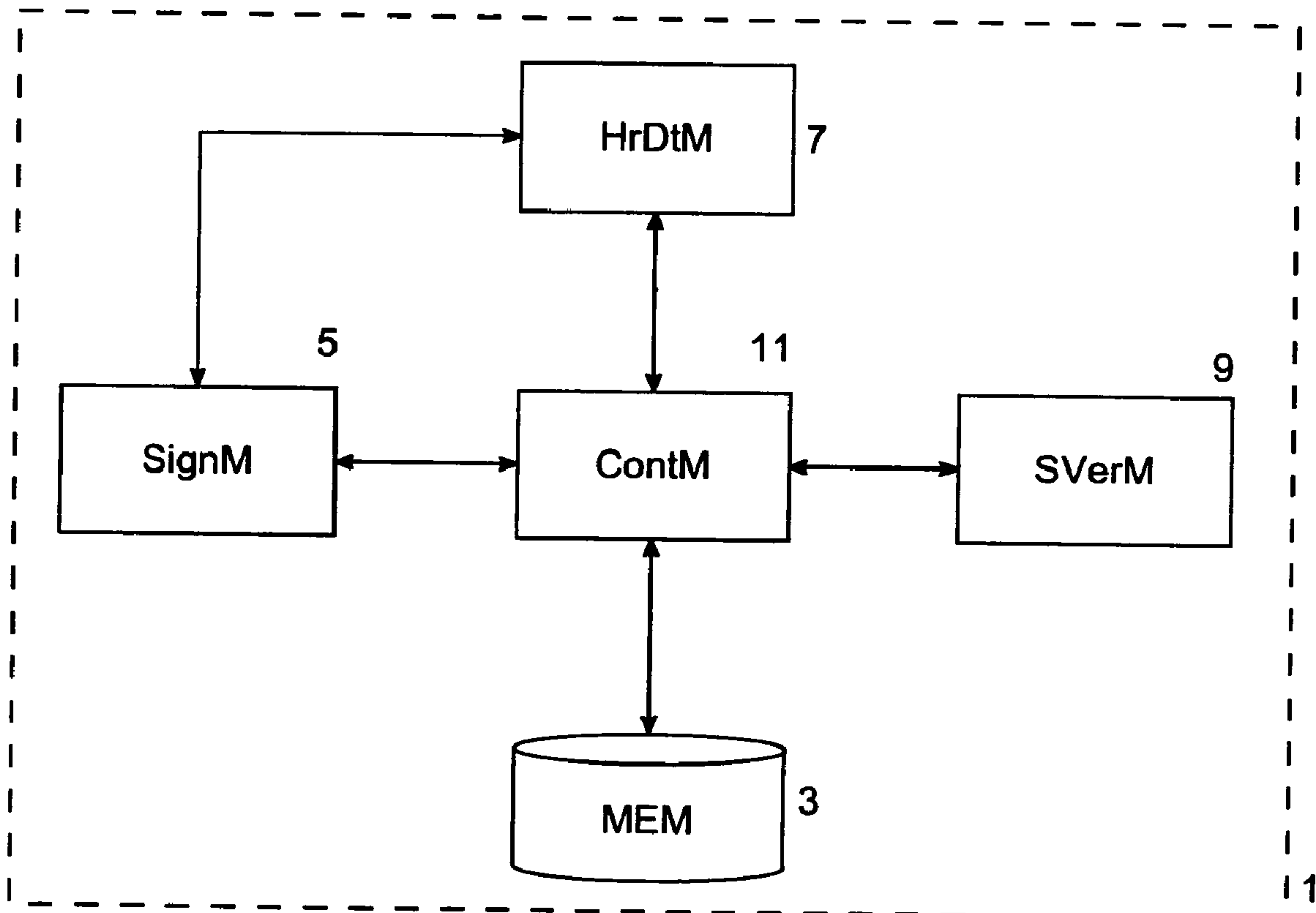




(86) **Date de dépôt PCT/PCT Filing Date:** 2009/04/21
(87) **Date publication PCT/PCT Publication Date:** 2009/11/05
(45) **Date de délivrance/Issue Date:** 2017/10/31
(85) **Entrée phase nationale/National Entry:** 2010/10/19
(86) **N° demande PCT/PCT Application No.:** FR 2009/000471
(87) **N° publication PCT/PCT Publication No.:** 2009/133320
(30) **Priorité/Priority:** 2008/04/25 (FR0802333)

(51) **Cl.Int./Int.Cl. H04L 9/00** (2006.01)
(72) **Inventeur/Inventor:**
ROUSSEAU, GUILLAUME, FR
(73) **Propriétaire/Owner:**
UNIVERSITE PARIS DIDEROT-PARIS 7, FR
(74) **Agent:** ROBIC

(54) **Titre : DISPOSITIF INFORMATIQUE DE GESTION TEMPORELLE DE DOCUMENTS NUMERIQUES**
(54) **Title: COMPUTER DEVICE FOR THE TIME-BASED MANAGEMENT OF DIGITAL DOCUMENTS**



(57) **Abrégé/Abstract:**

Dispositif pour gérer des documents numériques comprenant une mémoire (3) pour stocker un document numérique et une empreinte datée comprenant de premières signatures établies selon une première méthode et une valeur de temporalité, un

(57) Abrégé(suite)/Abstract(continued):

générateur de signatures (5) pour établir de secondes signatures selon une seconde méthode, un horodateur, un vérificateur de signatures (9) pour vérifier la conformité d'un document et d'une empreinte datée, et un superviseur (11) pour faire opérer le générateur de signatures (5) sur le document, faire opérer le vérificateur de signatures (9) sur le document et son empreinte datée, et, en cas de conformité, faire opérer l'horodateur (7) sur les secondes signatures avec la valeur de temporalité.

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international



(10) Numéro de publication internationale
WO 2009/133320 A1

(43) Date de la publication internationale
5 novembre 2009 (05.11.2009)

(51) Classification internationale des brevets :
H04L 9/00 (2006.01)

(21) Numéro de la demande internationale :
PCT/FR2009/000471

(22) Date de dépôt international :
21 avril 2009 (21.04.2009)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
0802333 25 avril 2008 (25.04.2008) FR

(71) Déposant (pour tous les États désignés sauf US) :
UNIVERSITE PARIS DIDEROT - PARIS 7 [FR/FR];
5, rue Thomas Mann, F-75205 Paris 13 (FR).

(72) Inventeur; et

(75) Inventeur/Déposant (pour US seulement) :
ROUSSEAU, Guillaume [FR/FR]; 24, avenue Félix
Faure, F-75015 Paris (FR).

(74) Mandataire : **PLAÇAIS, Jean-Yves**; Cabinet Netter, 36,
avenue Hoche, F-75008 Paris (FR).

(81) États désignés (sauf indication contraire, pour tout titre
de protection nationale disponible) : AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,
CA, CH, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG,
SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) États désignés (sauf indication contraire, pour tout titre
de protection régionale

WO 2009/133320 A1



de signatures (9) pour vérifier la conformité d'un document et d'une empreinte datée, et un superviseur (11) pour faire opérer le générateur de signatures (5) sur le document, faire opérer le vérificateur de signatures (9) sur le document et son empreinte datée, et, en cas de conformité, faire opérer l'horodateur (7) sur les secondes signatures avec la valeur de temporalité.

Dispositif informatique de gestion temporelle de documents numériques

5

L'invention a trait à un dispositif de gestion temporelle de documents numériques.

10

Les documents de type numériques peuvent présenter des contenus divers, tels que de la musique, du texte, des images, de la vidéo, ou encore du code source.

15

Il arrive qu'on veuille établir une comparaison de documents numériques, en particulier par rapport à leur contenu.

20

Dans le cas où deux documents numériques sont mémorisés sous la forme de deux fichiers informatiques distincts, de multiples comparaisons peuvent être effectuées, plus ou moins directement.

25

La plupart des interfaces graphiques des systèmes d'exploitation actuels indiquent par exemple la quantité de mémoire nécessaire au stockage d'un fichier particulier. Ces interfaces graphiques affichent également des dates de dernière modification, de création, et/ou de dernier accès, telles que stockées dans le fichier lui-même.

30

Il s'agit là d'une comparaison grossière et peu fiable : deux documents aux contenus différents peuvent avoir été créés à une même date et présenter des tailles égales.

35

Des comparaisons plus fiables peuvent être faites en comparant deux fichiers informatiques octet par octet. De nombreux outils informatiques du commerce offrent cette possibilité.

Cependant, il s'agit là d'une comparaison de fichiers informatiques entre eux, et non d'une comparaison de leurs contenus : ces outils se contentent le plus souvent d'établir si les fichiers informatiques comparés sont identiques ou
5 non.

Par exemple, lorsque les fichiers de deux documents issus de logiciels de traitement de texte différents, ou de deux versions différentes d'un même logiciel de traitement de
10 texte, sont comparés, ces outils indiquent pratiquement dans tous les cas que ces fichiers sont différents. Pourtant, le texte contenu dans ces fichiers peut être identique, y compris en ce qui concerne sa mise en forme.

15 Certains logiciels permettent une comparaison du contenu de documents. C'est le cas par exemple de la plupart des logiciels de traitement de texte actuels.

Les possibilités de comparaison offertes par ces logiciels ne
20 sont cependant pas satisfaisantes.

La comparaison est en général limitée au cas où les fichiers correspondants ont été générés par le logiciel en question.

25 Elle est ensuite essentiellement manuelle, en sorte qu'elle devient vite fastidieuse lorsqu'il s'agit de comparer de multiples documents entre eux.

Elle est encore limitée au cas de deux fichiers informati-
30 ques, et devient inopérante lorsqu'un contenu, c'est-à-dire du texte, est physiquement réparti sur plusieurs fichiers informatiques distincts.

En outre, la comparaison est opérée sur l'ensemble du
35 contenu, ce qui engendre un traitement assez long dans le cas

de documents de taille importante et/ou de multiples comparaisons de documents deux à deux.

De plus, le résultat d'une telle comparaison se limite à une
5 mise en évidence des différences, ou des similarités, de contenu sans autre information pertinente quant à ce contenu.

Enfin, ces logiciels ne permettent pas une comparaison confidentielle des documents : deux auteurs qui voudraient
10 comparer leurs textes respectifs seraient obligés de se montrer ces derniers, ou, à tout le moins, de le faire voir à un tiers.

Autrement dit, la comparaison de fichiers textes, telle
15 qu'elle est offerte par les logiciels de traitement de texte du marché, impose une divulgation du contenu des documents. Et ceci peut être inacceptable, par exemple lorsque les documents en question concernent une oeuvre littéraire ou une partie d'un programme informatique.

20

Par ailleurs, il existe des dispositifs capables de réaliser des comparaisons de manière confidentielle. Par exemple, un dépôt auprès de l'IDDN permet d'obtenir une clé chiffrée unique, générée à partir d'un ou plusieurs fichiers corres-
25 pondant à un contenu. Le cas échéant, cette clé peut être comparée à une autre clé. L'opération de comparaison des documents se limite alors à la comparaison de courtes chaînes de caractères.

30 Ces dispositifs ne sont pas non plus satisfaisants.

D'abord, ils se contentent de conclure à l'identité de deux documents ou à leur différence, indépendamment de l'ampleur ou de la nature de cette différence. Typiquement, un simple
35 remplacement des marques de séparation de colonnes dans un

fichier texte, de tabulations en virgules par exemple, suffit à générer des chaînes de caractères différentes.

Ensuite, les dispositifs de ce type ne permettent pas de
5 s'intéresser à une partie seulement du contenu d'un document, ou plus généralement de s'intéresser à un document autrement que dans son ensemble. Par exemple, dans le cas où un document comprendrait tout ou partie du contenu d'un autre document, le traitement par les dispositifs de ce type se
10 limiterait à conclure à une différence des deux documents comparés. Or, il peut être souhaitable d'identifier une telle incorporation de contenu, en particulier lorsqu'il s'agit de revendiquer une partie de la propriété d'un document.

15 Enfin, la comparaison de chaînes de caractères impose que celles-ci aient été générées au moyen du même algorithme, ou à tout le moins d'algorithmes analogues, dans le sens où ces algorithmes doivent générer des signatures communes ou compatibles. Sinon, il n'est pas possible de conclure à une
20 différence de contenus à partir d'une différence de clés.

Or, au fil du temps, l'algorithme de chiffrement peut avoir été modifié, à multiples reprises.

25 Plus généralement, lorsqu'il s'agit de revendiquer une forme quelconque de propriété sur tout ou partie d'un document, il est nécessaire de comparer ces documents, en particulier quant à leur contenu, et d'obtenir des éléments de datation des contenus comparés.

30

Par exemple, si l'on souhaite revendiquer une partie de la création d'un logiciel, il est nécessaire de posséder le contenu correspondant du code source, la clé générée à partir de ce contenu et un élément de date prouvant notamment que ce
35 contenu n'a pas été généré a posteriori, à partir du logiciel compilé.

Il existe de nombreuses personnes, physiques ou morales, agissant en tant que tiers certificateurs. Ces personnes mettent en oeuvre des procédés consistant de manière générique à générer une archive à partir d'un ou plusieurs documents, à créer une clé unique à partir du contenu de ce document, et à attribuer à cette clé un élément d'horodatage, le plus souvent sous la forme d'un jeton, en rapport avec la date et l'heure auxquelles a été réalisé ledit procédé.

Là aussi, la façon dont est générée une archive à partir d'un ou plusieurs documents, l'algorithme utilisé pour générer la clé à partir de ladite archive et le procédé mis en oeuvre pour établir un jeton et la certification de ce jeton peuvent être amenés à changer au cours du temps.

Dans certain cas, ceci rend particulièrement délicat la revendication d'une quelconque propriété lorsqu'il s'agit de remonter à des politiques pouvant parfois dater de plusieurs années.

On se propose d'améliorer la situation.

L'invention vise un dispositif informatique de gestion temporelle de documents numériques du type comprenant une mémoire capable de stocker au moins un document numérique et une empreinte datée respective, ladite empreinte datée définissant une correspondance entre une ou plusieurs premières valeurs de signatures et au moins une valeur de temporalité, ces premières valeurs de signatures étant établies à partir de ce document numérique, conformément à une première méthode de signature, un générateur de signatures capable, sur présentation d'un contenu de document, d'établir une ou plusieurs secondes valeurs de signatures respectives, conformément à une seconde méthode de signature, un horodateur, comprenant une fonction d'élection de temporalité, capable d'établir une correspondance entre une ou

plusieurs valeurs de signature et une valeur-résultat de la fonction d'élection de temporalité, un vérificateur de signatures capable, sur présentation d'un contenu de document numérique et d'une empreinte datée, d'en vérifier la conformité mutuelle selon une ou plusieurs règles prédéterminées, un superviseur capable, sur présentation du document numérique et de son empreinte datée, de réaliser un traitement particulier. Ce traitement particulier consiste à faire opérer le générateur de signatures sur le document numérique pour obtenir une ou plusieurs secondes valeurs de signatures, à faire opérer le vérificateur de signatures sur le document numérique et l'empreinte datée, et, dans le cas où le document numérique est conforme à l'empreinte datée, à faire opérer l'horodateur avec au moins la valeur de temporalité du document numérique et certaines au moins des secondes valeurs de signatures afin de former une nouvelle empreinte datée comprenant de secondes valeurs de signatures.

Le dispositif selon l'invention permet de comparer des empreintes de documents, constituées de signatures, au lieu de comparer les documents eux-mêmes. Il en résulte d'abord une comparaison rapide et peu coûteuse en temps et en ressources de calcul. Incidemment cette comparaison n'est pas limitée à une comparaison de documents deux à deux. Il en découle ensuite une comparaison plus fine, en ce que plusieurs signatures peuvent être générées à partir d'un même document.

Le dispositif selon l'invention permet également de comparer des documents dont les empreintes ont été générées conformément à des méthodes de signature différentes. Ceci peut être fait en conservant le bénéfice de la date attribuée à une empreinte établie conformément à une ancienne méthode de signature, et ceci avec un niveau de confiance important.

Le dispositif selon l'invention permet de mettre en oeuvre une comparaison en terme d'identité ou de différence des contenus des documents. Mais ce dispositif permet surtout d'introduire un facteur temporel dans cette comparaison.

5 Autrement dit, il devient possible de dater les éléments communs ou les éléments par lesquels des documents diffèrent mutuellement. En particulier, des intégrations de portions de contenus d'un document dans un autre peuvent être mises en évidence. Dans ce cas, il est également possible d'identifier
10 un document d'origine et un document de destination, par comparaison des dates associées aux empreintes.

D'autres caractéristiques et avantages de l'invention apparaîtront à l'examen de la description détaillée ci-après,
15 et des dessins annexés sur lesquels :

- La figure 1 est un schéma fonctionnel d'un dispositif selon l'invention.
- 20 - La figure 2 est un organigramme illustrant le fonctionnement d'un contrôleur pour le dispositif de la figure 1.
- La figure 3 est un organigramme illustrant le fonctionnement d'un générateur de signatures pour le dispositif de la
25 figure 1, dans un premier mode de réalisation.
- La figure 4 est un tableau illustrant une empreinte temporalisée générée à l'aide du générateur de signatures illustré par la figure 3.
- 30 - La figure 5 est un organigramme illustrant le fonctionnement du générateur de signatures pour le dispositif de la figure 1, dans un troisième mode de réalisation.

- La figure 6 est un tableau illustrant une empreinte temporalisée générée à l'aide du générateur de signatures illustré par la figure 5.
- 5 - La figure 7 est un organigramme illustrant une variante de fonctionnement du contrôleur de la figure 1, le générateur de signatures étant conforme à son troisième mode de réalisation.
- 10 - La figure 8 est un organigramme illustrant le fonctionnement d'un horodateur pour le dispositif de la figure 1.
- La figure 9 est analogue à la figure 6, dans la variante de fonctionnement de la figure 7.
- 15 - La figure 10 est analogue à la figure 6, dans une autre variante de fonctionnement du contrôleur de la figure 1.
- La figure 11 est un tableau illustrant des empreintes temporelles générées à l'aide du générateur de signatures selon les premier et troisième modes de réalisation.
- 20 - La figure 12 est un organigramme illustrant un détail du fonctionnement du contrôleur de la figure 1.
- 25 - La figure 13 est un organigramme illustrant un autre détail du fonctionnement du contrôleur de la figure 1.
- La figure 14 est un tableau illustrant une empreinte temporalisée, générée à l'aide du contrôleur de la figure 1 fonctionnant conformément à la figure 13.
- 30 - La figure 15 est analogue à la figure 14 pour une variante de réalisation.

- La figure 16 est analogue à la figure 14 pour une autre variante de réalisation encore.
- La figure 17 est un organigramme illustrant le fonctionnement d'un générateur de signatures pour le dispositif de la figure 1 selon un autre mode de réalisation.
- La figure 18 est un diagramme illustrant également le fonctionnement du générateur de signatures dans le mode de réalisation de la figure 17.

Les dessins annexés pourront non seulement servir à compléter l'invention, mais aussi contribuer à sa définition, le cas échéant.

15

La figure 1 représente un dispositif informatique 1 selon l'invention pour une gestion temporelle de documents numériques.

20 Par document numérique, on entend ici tout ensemble cohérent de contenu sous forme numérique.

Un document numérique peut correspondre à un ou plusieurs fichiers informatiques, de type quelconque. Un fichier audio, un fichier vidéo et plus généralement tous fichiers multimédia, en format brut ou compressé selon un standard, sont des exemples de documents numériques, pour lesquels le dispositif informatique 1 peut être employé.

30 Les fichiers de type texte, qu'ils soient dans un format propre au logiciel ayant permis leur génération ou dans l'un des formats de fichier texte standardisés, constituent d'autres exemples de documents numériques.

35 Parmi ces fichiers de type texte, l'invention trouve un intérêt particulier avec les fichiers dits "sources", c'est-

10

à-dire comprenant une série d'instructions en un langage de programmation quelconque, destinées à être compilées en instructions exécutables par une machine de calcul, typiquement un ordinateur.

5

L'invention n'est cependant pas limitée à cette application particulière.

Plus généralement, le dispositif informatique 1 trouve un usage avantageux sur tout document numérique dont le contenu est susceptible de revêtir une forme quelconque de création, par exemple telle que régie par les dispositions légales sur le droit d'auteur.

15 En particulier, le dispositif informatique 1 se révèle tout à fait efficace lorsqu'il s'agit de revendiquer tout ou partie de la propriété d'un document, ce qui implique à un moment ou à un autre d'établir une forme de dépendance entre deux documents, ceci nécessitant notamment une robustesse du
20 dispositif informatique aux évolutions qui peuvent survenir au cours du temps.

Encore plus généralement, le dispositif informatique 1 peut être avantageusement appliqué à chaque fois qu'il existe un
25 intérêt quelconque à obtenir des informations fiables, en particulier de dates, relatives au contenu d'un document numérique.

Dans la suite de la présente description, à chaque fois qu'il sera fait référence à un document numérique, il conviendra de
30 considérer que ce document peut être physiquement constitué de plusieurs fichiers informatiques. Exceptionnellement, certaines mises en oeuvre particulières de l'invention pourront nécessiter qu'un document numérique individuel
35 corresponde à un unique fichier informatique. Cela sera alors clairement indiqué.

Par exemple, il sera régulièrement fait référence à un document numérique dans le cas d'un ensemble de fichiers informatiques de type "code source", constituant une version d'un logiciel ou plus généralement une étape dans un projet
5 en développement. Dans ce cas, deux documents distincts pourront être vus comme deux versions distinctes du projet.

Le terme "contenu" du document n'est pas nécessairement sans rapport avec le ou les fichiers informatiques qui le contiennent ou plus généralement son ou ses conteneurs. Le contenu
10 du document peut ainsi comprendre sa structure de stockage sous forme numérique, ou un ou plusieurs attributs de son conteneur. Le contenu d'un fichier informatique pourra comprendre le nom attribué à ce fichier informatique, en
15 particulier dans le cas où le nom de ce fichier est fortement lié au reste du contenu du document, par exemple lorsque le nom du fichier résulte d'une convention d'établissement de nom prenant en compte, par exemple, la date de génération dudit fichier. Le contenu du document peut également com-
20 prendre une structure hiérarchisée complexe. Par exemple, un fichier peut comprendre une archive, laquelle comprend des répertoires et une structure hiérarchisée de stockage, le tout formant, par exemple, un projet de développement de logiciel informatique. Le contenu de ce fichier-archive
25 comprend non seulement l'ensemble des codes-sources, mais également toute la structure hiérarchisée du stockage de ces sources.

La présente convention peut paraître contraire au langage informatique usuel, dans lequel le sens du mot "document"
30 résulte principalement de l'utilisation de ce mot dans l'interface graphique de certains systèmes d'exploitation.

Pourtant, l'acception du mot "document" dans le langage
35 courant est bien plus large et correspond tout-à-fait à la présente convention. D'autre part, il apparaît assez claire-

ment que le découpage d'un document numérique en un ou plusieurs fichiers informatiques répond le plus souvent à un choix arbitraire, auquel le dispositif selon l'invention est pratiquement insensible.

5

Dans certains cas, un document numérique pourra correspondre à une partie seulement d'un fichier informatique.

Le dispositif informatique 1 comprend une mémoire de stockage
10 3 capable, entre autres choses, de conserver des documents numériques, un générateur de signatures 5 capable, sur présentation d'un document numérique, d'établir une ou plusieurs signatures représentatives du contenu de ce document, un horodateur 7 capable d'associer une référence
15 temporelle à chaque signature qui lui est présentée selon des règles déterminées d'horodatage, et un vérificateur de signatures 9 capable de vérifier la conformité d'un contenu de document à une ou plusieurs signatures relatives à celui-ci.

20

La mémoire 3 peut être organisée à la manière d'une base de donnée, par exemple de type relationnelle. Elle peut être mise en oeuvre avec tous types de systèmes de fichiers, tels que FAT, NTFS, et sous tous systèmes d'exploitation, y
25 compris Unix.

Le dispositif informatique 1 comprend en outre un contrôleur 11, ou superviseur, capable d'interagir avec le générateur de signatures 5, l'horodateur 7, le vérificateur de signature 9
30 et la mémoire 3.

La figure 2 illustre le fonctionnement du contrôleur 11.

Un document numérique Di et une donnée de date associée Ti
35 sont présentés au contrôleur 11 au cours d'une étape 200.

La date T_i associée au document D_i est de préférence une date en rapport avec la création de ce document. En pratique, la date T_i peut être obtenue de différentes manières. Lorsque le document D_i correspond à un projet, la date T_i peut provenir
5 d'un serveur de stockage de sources, d'un outils de gestion de sources tel que CVS (pour "Concurrent Version System" ou système de versions concurrentes), ou plus simplement correspondre à une date de création du fichier informatique le plus récent/le plus ancien. Dans le cas d'un unique
10 fichier, cette date T_i peut être la date de création, telle que mémorisée à l'intérieur du fichier lui-même, une date certifiée par un tiers, ou encore une date de création de contenu, en particulier lorsque celui-ci est antérieur à la date du fichier informatique.

15

Dans une étape 202, le contrôleur 11 présente le document D_i au générateur de signatures 5, lequel renvoie une ou plusieurs signatures S_i représentatives du contenu du document D_i .

20

Avantageusement, la signature est statistiquement unique mais cohérente avec le contenu du document D_i . On entend par là que la probabilité que deux contenus distincts provoquent la génération de deux signatures identiques en valeur est aussi
25 réduite que possible, tout en assurant que deux contenus identiques provoquent une même valeur de signature.

Pour chacune des signatures S_i générées à l'étape 202, le contrôleur 11 appelle l'horodateur 7, lequel associe au moins
30 une référence de temporalité R_i à une valeur de signature, au cours d'une étape 204.

Par référence de temporalité, on entend ici une donnée qui a trait à une information de date, relative ou absolue, quant
35 à la valeur de signature en question, et plus précisément au contenu qui a permis de générer cette signature. Une référé-

rence de temporalité R_i peut prendre, de manière non limitative, la forme d'une date ou d'une correspondance à une date.

Par exemple, une référence de temporalité R_i peut prendre la
5 forme d'un identifiant de version dans une suite organisée de documents. Dans certains cas, une référence de temporalité R_i peut faire référence à une autre référence de temporalité, le cas échéant avec une information complémentaire, telle que "plus ancienne" ou "plus récente".

10

Le contrôleur 11 interagit ensuite avec la mémoire 3 pour y stocker, pour le document D_i , l'ensemble des signatures S_i générées et leur référence de temporalité, en correspondance mutuelle.

15

Cette correspondance entre un document D_i , l'ensemble des signatures S_i générées à partir de son contenu et les références de temporalités R_i associées auxdites signatures S_i est ici désigné "empreinte temporalisée du document D_i ".

20

La figure 3 illustre un premier mode de réalisation du générateur de signatures 5.

Dans ce mode de réalisation, une unique signature S_i est
25 générée à partir d'un document D_i .

Le générateur de signatures 3 est agencé de manière à appliquer une fonction filtre F_1 d'un premier type au document D_i , au cours d'une étape 300.

30

La fonction filtre F_1 est principalement agencée de manière que la signature S_i générée pour le document D_i soit robuste à tout changement minime.

35 On entend par là que la signature générée à partir du contenu du document D_i filtré doit être identique à une signature

générée à partir d'un document dont le contenu est estimé sensiblement identique à celui du document Di.

La notion de "changement minime" ou "d'identité des contenus" dépend en grande partie du type de document traité et de conventions adoptées.

Dans le cas d'un fichier de type source, il peut être convenu qu'une mise en forme du contenu du document ne constitue qu'un changement minime. La fonction filtre F1 peut alors être agencée en conséquence et, par exemple, appliquer une mise en forme prédéterminée au contenu avant génération de la signature. En variante, la fonction filtre F1 peut être agencée de manière à supprimer toute mise en forme.

15

La fonction filtre F1 peut également être agencée de manière à supprimer tous commentaires du fichier de type source, et/ou caractères étrangers à la sémantique du langage de programmation, et/ou caractères dépendants d'un système d'exploitation particulier, dans le cas où il est convenu que l'insertion de tels éléments ne modifie pas de manière significative le contenu du document Di.

20

La fonction filtre F1 peut encore être agencée de manière à renommer suivant une convention préétablie l'ensemble des variables et des fonctions décrites dans le fichier de type source, en sorte que la génération de signatures deviendra robuste à une opération de renommage de ces éléments.

25

D'autres exemples de modifications qui peuvent être jugées non significatives sont :

30

- des modifications concernant uniquement la mise en forme du contenu du document, telle que l'ajout de caractères d'espacement ou de lignes vierges dans un fichier de type texte,

35

- la simple réécriture du contenu, telle que le changement des noms de variables ou de fonctions, et/ou l'ajout ou la suppression de mentions de "copyright" (droit de copie), et/ou

5

- la modification du nom d'un ou plusieurs fichiers stockant le contenu, et plus généralement toutes modifications dans la structure de stockage de ce document telles que subdivision en un ou plusieurs fichiers, noms des répertoires de stockage
10 de ces fichiers, arborescence des répertoires et analogues.

Le cas échéant, la fonction filtre F1 peut être adaptée au type du document Di et/ou du système d'exploitation sur lequel ce document Di a été généré. Des fonctions filtre F1
15 différentes peuvent être prévues lorsqu'il s'agit de générer une signature représentative du contenu d'un document musical, vidéo ou d'image.

L'application de la fonction filtre F1 est avantageuse en ce
20 qu'elle améliore la robustesse des signatures générées. Cette application reste optionnelle.

Le générateur de signatures 5 est en outre agencé de manière à appliquer une fonction de hashage H1 d'un premier type au
25 contenu du document Di, après application éventuelle de la fonction filtre F1. La fonction de hashage H1 retourne une signature Si, statistiquement unique, représentative du contenu du document Di. En pratique, une telle signature peut prendre la forme d'une chaîne de caractères alphanumériques,
30 d'autres formes pouvant être envisagées.

La fonction de hashage H1 peut être réalisée de différentes façons.

Par exemple, cette fonction de hashage H1 peut mettre en oeuvre les algorithmes de cryptage MD5, SHA-1 ou SHA-256, et analogues.

- 5 Plus généralement, toute fonction capable d'établir à partir d'un contenu de document un identifiant en rapport avec ce contenu peut être utilisée en tant que fonction de hashage H1.
- 10 Avantageusement, on préférera des fonctions de hashage H1 telles, que la signature générée soit unique, ou plus exactement statistiquement unique. Ceci permet ultérieurement de comparer des signatures associées à des documents plutôt que le contenu de ces documents. Il en découle un gain de
- 15 temps de calcul considérable.

En pratique, des fonctions dites "irréversibles" ou "inviolables", telles que les fonctions établies sur la base d'algorithmes de cryptage, seront avantageusement utilisées.

20

- Les fonctions de hashage H1 possibles ne sont pas limitées aux seules fonctions de type cryptographique. Les fonctions susceptibles de donner d'autres informations pertinentes quant au contenu d'un document, telles que sa proximité en
- 25 terme de contenu à d'autres documents selon des conventions préétablies, peuvent être utilisées. De telles fonctions ne divulguent pas en effet le contenu du document, mais uniquement une certaine "proximité" à un autre document.

- 30 De telles fonctions permettent non seulement d'obtenir un identifiant numérique statistiquement unique, mais encore d'éviter que le contenu du document Di, ou un équivalent significatif de ce contenu, soit retrouvé à partir de sa signature, du moins dans la mesure d'efforts raisonnables.

35

L'emploi de telles fonctions autorise une large diffusion des signatures générées, avec un risque minime de divulgation des contenus. Ces derniers peuvent en effet revêtir un certain savoir-faire, en particulier lorsqu'il s'agit de fichiers de programmation.

Dans certains cas, il pourra cependant être préféré de ne divulguer qu'une partie des signatures générées, que certaines d'entre elles, par exemple les signatures générées à partir de contenus d'une taille supérieure à un seuil donné, et/ou que certains des attributs associés à une signature.

Il peut ainsi être mis en place ce l'on pourrait appeler une "politique de divulgation". Ceci peut être le cas en particulier lorsqu'un ensemble de documents constitue des versions successives d'un logiciel en développement. Dans ce cas en effet, la divulgation successive des signatures générées à partir de chacune des versions tend à fournir des informations supplémentaires, en sorte que les algorithmes de hashage et les fonctions filtre en deviennent, en termes de probabilité, nécessairement moins fiables. Ceci est tout particulièrement vrai dans les cas où, comme on le verra plus loin, on génère non pas une mais une pluralité de signatures à partir d'un même document, car, ce faisant, on multiplie les informations reliant les signatures entre elles. Plus la taille du contenu à partir duquel une signature a été générée est petite, moins sont robustes les fonctions filtre et de hashage utilisées aux attaques directes, c'est-à-dire visant à retrouver le contenu en question à partir de la signature par essais successifs. Toute indication de liaison entre les contenus dégrade également cette robustesse, car elle limite les essais nécessaires.

Selon une première variante du premier mode de réalisation, l'horodateur 7 est agencé de manière à attribuer à la signature Si sa date de génération en tant que référence de

temporalité Ri. Cette date de génération peut être obtenue du système d'exploitation, éventuellement corroborée par un serveur de temps, et peut prendre la forme d'un jeton d'horodatage.

5

Selon une seconde variante de ce premier mode de réalisation, l'horodateur 7 est agencé de manière à attribuer à la signature Si la date Ti associée au document Di en tant que référence de temporalité Ri. Cette attribution peut être conditionnée par l'obtention d'un élément probant accréditant cette date Ti, par exemple une déclaration d'un tiers certificateur, ou d'un jeton d'horodatage.

10

Selon une troisième variante de ce premier mode de réalisation, l'horodateur 7 est agencé de manière à attribuer à la signature Si la date Ti associée au document Di, si et seulement si cette date Ti est associée à un niveau de confiance acceptable, et à attribuer la date de génération de signature dans le cas contraire.

15

20

Quelque soit la variante de réalisation, l'horodateur 7 peut être agencé de manière à en outre appeler une fonction d'élection de temporalité pour la valeur de signature Si.

La fonction d'élection de temporalité peut être agencée de manière à vérifier l'existence de la signature Si dans la mémoire 3, et, dans le cas où cette signature existe, à attribuer en tant que référence de temporalité Ri, l'une de la nouvelle référence de temporalité Ri et de la référence temporelle déjà mémorisée. Par exemple, la plus ancienne de ces deux références de temporalité peut être attribuée à la signature Si. Ceci permet notamment de mettre en évidence une date de première apparition de la signature Si dans un ensemble de documents. Cette date de première apparition peut servir de base à l'identification d'intégration de contenu d'un document dans un autre.

25

30

35

La figure 4 illustre les empreintes temporalisées de documents D1 à D5 obtenue à l'aide du générateur de signatures 5 dans son premier mode de réalisation.

- 5 Les différentes signatures sont indiquées en colonne COL400 et les identifiants des différents documents en ligne ROW400. La correspondance entre une signature et un document particulier est indiquée par la présence d'un chiffre "1" encadré.
- 10 Par exemple, l'empreinte temporalisée du document D1 comprend la signature "165436" (présence du chiffre "1" dans la case COL401,ROW401).

Ici, les références temporelles R_i associées aux signatures
15 S_i ont été déterminées au moyen de la date T_i associée à chacun des documents D_i . Autrement dit, la valeur de la référence temporelle R_i n'est pas ici indiquée explicitement sur la figure, mais la correspondance entre une signature S_i et cette référence temporelle R_i se déduit de la présence
20 d'un chiffre "1".

Ceci peut être vu également comme l'attribution d'un numéro de document en tant que référence temporelle R_i , les documents D_i ayant été numérotés chronologiquement, par exemple
25 ici à partir des dates T_i qui leur ont été associées.

Ceci montre que la référence temporelle R_i ne prend pas nécessairement la forme d'une date. Dans certains cas, comme ici, la référence temporelle peut être relative.

30

La signature "165436" bénéficie de la date associée au document D1, le plus ancien des documents D_i , tandis que la signature "915528" bénéficie de la date du document D5, le plus récent des documents D_i .

35

La mémoire 3 est agencée de manière à maintenir pour chaque document D_i , ici les documents D_1 à D_5 , la signature S_i générée et la référence temporelle R_i associée à cette valeur de signature. Le cas échéant, la mémoire 3 peut être organisée de manière à maintenir une correspondance entre plusieurs documents, ici par exemple un lien est maintenu entre les empreintes des documents D_1 à D_5 .

Dans un second mode de réalisation du générateur de signatures 5, la génération d'empreintes temporalisées à partir de document D_i est réalisée hors du dispositif informatique 1.

Le générateur de signature 5 est agencé de manière à récupérer au minimum un identifiant de document D_i , une signature S_i générée à partir du document D_i , et une date associée à cette signature S_i .

Plusieurs documents peuvent être reçus simultanément. La mémoire 3 peut alors être agencée de manière à conserver les empreintes temporelles d'un ensemble de documents D_i liés entre eux.

En option, le module horodateur 7 peut appeler une fonction d'élection de temporalité pour chacune des signatures S_i afin d'établir une nouvelle référence de temporalité R_i à partir de références de temporalité associées à cette signature S_i dans la mémoire 3.

Dans une première variante de ce second mode de réalisation, la date associée à la signature S_i est considérée comme étant a priori valable. Ceci correspond à un niveau de confiance en la justesse de la date associée à la signature S_i assez faible. Ceci présente néanmoins l'avantage d'une mise en oeuvre assez simple du dispositif.

Dans une seconde variante de réalisation, le générateur de signatures 5 est agencé de manière à vérifier la validité de la date associée à la valeur de signature Si. Par exemple, le générateur de signatures 5 peut être agencé de manière à recevoir un jeton d'horodatage provenant d'un tiers horodatant fournissant des formes de stockage de confiance.

Dans ce cas, un jeton d'horodatage peut être associé de manière unique à la signature Si, une date étant attribuée à ce jeton. Lorsque la validité du jeton est vérifiée, on obtient un niveau de confiance dans l'association de la signature et de la date semblable au niveau de confiance accordé à l'émetteur dudit jeton. De multiples procédures de vérification existent, lesquelles procédures dépendent essentiellement de l'émetteur du jeton. Par exemple, le jeton, et éventuellement une date et/ou la signature Si, peuvent être présentés à un service tiers pour certification de l'association de la date et de la signature. Dans d'autres cas, l'émetteur du jeton peut faire connaître une clé publique qui lui est propre, laquelle clé peut être utilisée pour vérifier la consistance du jeton avec la signature Si et une valeur de date. Le jeton n'est pas nécessairement directement accessible au dispositif 1. Dans certains cas, seule une référence à un jeton d'horodatage, conservé auprès du tiers, pourra être accessible.

Des procédures de type différent peuvent également être mises en oeuvre pour vérifier la validité de la date associée à la signature Si.

30

La figure 5 illustre un troisième mode de réalisation du générateur de signatures 5.

Dans ce mode de réalisation, plusieurs signatures Si sont générées à partir d'un document Di. Ceci permet de mettre en oeuvre une comparaison plus fine de documents Di entre eux.

35

Le générateur de signatures 5 est agencé de manière à appliquer un filtre F2 d'un second type à tout document Di qui lui est présenté, au cours d'une étape 500.

- 5 Dans une étape 502, le générateur de signature 5 applique au contenu du document Di ainsi filtré une fonction de morcellement, capable d'extraire le contenu du document Di et de le scinder en plusieurs éléments, selon des règles prédéterminées.

10

La fonction de morcellement, et les règles conformément auxquelles cette fonction agit, peuvent être agencées de différentes manières.

- 15 Par exemple, dans le cas d'un fichier de type source, dont le contenu est rédigé selon un langage de programmation particulier, la fonction de morcellement peut être agencée de manière à extraire chacune des fonctions décrites dans le document Di. Les règles de morcellement peuvent alors être
20 établies sur la base d'une recherche d'expressions consacrées à la déclaration des objets de type fonction dans le langage de programmation en question.

- Dans le cas où un document est physiquement organisé en
25 plusieurs fichiers informatiques, y compris ayant entre eux des relations complexes, la fonction de morcellement peut être agencée de manière à individualiser ces fichiers informatiques, dans un premier temps au moins.

- 30 Étant donné que la notion de "contenu d'un document" est plus générale que les seules informations qu'il est possible d'afficher sur un ordinateur, tel que le texte contenu dans un fichier texte, l'image d'un fichier image ou un film contenu dans un fichier vidéo, la fonction de morcellement
35 peut être agencée de manière à agir sur des éléments non visualisables. Par exemple, la fonction de morcellement peut

- être agencée de manière à extraire la structure, ou l'arborescence, d'une archive, telle qu'une archive au format TAR.GZ par exemple, et plus généralement sur la structure de stockage du contenu d'un document numérique. La fonction de morcellement peut en outre être agencée de manière à agir sur des éléments de taille, ou "de granularité" différentes. Par exemple, dans le cas où un document Di représente un ensemble de fichiers sources, la fonction de morcellement peut être agencée de manière à d'abord découper les documents en fichiers, représentant ainsi un premier niveau de granularité, et ensuite à découper ledit fichier en fonctions. Autrement dit, le résultat de la fonction de morcellement appliquée au document Di a pour résultat un ensemble de fichiers et un ensemble de fonctions contenues dans ces fichiers. Autrement dit, il sera généré pour un fichier informatique une signature correspondante à ce fichier et une signature pour chacune des fonctions contenues dans ce fichier.
- 20 La fonction de morcellement peut ainsi être agencée de manière à découper plusieurs fois et de différentes manières un document Di, de manière non successive, chacune des opérations de découpe fournissant un ensemble d'éléments à signer.
- 25 Dans une étape 504, le générateur de signatures 5 débute une boucle sur chacune des parties du contenu du document Di issues de l'étape 502.
- 30 Cette boucle débute par l'application d'une fonction filtre F3 d'un troisième type, au cours d'une étape 506, et se poursuit par l'application d'une fonction de hashage H2 d'un second type, au cours d'une étape 508.
- 35 La fonction filtre F2 vise avant tout à rendre le résultat de la fonction de morcellement le plus robuste possible.

Autrement dit, cette fonction filtre F2 est agencée de manière que deux documents composés de manière semblable soient découpés de la même manière. Aussi la fonction filtre F2 peut-elle être établie en rapport avec la fonction de morcellement.

Dans le cas de fichiers de type sources, la fonction filtre F2 peut être agencée de manière à mettre en forme le contenu du document, conformément à une convention de présentation, tandis que la fonction de morcellement est agencée de manière à découper en fonction de la présentation en question.

Pour l'essentiel, la fonction filtre F3 répond à des impératifs analogues à ceux de la fonction filtre F1, en particulier concernant la robustesse des signatures générées.

La fonction filtre F3 peut être adaptée en fonction du type de contenu de la partie découpée : des filtres différents peuvent être mis en oeuvre suivant que les parties découpées correspondent à des fonctions, des données, des parties d'images ou de pièces musicales.

Il est fait référence ici à une unique fonction de hashage H2 aux seules fins de simplification. En pratique, plusieurs fonctions de hashage différentes pourront être mises en oeuvre, celles-ci pouvant, par exemple, être adaptées selon le contenu de la partie à traiter.

Selon une première variante de ce troisième mode de réalisation, l'horodateur 7 est agencé de manière à associer la date T_i associée au document D_i en tant que référence de temporalité R_i à chacune des signatures S_i générées pour ce document D_i , dans un premier temps au moins.

Ceci peut être fait au moyen d'une fonction d'élection de temporalité qui établit la date T_i en tant que référence

Dans une étape 702, l'horodateur 7 coopère avec la mémoire 3 pour déterminer si la signature Si est déjà stockée dans cette mémoire. En variante, la recherche de la signature Si peut être limitée à des documents Dj dont les empreintes sont
5 mémorisées en rapport avec le document Di en question.

Si oui, alors l'horodateur 7 est agencé de manière à appeler une fonction d'élection temporalité avec l'ensemble des dates associées à ladite signature dans la mémoire 3 et la date Ti
10 associée au document Di, dans une étape 706. La fonction d'élection de temporalité retourne une référence de temporalité Ri, calculée à partir de ces dates, qui sera mémorisée en correspondance avec la signature en question.

15 Sinon, la date Ti associée au document Di est établie en tant que référence de temporalité Ri de cette signature, au cours d'une étape 706.

Ici, la fonction d'élection de temporalité travaille sur
20 l'ensemble des empreintes temporalisées déjà stockées dans la mémoire 3 pour attribuer aux signatures Si calculées à partir d'un nouveau document Di des références temporelles Ri potentiellement calculées à partir des dates Ti de documents Di déjà traités.

25

Ceci permet d'établir une empreinte globale pour un ensemble de documents, dans laquelle un ensemble de signatures est recensé, chaque signature ayant une référence de temporalité associée.

30

Ceci permet également d'actualiser les références temporelles Ri associées aux signatures Si au fur et à mesure que des documents Di sont traités par le dispositif informatique 1. Il devient possible de créer une bibliothèque d'empreintes
35 temporalisées servant à la comparaison d'une pluralité de

documents, y compris futurs, sans stocker les documents eux-mêmes.

La figure 8 illustre un exemple de réalisation de la fonction
5 d'élection de temporalité.

Dans une étape 800, la fonction d'élection de temporalité vérifie si la date T_i est plus ancienne que la référence de temporalité R_i associée à la signature S_i dans la mémoire 3.
10

Si oui, alors la référence de temporalité R_i prend la valeur de la date T_i associée au document D_i , dans une étape 802.

Sinon la référence de temporalité R_i demeure inchangée (étape
15 804).

Dans ce mode de réalisation, la référence de temporalité R_i est déterminée en tant que plus ancienne date de présence de la signature S_i dans l'ensemble des documents D_i traités.
20

Le cas échéant, la fonction d'élection de temporalité peut être agencée de manière à prendre en compte d'autres critères, tels qu'un indice de fiabilité de la date T_i , par exemple.
25

La figure 9 illustre les empreintes temporalisées de documents D_1 à D_{18} établies conformément à la seconde variante du troisième mode de réalisation.

30 La colonne COL901 rassemble les signatures générées à partir des documents D_1 à D_{18} . Ces documents D_1 à D_{18} sont ordonnés chronologiquement dans la ligne ROW901, grâce par exemple à leur date T_i associée.

35 La présence d'un chiffre encadré à l'intersection d'une colonne correspondant à un document D_i et d'une ligne

1/18

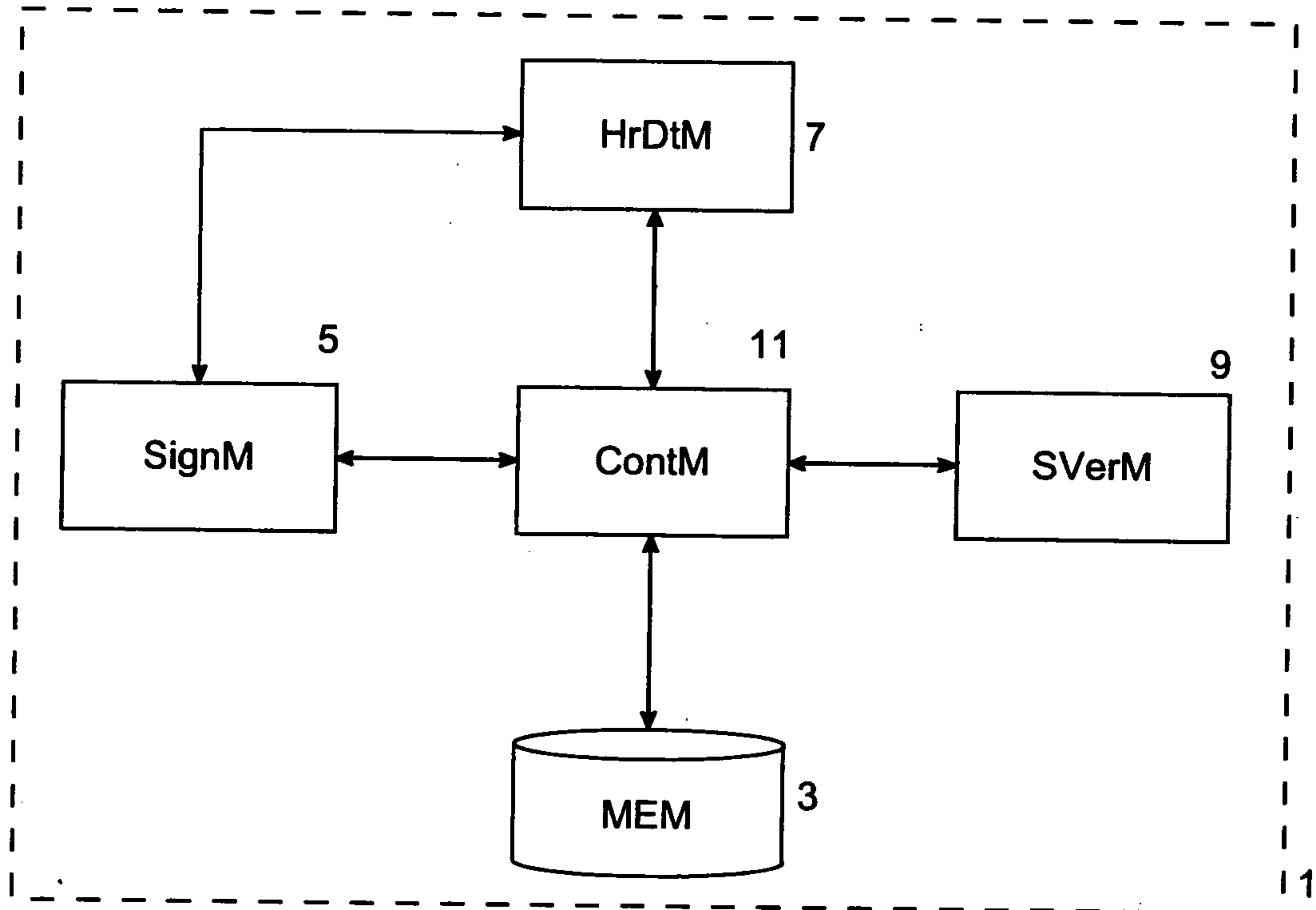


Fig. 1

2/18

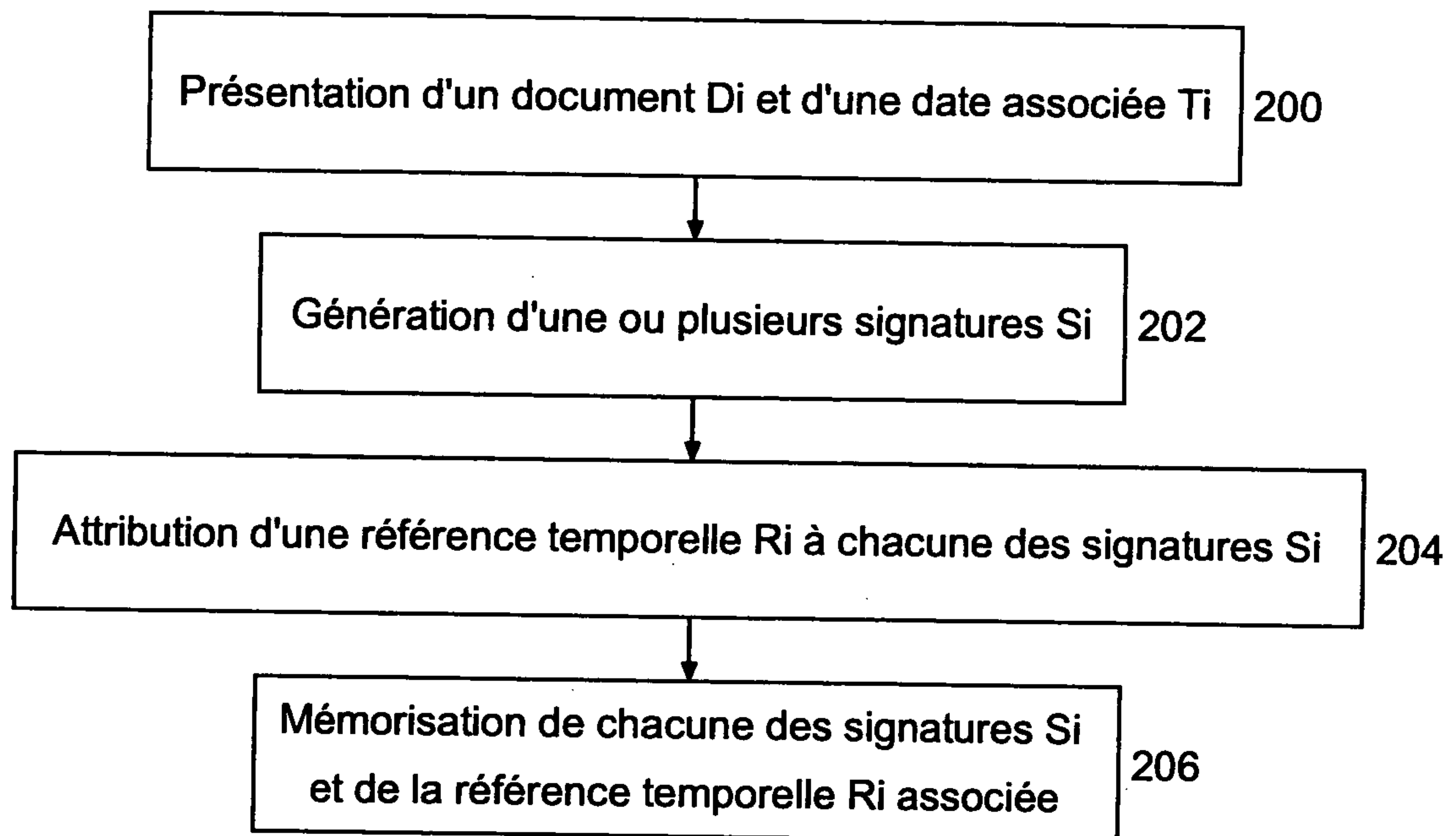


Fig. 2

3/18

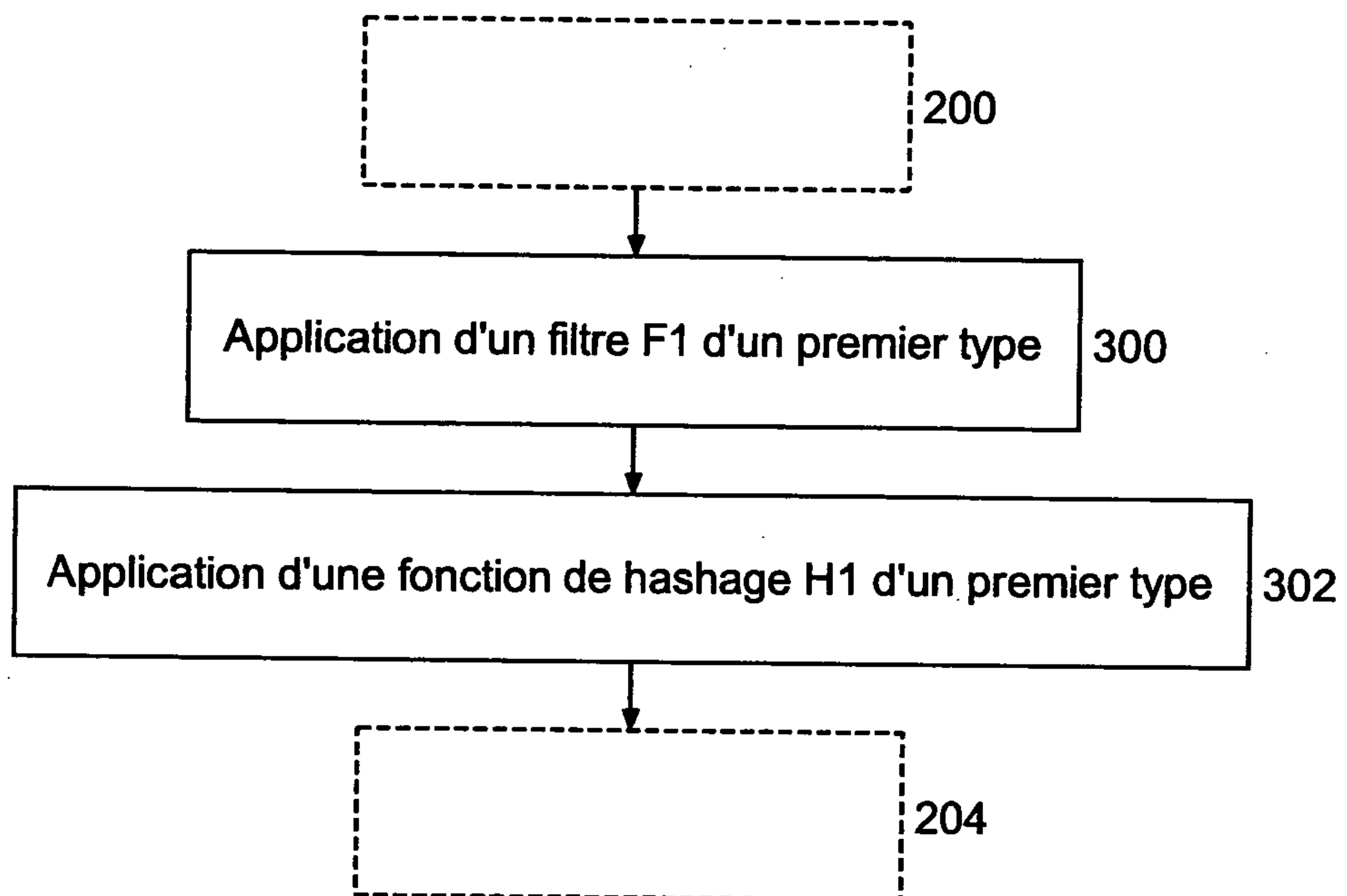


Fig. 3

