



- (51) **International Patent Classification:**
G06F 13/38 (2006.01) *G06F 13/40* (2006.01)
- (21) **International Application Number:**
PCT/US2015/029608
- (22) **International Filing Date:**
7 May 2015 (07.05.2015)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (71) **Applicant:** HEWLETT PACKARD ENTERPRISE DEVELOPMENT LP [US/US]; 11445 Compaq Center Drive West, Houston, TX 77070 (US).
- (72) **Inventors:** DONAGHY, Dave; Longdown Avenue Stoke Gifford, Gifford Bristol BS34 8QZ (GB). SIMPSON, Ben; Longdown Avenue Stoke, Gifford Bristol BS34 8QZ (GB). BUTT, John; Longdown Avenue Stoke, Gifford Bristol BS34 8QZ (GB).
- (74) **Agents:** DAS, Manav et al.; Hewlett Packard Enterprise, 3404 E. Harmony Road, Mail Stop 79, Fort Collins, CO 80528 (US).
- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,

[Continued on next page]

(54) **Title:** INFORMATION CONTENT OF A DATA CONTAINER

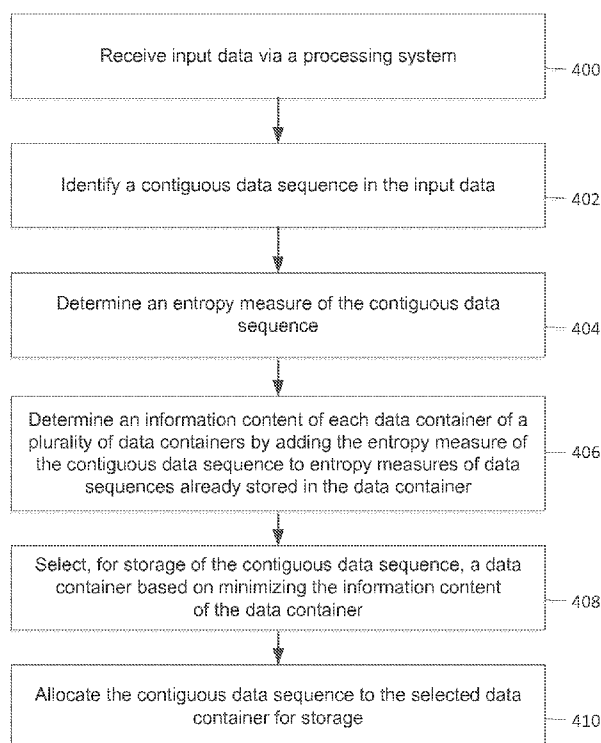


Fig. 4

(57) **Abstract:** Minimizing information loss in a data container is disclosed. One example is a system including a data processor, an entroposcope module, and a distribution processor. The data processor receives input data and identifies a contiguous data sequence in the input data. The entroposcope module determines an entropy measure of the contiguous data sequence. The distribution processor determines an information content of each data container of a plurality of data containers by adding the entropy measure of the contiguous data sequence to entropy measures of data sequences already stored in the data container, and selects, for storage of the contiguous data sequence, a data container based on minimizing the information content of the data container.



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).

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

Published:

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

Declarations under Rule 4.17:

— *as to the identity of the inventor (Rule 4.17(i))*

INFORMATION CONTENT OF A DATA CONTAINER

Background

[0001] Data containers are generally utilized to store data that may need to be recovered at another time. Data loss may be caused due to various actions, intentional or accidental. Various approaches may be applied to minimize such data loss.

Brief Description of the Drawings

[0002] Figure 1 is a functional block diagram illustrating one example of a system for minimizing information loss in a data container.

[0003] Figure 2 is a flow diagram illustrating one example of selecting a data container to minimize information loss.

[0004] Figure 3 is a block diagram illustrating one example of a computer readable medium for minimizing information loss in a data container.

[0005] Figure 4 is a flow diagram illustrating one example of a method for minimizing information loss in a data container.

Detailed Description

[0006] Data storage devices may be typically utilized to store data whose recovery might be required at some future time. For example, deduplication measures may be applied across multiple input streams and across wide-ranging time periods. Also, for example, standard compression may be utilized for contiguous sequences of bytes inside individual streams, typically sequences of 10M or so. Generally, deduplication measures may be applied to data being stored before compression techniques are applied. Each such data entity, compressed or otherwise, may be stored in a single container file. In some instances, there may be a one-many mapping between compressed data entities and container files.

[0007] Certain actions, malicious and/or accidental, may cause data loss to occur via elimination and/or modification of entities used by storage devices to store items of data valuable to system users. Existing methods for data

protection aim to minimize a chance of data loss. However, such methods may generally not account for actual content of stored data segments.

[0008] In many instances, actual value of an individual data item to a user may not be related to a size of the data item as measured by the storage system. For example, one item may include several billions of bytes of data, many of which may be utilized as padding to enforce a uniform size between comparable items. In such a case, it may be possible that a fraction of the data may be useful; for example, a million bytes of data may be of use to a user, whereas the remaining billions may be redundant.

[0009] In such cases, a measure of information content of the data items may provide a higher measure for usefulness of a first collection of items and a lower measure for usefulness of a second collection of items. Applying such information-theoretic measures to a data item may allow a storage system to minimize information loss associated to a data loss event. For example, given a complex (e.g., many-many) mapping between items of interest to users and items maintained by a storage system, storage containers for user items may be selected in such a way that each storage container contains a similar amount of information. Such allocation of data for storage may prevent a scenario where a single storage container has significantly more information content than its peers, and may thereby reduce the impact of information loss related to its data loss.

[0010] As described in various examples herein, minimizing information loss in a data container is disclosed. One example is a system including a data processor, an entroposcope module, and a distribution processor. The data processor receives input data and identifies a contiguous data sequence in the input data. The entroposcope module determines an entropy measure of the contiguous data sequence. The distribution processor determines an information content of each data container of a plurality of data containers by adding the entropy measure of the contiguous data sequence to entropy measures of data sequences already stored in the data container, and selects, for storage of the contiguous data sequence, a data container based on minimizing the information content of the data container.

[0011] In the following detailed description, reference is made to the accompanying drawings which form a part hereof, and in which is shown by way of illustration specific examples in which the disclosure may be practiced. It is to be understood that other examples may be utilized, and structural or logical changes may be made without departing from the scope of the present disclosure. The following detailed description, therefore, is not to be taken in a limiting sense, and the scope of the present disclosure is defined by the appended claims. It is to be understood that features of the various examples described herein may be combined, in part or whole, with each other, unless specifically noted otherwise.

[0012] Figure 1 is a functional block diagram illustrating one example of a system 100 for minimizing information loss in a data container. System 100 includes a data processor 104, an entroposcope module 108, and a distribution processor 110. The data processor 104 receives input data 102 and identifies a contiguous data sequence 106 in the input data 102. For example, input data 102 may include a stream of data, and the data processor 104 may identify contiguous data sequences of bytes in the stream of data. In some examples, the contiguous data sequences may be sequences of size approximately 10M. Generally, the term "contiguous" as used herein, describes objects that may be adjacent to one another, without necessarily being adjacent to one another. As used with reference to input data 102, contiguous data may be data that may be moved, and/or stored as a block of data, with minimal gaps between data objects. Actual determination of what is contiguous may depend on the type and size of data, storage environments, temporal factors (e.g., when data is generated, received, and/or to be stored), and/or spatial factors (e.g., where data is generated, received, and/or to be stored).

[0013] The entroposcope module 108 may determine an entropy measure of the contiguous data sequence 106. Generally, the entropy measure is a measure of an average amount of information in the contiguous data sequence 106. As described herein, data to be stored is often compressed. In some examples, the entropy measure of the contiguous data sequence 106 is a size of the compressed contiguous data sequence. The entroposcope module 108 may

associate an entropy measure to individual contiguous sequences of bytes of data in the input data 102. Generally, the entropy measure is a numerical quantity whose value may increase as information content increases. The value of the entropy measure may not generally increase as a function of a length of the data sequence to which it is associated.

[0014] The distribution processor 110 may be communicatively linked to a plurality of data containers (e.g., Data Container 1 112(1), Data Container 2 112(2), ..., Data Container X 112(x)). The plurality of data containers are data storage systems. In some examples, deduplication measures may be performed to data being stored. Generally, deduplication is performed to remove and/or reduce the extent to which individual data sequences overlap with one another in terms of data content. Generally, deduplication measures may be performed across multiple streams of data, and across wide-ranging time periods. Accordingly, overall entropy calculations may be more accurately indicative of an information content of a data container.

[0015] In some examples, data being stored may be compressed after deduplication measures have been performed. Standard compression may be applied to contiguous data sequences in individual streams of data.

[0016] In some examples, once deduplication and compression have been performed, the contiguous data sequence may be stored in a data container such as, for example, Data Container 1 112(1), Data Container 2 112(2), ..., Data Container X 112(x), with a one-to-many mapping between compressed data sequences and data containers.

[0017] The distribution processor 110 stores items of interest to users (e.g., files, data stream and so forth) by distributing the data that contained in such items over data containers on some internal storage medium. Generally, such internal data containers are liable to corruption and loss in malicious or accidental failure scenarios. Because of a complex mapping between user items and data containers, the loss of one such data container may result in a loss of an arbitrary amount of data from many user items.

[0018] Generally, when individual data sequences are stored internally by the system 100, data sequences may be assigned to data containers using a

selection process that ensures that a maximum information content associated to a data container is a minimum, when compared to all other possible assignments of sequences to containers. Accordingly, loss of any one data container may result in minimal data loss in terms of loss of information content.

[0019] The distribution processor 110 determines an information content of each data container of a plurality of data containers (e.g., Data Container 1 112(1), Data Container 2 112(2), ..., Data Container X 112(x)) by adding the entropy measure of the contiguous data sequence to entropy measures of data sequences already stored in the data container. In some examples, the information content of a data container is a sum of entropy measures of all data sequences stored in the data container. For example, the information content of a data container may be a sum of sizes of all compressed data sequences stored in the data container.

[0020] The distribution processor 110 selects, for storage of the contiguous data sequence 106, a data container based on minimizing the information content of the data container. For example, the distribution processor 110 may compare the determined information content of each data container of the plurality of data containers (e.g., Data Container 1 112(1), Data Container 2 112(2), ..., Data Container X 112(x)) by adding the entropy measure of the contiguous data sequence 106 to the entropy measures of data sequences already stored in the data container. Accordingly, at least one data container of the plurality of data containers (e.g., Data Container 1 112(1), Data Container 2 112(2), ..., Data Container X 112(x)) may be determined to have a minimum information content. Such a data container, say Data Container 1 112(1), may be selected for storage of the contiguous data sequence 106. The term “minimize”, as used herein, generally refers to “reducing something to a least amount”. In some examples, minimizing may mean determining a smallest of a plurality of comparable numerical quantities.

[0021] Figure 2 is a flow diagram illustrating one example of selecting a data container to minimize information loss. Various components described in Figure 2 may be components in system 100 described with reference to Figure 1. For example, the data processor may be the data processor 104, the

entroposcope module may be the entroposcope module 108, and the distribution processor may be the distribution processor 110.

[0022] At 200, a contiguous data sequence, S , is identified by the data processor. At 202, the entroposcope module determines an entropy measure for the sequence S as a compressed size of the sequence. For example, the entropy measure for the sequence S may be determined to be a million bytes. At 204, the distribution processor determines the information content of each data container. For example, an existing information content for data container 1 206(1) may be 40.4 million bytes. Likewise, a current information content for data container 2 206(2) may be 35 million bytes, and a current information content for data container 3 206(3) may be 55 million bytes. In some examples, the distribution processor may retrieve such current information content data from a lookup table. Based on such data, the distribution processor may determine the information content of each data container by adding the entropy measure of the sequence S to the current information content for each data container. Accordingly, the distribution processor may determine Information Content 1 (for Data Container 1 206(1)) to be $40.4 + 1 = 41.4$ million bytes; Information Content 2 (for Data Container 2 206(2)) to be $35 + 1 = 36$ million bytes; and Information Content 3 (for Data Container 3 206(3)) to be $55 + 1 = 56$ million bytes. Next, the distribution processor may determine the minimum of Information Content 1, Information Content 2, and Information Content 3, to be 36 million bytes. Accordingly, Data Container 2 206(2), corresponding to minimum information content, Information Content 2, may be selected by the distribution processor as the data container to store the contiguous data sequence, S .

[0023] In some examples, a selected data container may not be available for storage. For example, the selected data container may have attained its maximum allowable storage capacity. In such instances, another data container may be selected. For example, Data Container 1 206(1) has an information content smaller than Data Container 3 206(3), and the distribution processor may select Data Container 1 206(1) for storage. In some examples, when a data container is no longer available for data storage, such information may be made available to the distribution processor, and such unavailable data container may

not be considered for storage. For example, the distribution processor may not retrieve a current information content data for an unavailable data container. Also, for example, the distribution processor may not determine an information content data for an unavailable data container.

[0024] Referring again to Figure 1, in some examples, the distribution processor 110 may further allocate the contiguous data sequence 106 to the selected data container, e.g., Data Container 1 112(1), for storage. In some examples, such allocation may include providing write instructions to the selected data container. In some examples, such allocation may include further deduplication, compression, segmentation, and/or other data storage tasks to be applied to the contiguous data sequence 106 prior to actual storage.

[0025] In some examples, the selected data container, e.g., Data Container 1 112(1), stores the allocated contiguous data sequence, based, for example, on instructions received from the distribution processor 110. In some examples, the selected data container, e.g., Data Container 1 112(1), may update its information content by adding the entropy measure of the allocated contiguous data sequence 106 to entropy measures of data sequences previously stored in the selected data container, e.g., Data Container 1 112(1).

[0026] The components of system 100 may be computing resources, each including a suitable combination of a physical computing device, a virtual computing device, a network, software, a cloud infrastructure, a hybrid cloud infrastructure that may include a first cloud infrastructure and a second cloud infrastructure that is different from the first cloud infrastructure, and so forth. The components of system 100 may be a combination of hardware and programming for performing a designated visualization function. In some instances, each component may include a processor and a memory, while programming code is stored on that memory and executable by a processor to perform a designated function.

[0027] For example, the entroposcope module 108 may be a combination of hardware and programming for determining an entropy measure of the contiguous data sequence 106. For example, the entroposcope module 108 may include programming to identify a compressed size of the contiguous data

sequence 106. The entroposcope module 108 may include hardware to physically store, for example, the determined entropy measure of the contiguous data sequence 106. For example, the determined entropy measure of the contiguous data sequence 106 may be stored in a lookup table to be optionally accessed by the distribution processor 110, and/or the plurality of data containers (e.g., Data Container 1 112(1), Data Container 2 112(2), ..., Data Container X 112(x)).

[0028] Likewise, the distribution processor 110 may be a combination of hardware and programming for determining an information content, and for selecting an appropriate data container. For example, the distribution processor 110 may include programming to determine the information content of each data container of the plurality of data containers (e.g., Data Container 1 112(1), Data Container 2 112(2), ..., Data Container X 112(x)) by adding the entropy measure of the contiguous data sequence to entropy measures of data sequences already stored in the data container. The distribution processor 110 may include hardware to physically store, for example, an updated lookup table of information content associated with each data container. Also, for example, the distribution processor 110 may include software programming to dynamically interact with the other components of system 100.

[0029] Generally, the components of system 100 may include programming and/or physical networks to be communicatively linked to other components of system 100. In some instances, the components of system 100 may include a processor and a memory, while programming code is stored and on that memory and executable by a processor to perform designated functions.

[0030] Generally, input data 102 may be received via computing devices. A computing device, as used herein, may be, for example, a web-based server, a local area network server, a cloud-based server, a notebook computer, a desktop computer, an all-in-one system, a tablet computing device, a mobile phone, an electronic book reader, or any other electronic device suitable for provisioning a computing resource to perform a unified visualization interface. The computing device may include a processor and a computer-readable storage medium.

[0031] Figure 3 is a block diagram illustrating one example of a computer readable medium for minimizing information loss in a data container. Processing system 300 includes a processor 302, a computer readable medium 308, input devices 304, and output devices 306. Processor 302, computer readable medium 308, input devices 304, and output devices 306 are coupled to each other through a communication link (e.g., a bus).

[0032] Processor 302 executes instructions included in the computer readable medium 308. Computer readable medium 308 includes data receipt instructions 310 to receive input data via a processing system.

[0033] Computer readable medium 308 includes sequence identification instructions 312 to identify a contiguous data sequence in the input data.

[0034] Computer readable medium 308 includes entropy measure determination instructions 314 to determine an entropy measure of the contiguous data sequence.

[0035] Computer readable medium 308 includes information content determination instructions 316 to determine an information content of each data container of a plurality of data containers by adding the entropy measure of the contiguous data sequence to entropy measures of data sequences already stored in the data container.

[0036] Computer readable medium 308 includes data container selection instructions 318 to select, for storage of the contiguous data sequence, a data container based on minimizing the information content of the data container.

[0037] Computer readable medium 308 includes data storage instructions 320 to store the allocated contiguous data sequence in the selected data container.

[0038] Input devices 304 include a keyboard, mouse, data ports, and/or other suitable devices for inputting information into processing system 300. In some examples, input devices 304, such as a computing device, are used by the interaction processor to receive input data. Output devices 306 include a monitor, speakers, data ports, and/or other suitable devices for outputting information from processing system 300. In some examples, output devices 306 are used to provide an information content of a data container.

[0039] As used herein, a “computer readable medium” may be any electronic, magnetic, optical, or other physical storage apparatus to contain or store information such as executable instructions, data, and the like. For example, any computer readable storage medium described herein may be any of Random Access Memory (RAM), volatile memory, non-volatile memory, flash memory, a storage drive (e.g., a hard drive), a solid state drive, and the like, or a combination thereof. For example, the computer readable medium 308 can include one of or multiple different forms of memory including semiconductor memory devices such as dynamic or static random access memories (DRAMs or SRAMs), erasable and programmable read-only memories (EPROMs), electrically erasable and programmable read-only memories (EEPROMs) and flash memories; magnetic disks such as fixed, floppy and removable disks; other magnetic media including tape; optical media such as compact disks (CDs) or digital video disks (DVDs); or other types of storage devices.

[0040] As described herein, various components of the processing system 300 are identified and refer to a combination of hardware and programming configured to perform a designated visualization function. As illustrated in Figure 3, the programming may be processor executable instructions stored on tangible computer readable medium 308, and the hardware may include processor 302 for executing those instructions. Thus, computer readable medium 308 may store program instructions that, when executed by processor 302, implement the various components of the processing system 300.

[0041] Such computer readable storage medium or media is (are) considered to be part of an article (or article of manufacture). An article or article of manufacture can refer to any manufactured single component or multiple components. The storage medium or media can be located either in the machine running the machine-readable instructions, or located at a remote site from which machine-readable instructions can be downloaded over a network for execution.

[0042] Computer readable medium 308 may be any of a number of memory components capable of storing instructions that can be executed by Processor 302. Computer readable medium 308 may be non-transitory in the sense that it does not encompass a transitory signal but instead is made up of

one or more memory components configured to store the relevant instructions. Computer readable medium 308 may be implemented in a single device or distributed across devices. Likewise, processor 302 represents any number of processors capable of executing instructions stored by computer readable medium 308. Processor 302 may be integrated in a single device or distributed across devices. Further, computer readable medium 308 may be fully or partially integrated in the same device as processor 302 (as illustrated), or it may be separate but accessible to that device and processor 302. In some examples, computer readable medium 308 may be a machine-readable storage medium.

[0043] Figure 4 is a flow diagram illustrating one example of a method for minimizing information loss in a data container.

[0044] At 400, input data may be received via a processing system.

[0045] At 402, a contiguous data sequence may be identified in the input data.

[0046] At 404, an entropy measure of the contiguous data sequence may be determined.

[0047] At 406, an information content of each data container of a plurality of data containers may be determined by adding the entropy measure of the contiguous data sequence to entropy measures of data sequences already stored in the data container.

[0048] At 408, a data container may be selected for storage of the contiguous data sequence, the selection based on minimizing the information content of the data container.

[0049] At 410, the contiguous data sequence may be allocated to the selected data container for storage.

[0050] In some examples, the method may further include storing the allocated contiguous data sequence in the selected data container. In some examples, the method may further include updating the information content by adding the entropy measure of the allocated contiguous data sequence to entropy measures of data sequences previously stored in the selected data container.

[0051] In some examples, the method may further include compressing the contiguous data sequence. In some examples, the method may further include performing a deduplication of the contiguous data sequence prior to the compressing of the contiguous data sequence. In some examples, the entropy measure of the contiguous data sequence may be a size of the compressed contiguous data sequence.

[0032] Examples of the disclosure provide a generalized system for minimizing information loss in a data container. The generalized system minimizes the impact in terms of information loss of malicious or accidental data loss by smoothing the distribution of data across multiple exposed locations based on the data's entropic information content.

[0033] Although specific examples have been illustrated and described herein, a variety of alternate and/or equivalent implementations may be substituted for the specific examples shown and described without departing from the scope of the present disclosure. This application is intended to cover any adaptations or variations of the specific examples discussed herein. Therefore, it is intended that this disclosure be limited only by the claims and the equivalents thereof.

CLAIMS

1. A system comprising:
 - a data processor to receive input data and identify a contiguous data sequence in the input data;
 - an entroposcope module to determine an entropy measure of the contiguous data sequence;
 - a distribution processor to:
 - determine an information content of each data container of a plurality of data containers by adding the entropy measure of the contiguous data sequence to entropy measures of data sequences already stored in the data container, and
 - select, for storage of the contiguous data sequence, a data container based on minimizing the information content of the data container.
2. The system of claim 1, wherein the distribution processor is to further allocate the contiguous data sequence to the selected data container for storage.
3. The system of claim 1, wherein the selected data container is to store the allocated contiguous data sequence.
4. The system of claim 3, wherein the selected data container is to update the information content by adding the entropy measure of the allocated contiguous data sequence to entropy measures of data sequences previously stored in the selected data container.
5. The system of claim 1, further comprising a data compressor to compress the contiguous data sequence.
6. The system of claim 5, wherein the data processor is to further perform a deduplication of the contiguous data sequence prior to compression.

7. The system of claim 5, wherein the entropy measure of the contiguous data sequence is a size of the compressed contiguous data sequence.

8. A method comprising:
receiving input data via a processing system;
identifying a contiguous data sequence in the input data;
determining an entropy measure of the contiguous data sequence;
determining an information content of each data container of a plurality of data containers by adding the entropy measure of the contiguous data sequence to entropy measures of data sequences already stored in the data container;
selecting, for storage of the contiguous data sequence, a data container based on minimizing the information content of the data container; and
allocating the contiguous data sequence to the selected data container for storage.

9. The method of claim 8, further comprising storing the allocated contiguous data sequence in the selected data container.

10. The method of claim 9, further comprising updating the information content by adding the entropy measure of the allocated contiguous data sequence to entropy measures of data sequences previously stored in the selected data container.

11. The method of claim 8, further comprising compressing the contiguous data sequence.

12. The method of claim 11, further comprising performing a deduplication of the contiguous data sequence prior to the compressing of the contiguous data sequence.

13. The method of claim 11, wherein the entropy measure of the contiguous data sequence is a size of the compressed contiguous data sequence.

14. A non-transitory computer readable medium comprising executable instructions to:

receive input data via a processing system;

identify a contiguous data sequence in the input data;

determine an entropy measure of the contiguous data sequence;

determine an information content of each data container of a plurality of data containers by adding the entropy measure of the contiguous data sequence to entropy measures of data sequences already stored in the data container;

select, for storage of the contiguous data sequence, a data container based on minimizing the information content of the data container; and

store the allocated contiguous data sequence in the selected data container.

15. The non-transitory computer readable medium, further comprising instructions to compress the contiguous data sequence.

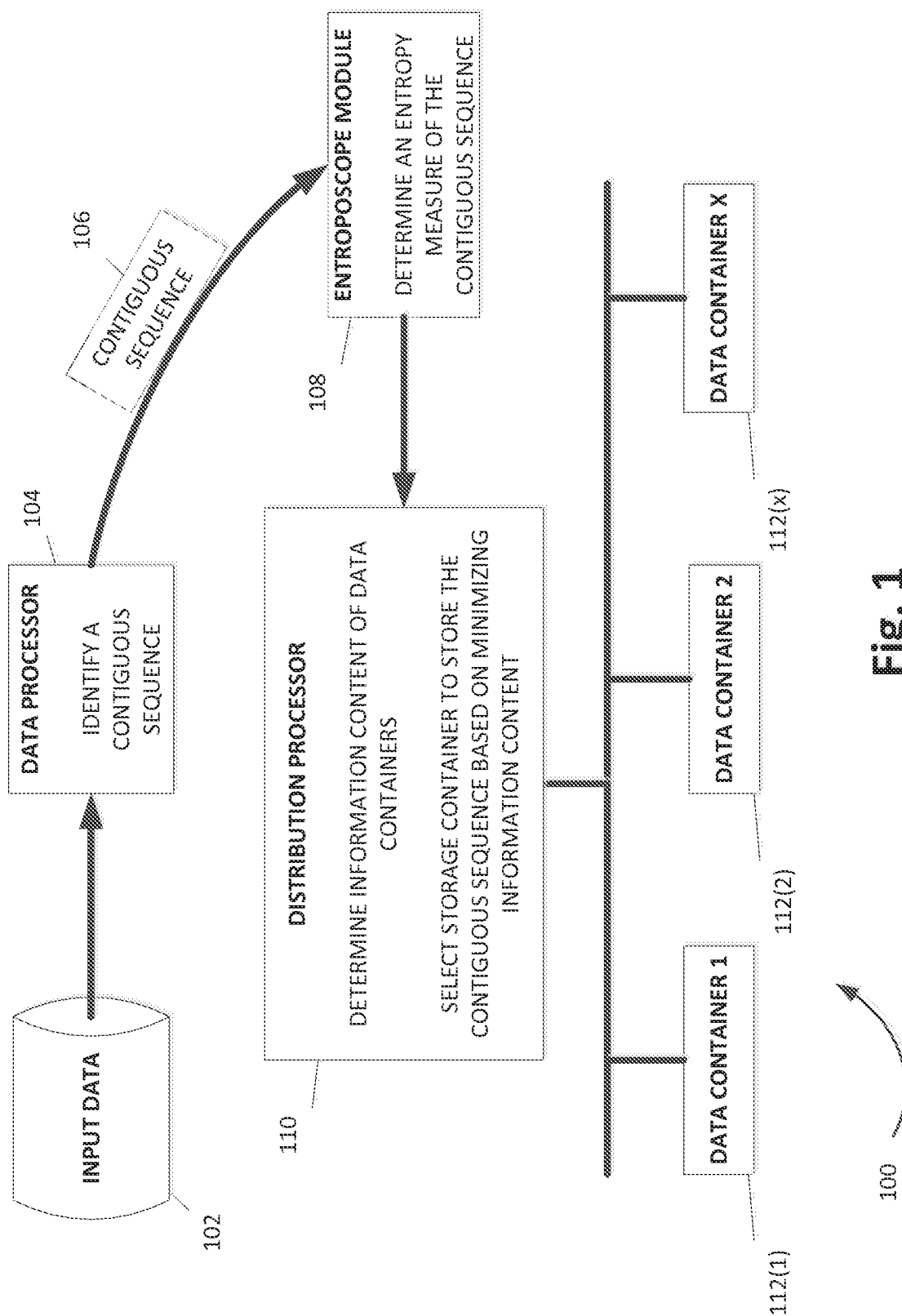


Fig. 1

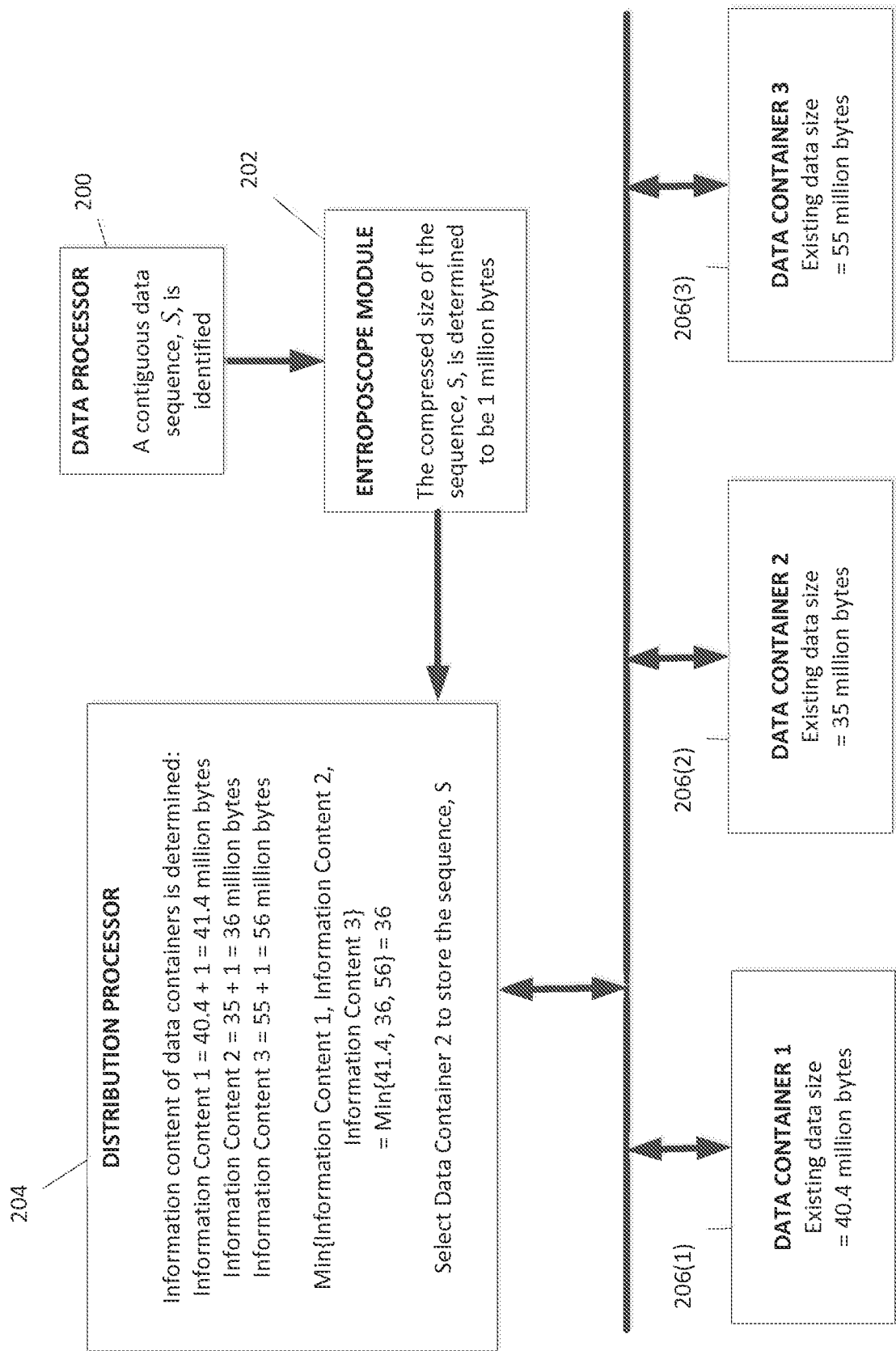


Fig. 2

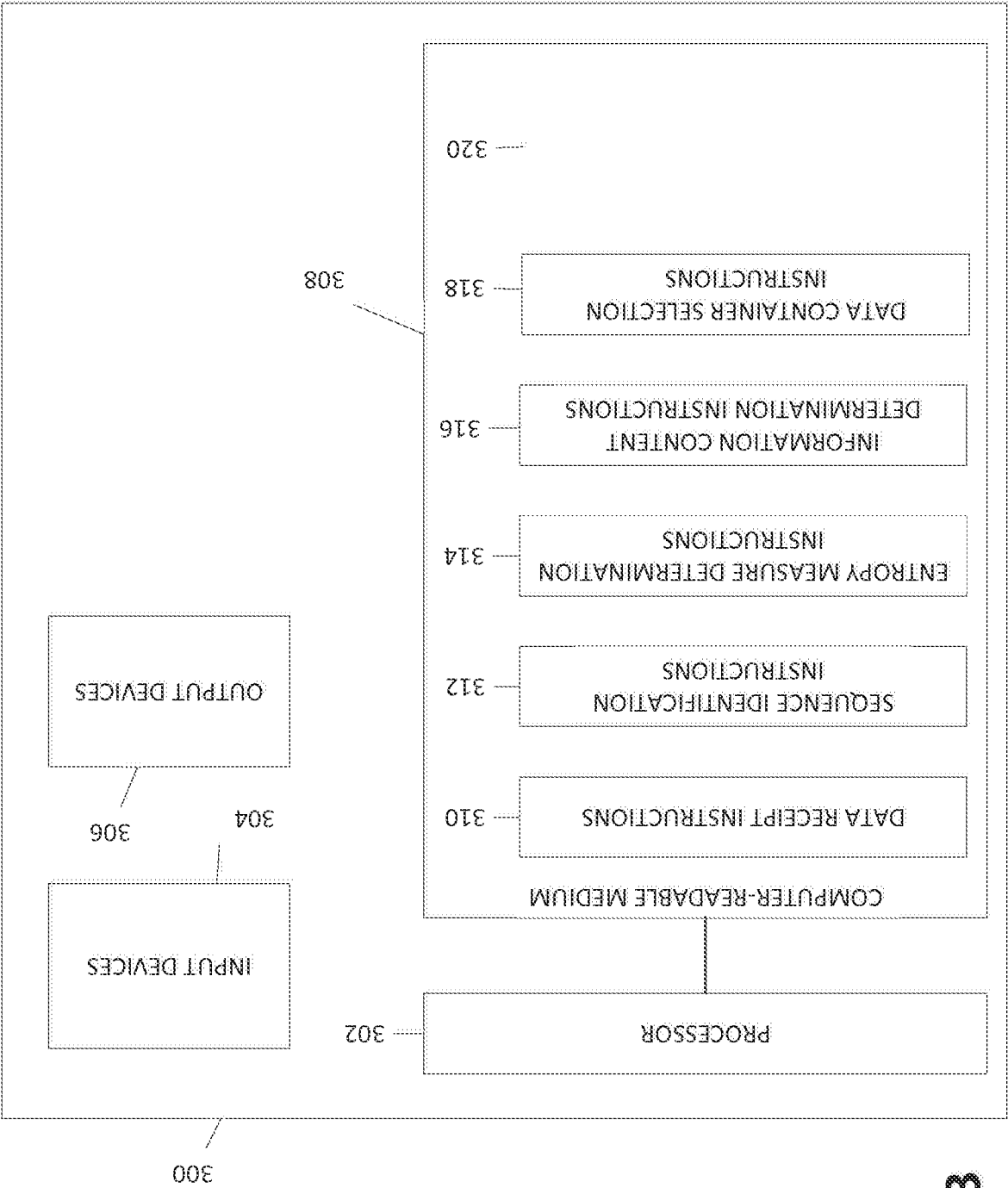


Fig. 3

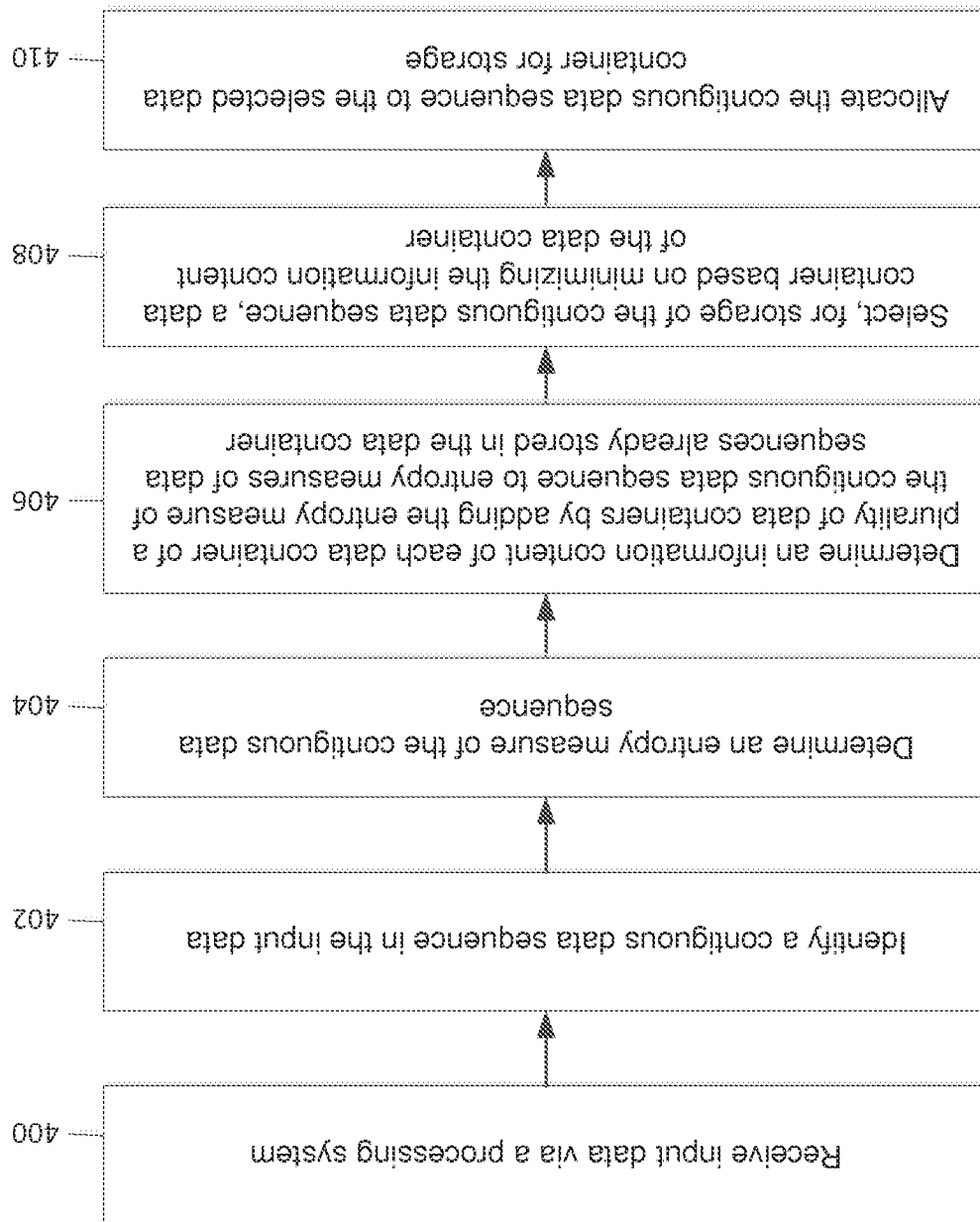


Fig. 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2015/029608**A. CLASSIFICATION OF SUBJECT MATTER****G06F 13/38(2006.01)i, G06F 13/40(2006.01)i**

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 13/38; G06F 15/00; G06F 17/00; H04N 7/167; H04N 7/26; H04L 1/08; H03M 7/40; G06F 13/40

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS(KIPO internal) & Keywords: contiguous data sequence, entropy measure, information content, data container, data deduplication, data compression, and similar terms.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2003-0090398 A1 (NENAD RIJAVEC) 15 May 2003 See paragraphs [0003], [0031], [0038], [0042], [0057], [0087], and [0102].	1-15
A	US 2011-0013701 A1 (XAVIER HENOCQ et al.) 20 January 2011 See paragraphs [0001], [0018]-[0021], [0040], [0102], and [0294].	1-15
A	US 2013-0289756 A1 (BARBARA RESCH et al.) 31 October 2013 See paragraphs [0047], [0050], [0063], and [0107].	1-15
A	US 2011-0075217 A1 (TIM PREBBLE) 31 March 2011 See paragraphs [0010], [0016], [0020], and [0040].	1-15
A	US 2014-0376720 A1 (FRANCIS YEE-DUG CHAN et al.) 25 December 2014 See paragraphs [0008], [0010], [0055], and [0065].	1-15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

05 February 2016 (05.02.2016)

Date of mailing of the international search report

05 February 2016 (05.02.2016)

Name and mailing address of the ISA/KR

International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon Metropolitan City, 35208,
Republic of Korea

Facsimile No. +82-42-472-7140

Authorized officer

NHO, Ji Myong

Telephone No. +82-42-481-8528



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2015/029608

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2003-0090398 A1	15/05/2003	US 2003-0090706 A1 US 2003-0090709 A1 US 6653954 B2 US 7043077 B2 US 7068398 B2	15/05/2003 15/05/2003 25/11/2003 09/05/2006 27/06/2006
US 2011-0013701 A1	20/01/2011	EP 2285122 A1 EP 2285122 B1 US 8462854 B2	16/02/2011 13/11/2013 11/06/2013
US 2013-0289756 A1	31/10/2013	CN 103999150 A EP 2659480 A1 EP 2659481 A1 EP 2659482 A1 EP 2659483 A1 EP 2791935 A1 EP 2793223 A1 JP 2015-505992 A US 2013-0275421 A1 US 2013-0282388 A1 US 2013-0287214 A1 US 2014-0330556 A1 WO 2012-089313 A1 WO 2012-091935 A1 WO 2012-091935 A9 WO 2012-091936 A1 WO 2012-091938 A1 WO 2013-090207 A1	20/08/2014 06/11/2013 06/11/2013 06/11/2013 06/11/2013 22/10/2014 22/10/2014 26/02/2015 17/10/2013 24/10/2013 31/10/2013 06/11/2014 05/07/2012 05/07/2012 26/09/2013 05/07/2012 05/07/2012 20/06/2013
US 2011-0075217 A1	31/03/2011	US 8782094 B2	15/07/2014
US 2014-0376720 A1	25/12/2014	CA 2782825 A1 EP 2507995 A1 EP 2507995 A4 JP 2013-513298 A JP 5723888 B2 US 2011-135090 A1 US 8781122 B2 WO 2011-068668 A1	09/06/2011 10/10/2012 09/07/2014 18/04/2013 27/05/2015 09/06/2011 15/07/2014 09/06/2011