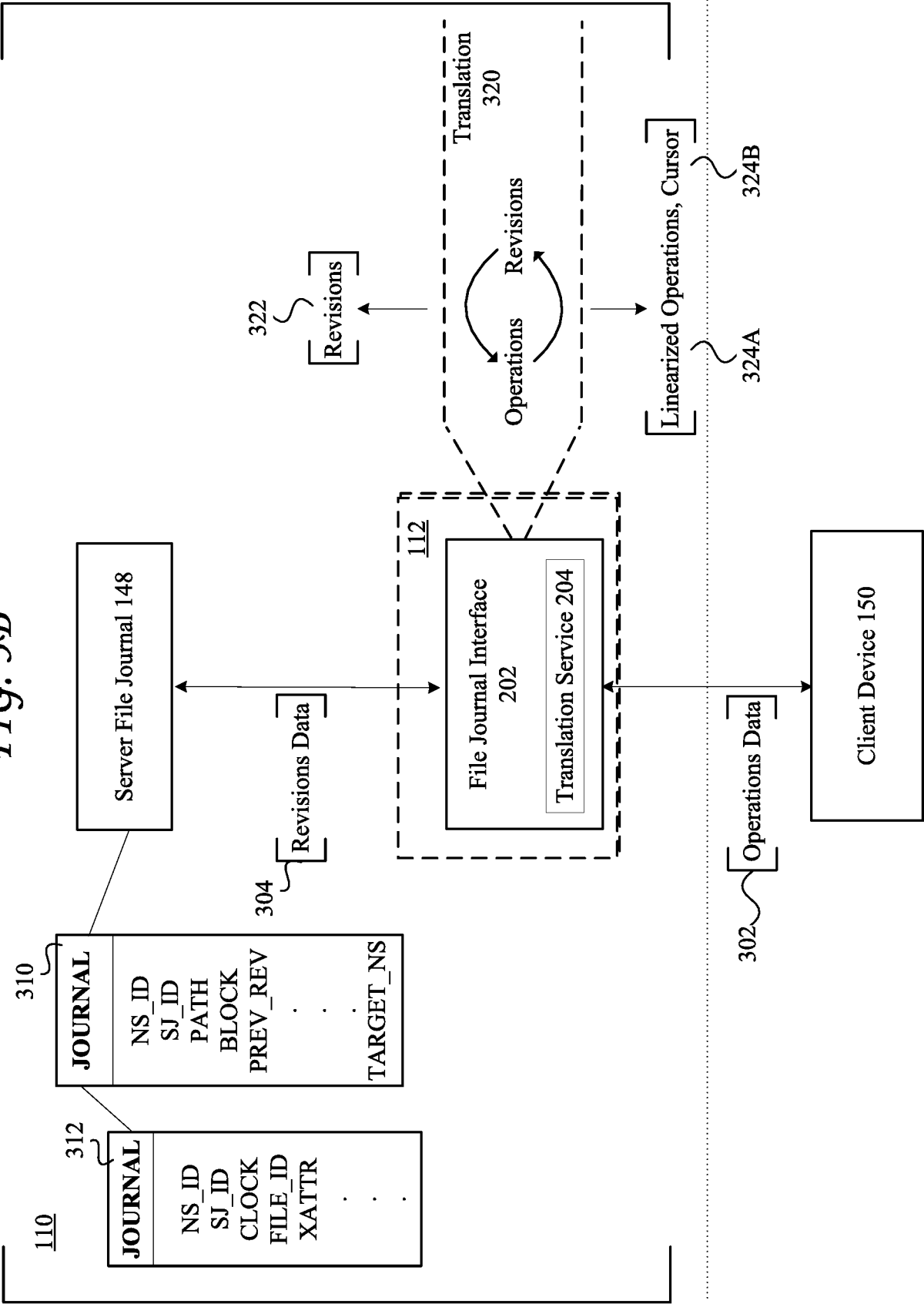
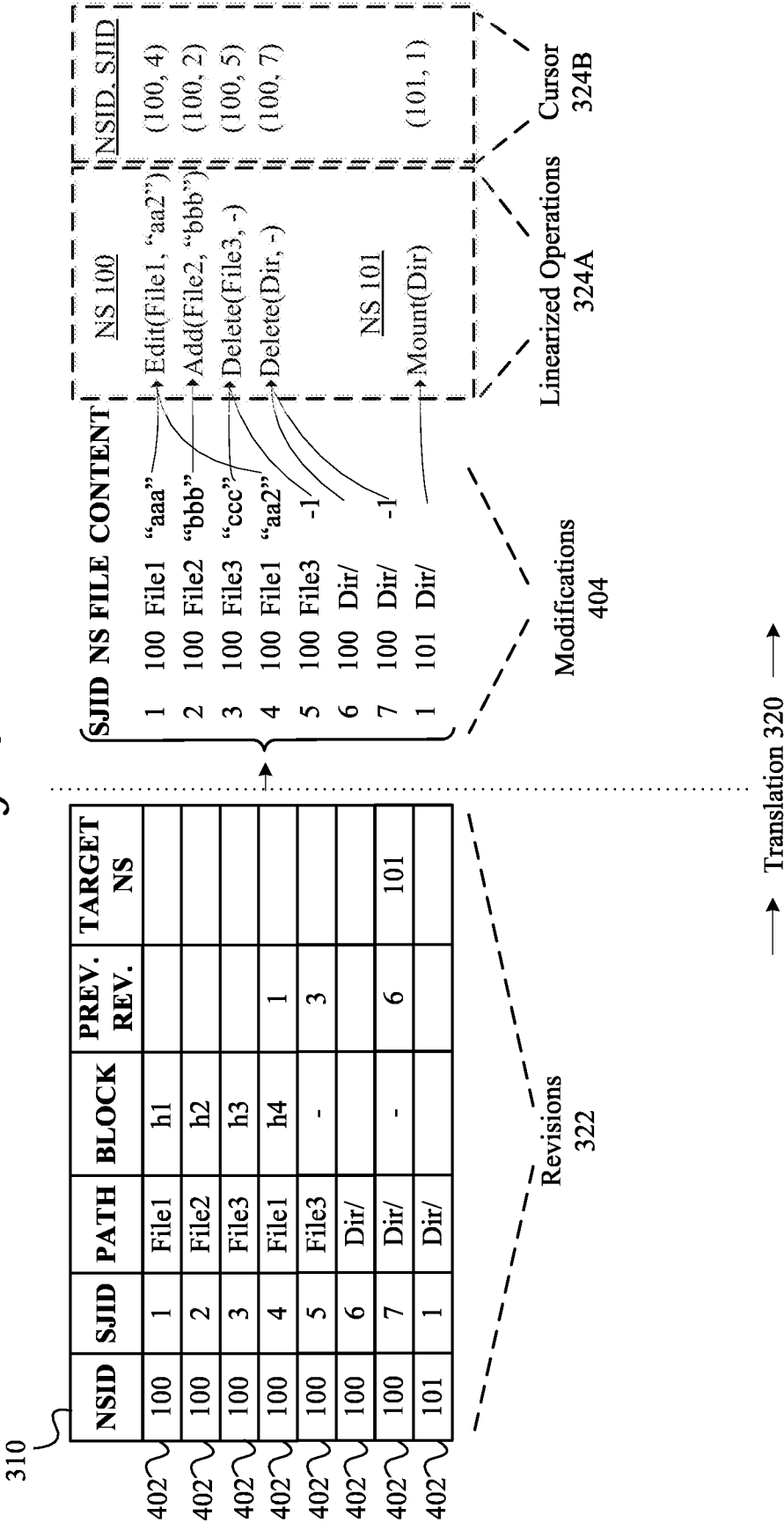


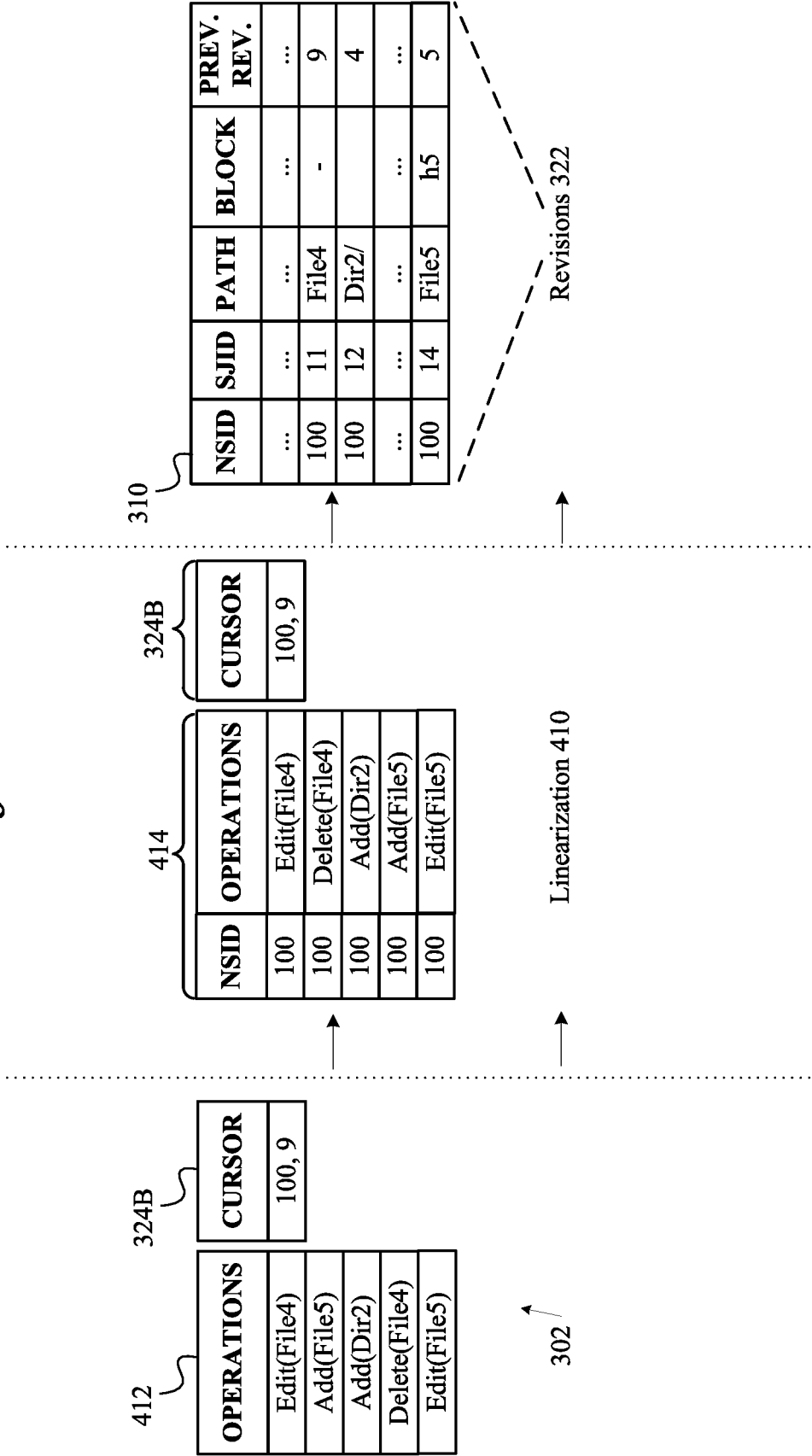
FIG. 4A



Server File Journal 148

File Journal Interface 202

FIG. 4B



Client Device 150

File Journal Interface 202

Server File Journal 148

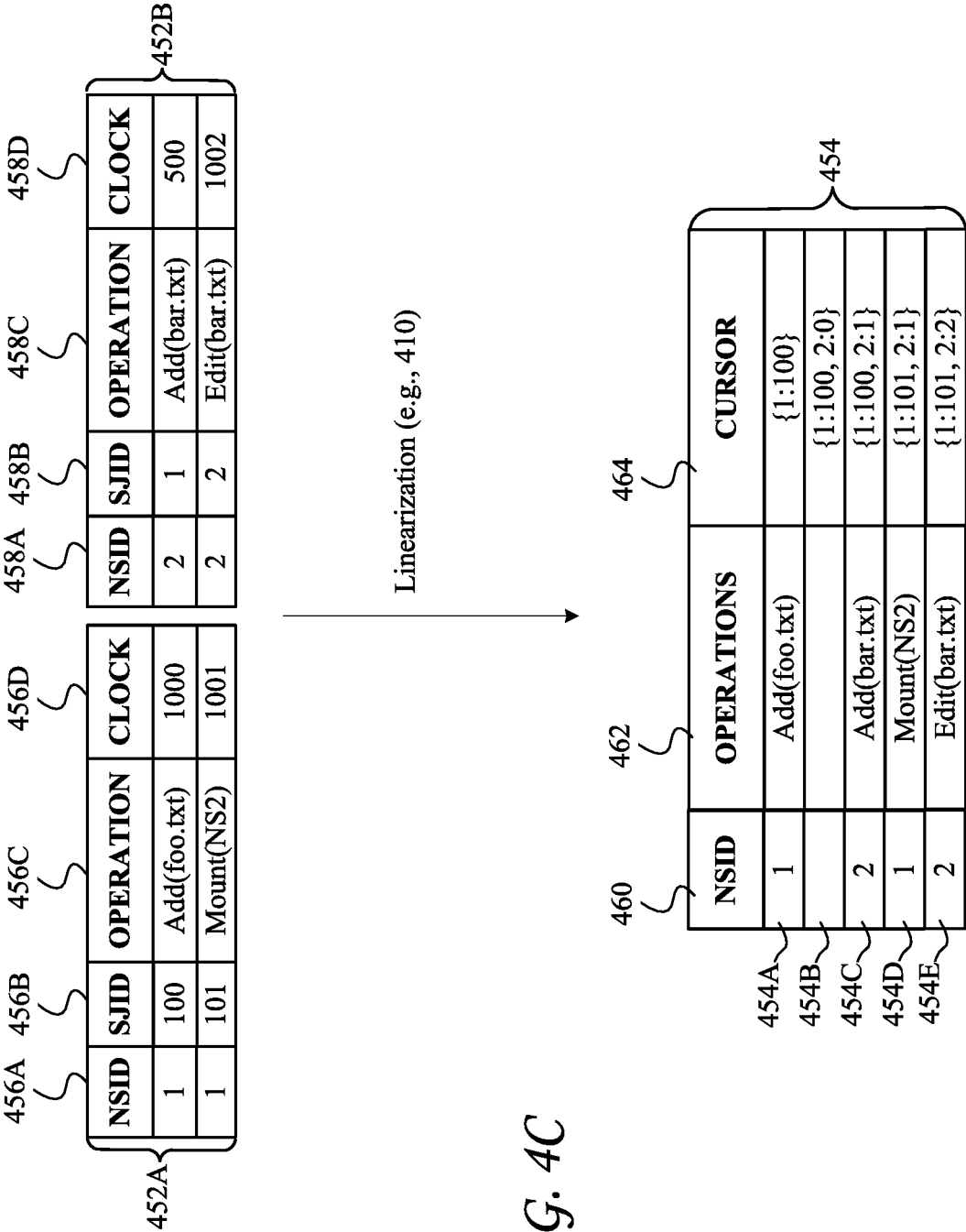


FIG. 4C

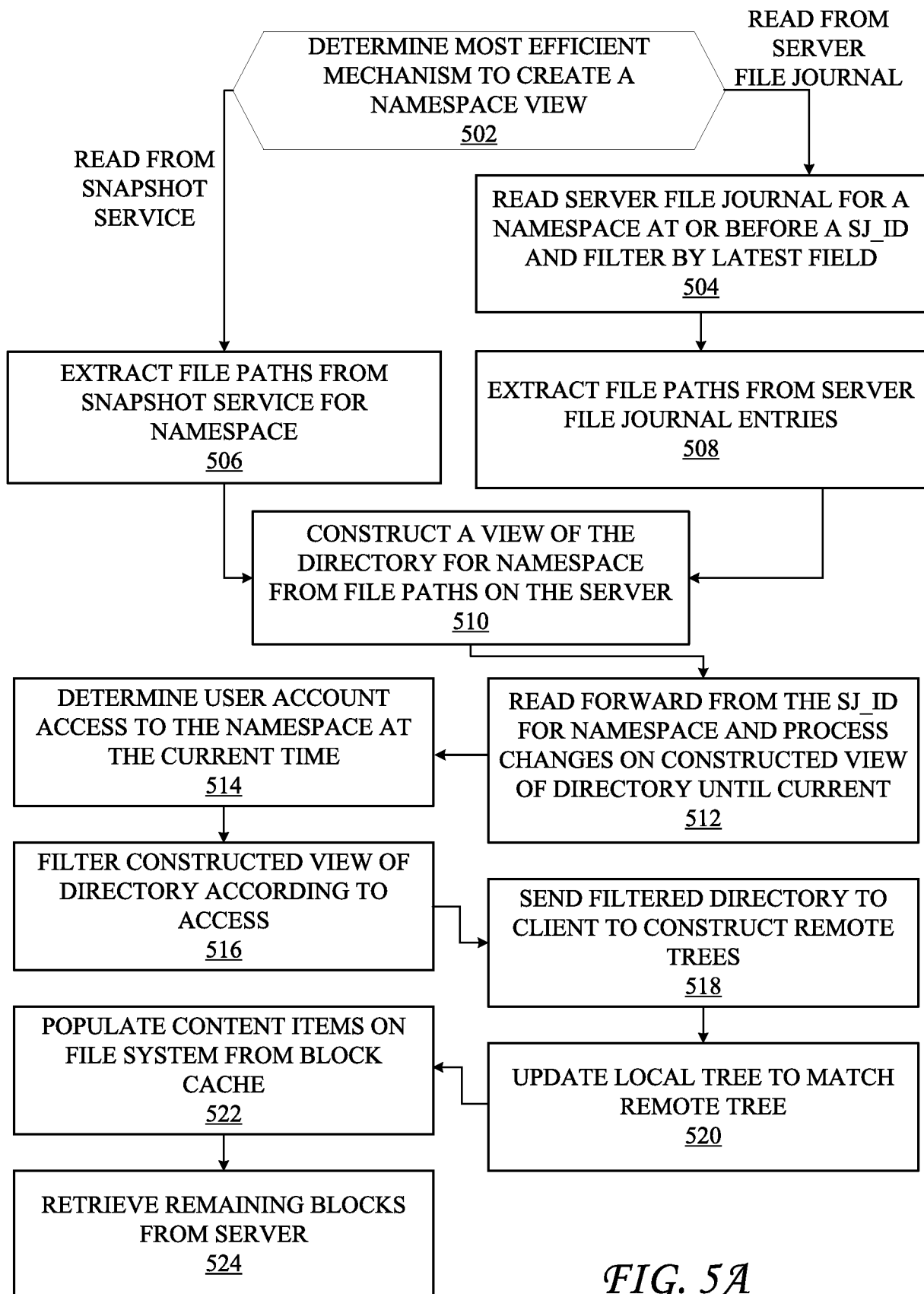


FIG. 5A

From Mount Table:

```
mounts for ns=directory
directory/user_1
directory/design
directory/user_handbook
directory/finance
directory/marketing
```

From Namespace View Builder:

```
ns=user_1
paths=
user_1/shared_with_user_Z
user_1/unshared.tx
user_1/
```

User_1's View
530

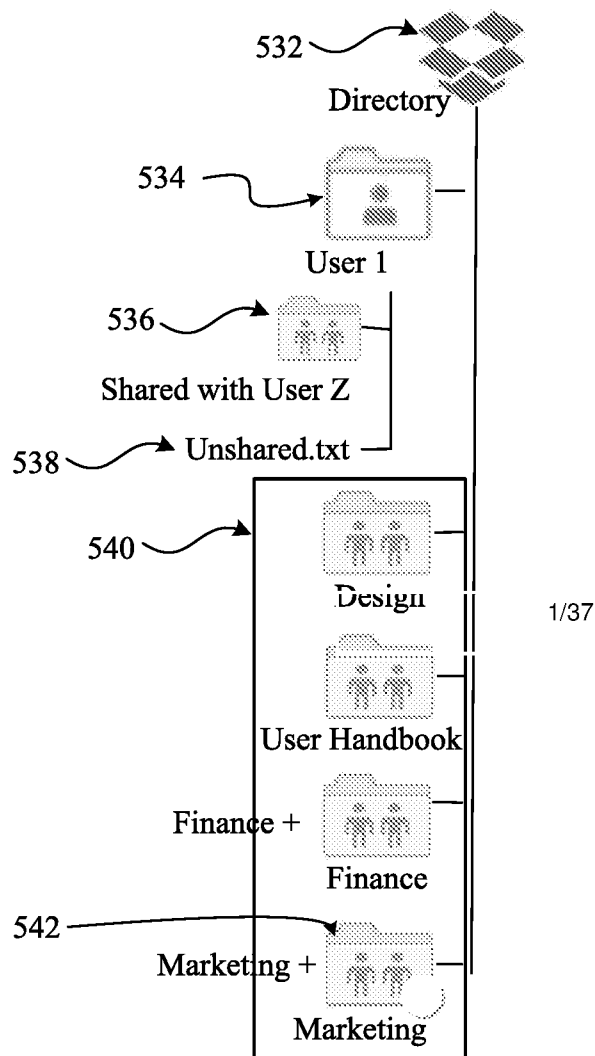


FIG. 5B

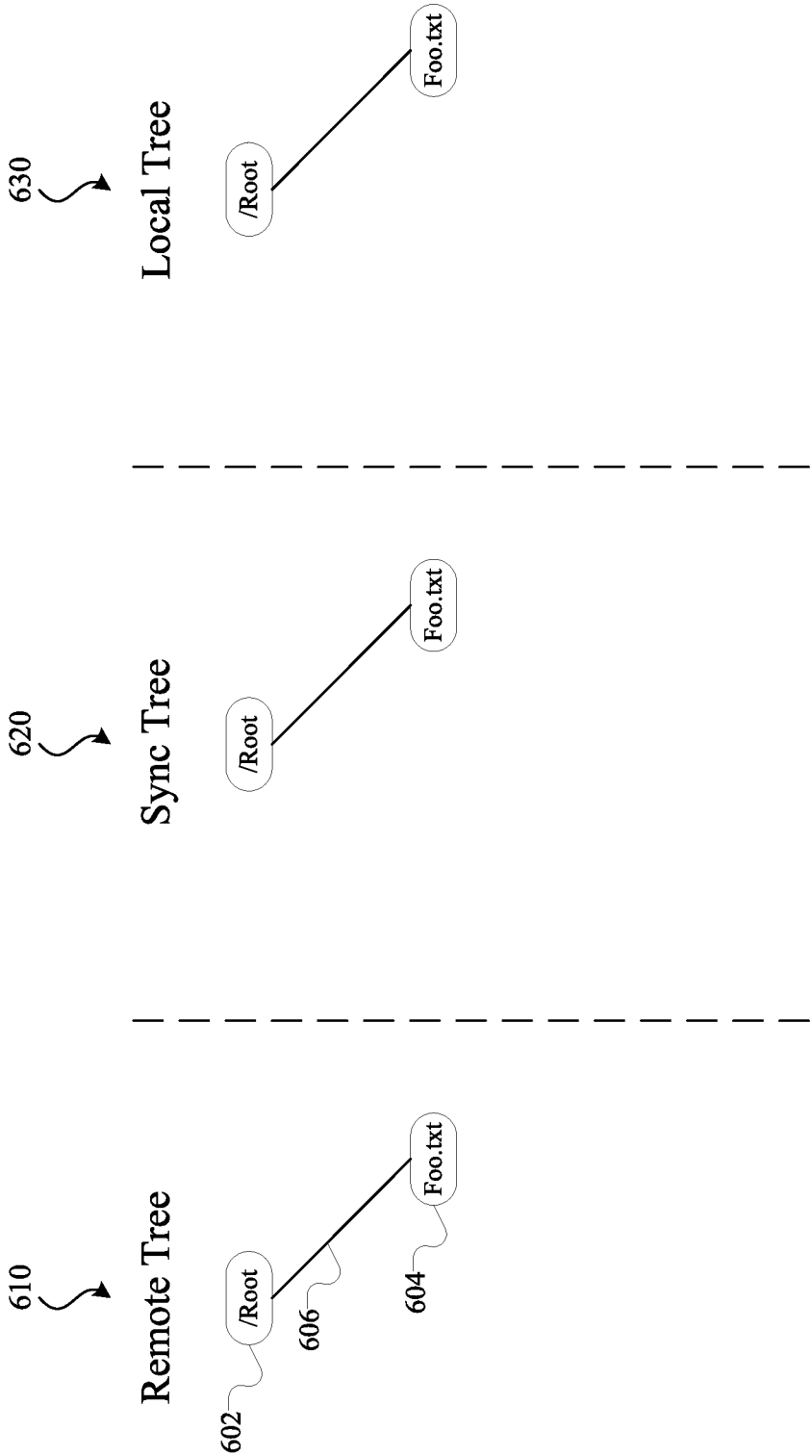


FIG. 6A

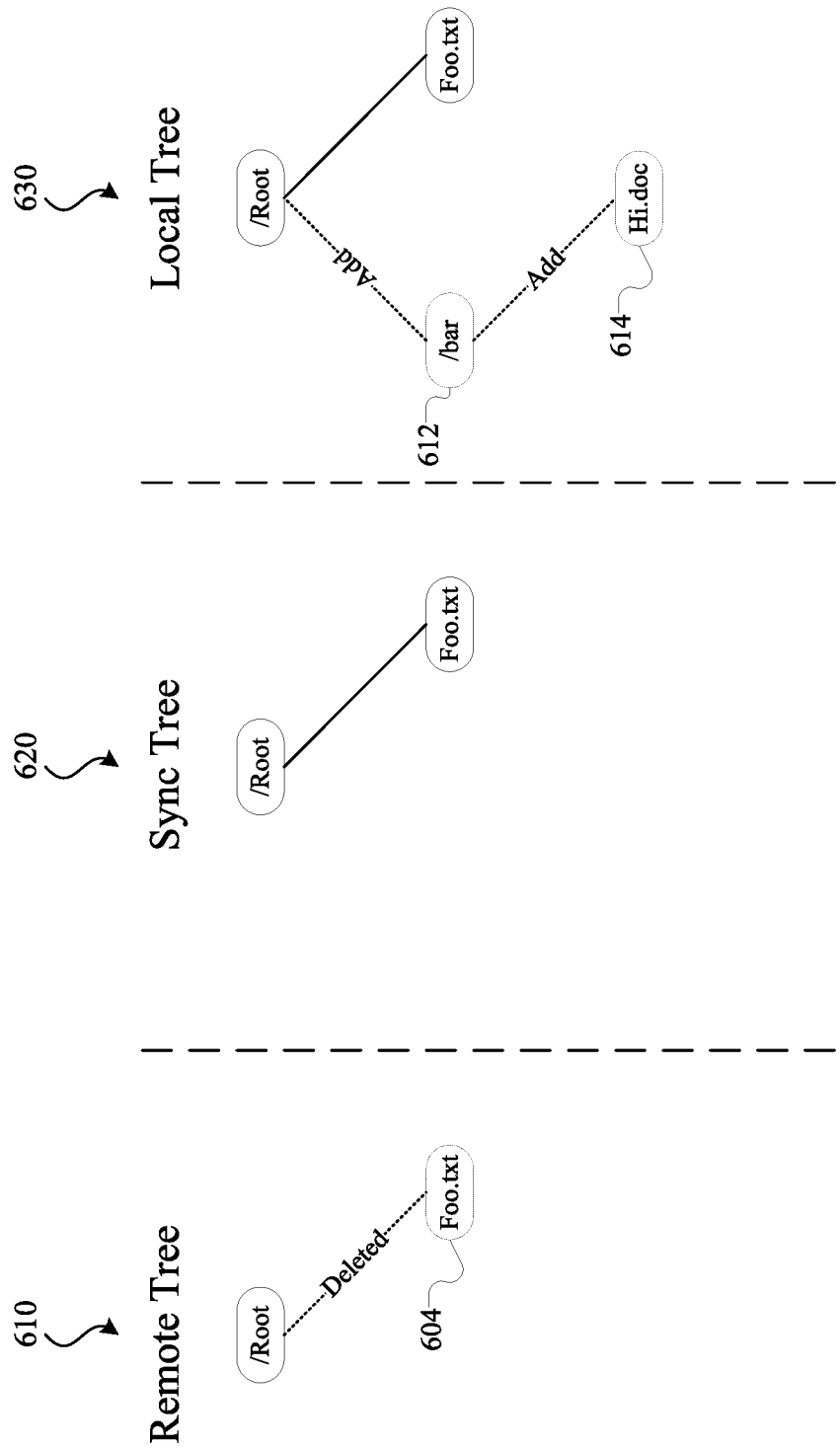


FIG. 6B

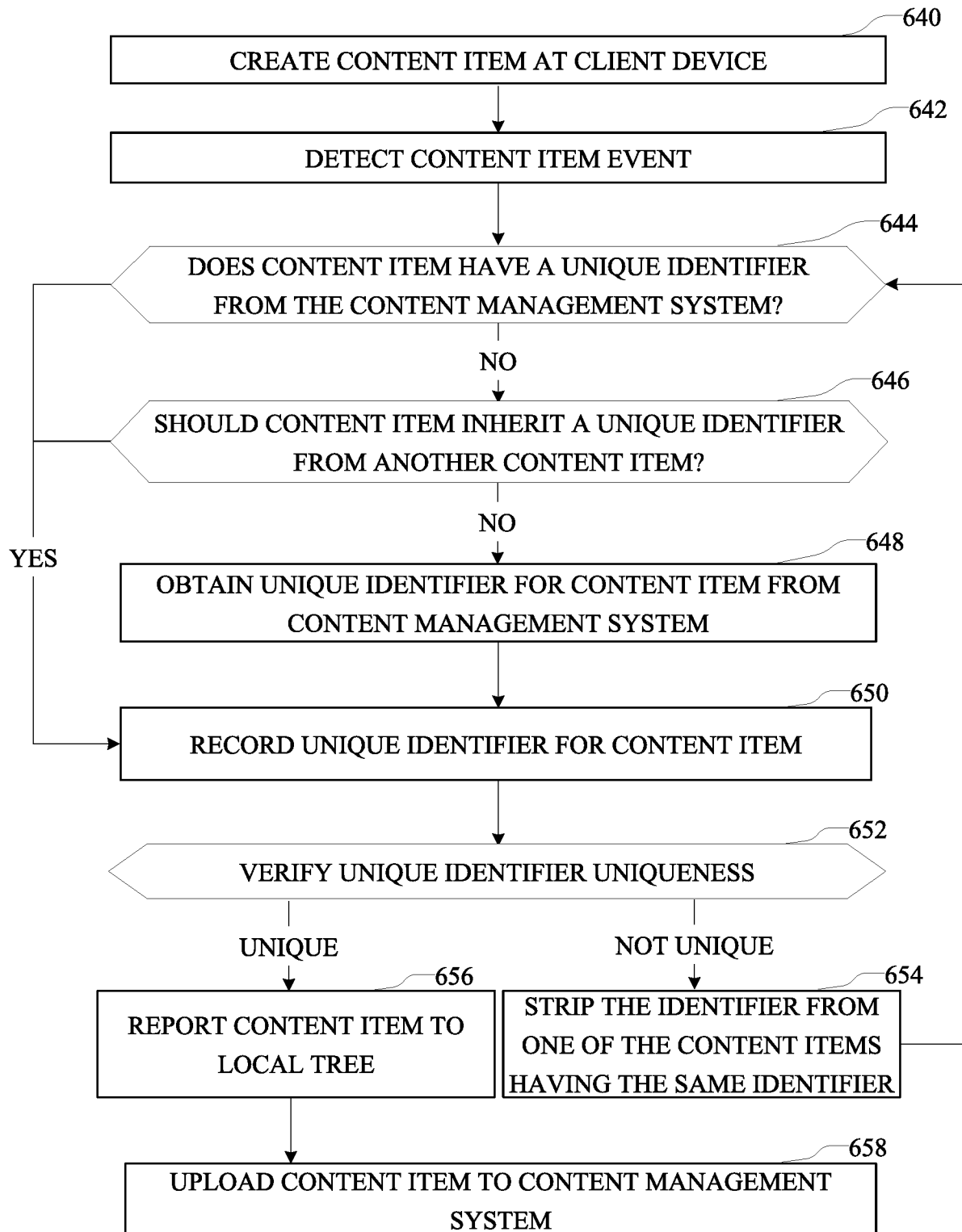


FIG. 6C

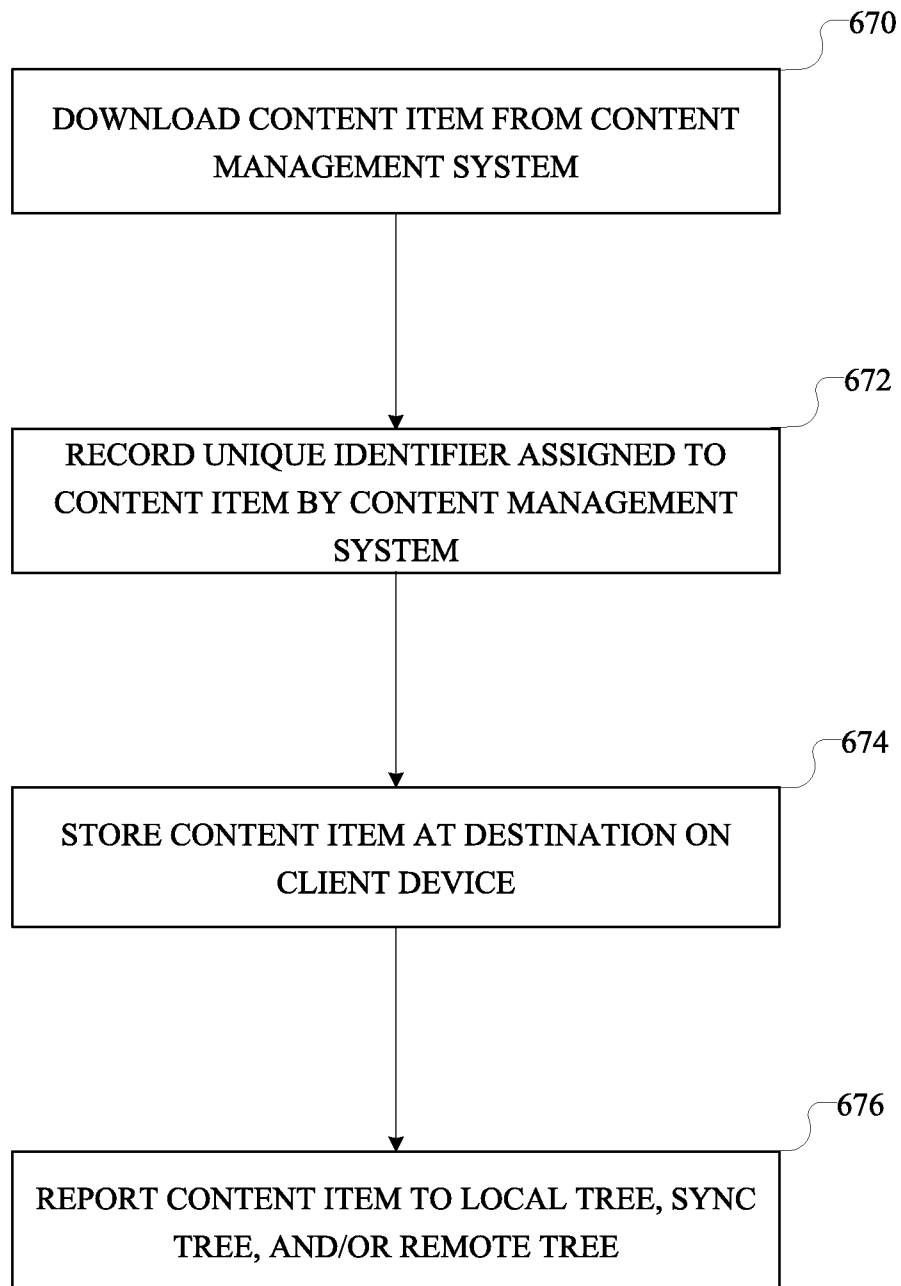
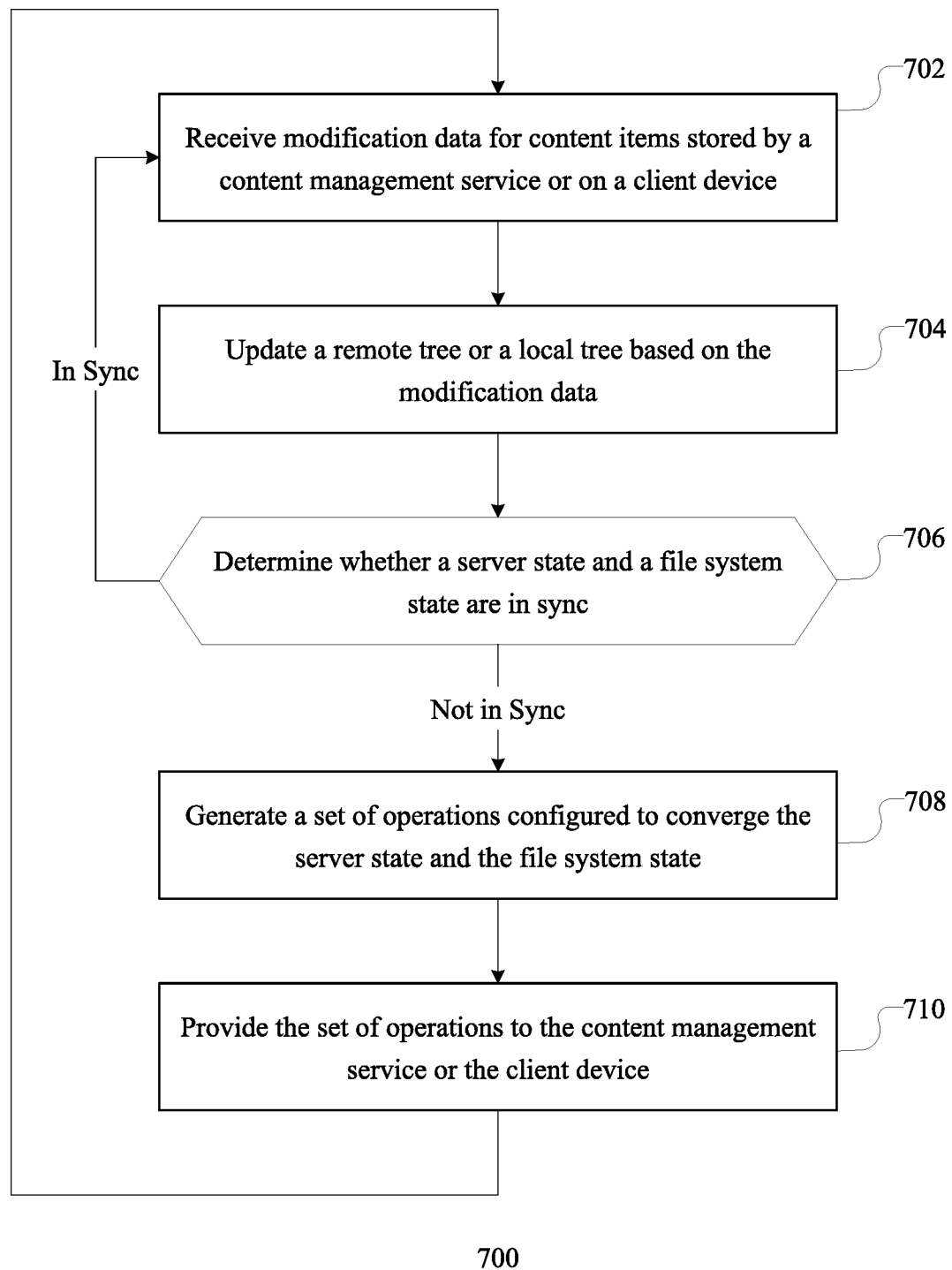
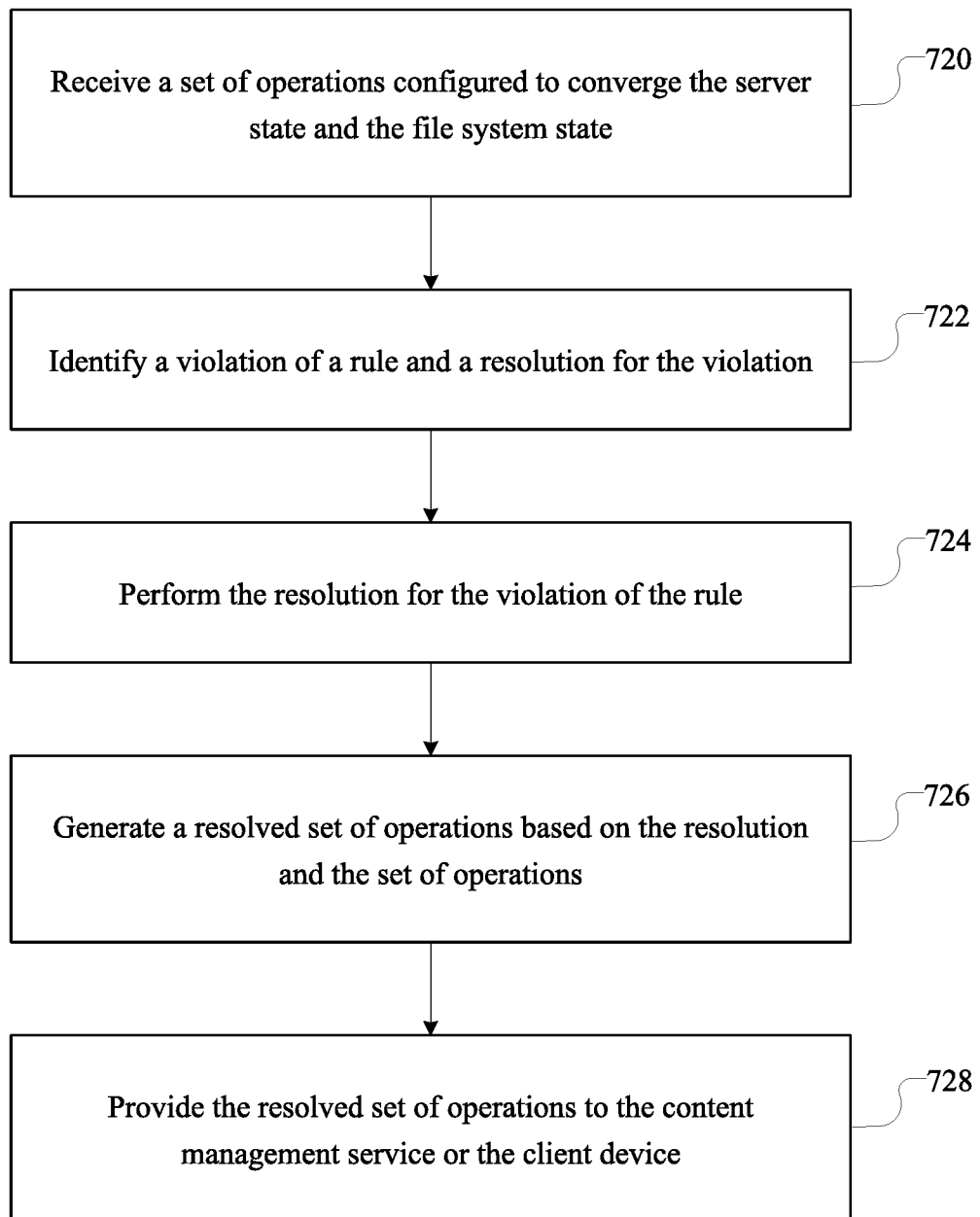


FIG. 6D

*FIG. 7A*

*FIG. 7B*

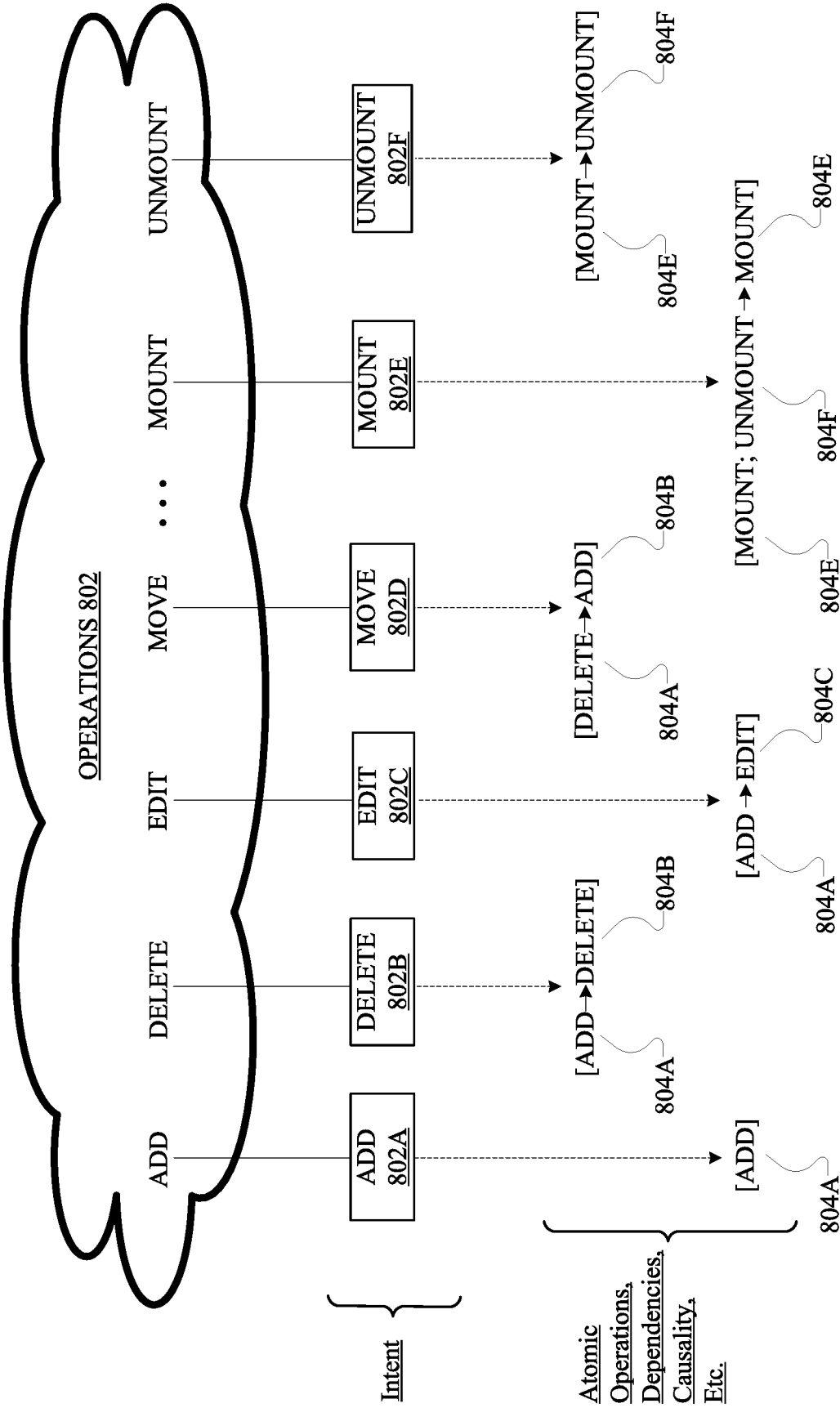
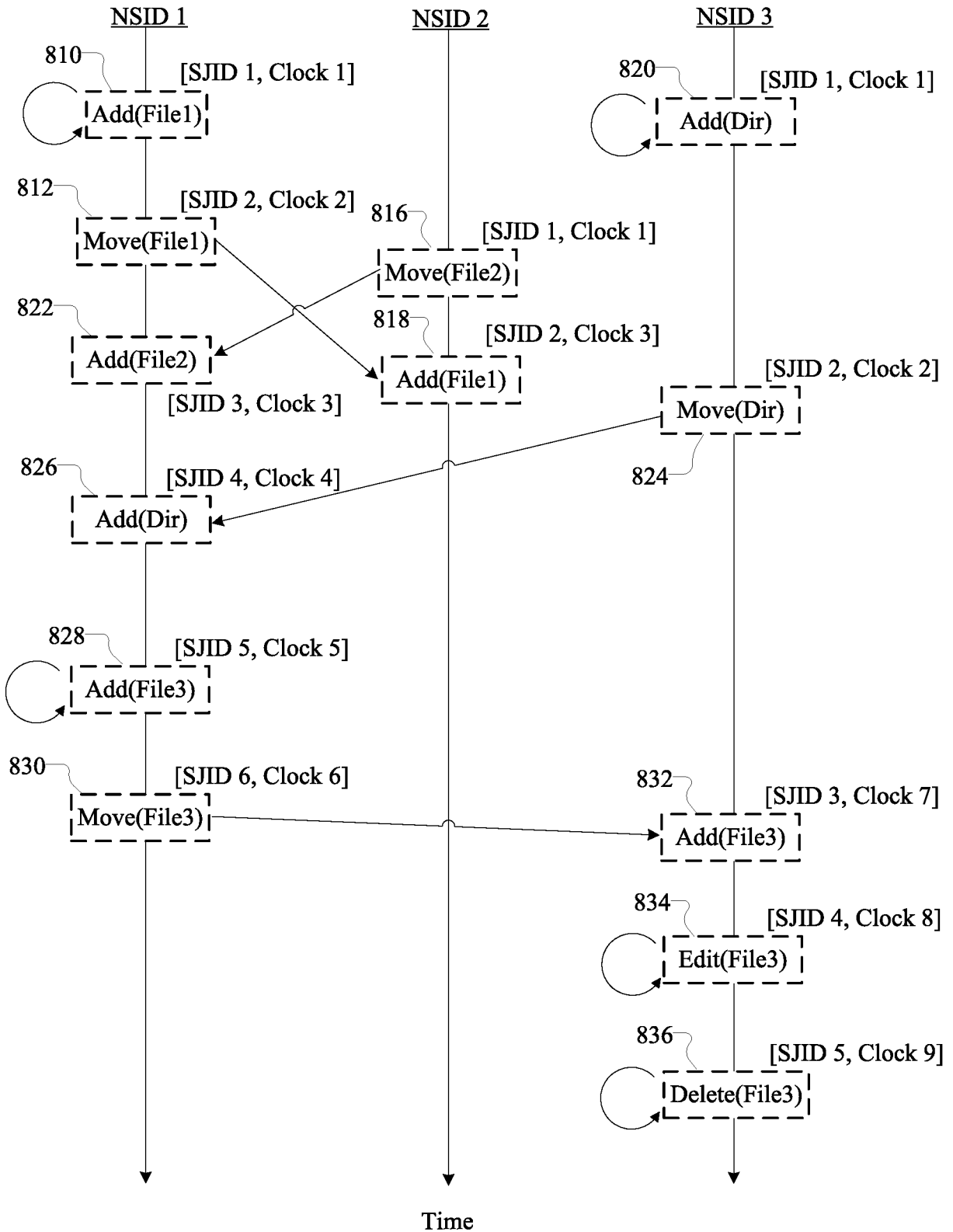


FIG. 8A

FIG. 8B



900

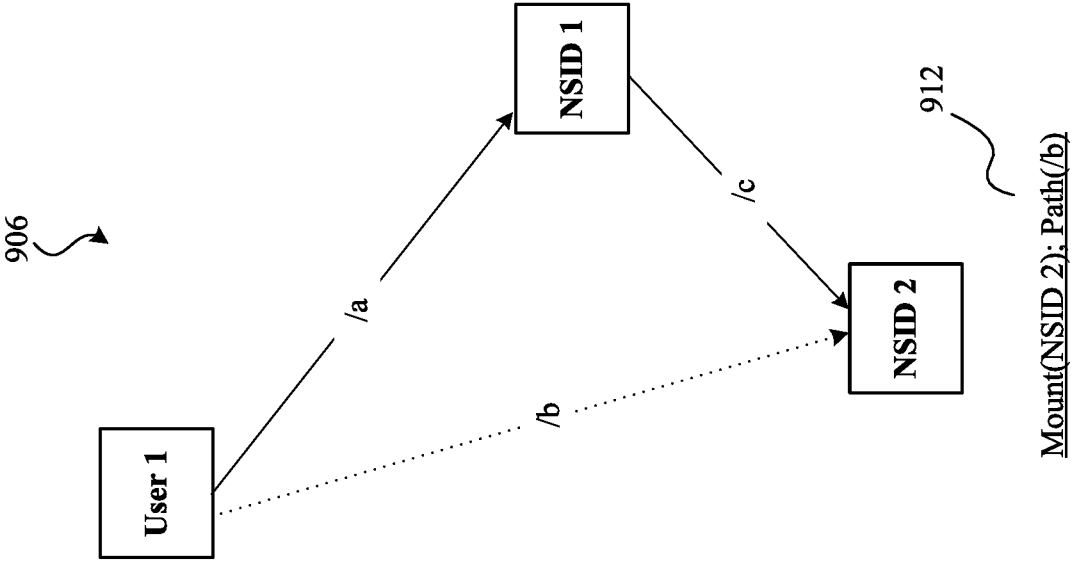
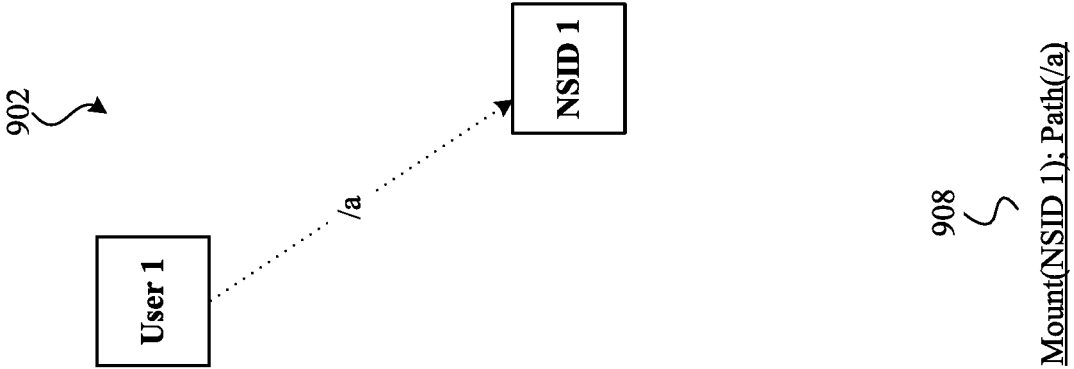
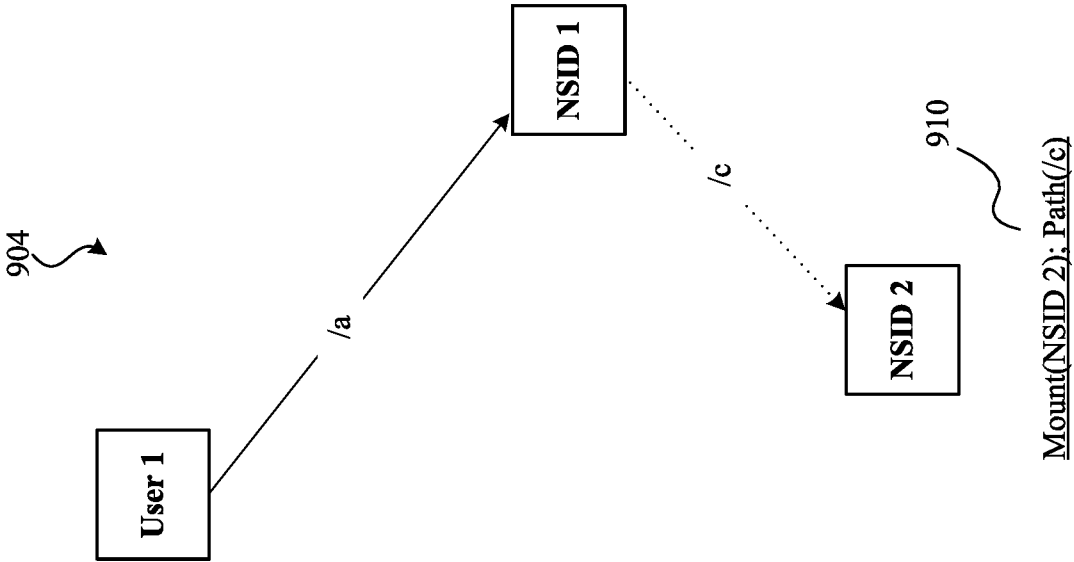


FIG. 9A



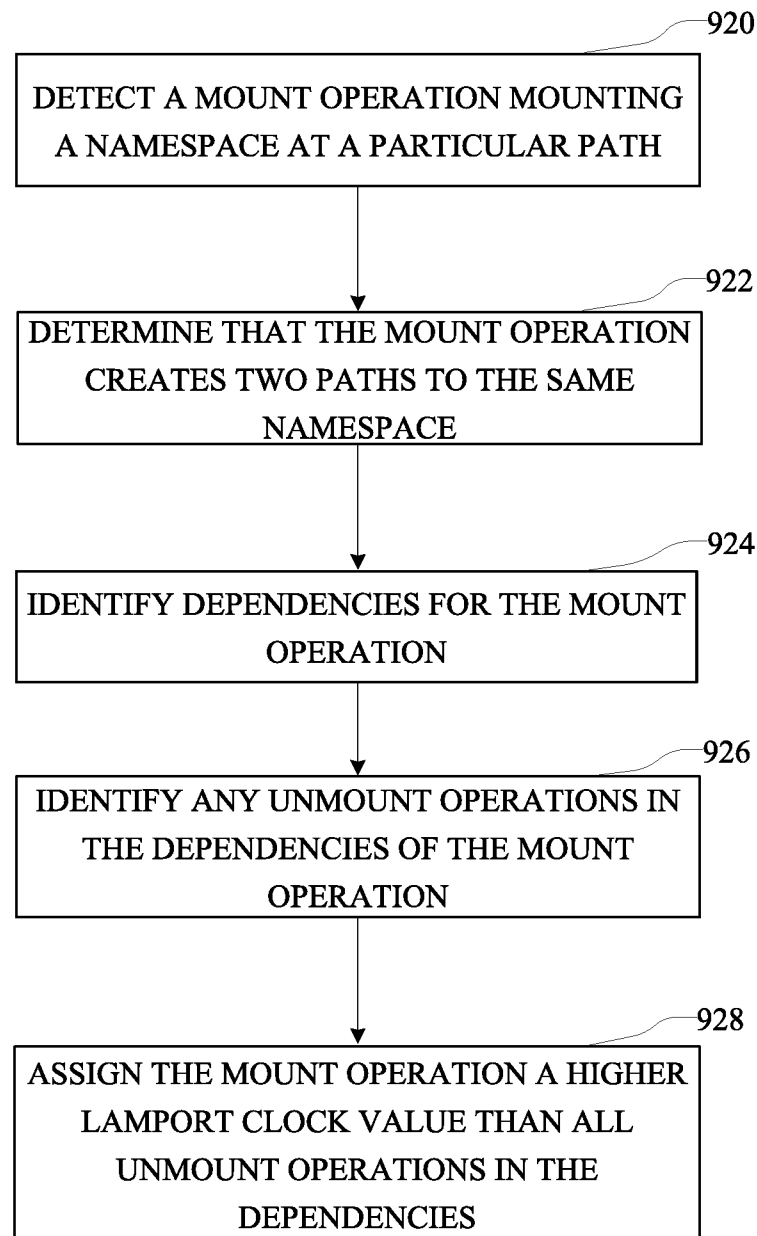
*FIG. 9B*

FIG. 9C

930

NSID 1	
OPERATION	CLOCK
Mount(/a, NS1)	1

908

910

NSID 2	
OPERATION	CLOCK
Mount(/c, NS2)	0
Unmount(/c, NS2)	1
Mount(/b, NS2)	2

932

934

936

912

914

FIG. 9D

940

NSID	OPERATIONS	CURSOR
1	Mount(/a, NS1)	{1:1}
2	Mount(/c, NS2)	{1:1, 2:0}
2	Unmount(/c, NS1)	{1:1, 2:1}
2	Mount(/b, NS1)	{1:1, 2:2}

908

910

912

914

942

944

946

948

FIG. 10A

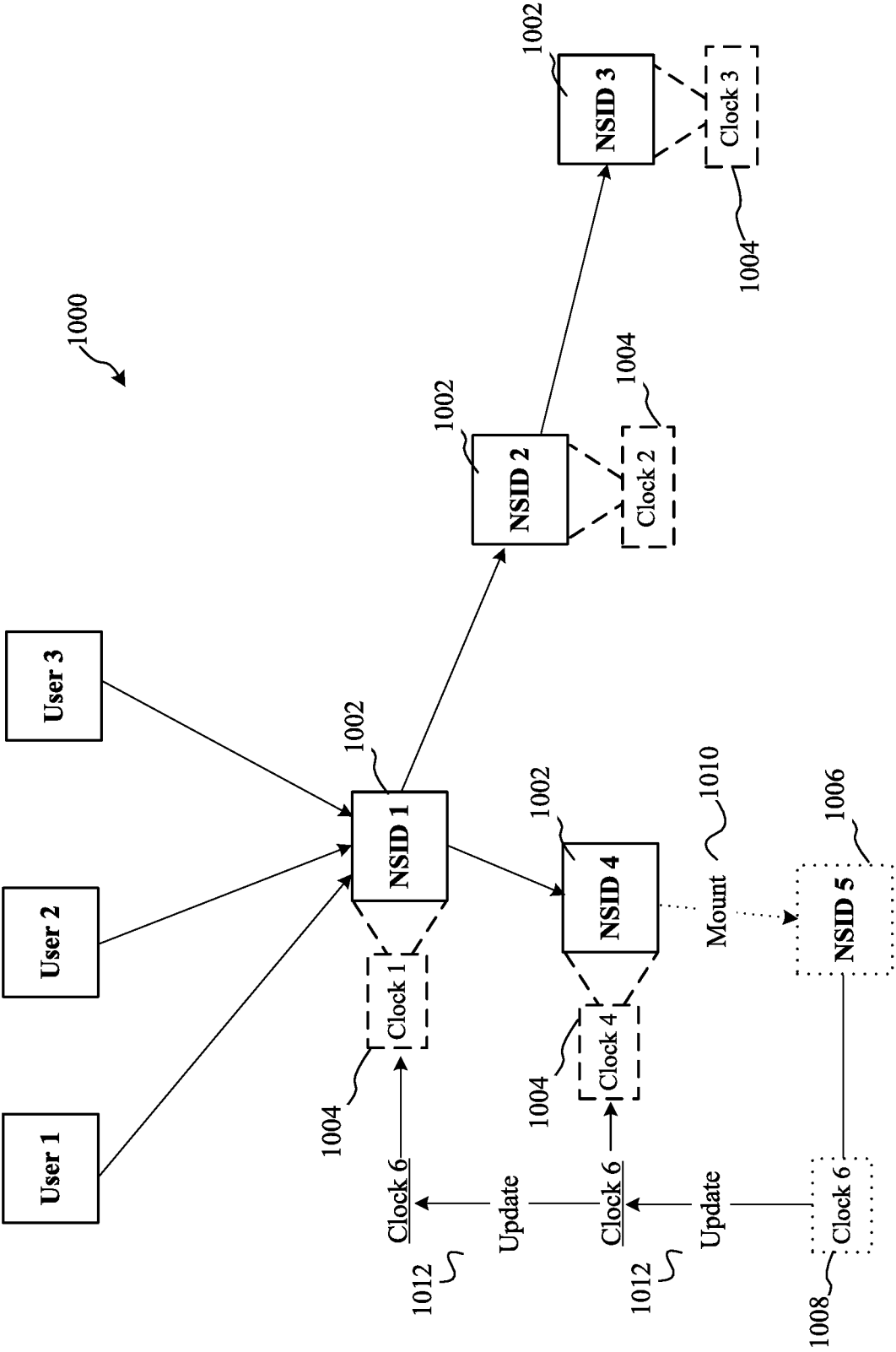


FIG. 10B

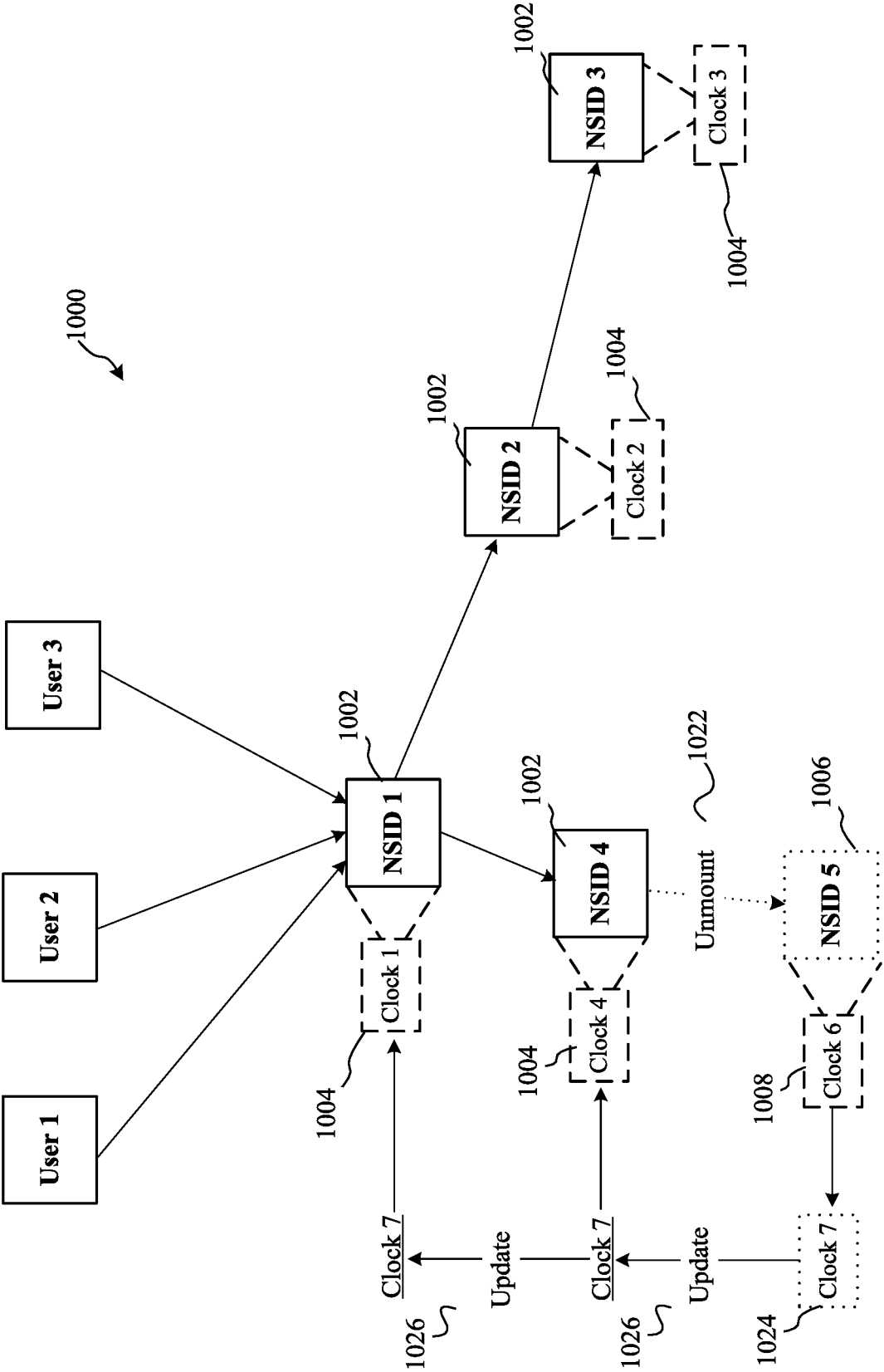


FIG. 11A

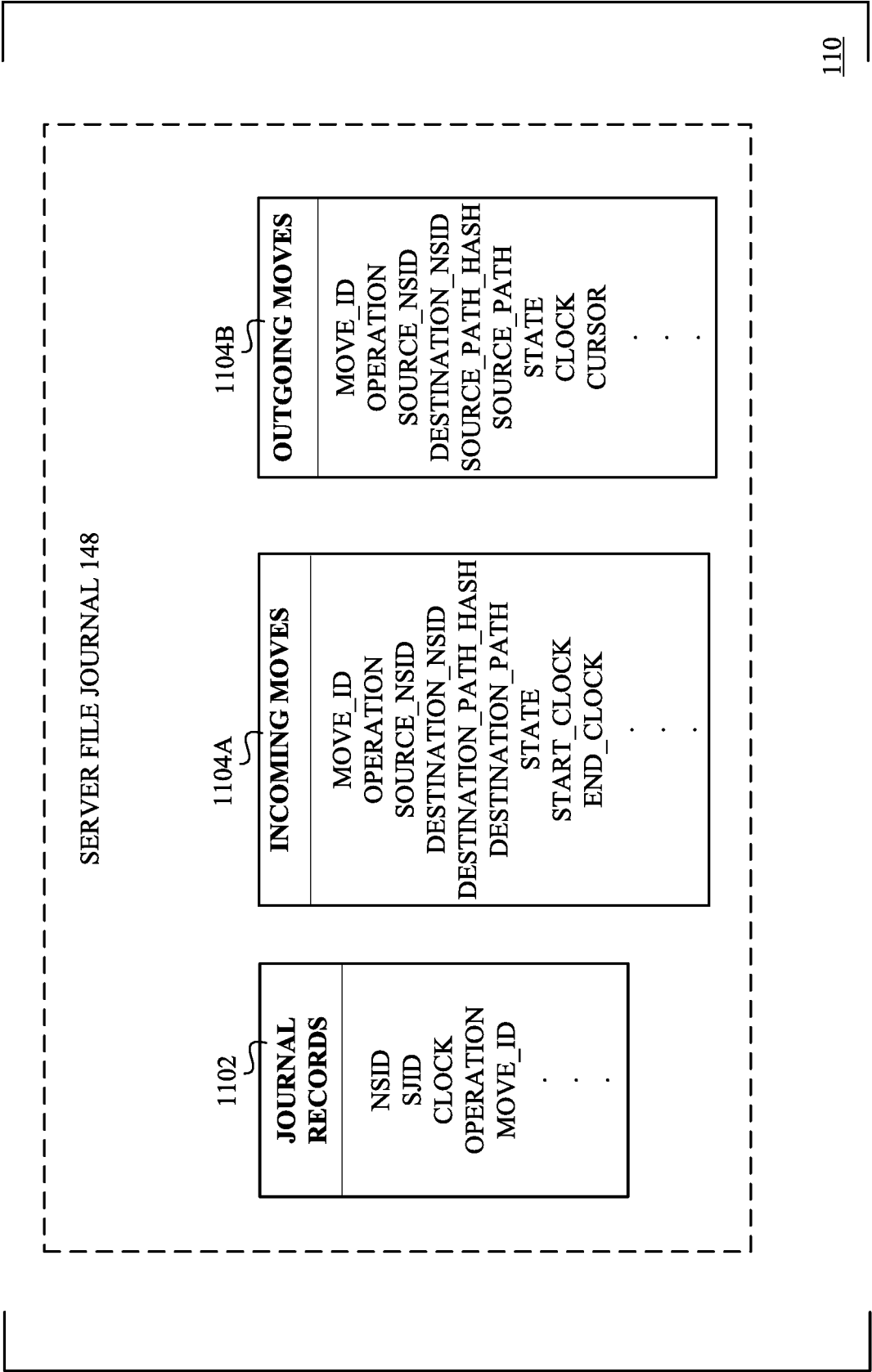
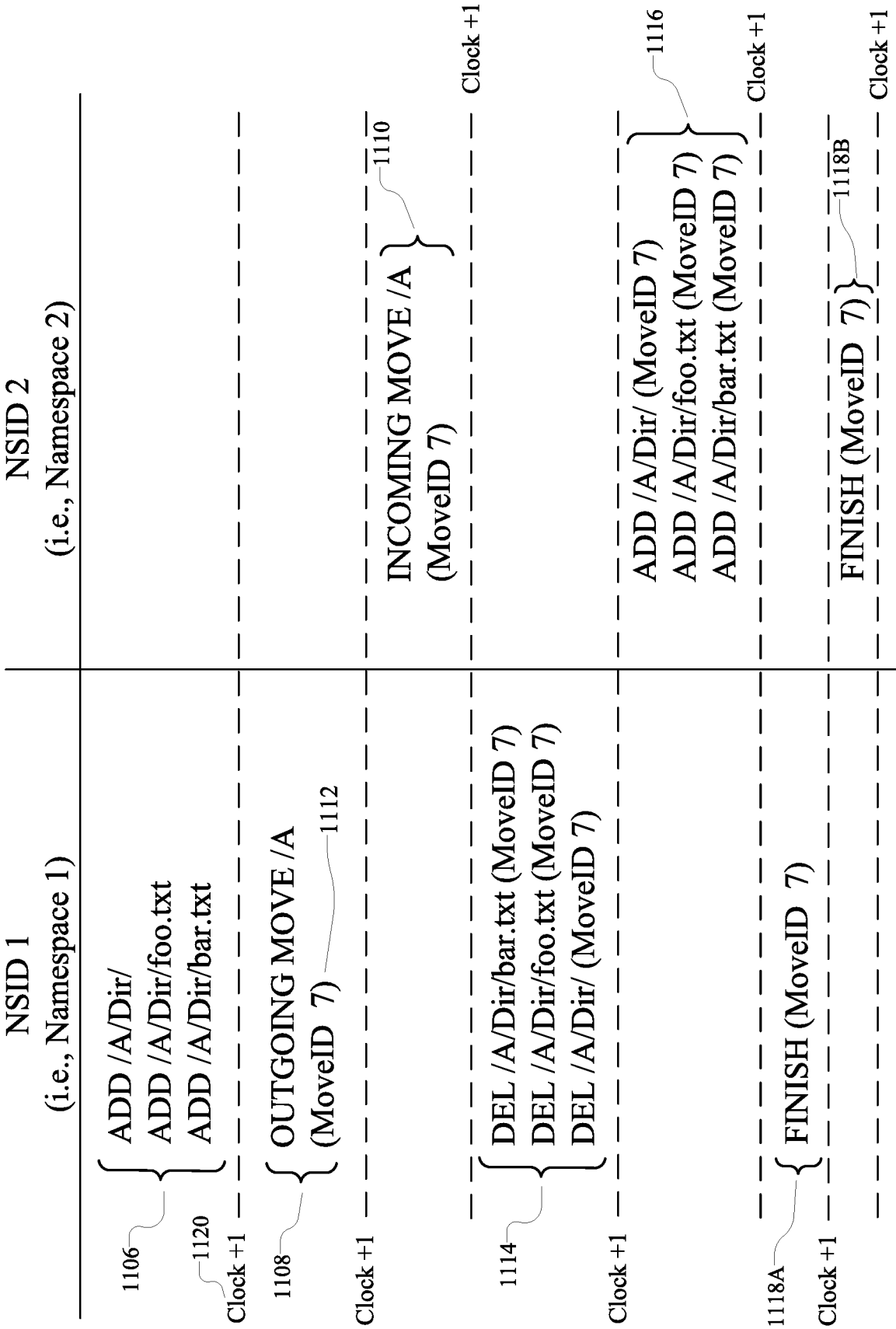


FIG. 11B



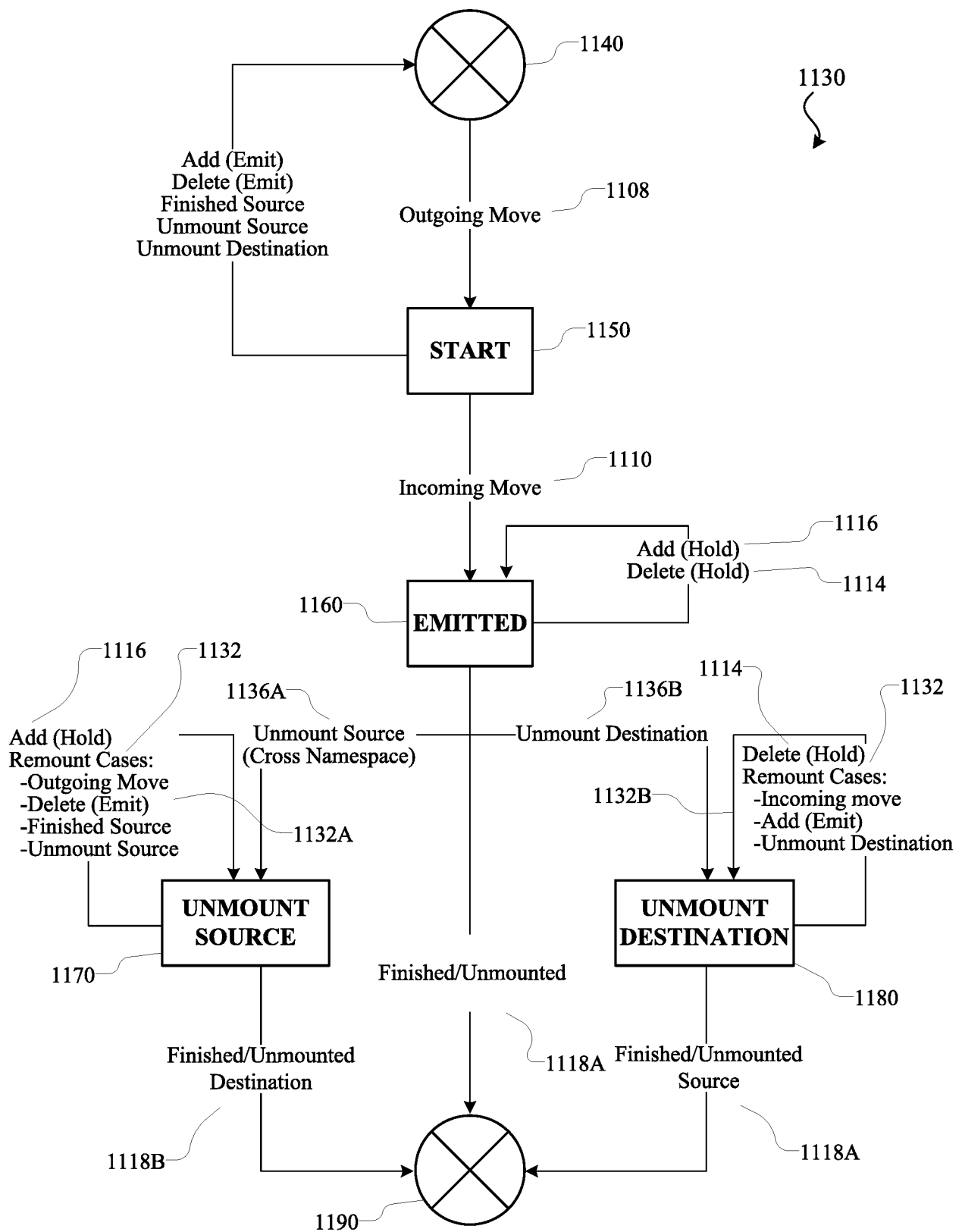
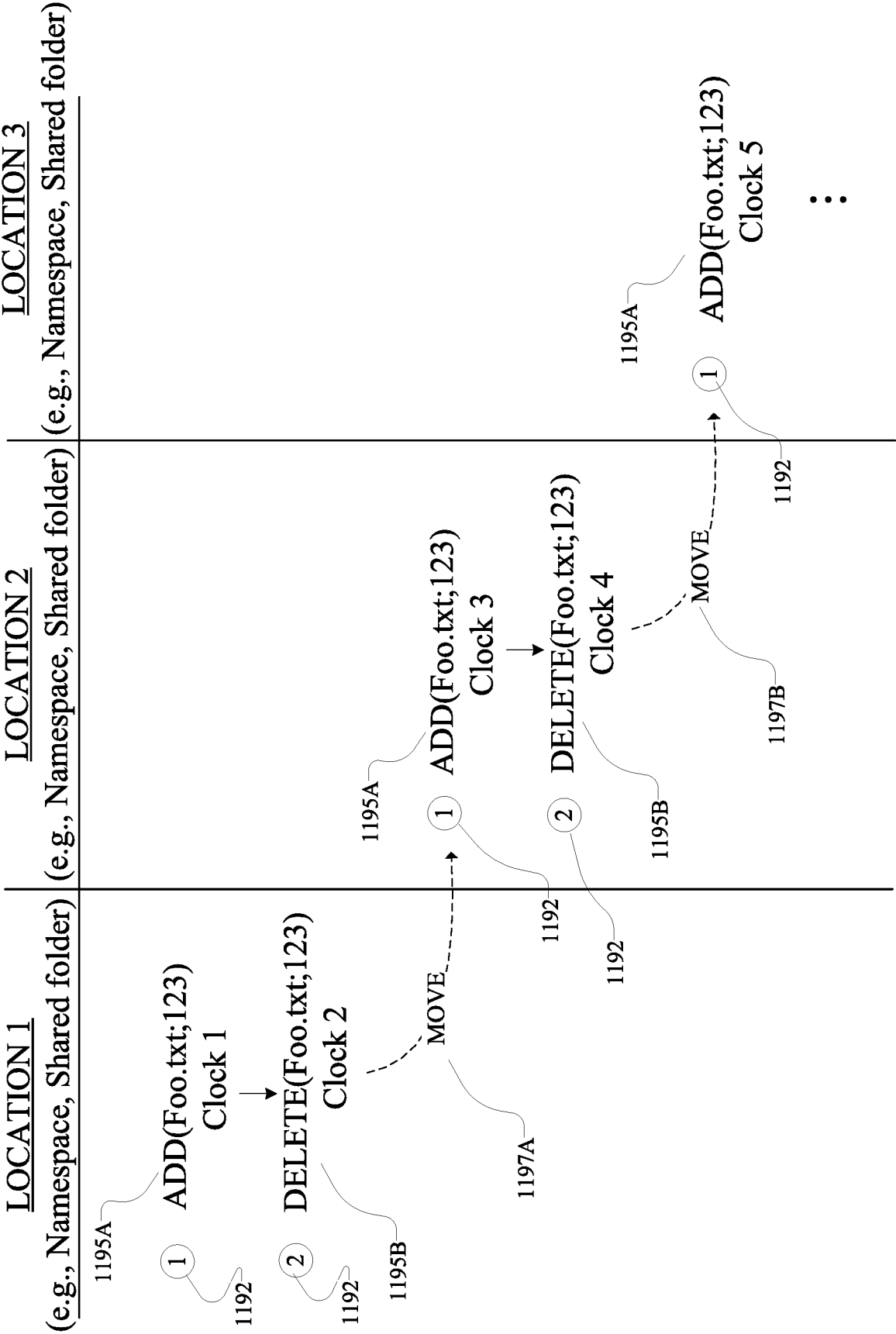


FIG. 11C

FIG. 11D



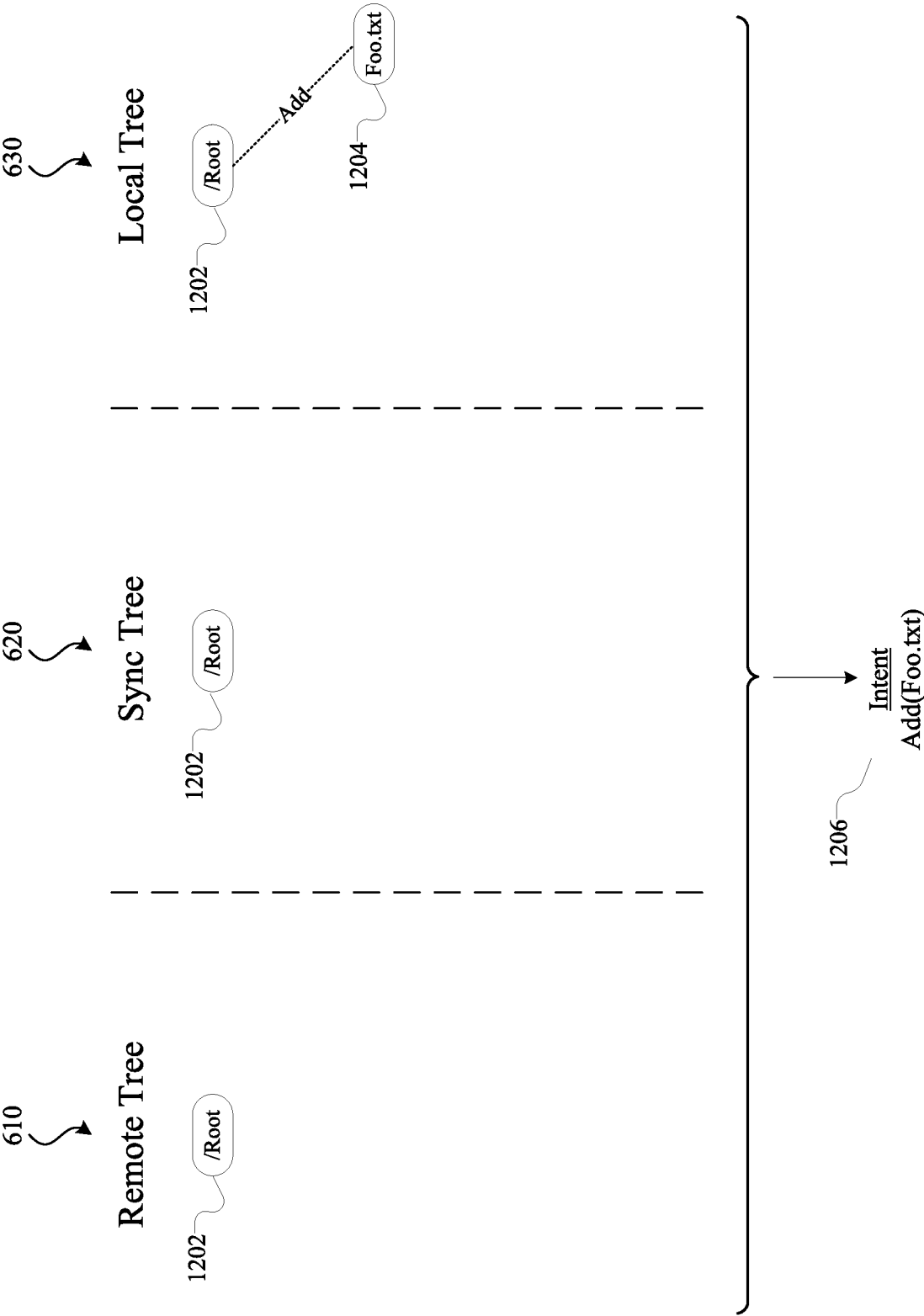


FIG. 12

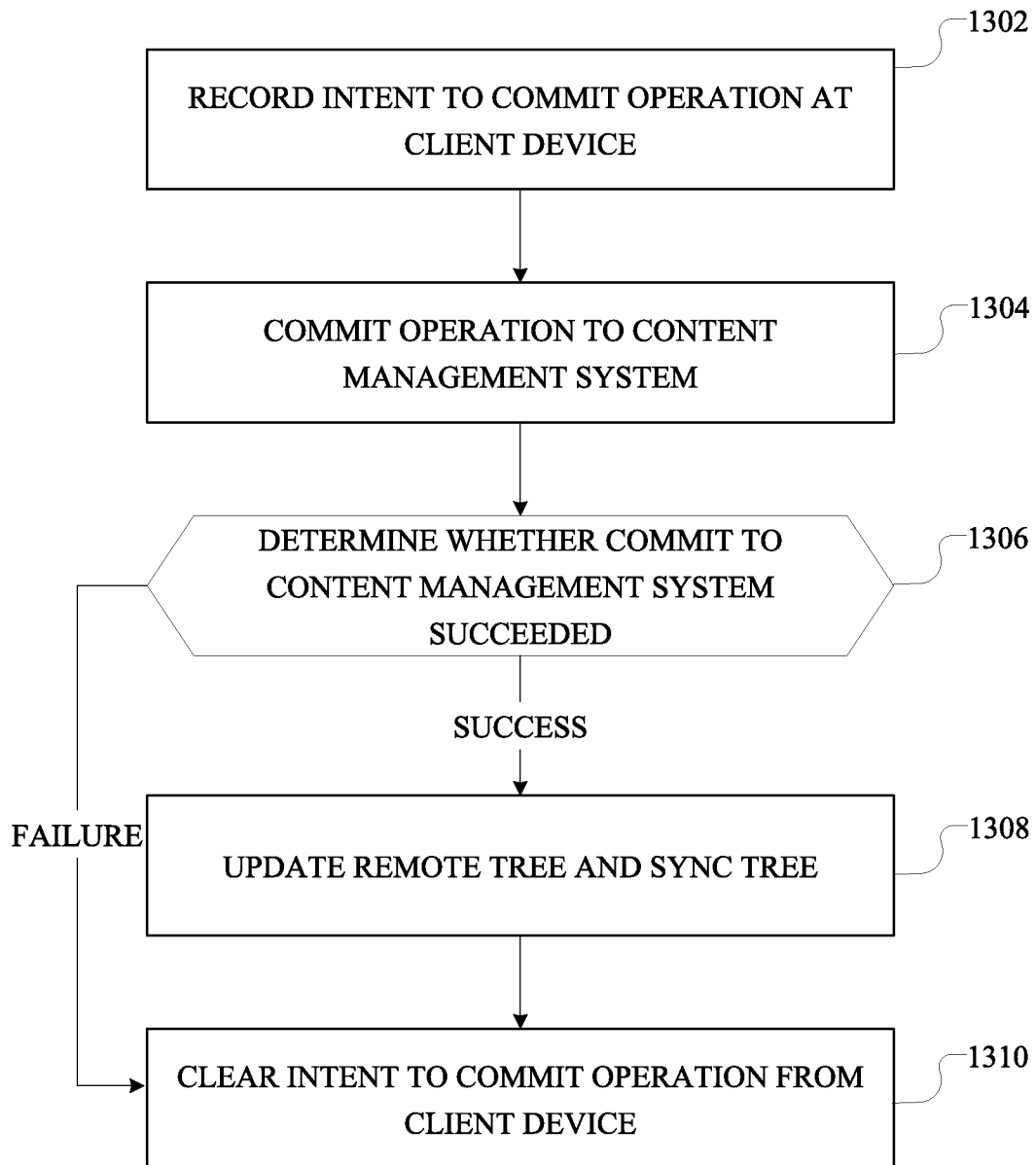


FIG. 13A

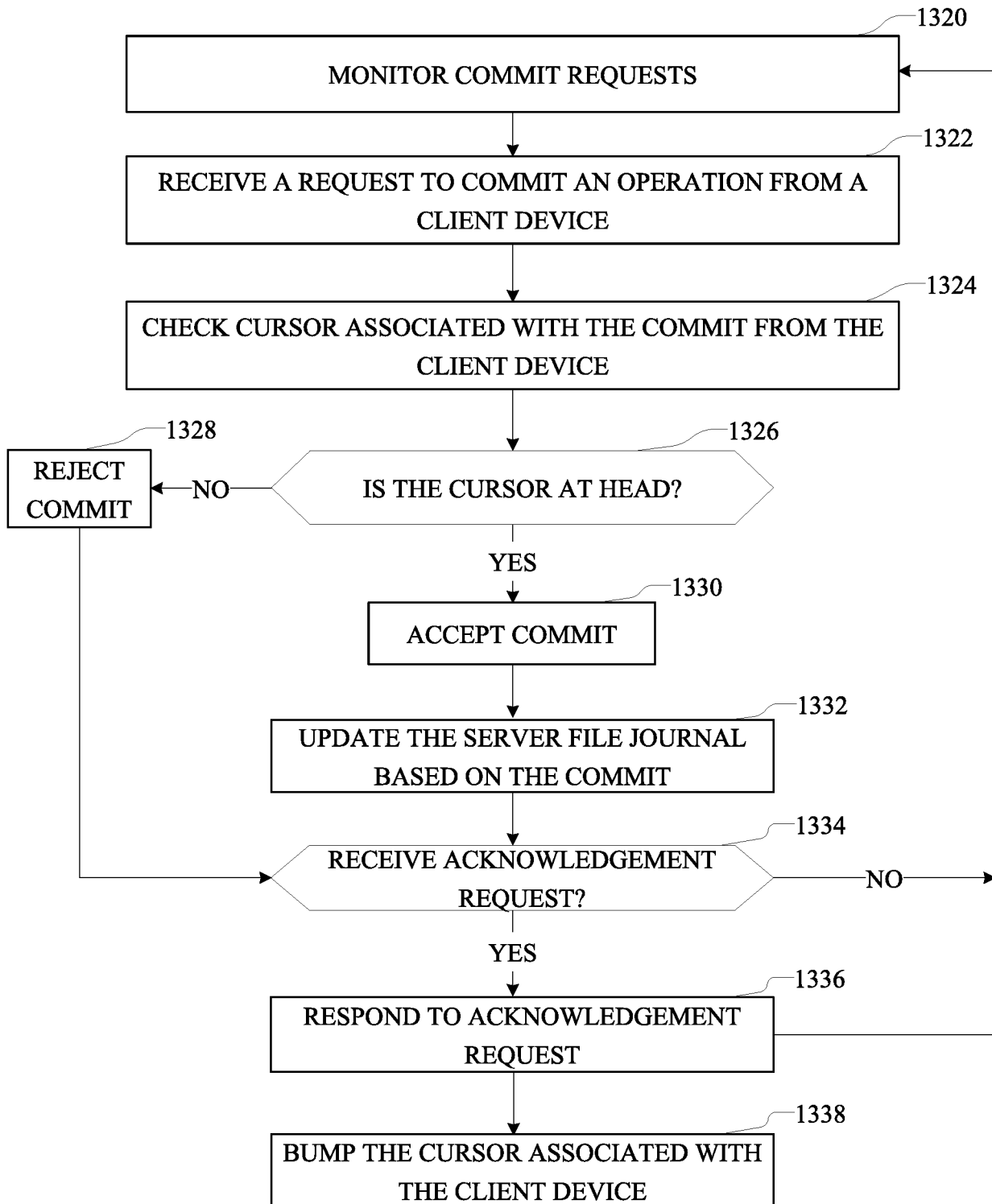


FIG. 13B

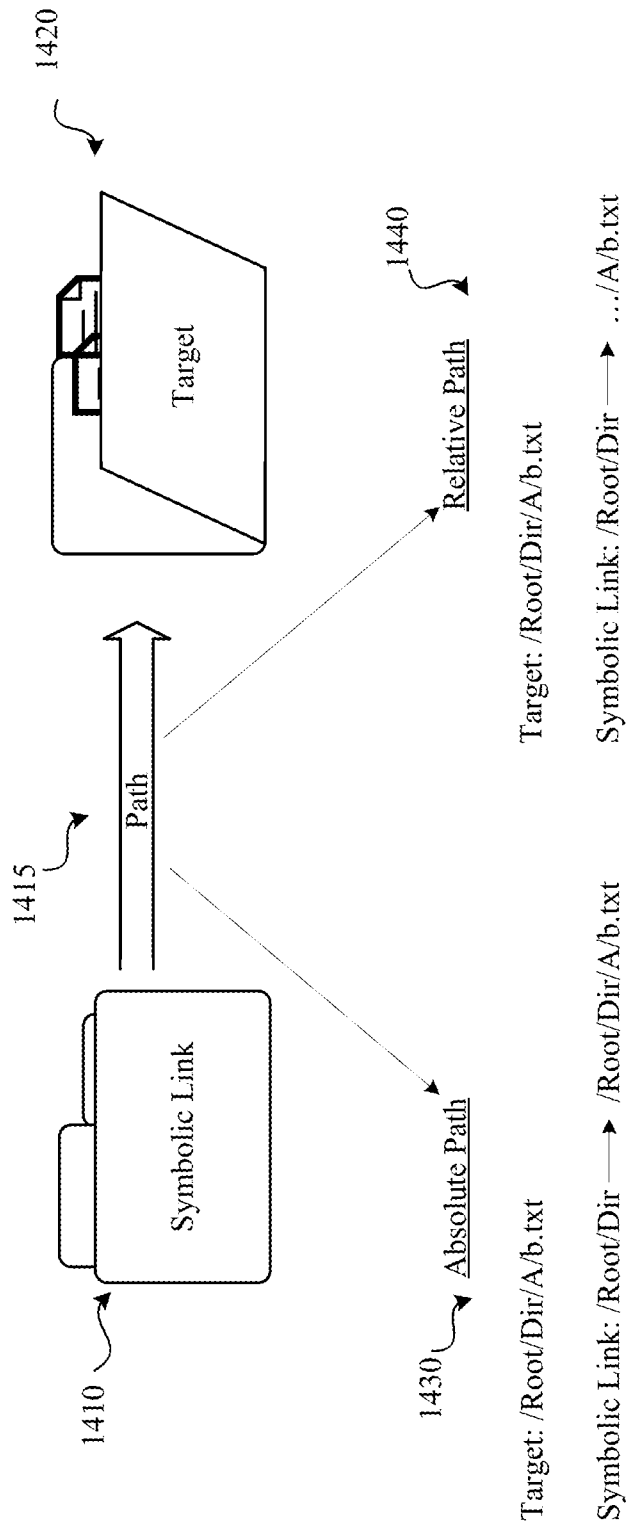


FIG. 14

TARGET AND SYMBOLIC LINK ARE CONTAINED IN THE SAME NAMESPACE

1500 ↪

OPTIONS 1505	RELATIVE PATH 1440	ABSOLUTE PATH 1430
<u>Option 1</u> Synchronize Sym. Link as a Sym. Link	- Maintains data integrity - Consistency across platforms - Inconsistencies across hosts for targets that have synchronization disabled	- Maintains data integrity - Consistency across platforms - Inconsistencies across hosts for targets that have synchronization disabled - Inconsistencies across hosts based on incorrect path at a host
<u>Option 2</u> Follow Sym. Link to Target and Synchronize Target	- Corrupts content items that rely on sym. Link (e.g., Application Bundles) - Consistent behavior - Synchronization asymmetry - Potential security problems	- Corrupts certain content items - Consistency across platforms - Synchronization asymmetry - Potential security problems
<u>Option 3</u> Ignore Sym. Link	- Consistency across platforms - Consistency across hosts - Corrupts content items that rely on sym. Link (e.g., Application Bundles)	- Consistency across platforms - Consistency across hosts - Corrupts content items that rely on sym. Link (e.g., Application Bundles)

↑
Symbolic Link

FIG. 15A

1510 TARGET AND SYMBOLIC LINK ARE CONTAINED ON DIFFERENT NAMESPACES

OPTIONS 1505	RELATIVE PATH 1440	ABSOLUTE PATH 1430
Option 1 Synchronize Sym. Link as a Sym. Link	- Maintains data integrity - Consistency across platforms - Inconsistency problems across hosts for targets that have synchronization disabled	- Maintains data integrity - Consistency across platforms - Inconsistency problems across hosts for targets that have synchronization disabled - Inconsistency problems across hosts based on incorrect path at a host
Option 2 Follow Sym. Link to Target and Synchronize Target	- Corrupts content items that rely on sym. link - Consistency across platforms - Potential security problems - Synchronization asymmetry - Potential inconsistencies	- Corrupts content items that rely on sym. link - Consistency across platforms - Synchronization asymmetry - Potential security problems - Potential inconsistencies
Option 3 Ignore Sym. Link	- Consistency across platforms - Consistency across hosts - Corrupts content items that rely on sym. link (e.g., Application Bundles)	- Consistency across platforms - Consistency across hosts - Corrupts content items that rely on sym. link (e.g., Application Bundles)

Symbolic Link

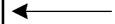
FIG. 15B

INTERNAL TARGET

1520



OPTIONS 1505	RELATIVE PATH 1440	ABSOLUTE PATH 1430
<u>Option 1</u> Synchronize Sym. Link as a Sym. Link	- Maintains data integrity - Consistency across platforms - Inconsistencies across hosts for targets that have synchronization disabled - Potential inconsistency for sym. links spanning multiple namespaces	- Maintains data integrity - Consistency across platforms - Inconsistencies across hosts for targets that have synchronization disabled - Inconsistencies across hosts based on incorrect path at a host - Potential inconsistency for sym. links spanning multiple namespaces
<u>Option 2</u> Follow Sym. Link to Target and Synchronize Target	- Corrupts content items that rely on sym. link (e.g., Application Bundles) - Consistency across platforms - Synchronization asymmetry - Potential inconsistency for sym. links spanning multiple namespaces	- Corrupts content items that rely on sym. link (e.g., Application Bundles) - Consistency across platforms - Synchronization asymmetry - Potential security problems - Potential inconsistency for sym. links spanning multiple namespaces
<u>Option 3</u> Ignore Sym. Link	- Consistency across platforms - Consistency across hosts - Corrupts content items that rely on sym. link (e.g., Application Bundles)	- Consistency across platforms - Consistency across hosts - Corrupts content items that rely on sym. link (e.g., Application Bundles)



Symbolic Link

FIG. 15C

EXTERNAL TARGET

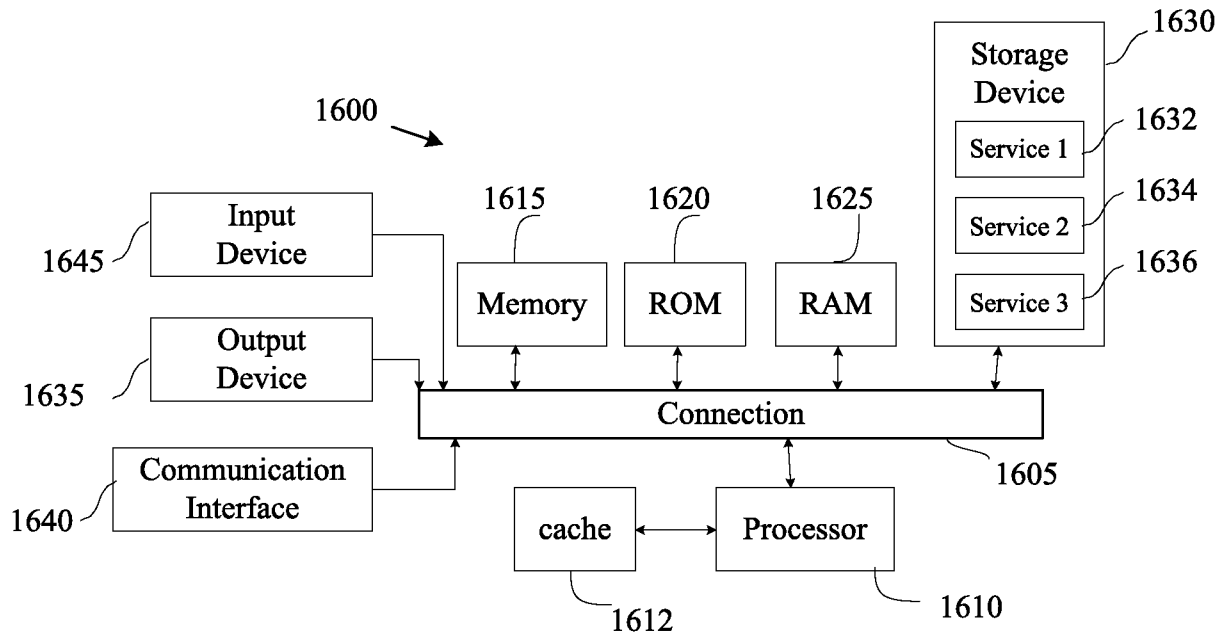
1530 ~

OPTIONS 1505	RELATIVE PATH 1440	ABSOLUTE PATH 1430
<u>Option 1</u> Synchronize Sym. Link as a Sym. Link	- Maintains data integrity - Inconsistencies across platforms - Invalid sym. links	- Maintains data integrity - Inconsistencies across platforms - Inconsistencies across hosts based on incorrect path at a host - Invalid sym. links
<u>Option 2</u> Follow Sym. Link to Target and Synchronize Target	- Corrupts content items that rely on sym. link (e.g., Application Bundles) - Synchronization asymmetry - Inconsistencies across platforms	- Corrupts content items that rely on sym. link (e.g., Application Bundles) - Inconsistencies across platforms - Synchronization asymmetry - Potential security problems
<u>Option 3</u> Ignore Sym. Link	- Consistency across platforms - Consistency across hosts - Corrupts content items that rely on sym. link (e.g., Application Bundles)	- Consistency across platforms - Consistency across hosts - Corrupts content items that rely on sym. link (e.g., Application Bundles)

↑
Symbolic Link

FIG. 15D

FIG. 16



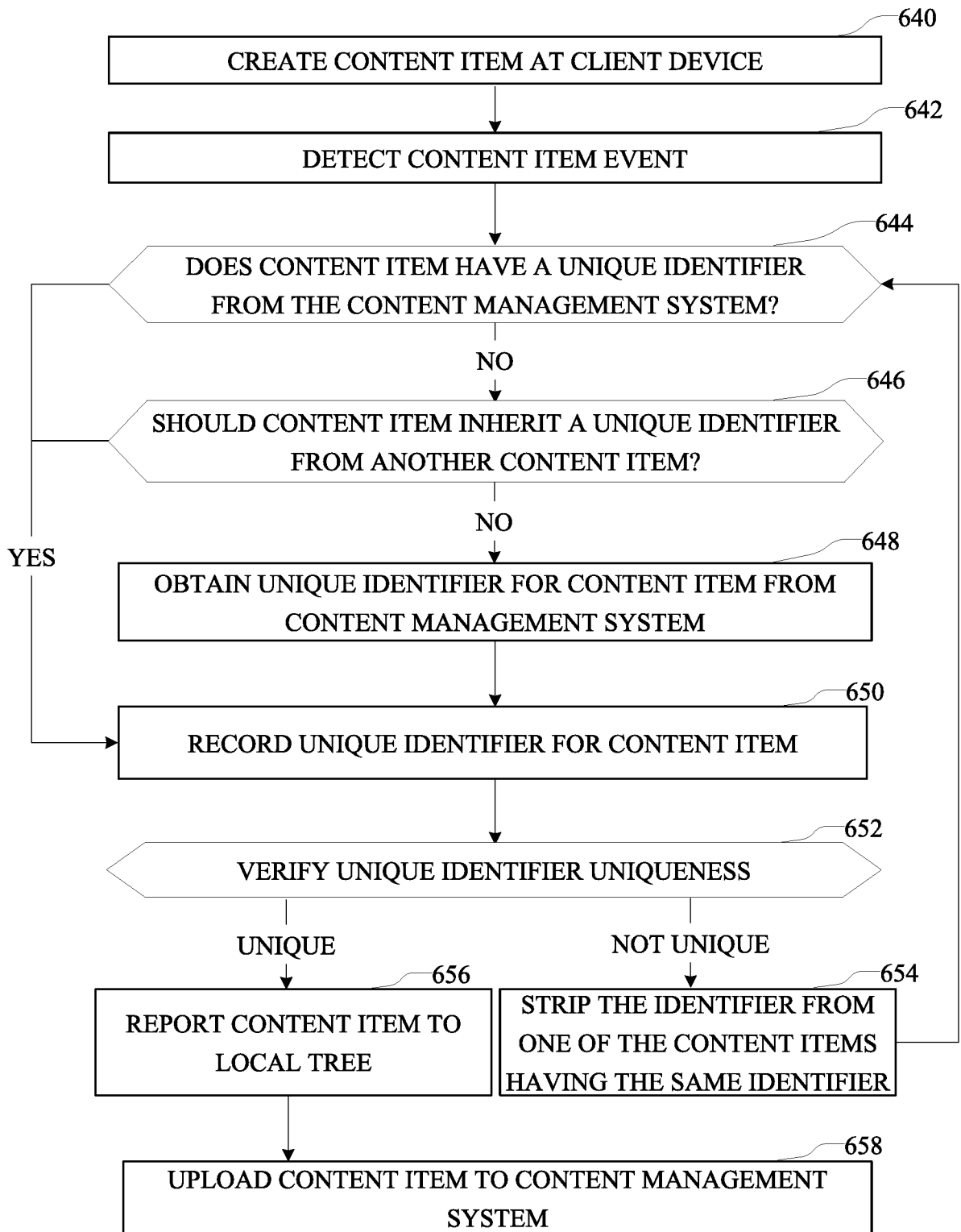


FIG. 6C