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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 December 2011 (22.12.2011)

PCT

(10) International Publication Number
WO 2011/157708 A1

(51) International Patent Classification:
G06F 21/00 (2006.01)

(21) International Application Number:
PCT/EP2011/059846

(22) International Filing Date:
14 June 2011 (14.06.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10075257.5 14 June 2010 (14.06.2010) EP
10 2010 023 894.5 16 June 2010 (16.06.2010) DE

(71) Applicant *(for all designated States except US)*: **FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V.** [DE/DE]; Hansastraße 27c, 80686 München (DE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BREITENSTROM, Christian** [DE/DE]; Kavalierstraße 19A, 13357 Berlin (DE). **SCHÜRMANN, Gerd** [DE/DE]; Dahmestraße 4a, 15738 Zeuthen (DE). **POPESCU-ZELETIN, Radu** [DE/DE]; Wissmannstraße 21, 14193 Berlin (DE). **KLESSMANN, Jens** [DE/DE]; Bellermannstraße 78, 13357 Berlin (DE). **PENSKI, Andreas** [DE/DE]; Max-Hagen-Weg 13a, 17491 Greifswald (DE).

(74) Agent: **GROSS, Felix**; Patentanwälte Maikowski & Ninnemann, Postfach 15 09 20, 10671 Berlin (DE).

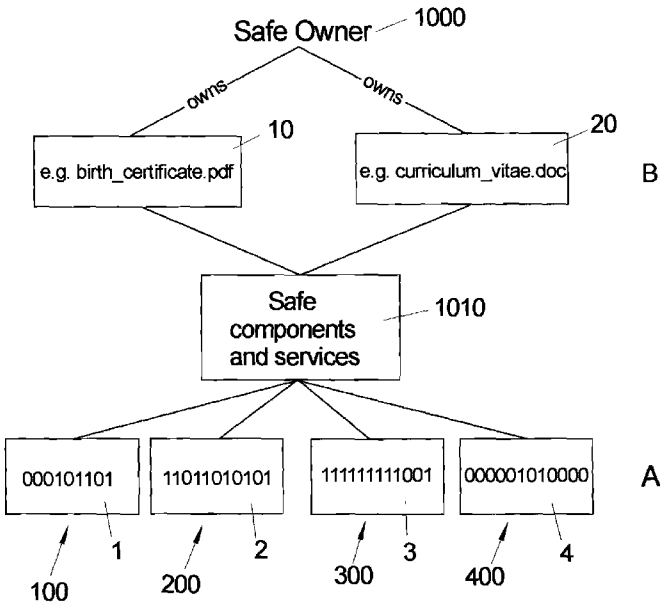
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD,

[Continued on next page]

(54) Title: METHODS AND SYSTEMS FOR SECURELY HANDLING DATASETS IN COMPUTER SYSTEMS

(57) Abstract: Methods and systems for safely handling at least one dataset, in particular a document (10, 20), particularly in a cloud computer environment are described, e.g. wherein a) the at least one dataset (10, 20) is partitioned into at least two dataset partitions (1, 2, 3, 4), b) the at least two dataset partitions (1, 2, 3, 4) are stored on and / or retrieved from at least two computer sites (100, 200, 300, 400) being part of, e.g., a cloud computer environment, so that on no computer site (100, 200, 300, 400) sufficient, in particular all, dataset partitions (1, 2, 3, 4) of the at least one dataset (10, 20) are present, in particular not present at the same time.

FIG 3



WO 2011/157708 A1

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, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

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

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

Methods and Systems for securely handling datasets in computer systems

Systems for collaborative sharing of datasets such as documents („cloud-ready“) are known in the art. However the trust in these cloud-based infrastructures to store personal, sensitive and /or confidential data is limited.

In the following a cloud computer environment is understood to be an environment where at least one cloud computer system is provided in addition to classical storage systems like individual computers, external storage media like USB sticks or smartcards etc.

Cloud computing describes in general IT services based on a network, in particular the Internet, and it typically involves the provision of dynamically scalable and often virtualized resources as a service over the network. Typical cloud computing providers deliver business applications online which are accessed from another web service or software like a web browser, while the software and data are stored on servers.

Solutions to prevent data leaks often get complex and the cloud customers have no chance to control the security measures taken by the cloud service provider to prevent insider/outsider attacks. The investment in data leak prevention tools and the necessary security consulting is often avoided.

The cloud customer can only rely on Service Level Agreements with the cloud provider based on certifications and regular security reviews by internal and trusted third parties. Even the cloud space providers have the problem to prove that every endeavor has been made to keep the customers data confidential. So they spend lots of money without being sure

or able to guarantee that their own administrators are not leaking confidential data.

Anonymous credential systems have been described by

Chaum, D. ("Security without identification: Transaction systems to make big brother obsolete", *Communications of the ACM 28(10),* pp1030-1044) and

Camenisch, J. and Herreweghen, E.V. ((2002) "Design and implementation of the idemix anonymous credential system", *CCS '02: Proceedings of the 9th ACM conference on Computer and communications security*, New York, NY, USA, ACM Press, pp 21-30.).

Secret sharing has been described by Shamir and Rabin (Rabin, M (1989), "Efficient dispersal of information for security, load balancing, and fault tolerance", *Journal of the ACM,* Vol.36, pp. 335—348).

Since the mid 90s technical infrastructures suitable for electronic document safes are discussed with different objectives in mind:

- performance and fault tolerance: (Paul, A. et al. (2007), "e-SAFE: An Extensible, Secure and Fault Tolerant Storage System", *Proc. IEEE Self-Adaptive and Self-Organizing Systems*,(SASO '07), IEEE Press, pp. 257-268, doi: 10.1109/SASO.2007.21,

- confidentiality in dispersed untrusted storage environments (Zhang, et al. (2008), „Towards A Secure Distribute Storage System", *Advanced Communication Technology*, (ICACT 2008), IEEE Press, Apr. 2008, pp. 1612-1617, doi:10.1109/ICACT.2008.4494090), (US-A 7,349,987), (Iyengar, A. et al. (1998), "Design and Implementation of a Secure Distributed Data Repository", *In Proc. of the 14th IFIP*

Internat. Information Security Conf., pp.123—135.) (Kubiatowicz et al. (2000), "Oceanstore: An architecture for global-scale persistent storage", *Proc. of the ninth international conference on Architectural support for programming languages and operating systems (ASPLOS '00)*, ACM SIGARCH Computer Architecture News, Dec. 2000, pp.190-201, ISSN:0163-5964), as distributed p2p systems (US Patent Application 20060078127), as smart card extension (US-A 7,206,847), as backup.

It is necessary to develop methods and systems for securely handling datasets, in particular files or documents, in a cloud computing environment. In following a number of embodiments are described which address this issue. A person skilled in the art will recognize that those embodiments can be combined to derive variations of the embodiments described.

One embodiment is a method for safely handling at least one dataset, in particular a document, in a cloud computer environment, wherein
a) the at least one dataset is partitioned into at least two dataset partitions,
b) the at least two dataset partitions are stored on and / or retrieved from at least two computer sites being part of a cloud computer environment, so that on no computer site all dataset partitions of the at least one dataset are present, in particular not present at the same time.

Storing datasets, e.g. sensitive documents, distributed over a cloud computer environment enhances the security of the individual documents. Access to an individual partition of the dataset would not compromise the security of the complete documents since the partition itself does not provide sufficient information to reconstruct the whole document.

Another embodiment is a method for safely handling at least one dataset, in particular a document, in a cloud computer environment, wherein a retrieval of a partitioned dataset to combine the at least one dataset comprises

a) an authorization by an authorized user, in particular an owner of the at least one dataset at a safe system,

b) on positive completion of the authorization, distribution information regarding the dataset partitions is automatically retrieved from the safe system,

c) all dataset partitions of the at least one dataset are retrieved from computer sites,

d) the dataset partitions are transformed into a logical representation of the at least one dataset.

Here a safe system stores the distribution-information of the datasets which are to be distributed to e.g. storage providers. The term “authorization” is meant in this and the following contexts that the user, in particular an owner, is in the possession of a right to do something.

A further embodiment is a method for safely handling at least one dataset, in particular a document, in a cloud computer environment, wherein a storing of a partitioned dataset derived from the at least one dataset comprises

a) an authorization by a user, in particular an owner of the at least one dataset at a safe system,

b) on positive completion of the authorization, distribution information regarding the dataset partitions is automatically transferred to the safe system,

c) all dataset partitions of the at least one dataset are automatically stored on computer sites as physical sub-representations of the at least one dataset, so that none of the computer sites comprises more than one instance of the dataset partitions, in particular not at the same time.

We use a distribution mechanism, especially a physical distribution mechanism of the data, in particular documents,

across multiple independent storage servers to increase the security of the data and documents and increase the trust in the Cloud environment.

Different embodiments are described in an exemplary way in the following figures,

Fig. 1 showing schematically a document with several document partitions and its distribution;

Fig. 2 showing schematically a document partitions distributed over several servers;

Fig. 3 showing schematically a logical and physical representation of data;

Fig. 4 showing schematically the interactions of a safe owner;

Fig. 5 showing an embodiment of a read mechanism involving a safe owner, a safe provider and storage providers;

Fig. 6 showing an actor model for an electronic safe system;

Fig. 7 showing an electronic life cycle of an safe system

Fig. 8 showing the collection of data involving a safe system owner and a safe user;

Fig. 9 showing a safe user infrastructure;

Fig. 10 showing the process of a safe user sending a release order request to a safe system.

In a first embodiment, the method and the system are used to improve the security of datasets 10, 20, in particular documents (10, 20).

A document 10, 20 is a special instance of a computer file, comprising blocks of arbitrary information or a resource for storing information, being available for a computer program. Therefore, documents 10, 20 or files are to be considered separate from programs, which can be used to, e.g., create, alter or store documents or files. A dataset 10, 20 is understood to be even more general than a file since it makes not inherent assumption about the internal structure. A dataset is understood to be a set of binary data which can be distinguished from other datasets.

Even though the embodiments described below are not limited to documents 10, 20, for the sake of simplicity the embodiments are described in the context of documents (10, 20), rather than datasets.

Before describing embodiments of methods and systems, a distinction in the representation of documents 10, 20 (or files) is given. For the sake of simplicity, the embodiment will be described in the context of documents 10, 20 only.

A logical representation of a document 10, 20 is a representation that enables a person 1000 with the correct authorizations for file usage on a computer system to, e.g., distribute, retrieve, interpret and process a physical distribution of the document 10, 20.

The physically distributed representation of a document 10, 20 contains distribution-information (e.g. a distribution tag 11, 12, 13) that enables the person 1000 authorized to access the document 10, 20 to transform the document 10, 20 back to its logical representation. The tag name 11, 12, 13, which can be the name of the file (or a part thereof), allows the identification of the document partitions 1, 2, 3, 4.

The document 10, 20 can be encrypted and then be subjected to a partition into the partitions 1, 2, 3, 4. The partitions can be stored at the storage providers under a number. The storing with the storage providers is performed together with the storing of a so called domain-pseudonym. This way the person 1000 authorized to access the document 10, 20 can read the documents 10, 20 without giving away the identity of the user, in particular the owner.

In the embodiment described here the physically distributed representation of the document 10, 20 is the physical partition of the document 10, 20 into document partitions 1, 2, 3, 4. Those document partitions 1, 2, 3, 4 can be distributed over different independent storage providers. A storage provider is a service to store arbitrary datasets, in particular documents 10, 20. It might be a professional IT service provider, a cloud provider, personal computers and / or mobile storage devices (USB cards, mobile phone etc).

Embodiments of the methods and systems for administering documents 10, 20 in distributed form is illustrated in Fig. 1, 2 and 3.

In Fig. 1 a document 10 is shown as one logical representation, as it exists on one particular computer system, e.g. a safe client or mobile device. The document 10 is divided into three different document partitions 1, 2, 3, each of the document partitions 1, 2, 3 is associated with a distribution tag 11, 12, 13 as distribution-information. The partitions 1, 2, 3 are stored in different computer sites 100, 200, 300 at storage providers.

If the individual document partitions 1, 2, 3 are known to a safe system 1010 (see Fig. 3, i.e. the distribution-information is known to the safe system 1010) they can be recombined to the document 10 in the correct order using the

distribution tags 11, 12, 13. The document partitions 1, 2, 3 are datasets comprising binary data.

In principle it is possible to keep the distribution-information 11, 12, 13 on the client and distribute the partitions 1, 2, 3 directly to the computer sites 100, 200, 300.

The safe system 1010 is a personalized collection of services to store, retrieve and manage the physical distribution representations of documents 10, 20. The safe service provides the authorized user, in particular the owner of sensitive documents the capability to transform her documents from physical to logical representation and vice versa.

The document partitions 1, 2, 3 can be arbitrarily chosen, i.e. the document 10 can be arbitrarily divided. The document partitions 1, 2, 3 can, but they do not have to have the same size. The number of document partitions 1, 2, 3 is two or greater than two so that a distribution over different sites (e.g. computer servers) 100, 200, 300, 400 (see Fig. 2, 3) can be achieved. One way to partition a document 10, 20 is the secret sharing algorithm of Rabin. The method of Rabin produces inherently partly-redundant partitions in a configurable manner.

It is not mandatory to store redundant parts (e.g. obtained by the secret sharing algorithm of Rabin) of documents partitions 1, 2, 3’ on different servers 100, 200, 300 but this increases the availability and therefore the safety of the document storage. In Fig. 2 each of the document partitions 1, 2, 3 have at least one redundant part. If the document 10, was e.g. partitioned by the algorithm of Rabin, two of the three partitions would be sufficient to recombine the document 10.

The partitioning is not bound to any specific algorithm, especially algorithms providing redundancy may be applied to increase the safety against (temporary or permanent) inaccessibility of document partitions or systems due to failures.

To ensure the trustworthiness of the cloud computer environment only the person 1000 holding a sufficiently large subset or all of the document partitions 1, 2, 3 can recombine the document partitions 1, 2, 3 to the one document 10. The concrete number of partitions required depends on the partitioning algorithm applied. In the simplest case of non-redundant, non-overlapping partitioning, if one of the document partitions 1, 2, 3 is missing and / or corrupted, the document 10 cannot be recombined. So if an internal or external attacker gets access to a computer system with one of the document partitions 1, 2, 3 (or one of its copies), the content of the complete document 10 cannot be known to him. If a portion is sufficiently small and/or carries several, non-continuous parts of document 10, even if unencrypted, having access to one document portion 1, 2, 3 would not provide meaningful information to an attacker.

In the case of the secret sharing algorithms of Rabin, a subset of the document partitions 1, 2, 3, 4 suffices to recombine the document 10, 20.

In a further embodiment at least one of the document partitions 1, 2, 3 is encrypted. This can mean that the original file 10, 20 has been encrypted, so that its partitions 1, 2, 3 obtained their encryption from the original document 10, 20. Alternatively or additionally the document partitions 1, 2, 3 can be encrypted separately prior to or after the distribution of the document partitions 1, 2, 3 to different servers 100, 200, 300, 400.

Before the mechanisms of the distribution are discussed, the advantage of the separation of responsibilities is described.

A first mechanism in order to achieve trustworthy cloud spaces is the separation of responsibilities. For most sensitive private or business documents 10, 20 the owner 1000 has to be in full control of the document 10, 20.

A storage provider is responsible for availability, consistency and confidentiality of the data. A storage provider can be the user using his own computer, a professional storage provider outside or a cloud provider in the future. Currently the documents 10, 20 are stored logically and physically at the same location. This - by nature - makes them vulnerable to attacks from inside and outside the storage location. This is one of the issues the embodiments shown here address.

The separation into a logical and a physical view of documents 10, 20 allows a better separation of concerns in the storage and retrieval of sensitive documents 10, 20 and a better mechanism against breaching attacks. The separation of the logical representation of the document 10, 20 and its physically distributed representation of the document 10, 20 over different independent sites is described in Fig. 1, 2, 3. The owner 1000 controls the access to his documents 10, 20 via the safe system 1010.

The safe system 1010 logically comprises (i.e. information about the storage locations of the documents 10, 20) all documents 10, 20 and their related management information. The safe system 1010 might work with documents 10, 20 in different formats.

In Fig. 3 a safe owner 1000 owns two documents 10, 20. Those documents 10, 20 can be e.g. a pdf-file and a doc-file. Naturally the safe owner 1000 can own any number of documents

10, 20 which can all have the same file format or which can have different file formats as shown in Fig. 3.

The safe owner 1000, e.g., hires a safe system 1010 provided by the safe provider. In other alternatives the safe owner 1000 can have more than one safe 1010 with more than one safe provider. The safe provider is an entity or organization that provides safe systems and is liable for the safe. The safe system 1010 provider and the safe owner 1000 can be the same entity.

The safe owner 1000 owns the contents of his safe 1010 or safes.

In Fig. 3 the physically distributed representation A shows four instances of data (indicated by the binary strings). These are documents partitions 1, 2, 3, 4 of the documents 10, 20 which are distributed over several sites 100, 200, 300, 400.

Only the logical view B of the documents 10, 20 makes the document partitions 1, 2, 3, 4 readable and processable. In this sense the cloud (represented by systems 100, 200, 300, 400 here) becomes a necessary infrastructure to provide the distribution infrastructure for the documents 10, 20 at the physical level for higher safety and / or security of the logical documents. The distribution of the document partitions 1, 2, 3, 4 does not decrease the safety and security of the documents by distributing them over the net; the distribution rather improves the safety and security by not storing the document partitions 1, 2, 3, 4 in one place.

The physical data of the documents 10 is distributed as document partitions 1, 2, 3, 4 across different physical storage providers 100, 200, 300, 400, the client computer, internal and external storage devices so that every physical

storage provider can only access the physical portion of the document 10, 20 he owns.

In a further embodiment at least for a subset of the at least two document partitions 1, 2, 3, 4 the storage site 100, 200, 300, 400 dynamically change. This can be achieved by the safe systems 1010 which dynamically can allocate new storage sites. This can be made according to a random (or pseudo-random) process. The dynamic shifting of the portioned document makes an unauthorized retrieval of the document 10, 20 even more difficult.

The transformation and the recovery mechanism is realized at the client side and controlled by the owner 1000 of the documents 10, 20.

External attacks on one hand can be successful only if a sufficiently large number of (in the simplest case all) physical parts of the data (i.e., document partitions 1, 2, 3, 4) is (are) online (reducing the time of a potential attack) and by breaching the independent storage sites of the document. Simultaneous attacks at different locations are in general very difficult if not impossible to achieve.

This required synchronity can be deliberately used by setting a schedule so that the distributed storage sites for the documents partitions 1, 2, 3, 4 are only online on certain times, known to the owner 1000 only. This is another example that distributing data (e.g. document partitions 1, 2, 3, 4) over a cloud can increase safety rather than decrease it.

Each storage provider is responsible only for the physical part of the document 10, 20 (i.e. the document partitions 1, 2, 3) which was assigned to him. He is not responsible for confidentiality of the document 10, 20, as each single physical document partition does not reveal any information

about the original document 10, 20 (logical view of the document).

If the document partitions are distributed to several storage providers, any insider attack (at any storage provider) cannot be successful since only parts of the physical data (i.e,. the some individual document partitions 1, 2, 3, 4) can be addressed.

The safe mechanism provides the trustworthy storage capability of sensitive documents 10. On top of this mechanism the usual encryption mechanisms can be used to increase the safety and / or security of the systems.

A safe client (e.g. a mobile device connectable to the safe system 1010) may encrypt or/and partition the document and store it at independent Storage Providers (see Fig. 2 and 3). The safe client may distribute the physical document representation based on different algorithms to independent Storage Providers.

This distribution and storage is performed at interconnected storage units as: mobile devices (USB, smartcards, mobile phones etc.), own computer, cloud(s). Connectivity between these storage devices and the safe client is a prerequisite. The confidentiality is achieved by physical document 10, 30 distributions, not by trust and regulation.

This is shown in Fig. 4. Here an owner 1000 is interacting with a mobile device 1001. As shown in Fig. 3 the document partitions 1, 2, 3, 4 are stored in a distributed way on sites 100, 200, 300, 400.

A mobile device 1001 becomes a safe owner key, increasing the user perception and responsibility towards trust and security in the cloud storage system. Actually the owner 1000 of the mobile device 1001 (e.g. mobile phone, mobile computer) can

temporarily have access to his documents 10, 20, i.e., access the document partitions 1, 2, 3, 4 from the cloud of the sites 100, 200, 300, 400, combine the document partitions 1, 2, 3, 4 to the document 10, 20 on the mobile device 1001, work with the documents 10, 20 and then transmit the document partitions 1, 2, 3, 4 back to the cloud. The documents 10, 20 would not be stored on the mobile device 1001, reducing the risk that sensitive documents 10, 20 are lost with the mobile device.

As mentioned above the logical information is split at the client side. That means, that the only location, where the logical data is available in plain text format (or in another human readable format), is on the safe client. Different mechanisms can be provided to secure the safe client.

Fig. 5 shows an embodiment of a read mechanism involving a safe owner (safe client), safe provider (with the safe system 1010) and storage providers.

The read process is started with an authorization of the owner 1000 at his or her safe system 1010 at the safe provider administering the safe system 1010. From the safe provider, the distribution information about the document partitions 1, 2, 3, 4 is obtained, the information is decrypted. In general it is advantageous to encrypt the stored data (e.g. distribution information) at the safe provider.

Now the individual document partitions 1, 2, 3, 4 are retrieved from the storage providers. This is a loop which lasts until the complete document 10 is restored to the safe client of the owner 1000, by first assembling the physical partitions of the documents 10.

Typically there are n Storage Providers involved. The degree of security of the safe concept increases with the number of independent storage providers for a certain document 10, 20.

The write interaction occurs in reverse order: first the safe client stores the physical partitions at independent storage providers, then she encrypts the Meta-information and at last she stores this encrypted Meta-information within her safe provided by the safe provider. The authorization (and authentication) makes sure, that only the legitimate owner 1000 of the safe system 1010 may access it.

In the following further aspects and embodiments of the safe handling of datasets 1, 2, 3, 4 are described, in particular with applications to electronic safes for process Oriented e-Government. The person skilled in the art will recognize, that the embodiments described above are applicable to the methods and systems described below and vice-versa.

Today highly available, scalable storage is very common and cheap and many “document safes” or collaborative document sharing systems are available on the market. However the trust in server based infrastructures to store personal, sensitive data is limited. Major data breaches in public and private sector service providers suggest that there is a need for a more structural answer to these problems.

Solutions to prevent data leaks often get complex and expensive on their own, so that public administrations tend to avoid the investment in security consulting and the necessary infrastructure. Even the operating of such complex infrastructures requires skilled and engaged personal which makes it expensive. To rely on organizational measures and the operating stuff means to be attackable to a certain degree.

Last but not least today's e-government solutions depend heavily on the experienced end user that is capable to keep his personal computer clean from viruses and malware. This seems at least questionable, as it is challenging to keep up with the development of new anti-virus software.
In the following a model for critical infrastructures is described that keeps personal, sensitive datasets 10, 20 confidential for a long time without relying on trustworthy IT solution providers, loyal personal or secure networks. The proposed infrastructure can be used to store documents 10. 20, e.g. the scan of a birth certificate as well as XML data, e.g. "place of residence" that can be used by the citizen in all kinds of application processes. The confidentiality of stored datasets 10, 20, the unobservability of communication and the unlinkability of user transactions are targeted.

The aspects of secure storage are separated and a concept of a trustworthy electronic safe system 1010 for data and documents is developed. In principle it is possible to partition the documents on a client device, e.g. a mobile device. Then the distribution-information is kept on the client device, while the document partitions 1, 2, 3, 4 are distributed (i.e. stored) at various storage providers. In a further embodiment, a safe system 1010 is introduced to keep the distribution-information (preferably encrypted) on a safe system provided by a safe provider, while the partitioning and / or recombination of the partition is still performed on the client device.

The safe system 1010 infrastructure requires multiple actors, the safe owner 1000, that manage her private data within the safe system 1010, the storage provider that stores small data blocks of content, the safe provider that stores some kind of directory information and the safe user that is getting insight into the owner's safe system 1010.

It should be noted that the person 1000 can be the safe owner, but it does not necessarily has to be the case.

The description starts with the key requirements structured in use cases and their explanation. Afterwards essential parts of a prototypical implementation for all of the actors and the protocol between them for a chosen scenario are described.

Process oriented e-government

Traditionally public administrations are structured along public duties. Many public authorities and their departments are tailored accordingly. This orientation often leads to procedures, which end at the organizational boundaries of individual authorities. In the age of IT-based processes this task oriented approach is too narrow. In many cases administrative processes are just one part of larger processes, cross-cutting through organizational limits and different levels. Especially businesses have to provide preliminary results in order to fulfill administrative requirements. Once public authorities have handled the requests, the businesses have to re-integrate them into their internal processes. Thus the necessary cooperation of public and private sector can function with less friction if there are less format mismatches. In the interest of a high-performing European Union a process oriented alignment of the public sector is thus necessary.

Process oriented e-government is the result of the paradigm shift from a task-oriented, regionally distributed paper based public management towards the collaborative cooperation of various public agencies and service providers.

User centric process management emphasizes the central role of the citizen in e-government processes. The European Services Directive (DIRECTIVE 2006/123/EC) strengthens the

position of applicants in e-government processes in order to promote growth and create jobs in the European Union. The directive demands transparency of even complex processes. The applicant should be able to understand and control the steps in the application process.

Electronic Safes as trustworthy e-government infrastructures

An electronic safe system 1010 for datasets and in particular documents 10, 20 offers the safe owner 1010 confidentiality and availability of information stored in it. The storage facilities are organized in a strongly decentralized manner (e.g. by the methods and systems described above), so the safe systems 1010 is not a conventional database, registry or something similar.

The digital analogy of a conventional safe is capable to keep the usage and the communicating parties confidential. The resulting communication traces are useless for any eavesdropper. The information stored in the safe system 1010 is assured not to be recoverable even on the long term without the safe owner's 1000 or her delegate's consent.

Electronic safes (or safe systems 1010) are essential e-government infrastructure components, as they reduce the repetitive data collections at the beginning of each workflow. Beside personal data the electronic safe stores results of administrative processes in terms of electronic certificates. The public administration gains on much higher data quality when these certificates are reused later on. So the electronic safe is a modern privacy enhancing component and on the same time a critical infrastructure that makes e-government processes more efficient.

Electronic safes promise to secure the availability and integrity of data and documents as a basis for the decision-

making processes of the administration. They make it legitimate to relief the public administrations of the archiving of data and documents of the applicant, which is desirable for economic and privacy reasons.

Proposed model

Separation of responsibilities

The first idea behind the concept of electronic safes is the separation of responsibilities.

For most users it seems easy to store their personal information on their own computer. In this case the user is full in control of her data but is responsible for availability i.e. regular backups and confidentiality of the data. This is getting more and more time consuming, because it includes keeping the personal computer free of viruses and other malware. At the same time the user will only be able to access the data when she has physically access to her computer.

Both reasons lead to a situation where the user is ready to give up some of her control over the data. She orders a professional service to store her data with some “trusted storage provider” and will be able to access the data whenever she has access to the Internet. This could include the usage of the data in e-government processes too.

As this scenario is desirable it fails often, because we have a single instance that is responsible for available, confidential storage and we have network connections that can be eavesdropped.

It is proposed to split and e.g. encrypt the dataset, in particular documents 10, 20 into many pieces of a “puzzle”

and store these puzzle-pieces with independent storage providers. Each storage provider is responsible only for the available storage of a single piece of the data. He is not responsible for confidentiality, as each single piece does not reveal any information about the original “plain text” (i.e. the dataset or document 10, 20).

Any eavesdropper will gain only puzzle-pieces (probably even encrypted) of information that are useless without the others. So the confidentiality is achieved by design not by trust and regulation.

Different actors and their roles are described in Fig. 6. The Safe Owner 1000 is the natural person, who collects her data within the Safe. The Safe Provider is an IT solution provider that operates the Safe.

Multiple Safe Providers on the market can follow some standard protocols. So the Safe Owner 1000 can easily access all Safes with the same Safe Client software. The Safe User is the party that might receive grants to access parts of the personal data of the Safe Owner. A Safe User is able to send data to the Safe Owner, without needing explicit grants from the Safe Owner to do so. It is assumed that there is a constant “safe-address” that is known to the Safe User. The Storage Provider is an IT service provider that offers highly available storage services. Following the insight, that security should not rely only on technological means but should be supported by complementary legal and organizational provisions (Schneier 2009), a notary is established that takes responsibility in cases, where the Safe Owner is not capable to act personally, e.g. the Safe Owner dies or she loses the Safe access etc. The Safe Owner can act as a Safe User in relation to another Safe, as well as the Safe User might have its own Safe.

In a scenario related to e-government we consider public administrations as Safe Users. As described above, a Safe Owner might have multiple Safes with different Safe Providers. The Safe Infrastructure comprehends the Safe Client application, the Safe Provider web services, the Storage Provider web services and the anonymity services each with the underlying web container, database, trusted operating system and hardware.

Anonymous communication

Even with the separation of responsibilities outlined above an eavesdropping adversary would be able to infer communication patterns and communicating parties. Collaborating Storage Providers could use their combined knowledge to find the parts belonging together. Given some encryption used they would have to wait until the algorithm or the key size is weak enough to successfully decrypt the Safe Owner’s secrets. That’s why we need to make sure, that Storage Providers don’t know the origin of the incoming data. To prevent them to track the IP address we use available anonymity services.

However anonymous communication does not mean that everybody can use the storage services. Only the regular owner and the authorized user of the data will be able to retrieve the data. Communicating parties that use the Electronic Safe can verify each other’s identity. An accounting based on some subscription model is adopted. After all, the model is ready to be integrated into e-government processes.

Related work

The base technologies like anonymous credential systems (Chaum 1985) (Camenish 2002) and secret sharing (Rabin 1989) have been invented long time ago. Since the mid 90s technical infrastructures suitable for electronic document safes are

discussed with different objectives in mind: performance and fault tolerance (Paul 2007), confidentiality in dispersed untrusted storage environments (Zhang 2008) (Redlich 2008) (Iyengar 1998) (Kubiatowicz 2000), as distributed p2p systems (Cacayorin 2004), as smart card extension (Albert Jr. 2000), as backup etc. Document storage services come in various flavors: as collaborative environment, as file sharing facility, or as part of web-conferencing tools. There are huge document safe infrastructures deployed, e.g. the Austrian cyberdoc system, that connects notary's offices all over Austria or the Danish e-Boks system, that facilitates communication between Danish citizens, companies and public authorities - to name only two of them. None of these systems however provides the targeted privacy, availability and confidentiality on the long term. With the concepts presented here the Safe is incorporated into e-government processes while preserving the privacy keeping properties of the conventional safe. Cheap storage services are used - both to make it easy to come up with a new conformant storage service - so to enhance the secure base of independent Storage Providers and to make the safe operating more attractive for Safe Owners.

Requirements analysis

The history of electronic signature cards shows that it is a cumbersome task to roll out fundamental e-government infrastructures. We have to consider the whole life cycle of the electronic Safe, create real benefits for the Safe Owner, evaluate the organizational needs of public administration as Safe Users, make the model attractive for potential Safe Providers to name only some of the tasks. In particular the relation between real benefit for the Safe Owner and attractiveness for the Safe User to incorporate the electronic Safe into its processes is a chicken or the egg causality dilemma.

For the sake of brevity we concentrate on the requirements that directly have an impact on the Safe User's protocol and the integration into e-government processes. We use the key words as in RFC 2119 defined.

Trust model

Trustworthy infrastructures require fundamental measures, not only on a technical but also on the organizational and legal level. The underlying assumptions are:

- The IT-service providers are capable to provide highly available storage services,
- On the long term data losses within IT-service providers cannot be excluded,
- On the long term data breaches within IT-service providers from outside or inside will occur,
- It is possible, that different IT-service providers cooperate to recover the content of safes against the consent of a Safe Owner,
- There is a certain chance that there will be some kind of a superior instance which tries to control the independent IT service providers.

Finally: the Safe Owner is not assumed to be an IT affine responsible person who manages to keep her computer clean from viruses and malware.

Life-cycle of an electronic Safe

From the citizen's point of view the life cycle (see use cases in Fig. 7) of an electronic Safe starts and ends with the contract to a Safe Provider. Nevertheless there are other activities concerning the reliable transfer of Safe Services. Comparable to certificate service providers (CSP) the Safe Providers are liable to transfer their existing Safes to another Safe Provider when they close down their business.

The allocation of an electronic Safe implies the registration of the citizen with the Safe Provider. This may be achieved by personal identification or online with strong authentication, i.e. national ID-card. The Safe Owner buys the services around the electronic Safe from the Safe Provider. The Safe Provider issues personalized certificates and sends the certificates on some specialized hardware and the corresponding initialization passwords to the Safe Owner. Here the same security considerations apply as for the rollout of signature cards. The Safe Provider is responsible to allocate the reliable IT-resources in his own administrative domain as well as in the domains of the available Storage Providers.

The operating of the Safe comprises all the transactions of the Safe Owner, the storage of valuable personal information, the access of other Safe Users to released data and the information of the Safe Owner in case of incoming release order requests. The maintenance of the network of Safe Providers and Storage Providers is a substantial part of that position.

The locking and unlocking of an electronic Safe occurs when the Safe Owner has lost his credentials or does not pay the agreed subscription fees. In this case the Safe Owner will not be able to manage her data anymore but other Safe Users can still access the data previously released.

The Safe Owner might give up her electronic Safe. If she used the electronic Safe in e-government transactions there will be some other Safe Users, i.e. public authorities, relying on the accessibility of the released data. These data have to be stored until the retention period is expired.

After the expiration date the data are deleted automatically. There have to be mechanisms to extend the retention period with consent of the Safe Owner and it is an open topic what

to do if the Safe Owner has given up the electronic Safe in the meantime.

Essential use cases

Besides from the base functionality like storing and retrieving datasets 10, 20 it is interesting to discuss how the electronic Safe is integrated into e-government processes.

We start with a so called “release order request”, see Fig. 8, that the Safe User, i.e. public authority, sends to the electronic Safe. The release order request contains the specification of all the information that is necessary for a particular administrational service. To give an example: suppose the Safe Owner applies online for a dedicated parking permit. For this administrational service the applicant has to provide an official statement about his current place of residence that is not older than 3 months. In this case the release order request contains exactly one document request. Typically release order requests contain lists of data groups, like home address, birth information, not the data items like street, number, month of birth etc..

The data groups are structured around the conventional forms that we use today to apply for administrational services. The use case of an incoming release order request might include the information to the Safe Owner about a new waiting message by SMS or other convenient channel.

Next step is an activity of the Safe Owner, who receives the release order request in her Safe Owner Client, evaluates its requesting party and the underlying context, checks which data groups they request and grants access to this bundle of data.

The grants can be withdrawn, as long as the Safe User is not working with these data. In most cases the Safe User - the public administration - will retrieve these data and check the completeness of the released data.

If the data are well prepared enough to base a decision on, the public authority requires that these data remain unchanged for documentation reasons until a certain retention period is expired. For this to happen the Safe User locks the released data.

A special use case is the combination of several successive release order/grant access communications. When the Safe Owner initiates a more complex administrative process, that collects data dependent on the data received in a previous step, then the “online session” is the intended convenience functionality. Here the Safe Owner gives kind of a repeated access grant to a specified administrative process during this session.

While we focus here on the Safe User communication and the incorporation into e-government processes, please find a list of other relevant use cases in (Albert Jr. 2000).

Key requirements related to the integration in e-government processes

Security related key requirements summarized

REQ_S01: The Safe Provider and the Storage Providers MUST NOT be able to infer the communication partners of the Safe Owner or the exchanged data.

REQ_S02: An eavesdropper MUST be prevented to get any knowledge about the stored or retrieved content. It MUST NOT be possible to store the exchanged encrypted messages until their decryption is possible to recover the safe content.

REQ_S03: It MUST NOT be possible for an adversary to retrieve any data that belong to another Safe Owner or any data that is part of a release order request.

REQ_S04: The Safe Owner SHOULD be able to grant permissions to save, send, copy, print her data. Therefore the owner SHOULD get to know, whether the requesting system is a system that can enforce these permissions (mutual attestation of the involved systems).

REQ_S05: The Safe Owner and the Safe User MUST be provided a way to discover, when the Safe Client and the Safe Infrastructure they are working with, is in an untrustworthy state (trusted display problem).

REQ_S06: The Safe Owner and the Safe User MUST be able to use the Safe Infrastructure with knowledge (convenient password) and some property that is easy to take with them and difficult to exchange without notice of the user.

Organizational

REQ_O01: The safe MUST to be designed in such a way that multiple steps of safe incorporation are possible. First step is the complementary usage of safes to facilitate data collection via conventional forms. Second step SHOULD be the incorporation into process infrastructures. Third step MAY be the exclusive access via trusted clients.

REQ_O02: As nationwide public key infrastructures are not yet available, the safe protocols MUST be formed to recognize that. So Release Order Requests MUST transmit the public key of the Safe User and the Safe Provider MUST provide a service to retrieve the public key of the Safe Owner. It MUST be possible to validate signatures of Safe Users offline.

REQ_O03: The released data MUST remain unchanged, until the retention period is expired.

REQ_O04: The Safe Owner MUST be able to have some backup, in case she loses the access media.

Technical requirements

REQ_T01: As most service oriented architecture building blocks rely on web services, the protocols used SHOULD be built upon the web services stack and comply to the WS-I recommendations.

Business requirements

REQ_B01: Multiple clients MUST be usable, e.g. personal computer, kiosk system, mobile phone.

REQ_B02: Standards and open systems MUST be used wherever possible to create a broad community of Safe Providers.

REQ_B03: The XML data structures to be stored in the safe MUST be extensible, they will result from and be used in forms of all kinds. The design MUST allow to map available to required data and to incorporate existing schemata.

REQ_B04: The response time of the Safe Client application SHOULD compare to current web applications (search engines). The initial time to open up the safe and the loading of large documents SHOULD each require not more than a minute.

REQ_B05: The safe MUST allow the versioning of all data and documents stored in it.

Proposed Solution

Systems architecture conceptual view

Based on the requirements outlined above a prototype is described that supports an application for a place in an after-school care club. The conventional application form requires data about the school kids, the parents, their address and many more information up to their financial situation. The form comprises multiple pages so it is well understood that the electronic safe creates a real advantage over the conventional form in an e-government portal.

To integrate the electronic safe into current e-government portals we assume the base components shown in Fig. 9.

Additionally to common infrastructure components like a case management system and the archive storing case data and documents, we find a modern process management solution connected to a portal, that the citizen may use as entry point to initialize administrational services. The process management incorporates services provided by electronic safe integration components.

Trusted viewer

There is a certain demand to avoid dispersal of sensitive, private information across various IT systems. However, for current e-Government scenarios citizens have to provide all necessary data as unprotected clear text even if they are just used for visa. The future usage of these data is beyond citizens' control. The Trusted Viewer is dedicated hard- and software that permits the Safe User to view Safe content while preserving Safe-Owners control over their data. According to REQ_S04, it disallows any data manipulation, duplication or storage without consent of the Safe Owner. Since it operates on a trusted computing infrastructure the Trusted Viewer can provide proof of a well-defined state and that it will perform as expected. Our Safe will deliver the Safe-Owners data only after verifying that proof. So, the

Safe-Owner is in control of her sensitive data which does not disperse around. Seen from the process modeling view, the Trusted Viewer is a desktop where the public officer executes human interaction tasks that require working with real sensitive data.

An alternate option is to separate the case data from identifying items and to use pseudonyms instead. The Safe Owner is able to proof the ownership of a pseudonym. That's why it is reasonable to decide applications based on pseudonyms in favor of Safe Owner's privacy. However the Trusted Viewer is the preferable way, because it is difficult to rule whether data are identifying or not.

Safe Request Management

As each of the data retrievals includes multiple requests to the Safe Provider and the Storage Providers this complex protocol is encapsulated within the Safe request management. The Safe request management is used by each of the clients, i.e. the Trusted Viewer(s) and the Rule Engine.

Rule Engine

The Rule Engine is capable to execute simple tests on the data retrieved from the electronic safe. A certain amount of incoming applications could be pre-evaluated by this rule engine without any human involvement. This could potentially ease the human interaction tasks later on. Additionally it might be viable to let the rule engine work on the real sensitive data and to blind these data in following steps where humans are involved.

Certificate Issuance

We work with two types of certificates: X.509 certificates and idemix certificates.

The X.509 certificates are used throughout the communication, whenever the communicating parties need to know each other, e.g. to sign release order requests or to hide the communicating parties from the Safe Provider by encrypting the requests and their answers.

The idemix certificates are used whenever the service provider needs to authorize a client without revealing her identity. Such a case occurs when the Safe Owner wants to store some data with the Storage Provider. The Safe Owner has to proof that he has the right to store some data ("I have paid the subscription fee!") without telling who she is. The Safe User needs a Certificate Issuance to be able to issue role based certificates. If the citizen releases some data, these data are encrypted for a particular receiver. The receiver might be a natural person or a role, e.g. the "Single Point of Contact" mentioned in the European Services Directive.

Data structures

The model of the data follows the idea, that the valuable information, e.g. a document, is split into separate chunks of bytes that are stored in little "buckets" with different Storage Providers. The information, which chunks belong to which document, is encrypted and stored with the Safe Provider. The buckets stored with the Storage Provider are secured against unauthorized access by cryptographic means, so called idemix domain pseudonyms.

The information model in its central classes is documented in (Penski 2010).

Safe User Communication protocols

The splitting of each information into chunks of bytes that are stored with independent Storage Providers involves even the release order requests. The diagram in Fig. 10 abstracts from necessary authentication and authorization procedures with Safe Provider and Storage Providers. Starting point is the Public Officer that works on a certain application of a person who owns an electronic Safe. The officer specifies in his Trusted Viewer which information is needed and which Safe should be used. The Safe Request Management will first retrieve the public key of the Safe Owner. This key is used later on to encrypt the information, where the release order request is stored and how the Safe Owner client is able to reassemble the information. Then the release order is split into several parts that are subsequently stored with independent Storage Providers. Each of these steps requires a new authorization sequence because the Safe Request Management remains completely anonymous in relation to the Storage Provider. The Storage Provider is not able to discover whether two subsequently arriving requests are sent from the same Safe User. As soon as all parts are stored in buckets, the list of all buckets, ordered in their correct sequence is encrypted with the public key of the Safe Owner. The encrypted bucket list is stored with the Safe Provider. Finally, as a convenience service the Safe Provider informs the Safe Owner that “something” is arrived.

This is only a very small part of the entire communication protocol, however the underlying principle should be visible: all communicating parties exchange lists of buckets that contain pointers to the real information. If the Safe Provider is capable to break the encryption of this information (probably after a period of six to ten years) he is not in the position to have access on the buckets stored with the Storage Providers. Since the buckets are relocated from time to time, the information gained from breaking the encryption is useless.

Conclusion

We presented embodiments to incorporate sensitive personal information in modern e-government processes. This was focused on the Safe User infrastructure.

The risk of data leakage caused by the operating staff of a single online storage provider is accomplished by the separation of responsibilities and the dispersal of data to independent Storage Providers. The risk of unauthorized access to the Safe content is minimized by domain pseudonyms stored together with the data. As the weakest point is the client by which the sensitive information is displayed, the Trusted Viewer and the Safe Client of the Safe Owner are secured by trusted computing mechanisms (Trusted Computing Group, Incorporated 2007). The Risk of a superior instance, controlling all the Storage Providers and the Safe Providers can be best reduced by an easy way to offer and incorporate new independent Storage Services, which are not controllable by a superior instance.

As a result we achieved a trustworthy decentralized e-government infrastructure component by splitting responsibilities on multiple organizational units and by combining privacy enhancing technologies with trusted computing and information dispersal.

References

Albert Jr., W.P. and Kotzin M. 2000, “SMART CARD WITH BACK UP”, United States patent number 7206847, Filing date 22nd Mai 2000, Issue date 17th Apr.2007.

Cacayorin, P. (2004), “Dispersed data storage using cryptographic scrambling description” [online], http://www.freshpatents.com/Dispersed-data-storage-using-cryptographic-scrambling-dt20060413ptan20060078127.php, USPTO Patent Application 20060078127, Filing date 8 Oct. 2004.

Camenisch, J. and Herreweghen, E.V. (2002) “Design and implementation of the idemix anonymous credential system”, *CCS '02: Proceedings of the 9th ACM conference on Computer and communications security*, New York, NY, USA, ACM Press, pp 21-30.

Chaum, D. (1985) “Security without identification: Transaction systems to make big brother obsolete”, Communications of the ACM 28(10), pp1030-1044.

DIRECTIVE 2006/123/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 12 December 2006 on services in the internal market, available at: http://ec.europa.eu/internal_market/services/services-dir/index_en.htm

Iyengar, A., Cahn, R. Garay, J.A., Jutla, G. (1998), “Design and Implementation of a Secure Distributed Data Repository”, In Proc. of the 14th IFIP Internat. Information Security Conf., pp.123—135.

Kubiatowicz, Bindel, Chen, Eaton, Geels, Gummadi, Rhea, Weatherspoon, Weimer, Wells, and Zhao (2000), “Oceanstore: An architecture for global-scale persistent storage”, Proc. of the ninth international conference on Architectural support for programming languages and operating systems (ASPLOS '00), ACM SIGARCH Computer Architecture News, Dec. 2000, pp.190-201, ISSN:0163-5964 .

Paul, A., Agarwala, S. and Ramachandran, U. (2007), “e-SAFE: An Extensible, Secure and Fault Tolerant Storage System”, Proc. IEEE Self-Adaptive and Self-Organizing Systems,(SASO '07), IEEE Press, , pp. 257-268, doi: 10.1109/SASO.2007.21

Penski, A. and Breitenstrom, C. (2010) in Press, “Electronic safes: information model and security approach”, Proc. of the 2nd International Conference on Networks Security, Wireless Communications and Trusted Computing (*NSWCTC* 2010)

Rabin, M (1989), “Efficient dispersal of information for security, load balancing, and fault tolerance”, *Journal of the ACM*, Vol.36, pp. 335—348.

Schneier, B. (2009) “Crypto-Gram-Newsletter”, [online], http://www.schneier.com/crypto-gram-0901.html.

Redlich, R.M. and Nemzow, M.A. (2008), “DATA SECURITY SYSTEM AND METHOD WITH PARSING AND DISPERSION TECHNIQUES”, United States patent number 7349987, Filing date 23 May 2002, Issue date 25 Mar 2008.

Trusted Computing Group, Incorporated (2007), “TCG Software Stack (TSS) Specification Version 1.2”, http://www.trustedcomputinggroup.org/resources/tcg_software_stack_tss_specification.

Zhang, Zhang, Xian, Chen, Feng, (2008), „Towards A Secure Distribute Storage System”, *Advanced Communication Technology*, (ICACT 2008), IEEE Press, Apr. 2008, pp. 1612-1617, doi: 10.1109/ICACT.2008.4494090

Claims

1. Method for safely handling at least one dataset, in particular a document (10, 20), particularly in a cloud computer environment, wherein
a) the at least one dataset (10, 20) is partitioned into at least two dataset partitions (1, 2, 3, 4),
b) the at least two dataset partitions (1, 2, 3, 4) are stored on and / or retrieved from at least two computer sites (100, 200, 300, 400) being part of, e.g., a cloud computer environment, so that on no computer site (100, 200, 300, 400) sufficient, in particular all, dataset partitions (1, 2, 3, 4) of the at least one dataset (10, 20) are present, in particular not present at the same time.

2. Method for safely handling at least one dataset, in particular a document (10, 20), in, e.g., a cloud computer environment, wherein a retrieval of a partitioned dataset (1, 2, 3, 4) to combine the at least one dataset (10, 20) comprises
a) an authorization by a user, in particular an owner (1000) of the at least one dataset (10, 20) at a safe system (1010),
b) on positive completion of the authorization, automatic retrieval of distribution information regarding the dataset partitions (1, 2, 3, 4) from the safe system (1010),
c) retrieval of all dataset partitions (1, 2, 3, 4) of the at least one dataset (10, 20) from computer sites (100, 200, 300, 400),
d) transformation of the dataset partitions (1, 2, 3, 4) into a logical representation of the at least one dataset (10, 20).

3. Method for safely handling at least one dataset, in particular a document (10, 20), in, e.g., a cloud computer environment, wherein a storing of a partitioned dataset (1, 2, 3, 4) derived from the at least one dataset (10, 20) comprises
a) an authorization by an owner (1000) of the at least one dataset (10, 20) at a safe system (1010),
b) on positive completion of the authorization, automatic transfer of distribution information regarding the dataset partitions (1, 2, 3, 4) to the safe system (1010),
c) automatic storing all dataset partitions (1, 2, 3, 4) of the at least one dataset (10, 20) on computer sites (100, 200, 300, 400) as physical sub-representations of the at least one dataset (10, 20), so that none of the computer sites (100, 200, 300, 400) comprises more than one instance of the dataset partitions (1, 2, 3, 4), in particular not at the same time.

4. Method according to at least one of the preceding claims, wherein the methods of claim 1, 2 and / or 3 are executed on a system under control of the owner (1000) of the dataset (10, 20), in particular on a mobile device (1001).

5. Method according to at least one of the preceding claims, wherein the distribution-information regarding the document partitions (1, 2, 3, 4) is stored in a safe system (1010).

6. Method according to at least one of the preceding claims, wherein the at least one dataset (10, 20) is partitioned and / or the document (10, 20) is recombined on a client under control of the owner (1000) of the at least one dataset (10, 20).

7. Method according to at least one of the preceding claims, wherein the at least one dataset (10, 20) is automatically partitioned according to a predefined set of rules, in particular into dataset partitions (1, 2, 3, 4) of equal size or according to a secret sharing algorithm of Rabin.

8. Method according to any of the preceding claims, wherein each dataset partition (1, 2, 3, 4) is automatically associated with an distribution-information, in particular a distribution tag (11, 12, 13).

9. Method according to any of the preceding claims, wherein the dataset partitions (1, 2, 3, 4) are at least partially automatically encrypted.

10. Method according to any of the preceding claims, wherein at least for a subset of the at least two dataset partitions (1, 2, 3, 4) the storage site (100, 200, 300, 400) dynamically changed.

11. Method according to any of the preceding claims, wherein communication with the safe system (1010) and / or storage provider is at least partially anonymous.

12. Method according to any of the preceding claims, wherein the distribution information is safely transmitted to a further user of the dataset (10, 20).

13. Method according to any of the preceding claims, wherein at least one user of a safe system (1010) is a public administration.

14. System for safely handling at least one dataset, in particular a document (10, 20), in, e.g., a cloud computer environment, with

a) a partition means to partition the at least one dataset (10, 20) into at least two dataset partitions (1, 2, 3, 4),

b) a safe system (1010) for storing / and or retrieving the at least two dataset partitions (1, 2, 3, 4) from and / or to at least two computer sites (100, 200, 300, 400) being part of, e.g., a cloud computer environment, so that none computer site (100, 200, 300, 400) comprises instances of all dataset partitions (1, 2, 3, 4) of the at least one dataset (10, 20), in particular not at the same time.

15. System for safely handling at least one dataset, in particular a document (10, 20), in, e.g., a cloud computer environment, with retrieval means for a partitioned dataset (1, 2, 3, 4) to combine the at least dataset (10, 20),

characterized by,

an authorization means for a user, in particular an owner (1000) of the at least one dataset (10, 20) at a safe system (1010) and a retrieval means which, on positive completion of the authorization, uses distribution information regarding the dataset partitions (1, 2, 3, 4) to automatically retrieve from the, e.g., cloud computer environment all dataset partitions (1, 2, 3, 4) of the at least one dataset (10, 20) from computer sites (100, 200, 300, 400), and the dataset partitions (1, 2, 3, 4) are transformed into a logical representation of the at least one dataset (10, 20).

16. System for safely handling at least one dataset, in particular a document (10, 20), in, e.g., a cloud computer environment with storing means for a

partitioned dataset (1, 2, 3, 4) derived from the at least one dataset (10, 20),

characterized by,

an authorization means for an owner (1000) of the at least one dataset (10, 20) at a safe system (1010) and a storing means which, on positive completion of the authorization, uses automatically generated distribution information regarding the dataset partitions (1, 2, 3, 4) to automatically store all dataset partitions (1, 2, 3, 4) of the at least one dataset (10, 20) on computer sites (100, 200, 300, 400) as physical representations of the at least one dataset (10, 20), so that none of the computer sites (100, 200, 300, 400) comprises more than one instance of the dataset partitions (1, 2, 3, 4), in particular not at the same time and to store the distribution information at the safe system (1010).

FIG 1

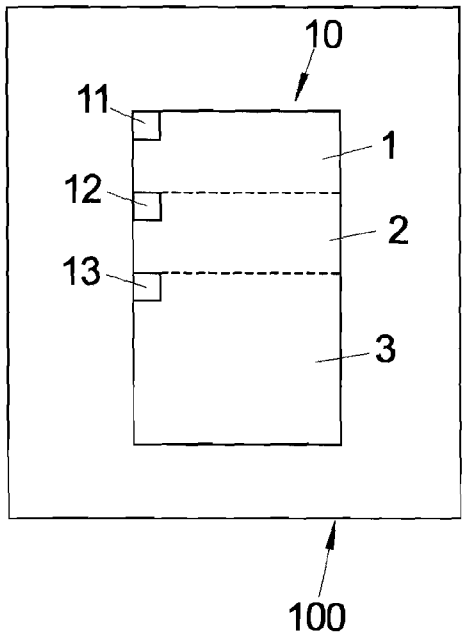


FIG 2

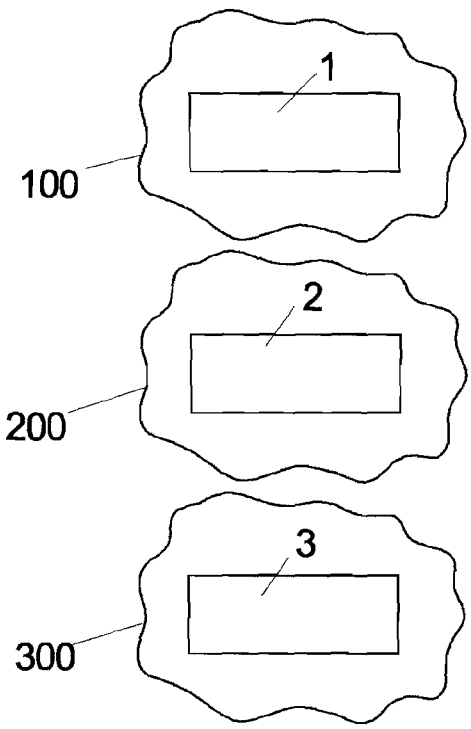


FIG 3

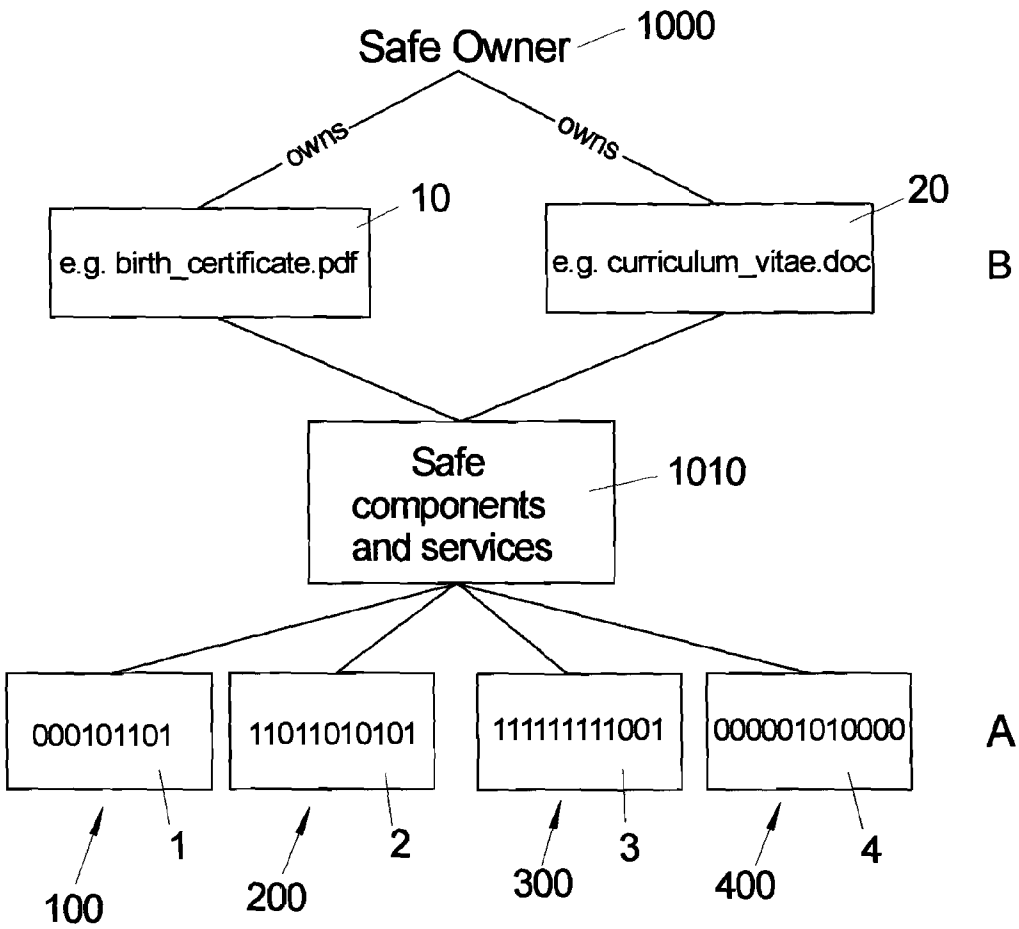

Safe Owner
1000
owns
owns
10
20
e.g. birth_certificate.pdf
e.g. curriculum_vitae.doc
B
Safe components and services
1010
000101101
11011010101
111111111001
000001010000
A
1
2
3
4
100
200
300
400

FIG 4

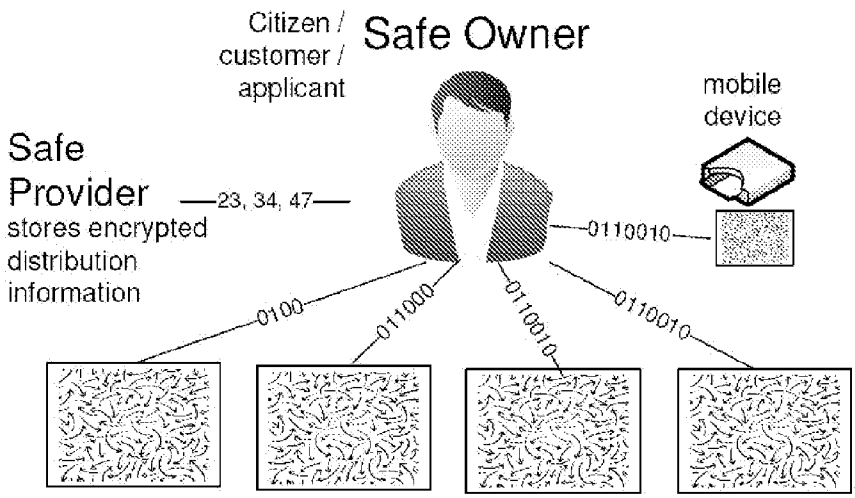

Citizen / customer / applicant
Safe Owner
mobile device
Safe Provider
23, 34, 47
stores encrypted distribution information
0110010
0100
011000
0110010
0110010

FIG 5

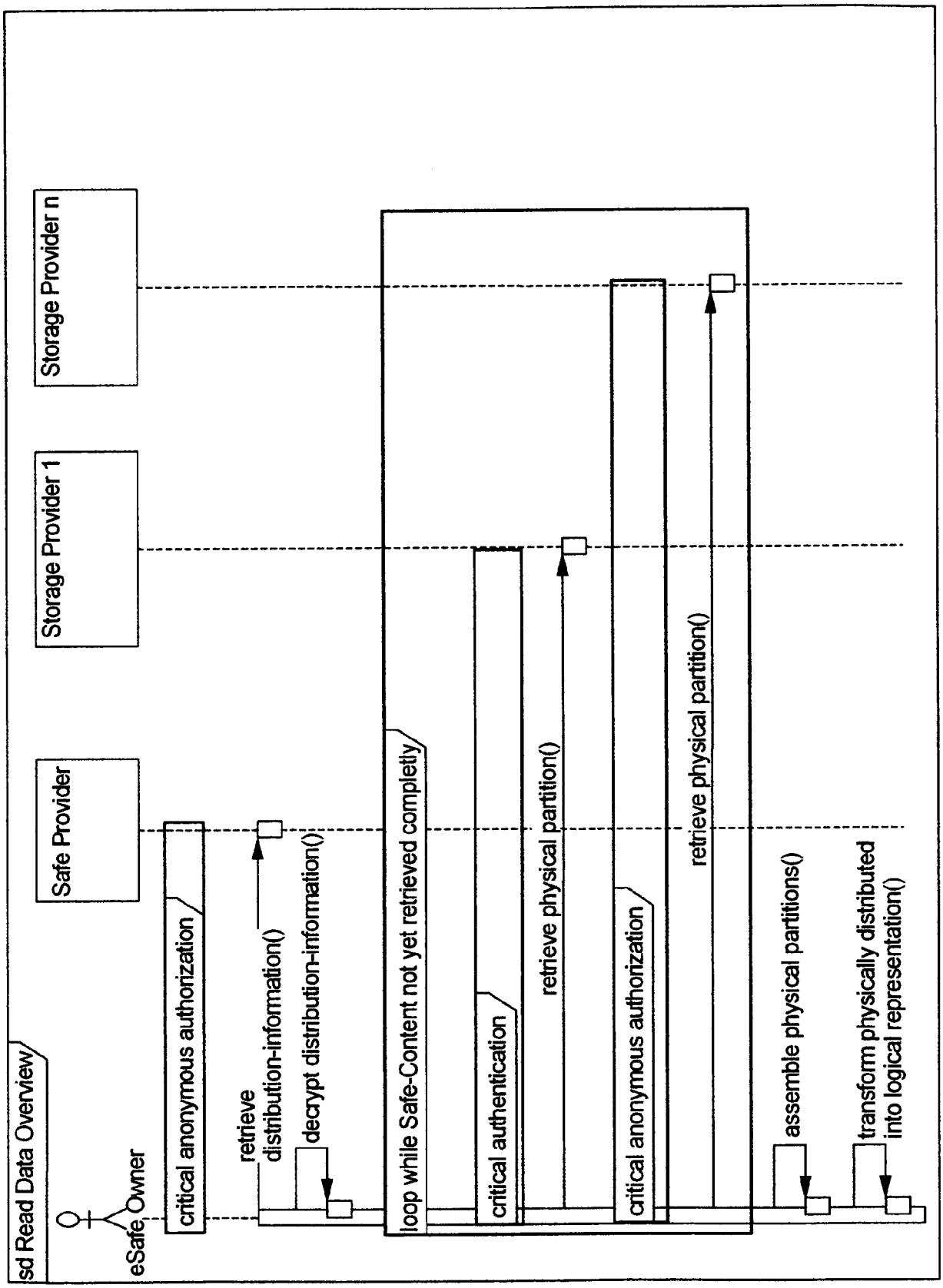

sd Read Data Overview
eSafe Owner
Safe Provider
Storage Provider 1
Storage Provider n
critical anonymous authorization
retrieve distribution-information()
decrypt distribution-information()
loop while Safe-Content not yet retrieved completly
critical authentication
retrieve physical partition()
critical anonymous authorization
retrieve physical partition()
assemble physical partitions()
transform physically distributed into logical representation()

Fig. 6

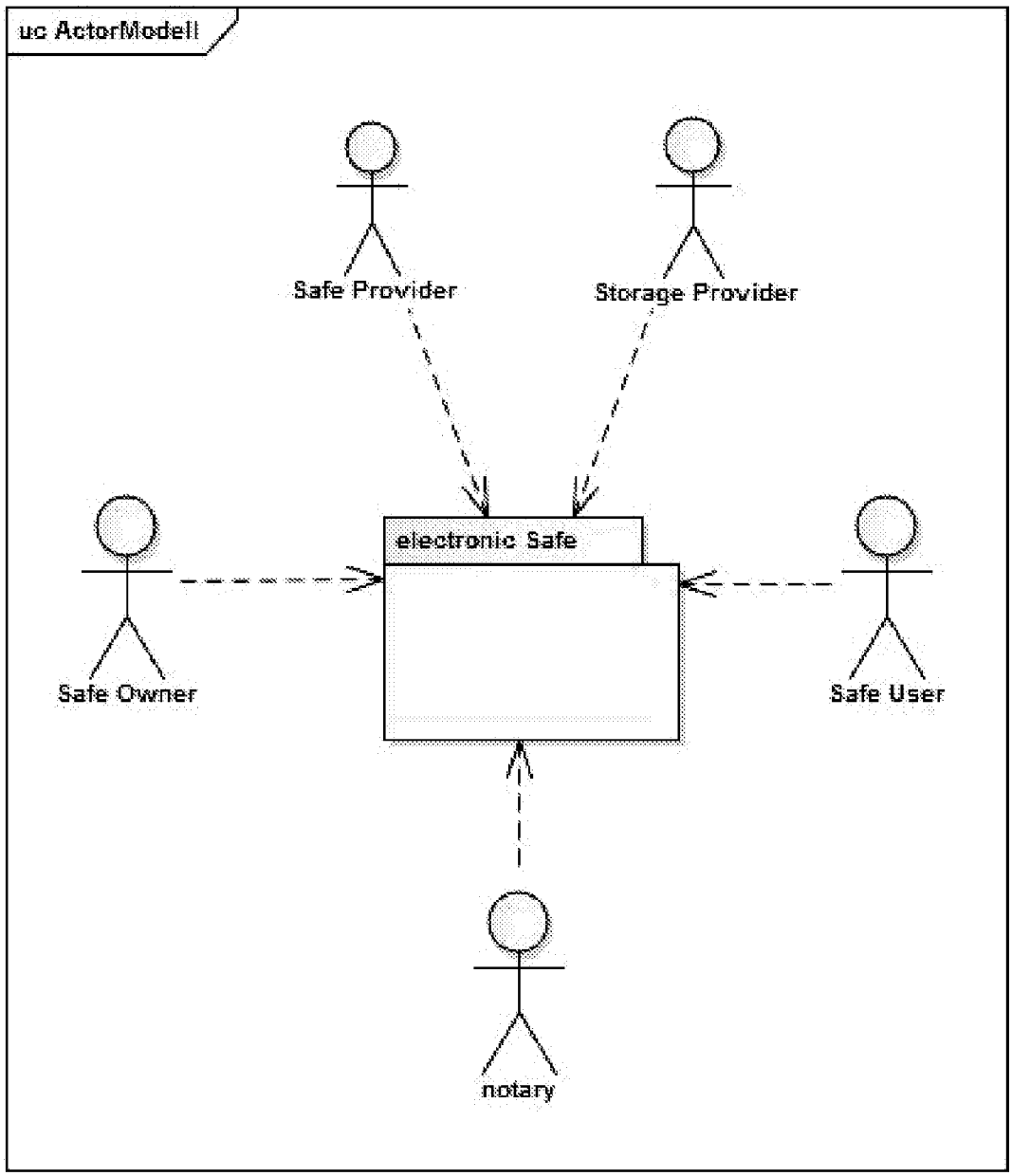

uc ActorModell
Safe Provider
Storage Provider
electronic Safe
Safe Owner
Safe User
notary

Fig. 7

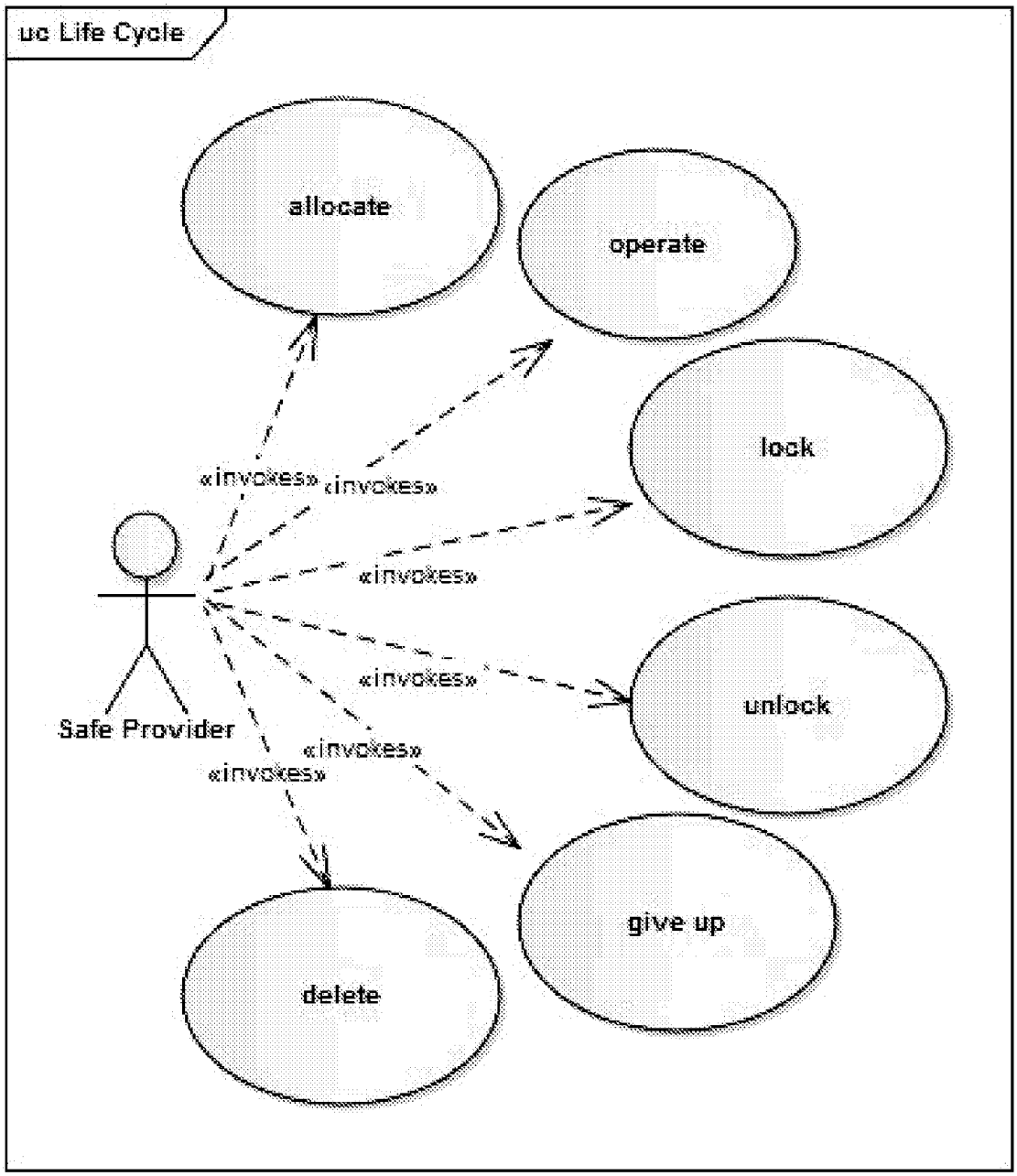

uc Life Cycle
allocate
operate
lock
unlock
give up
delete
Safe Provider
«invokes»
«invokes»
«invokes»
«invokes»
«invokes»
«invokes»

FIG 8

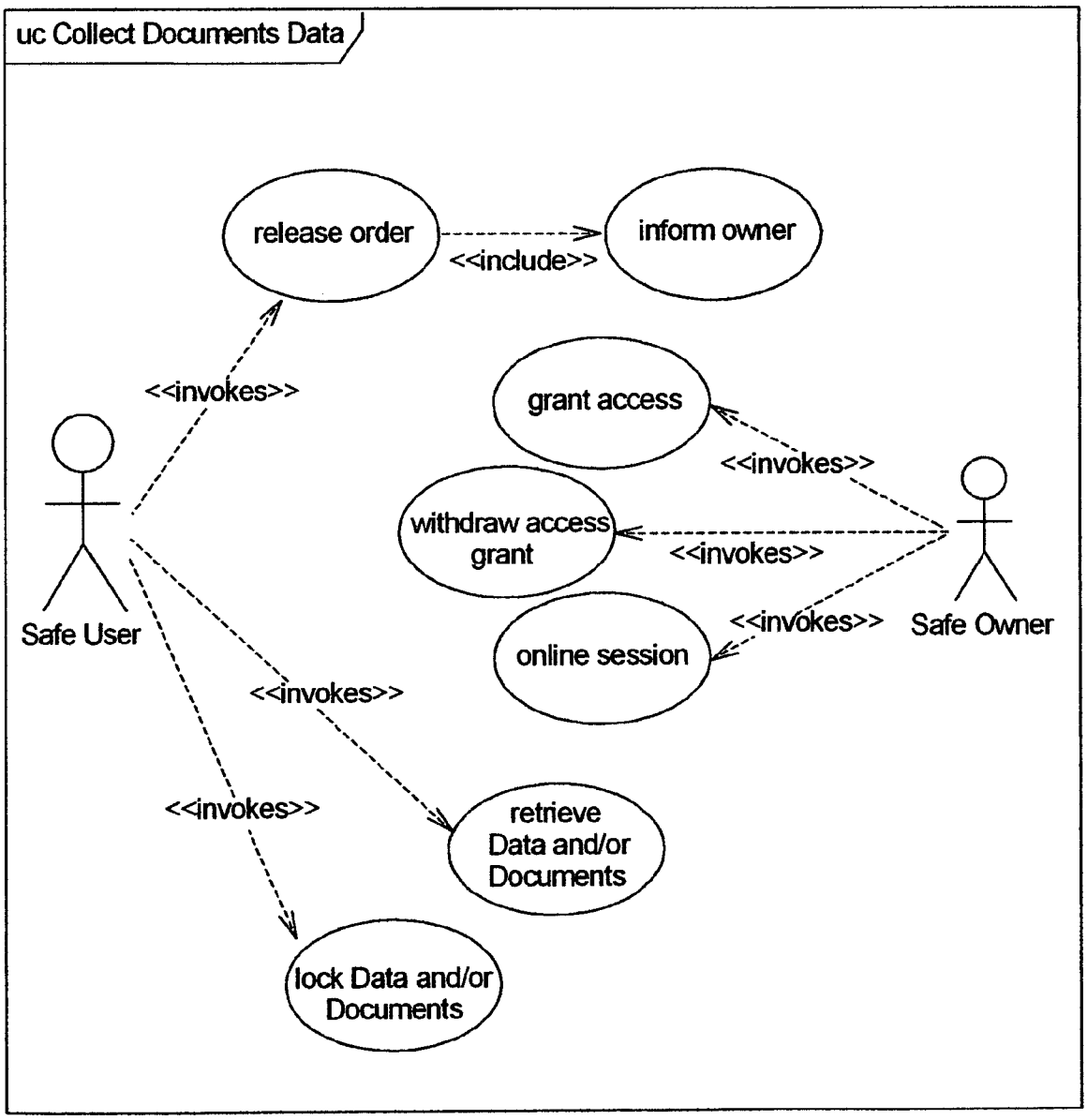

uc Collect Documents Data
release order
inform owner
<<include>>
<<invokes>>
grant access
<<invokes>>
withdraw access grant
<<invokes>>
Safe User
Safe Owner
<<invokes>>
online session
<<invokes>>
<<invokes>>
retrieve Data and/or Documents
lock Data and/or Documents

FIG 9

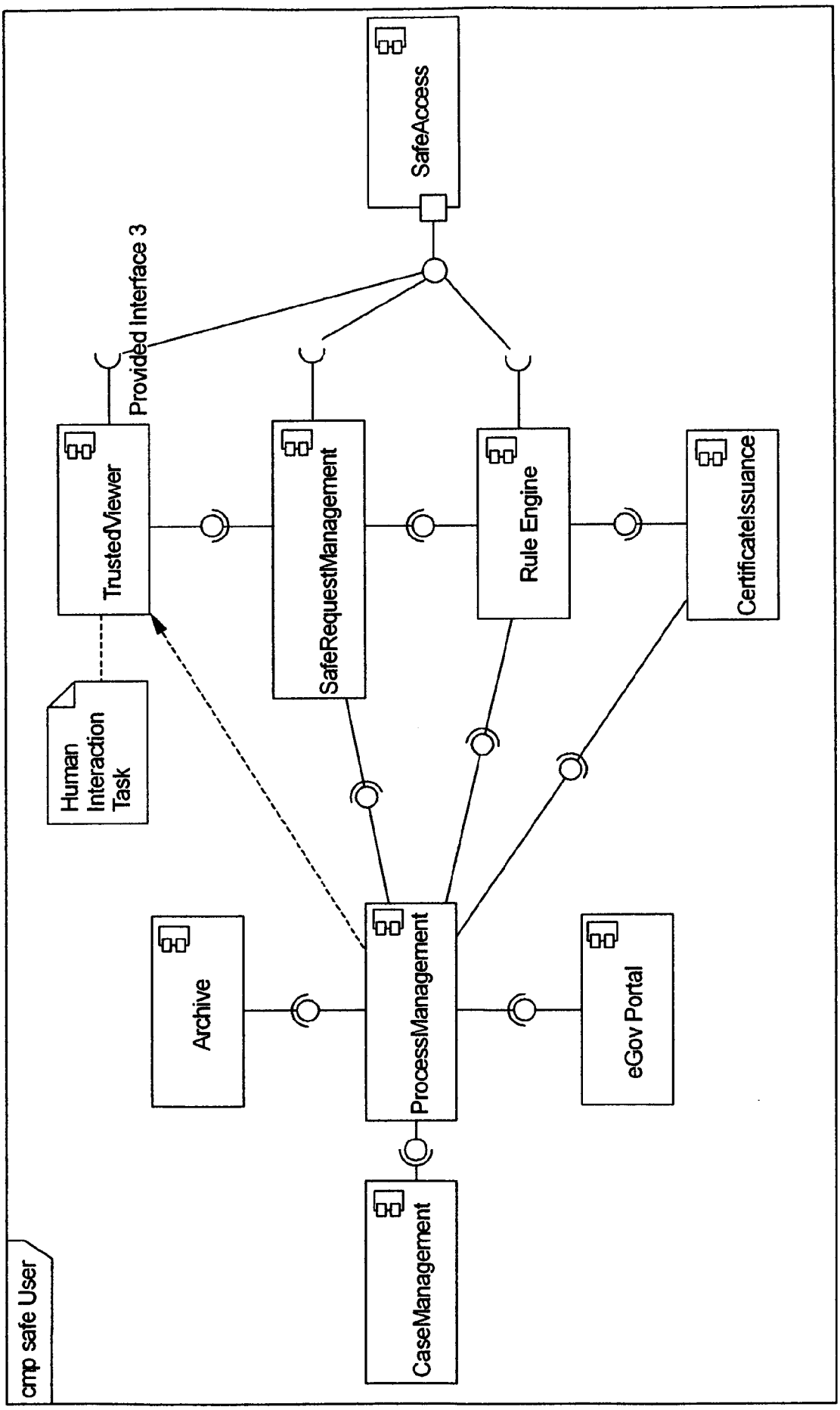
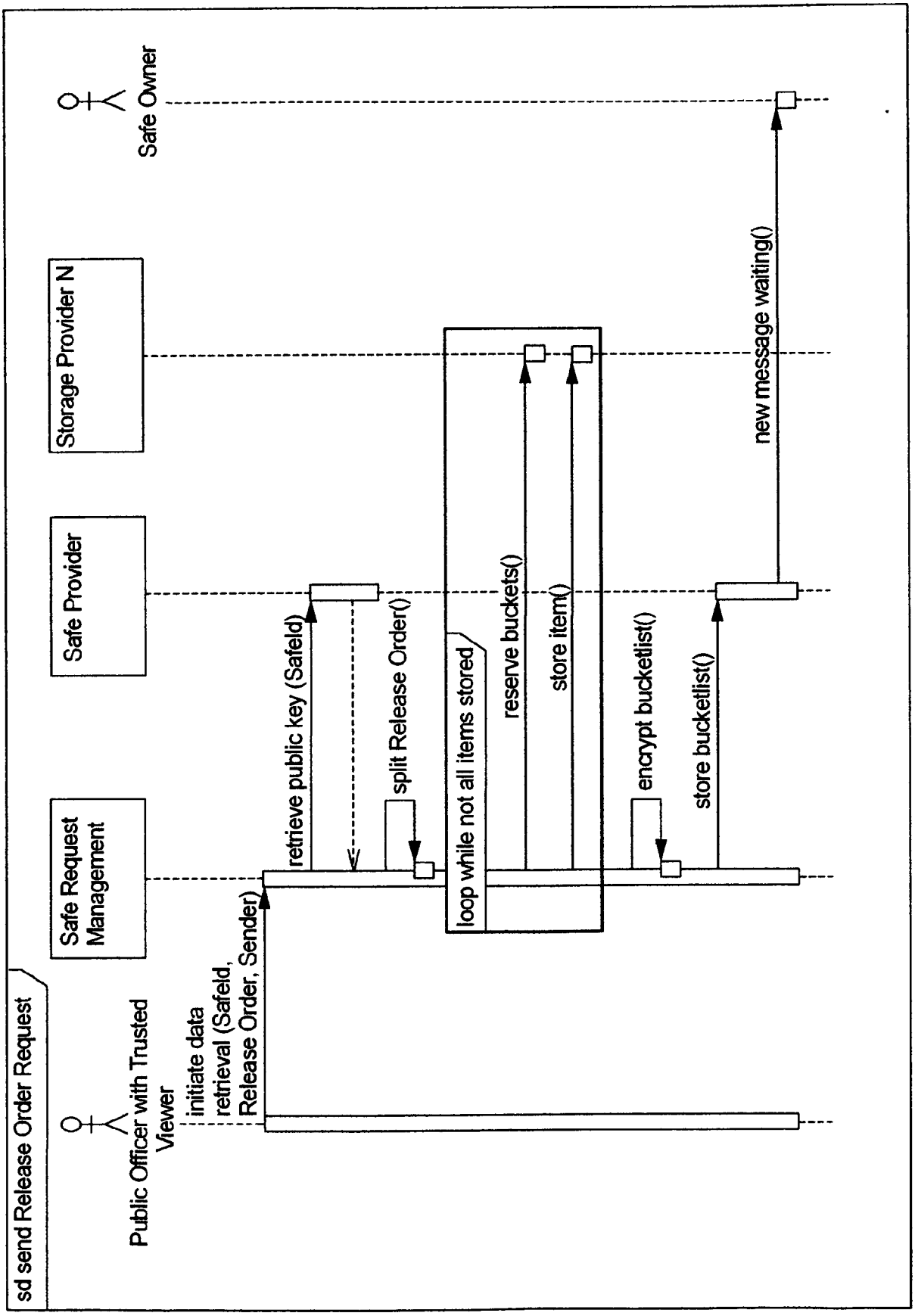

cmp safe User
SafeAccess
Provided Interface 3
TrustedViewer
SafeRequestManagement
Rule Engine
CertificateIssuance
Human Interaction Task
Archive
ProcessManagement
eGov Portal
CaseManagement

FIG 10


sd send Release Order Request
Public Officer with Trusted Viewer
Safe Request Management
Safe Provider
Storage Provider N
Safe Owner
initiate data retrieval (Safeld, Release Order, Sender)
retrieve public key (Safeld)
split Release Order()
loop while not all items stored
reserve buckets()
store item()
encrypt bucketlist()
store bucketlist()
new message waiting()

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2011/059846

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F21/00
ADD.

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

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2008/060085 A1 (SAMZELIUS JAN [US] ET AL) 6 March 2008 (2008-03-06) abstract; figure 1 paragraph [0018] paragraph [0020] paragraph [0027] paragraph [0035] -----	1-16
X	WO 2007/133791 A2 (KANE RICHARD [US]) 22 November 2007 (2007-11-22) abstract paragraph [0011] claims 1-24 -----	1-16
X	US 2005/240749 A1 (CLEMO GARY [GB] ET AL) 27 October 2005 (2005-10-27) abstract; figures 1-10 paragraph [0001] - paragraph [0024] -----	1-16

☐ 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 document 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	Date of mailing of the international search report
24 October 2011	03/11/2011

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Harms, Christoph

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2011/059846

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2008060085 A1	06-03-2008	NONE	
WO 2007133791 A2	22-11-2007	NONE	
US 2005240749 A1	27-10-2005	GB 2412760 A	05-10-2005
		JP 2005293592 A	20-10-2005
		US 2007271349 A1	22-11-2007