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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 October 2006 (19.10.2006)

PCT

(10) International Publication Number
WO 2006/109243 A2

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

(21) International Application Number:
PCT/IB2006/051105

(22) International Filing Date: 11 April 2006 (11.04.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
05102981.7 15 April 2005 (15.04.2005) EP

(71) Applicant *(for all designated States except US)*: **KONINKLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **HAITSMA, Jaap, A.** [NL/NL]; c/o Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agents: **ENGELFRIET, Arnoud, P.** et al.; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

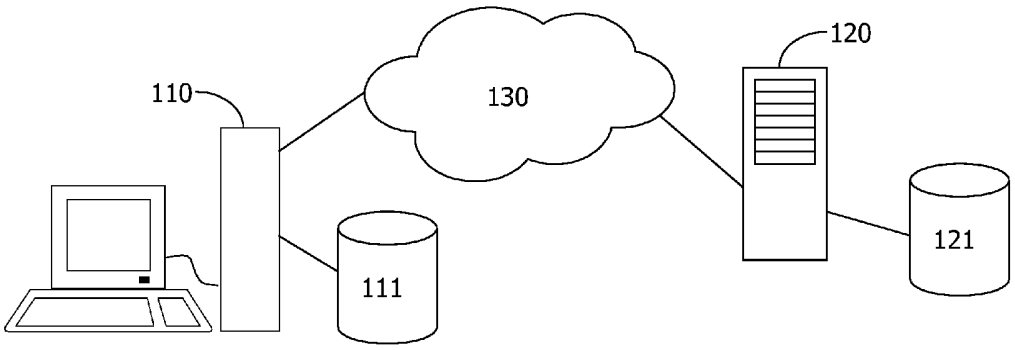
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: IMPROVED FILE MANAGEMENT



(57) Abstract: A system for file management, comprising copying means for copying a content item from a first storage location to a second storage location, characterized by fingerprinting means for determining a first robust fingerprint for the content item, by determining means for determining whether a second content item with a second robust fingerprint that is substantially identical to the first robust fingerprint is present on the second storage location, and by the copying means being configured to copy only upon a negative determination.

WO 2006/109243 A2

Improved file management

The invention relates to a system for file management, comprising copying means for copying a content item from a first storage location to a second storage location.

Today more and more content from multiple sources and an increasing amount of storage capacity is available to consumers. Devices such as the Apple iPod enable users to carry their content with them. And Web-based mail service such as Google's Gmail and Yahoo! Mail enable users to store up to one gigabyte at a remote location, so that it can be accessed from any location. To use such devices and services, the consumer will have to copy files quite frequently from one location (e.g. his PC) to another (e.g. a peripheral device like the iPod or a remote storage facility like a Gmail account).

There are many problems associated with such copying. To name but a few: the user may accidentally copy the same file, or different files with the same content, to the destination location; the metadata associated with a particular content item may get lost; rights associated with with a particular content item may get lost.

The invention provides an improved method of copying. The system according to the invention is characterized by fingerprinting means for determining a first robust fingerprint for the content item, by determining means for determining whether a second content item with a second robust fingerprint that is substantially identical to the first robust fingerprint is present on the second storage location, and by the copying means being configured to copy only upon a negative determination.

Hence the invention prevents the needless copying of content items that are already present on the second storage location. By employing robust fingerprints, duplicate content (as opposed to duplicate files) can be detected. This is preferred over checking only for duplicate files, as it may easily occur that duplicate content is available in multiple different files that are encoded in different formats, or at different quality levels. Such

duplicates cannot be detected by comparing files, as the files are different even though the content is substantially similar.

In an embodiment the means for copying are configured to delete the content item from the first storage location subsequent to a successful copying to the second storage location or subsequent to the negative determination. This realizes a 'move' operation, whereby the content item no longer resides on the first location afterwards.

In an embodiment the the first storage location is comprised in a personal computer and the second storage location is comprised in a peripheral device operable to communicate with the personal computer. Such a peripheral device might be the above-mentioned Apple iPod or similar music or video playback device, or a data storage device such as a USB memory stick or a (re)writable CD or DVD. Alternatively the first location could be a hard disk in a client system, and the second location could be a storage facility such as a Web-based e-mail system, Web hosting area or a remote fileserver.

The first robust fingerprint may be computed when the copying is initiated. Alternatively, the first robust fingerprint may be retrieved from a storage location, e.g. from the first storage location. By computing these fingerprints in advance, the copying operation can complete faster at the cost of storing potentially many fingerprints.

In an embodiment the determining means are configured to determine whether the second content item is present in an authorized area of the second storage location. Especially in situations where the second storage location is a storage facility used by multiple users, it makes sense to only check for duplicate content in authorized areas, since content in other areas may not be accessed.

In an embodiment the copying means are configured to record, upon a positive determination, an authorization to access the second content item. This is a quick and effective way to realize that the user can access the content item from the second location after initiating the copying operation. The content in question already resides on the second location. Hence only an authorization to access this content needs to be recorded.

In a variation of this embodiment the copying means are configured to derive the authorization from an authorization associated with the first content item. For instance, consider the situation where the first location is a PC and the second location is a peripheral device such as an iPod. Now if the first content item is authorized for playback only but not transmission over a network, the second content item would receive this same authorization.

In a variation of this embodiment the copying means are configured to derive the authorization from authentication data provided in conjunction with the copying

operation. For instance, the user may have to log into the fileserver to which he wants to copy the file. His authenticaton data, e.g. a username and password or a biometric identification, can then be used at the fileserver to determine what kind of authorization needs to be granted.

Further advantageous embodiments are set out in the dependent claims.

These and other aspects of the invention will be apparent from and elucidated with reference to the embodiments shown in the drawing, in which:

Fig. 1 schematically shows a first embodiment;

Fig. 2 schematically shows further embodiments; and

Fig. 3 schematically shows another embodiment.

Throughout the figures, same reference numerals indicate similar or corresponding features. Some of the features indicated in the drawings are typically implemented in software, and as such represent software entities, such as software modules or objects.

Fig. 1 schematically shows a first embodiment of the invention. A first device, shown as a personal computer 110, comprises a first storage medium 111. A second device, shown as a network file server 120, comprises a second storage medium 121. The devices 110 and 120 are connected via a network 130. This network 130 could be a local area network or a wide-area network such as the Internet. Alternatively a direct connection between the devices 110, 120 may be used. The connection or access to the network 130 may be wired or wireless.

The device 110 comprises copying means for copying a content item from the first storage location 111 to the second storage location 121. Such means by themselves are well known in the art. They may comprise e.g. an FTP or SSH client, a Web browser or support for accessing remote filesystems. Typically these means are provided as software.

In the situation illustrated in Fig. 1, copying usually takes place using file transfer protocols such as FTP, SSH, or HTTP but a content item can also be copied to a remote location by sending it via e-mail. For example, the file can be attached and sent to an e-mail address. Using a web-based interface, the recipient (who may be the same person who sent the e-mail) can select the attached file and store it on the second storage location.

The term content item is used generally to refer to items such as music, speeches, movies, animations, videoclips for music, ringtones, spoken books.

In addition, the device 110 comprises a fingerprinting module that can compute a first robust fingerprint for the content item that is to be copied.

Although of course any method for computing a robust fingerprint can be used, one method for computing a robust fingerprint is described in international patent application WO 02/065782 (attorney docket PHNL010110). This method generates robust fingerprints for multimedia content such as, for example, audio clips, where the audio clip is divided in successive (preferably overlapping) time intervals. For each time interval, the frequency spectrum is divided in bands. A robust property of each band (e.g. energy) is computed and represented by a respective fingerprint bit.

Multimedia content is thus represented by a fingerprint comprising a concatenation of binary values, one for each time interval. The fingerprint does not need to be computed over the whole multimedia content, but can be computed when a portion of a certain length has been received. There can thus be plural fingerprints for one multimedia content, depending on which portion is used to compute the fingerprint over.

Further, video fingerprinting algorithms are known, e.g. from the following disclosure: Job Oostveen, Ton Kalker, Jaap Haitsma: "Feature Extraction and a Database Strategy for Video Fingerprinting". 117-128. IN: Shi-Kuo Chang, Zhe Chen, Suh-Yin Lee (Eds.): Recent Advances in Visual Information Systems, 5th International Conference, VISUAL 2002 Hsin Chu, Taiwan, March 11-13, 2002, Proceedings. Lecture Notes in Computer Science 2314 Springer 2002.

A first option is to compute the first robust fingerprint when the copying is initiated. A second option is to retrieve the first robust fingerprint from a storage location. This location could be the first storage medium 111. In this second option, pre-computed robust fingerprints are used. This saves time when the copying is initated, as no fingerprint needs to be computed at that time. However it does require storing fingerprints for the content items.

Having computed the first robust fingerprint, next a determination must be made whether a second content item with a second robust fingerprint that is substantially identical to the first robust fingerprint is present on the second storage location. This can be done in a variety of ways.

Again, a first option is to compute the second robust fingerprint when the copying is initiated. This now implies that robust fingerprints must be computed for all content items on the second storage medium, or at least a selected subset thereof.

This selected subset may be only those content items on the second storage medium which the user is authorized to access. On a multi-user system for example typically not all content items are accessible to all users. Considering all content items then is unnecessary. Only those content items which the user is authorized to access need to be considered. And so only fingerprints for those content items need to be compared against the first robust fingerprint.

A second option is to retrieve the second robust fingerprint from a storage location. This location could be the second storage medium 121. In this second option, pre-computed robust fingerprints are used. The pre-computed robust fingerprints may be computed when the content items in question are copied onto the second storage medium. They could be stored in a database or on the second storage medium.

If it is determined that there is a second content item on the second storage medium 121 that is substantially similar (but not necessarily identical) to the first content item, then the first content item does not need to be copied to the second storage medium 121.

One option now is to simply report that the content item already is present on the second storage medium 121, preferably with an identification of its filename and/or location on the second storage medium 121.

Another option is to record an authorization to access the second content item. This option is especially useful in the context of music lockers, to be discussed below. This authorization may be derived from an authorization associated with the first content item. For instance a permission to play back the first content item can be transferred to the second content item.

Another option is to create a reference to the second content item which carries an identifier for the first content item. If a substantially similar second content item is found to be present at the second location, then the first content item is not copied but a reference is made to the second content item, using the name or other identifying information of the first content item. This way, the user can locate the first content item by the original name, but the reference does not take up as much space as the content item itself. This option can be implemented e.g. using Unix-style 'hard links' or 'soft links' or using Windows-style 'shortcuts'.

If the first and second content items are found to be substantially similar, and one of the two items lacks metadata, then the lacking metadata could be copied from the other item. For instance, the first content item may have a title but no identification of the artist or album, whereas the second content item may have all this information. Then this information can be copied from the second to the first content item.

If the first and second content items are found to be substantially similar, and the first item is found to be of higher quality than the second item, then the copying might proceed anyway by replacing the second content item with the first. Higher quality may be determined by comparing number of channels, sampling rate, stereo vs. mono recording and so on.

Fig. 2 shows further embodiments of the present invention. The personal computer 110 as shown here can copy files to the second storage location 121 which is part of a peripheral device, here a palmtop computer 220. To this end the peripheral device 220 may be connected to the personal computer 110. In addition, the personal computer 110 can copy files to the second storage location 121 as part of a mobile telephone 240. Typically in such situations wireless data transfer is used.

Although shown in Fig. 2 as external to the devices 220, 240, it will be evident that the storage medium 121 can be located physically inside the devices.

Fig. 3 shows yet another embodiment of the invention as used in the context of a music locker service. Such a service offers the possibility to first upload one's content to a remote storage facility and consequently to access said content from any location. Typically the content is streamed to the location where the client is situated. With such services, users are offered a certain amount of storage, typically several gigabytes. So it is important to make sure users do not waste their storage with duplicate content.

In this embodiment, content is uploaded from the personal computer 110 to the server 120 where it is stored on the second storage medium 121. Later, the content can be accessed from any location, e.g. Internet radio 330. This access again occurs over the network 130, which typically is the Internet.

As explained above, when copying a first content item from the personal computer 110 to the server 120, a determination is made whether there is a second content item on the second storage medium 121 that is substantially similar (but not necessarily identical) to the first content item. If so, the first content item does not need to be copied. This achieves that the user always only has one copy of his content items on the music locker service.

Some music locker services may have a large amount of content items available themselves, to which they grant access if the user can prove he owns a legally purchased specimen of the content item. As noted above, this is where the option to record an authorization to access the second content item comes into play. By using the present invention, this proof is obtained if the robust fingerprint of the first content item is provided. Hence the user can be granted access to the second content item.

Optionally the determination of the robust fingerprint of the first content item could be accompanied by a determination of whether the first content item is legally acquired. For instance, this determination could be made by checking whether the first content item is read from a harddisk or from a compact disc or DVD disc.

The needed download capacity of the music locker provider can be reduced by locally caching a number of music files which the user plays often.

WO 2005/011281 (attorney docket PHNL030888) discloses a method and device for synchronising two or more signals. A fingerprint pair is generated on the basis of a segment of a first signal e.g. an audio signal and of a segment of a second signal e.g. a video signal at each synchronisation time point. The generated fingerprint pair(s) are stored in a database and communicated or distributed to a synchronisation device. During synchronisation, fingerprint(s) of the audio signal and fingerprint(s) of the video signal to be synchronised are generated and matched against fingerprints in the database. When a match is found, the fingerprints also determine the synchronisation time point, which is used to synchronise the two signals.

It should be noted that the above-mentioned embodiments illustrate rather than limit the invention, and that those skilled in the art will be able to design many alternative embodiments without departing from the scope of the appended claims. For instance, instead of directly exchanging the access control data between devices 101 and 105, it is also possible that the data is exchanged using an intermediary third party.

In the claims, any reference signs placed between parentheses shall not be construed as limiting the claim. The word "comprising" does not exclude the presence of elements or steps other than those listed in a claim. The word "a" or "an" preceding an element does not exclude the presence of a plurality of such elements.

The invention can be implemented by means of hardware comprising several distinct elements, and by means of a suitably programmed computer. In the device claim enumerating several means, several of these means can be embodied by one and the same item of hardware. The mere fact that certain measures are recited in mutually different

dependent claims does not indicate that a combination of these measures cannot be used to advantage.

CLAIMS:

1. A system for file management, comprising copying means for copying a content item from a first storage location to a second storage location, characterized by fingerprinting means for determining a first robust fingerprint for the content item, by determining means for determining whether a second content item with a second robust fingerprint that is substantially identical to the first robust fingerprint is present on the second storage location, and by the copying means being configured to copy only upon a negative determination.

2. The system of claim 1, in which the means for copying are configured to delete the content item from the first storage location subsequent to a successful copying to the second storage location or subsequent to the negative determination.

3. The system of claim 1, in which the first storage location is comprised in a personal computer and the second storage location is comprised in a peripheral device operable to communicate with the personal computer.

4. The system of claim 1, in which the fingerprinting means are configured to compute the first robust fingerprint when the copying is initiated.

5. The system of claim 1, in which the fingerprinting means are configured to retrieve the first robust fingerprint from a storage location.

6. The system of claim 1, in which the determining means are configured to determine whether the second content item is present in an authorized area of the second storage location.

7. The system of claim 1, in which the copying means are configured to record, upon a positive determination, an authorization to access the second content item.

8. The system of claim 7, in which the copying means are configured to derive the authorization from an authorization associated with the first content item.

9. The system of claim 7, in which the copying means are configured to derive the authorization from authentication data provided in conjunction with the copying operation.

10. The system of claim 1, in which the determining means are configured to compute the second robust fingerprint and to compare the computed second robust fingerprint with the first robust fingerprint.

11. The system of claim 1, in which the fingerprinting means are configured to retrieve the second robust fingerprint from a storage location.

12. The system of claim 11, in which the storage location is located on the first storage location.

13. A computer program product comprising instructions for causing a processor to operate as the system of claim 1.

FIG.1

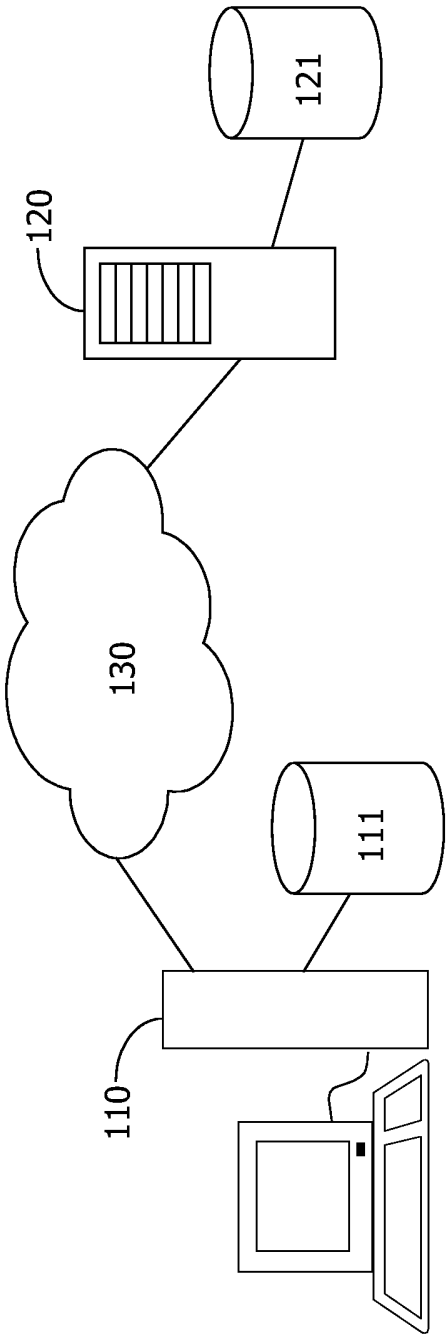
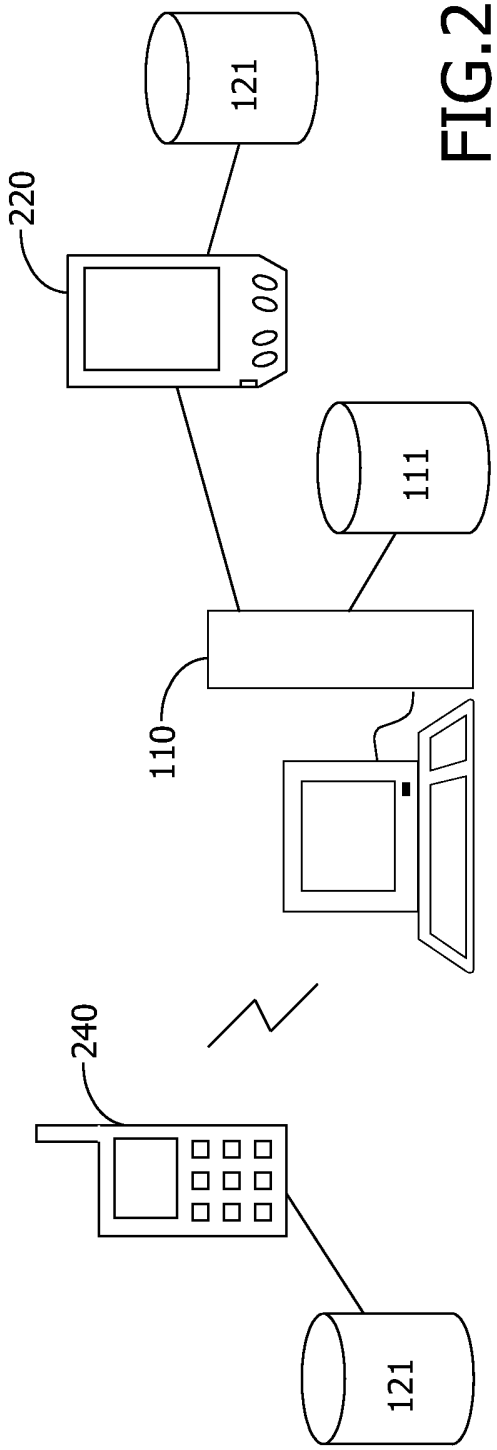


FIG.2



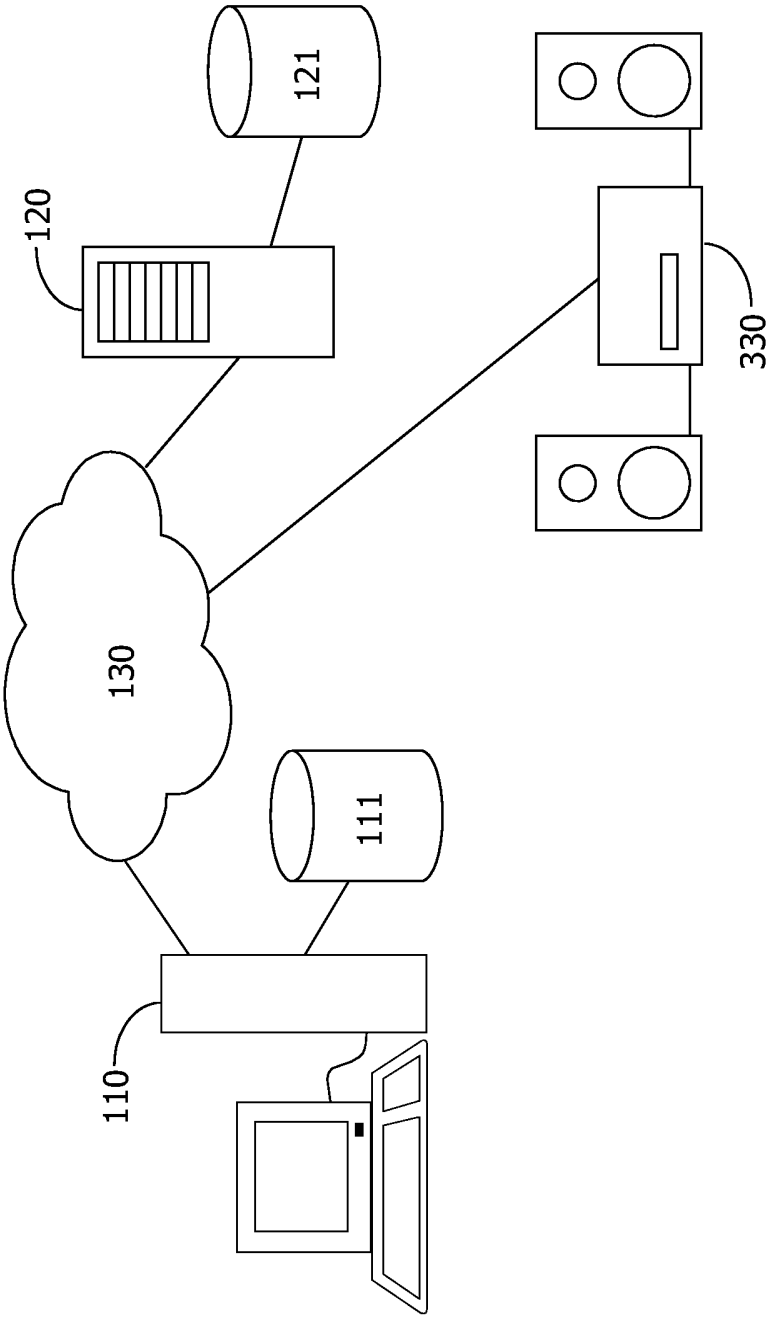


FIG. 3