Institut Mines-Télécom

Personnalisation de services : quelles technologies pour la préservation de la vie privée ?

Nesrine Kaâniche et Maryline Laurent

Livre électronique avril 2019

Chaire
Valeurs et Politiques
des Informations Personnelles

TELECOM ParisTech
TELECOM SudParis
Institut Mines-Télécom
Institut Mines-Télécom Business School

Les auteures

Maryline Laurent

Professeure en sciences informatiques à Télécom SudParis, cofondatrice de la Chaire Valeurs et Politiques des Informations Personnelles, responsable de l'équipe R3S du laboratoire CNRS UMR 5157 SAMOVAR et éditrice associée de plusieurs journaux scientifiques dont Annals of Telecommunications. Elle s'intéresse aux problématiques de sécurité et de vie privée dans les environnements : *cloud*, Internet des objets et gestionnaires d'identités. Elle a coédité plusieurs ouvrages dont « La gestion des identités numériques », Édition ISTE en 2015. Elle a publié plus d'une centaine de publications dans des journaux et conférences scientifiques de référence.

Nesrine Kaâniche

Chercheuse en sciences informatiques à Télécom SudParis, et membre de la Chaire Valeurs et Politiques des Informations Personnelles. Elle s'intéresse aux technologies de préservation de la vie privée, à la sécurité dans le *cloud* et aux applications de la cryptographie. Elle est co-auteure de plus de 25 publications dans des revues et ouvrages majeurs. Elle est relectrice régulière pour les revues Annals of Telecommunications et Security and Communication Networks.

Institut Mines-Télécom

CONTACTS

Maryline Laurent
maryline.laurent@telecom-sudparis.eu

Nesrine Kaâniche
nesrine.kaaniche@telecom-sudparis.eu

UMR 5157 SAMOVAR
Télécom SudParis
9 rue Charles Fourier
91011 EVRY
samovar.telecom-sudparis.eu

Chaire Valeurs et Politiques des Informations Personnelles
Télécom ParisTech - IMT
46 rue Barrault | F-75634 Paris Cedex 13
cvpip@imt.fr
www.informations-personnelles.org

À propos de ce document

Ce document est le compagnon de résultats scientifiques intitulés : Privacy Enhancing Technologies for solving the Privacy-Personalization Paradox: Taxonomy and Survey.

Destiné à un lectorat qui peut être novice ou avancé, il est proposé sous différents formats de lecture. Dans un contexte grandissant de personnalisation des services numériques, il permet à chacun de se situer dans les divers choix d'outils et de stratégies pour préserver la vie privée des utilisateurs. Il se focalise sur plusieurs aspects de la vaste question de la vie privée, sans en aborder cependant toutes les facettes.

Des technologies de plus en plus datavores de nos données personnelles

Le déploiement à grande échelle des technologies numériques, notamment le développement des infrastructures *cloud* et des réseaux sociaux, a augmenté singulièrement le volume de données personnelles collectées, traitées et conservées.

En 2018, ce volume de données produites quotidiennement était estimé[1] à 2.5 exabytes, 90% du volume créé l'ayant été pendant les deux années précédentes. Surtout, quels que soient les rapports qu'un individu entretient avec les technologies, il laisse des traces et des données personnelles, notamment lorsqu'il remplit un formulaire en ligne, utilise sa carte de crédit ou une carte de fidélité, ou passe un péage.

Et ce n'est pas tout ! Le nombre croissant d'objets connectés occupe une place de plus en plus courante, voire prépondérante, dans la vie quotidienne des utilisateurs. Cela va des téléphones mobiles débordant d'applications – s'appuyant sur la localisation, l'heure ou d'autres paramètres de contexte – aux applications domotiques, sans oublier les vêtements et accessoires collectant des données via des capteurs, pour monitorer le jour les séances de fitness, la nuit les troubles du sommeil, et toute sorte de mouvement ayant une forme intéressante à étudier.

Et si des technologies étaient aussi conçues pour minimiser la création de données personnelles ?

Des technologies pour préserver la vie privée

Avec l'augmentation galopante des usages du numérique, de plus en plus de personnes prennent conscience des risques de divulgation de données personnelles. Un certain nombre de mécanismes, réunis sous le vocable de ***Privacy Enhancing Technologies*** **(PETs)**, entendent apporter une réponse technique à ce problème. Construits selon les principes de *Privacy by Design*, ils permettent l'émergence de services ayant comme objectif de protéger efficacement la vie privée, dans des domaines et dans des environnements variés.

Des services personnalisés de plus en plus invasifs

En parallèle, une classe de services, dite ***services personnalisés***, est apparue.

Chacun vit dans un environnement hyper-connecté générant, captant ou traitant continuellement de larges corpus de données. Les utilisateurs se voient alors proposer des recommandations ciblées, des films, des livres, des publicités, parfois hélas au prix d'atteintes à leur ***vie privée***. Les informations les concernant peuvent en effet faire l'objet de traitements inadéquats, menant à des publicités non désirées, des usurpations d'identité et des biais de discrimination.

Mais qu'est-ce que la vie privée ? On peut se référer au droit américain, lequel définit le « *right to privacy* » comme étant « le droit d'être laissé tranquille ». La Cour européenne des droits de l'homme se réfère pour sa part à la notion de « droit à l'autonomie personnelle » consistant à protéger, de façon positive, un droit pour la personne d'effectuer certains choix par elle-même. Cette préoccupation[2] est devenue à notre époque critique, notamment en raison des révélations successives concernant les abus de collecte et de traitement de données personnelles. Deux nous semblent particulièrement emblématiques : le programme d'espionnage américain NSA révélé par Edward Snowden en mai 2013 et le scandale Cambridge Analytica impliquant Facebook rendu public en avril 2018.

Ces services présentent à la fois des avantages et des inconvénients. Quel équilibre trouver ?

Une personnalisation couramment utilisée

Au gré de notre navigation sur Internet, ou lors de l'utilisation de services ou d'applications mobiles, nous sommes régulièrement amenés à nous authentifier. Cette *authentification* a pour objectif de «*prouver que nous sommes bien qui nous prétendons être*». Utilisé la plupart du temps pour nous accorder des droits supplémentaires (d'accès, de manipulation de données…), ce processus doit être sécurisé et digne de confiance; c'est pourquoi il repose sur des informations connues seulement de la personne concernée. Les technologies aidant, l'authentification n'est plus toujours créée par un acte formel et volontaire, comme entrer un mot de passe ou un code PIN.

Nous nous sommes habitués à l'établissement du lien entre notre identité numérique et nos données personnelles enregistrées et extraites de nos échanges. Les navigateurs ont intégré depuis longtemps des pisteurs (cookies…) qui transmettent des données personnelles aux prestataires de services, voire à des tiers, et qui permettent de pister l'individu. Une fois qu'une activité est associée à un individu précis, le prestataire peut enrichir son profil avec des données personnelles (ses préférences, ses centres d'intérêt…) et, grâce à des algorithmes de plus en plus performants, il peut lui apporter une information, un service ou un contenu ciblé. Parfois, la personnalisation est uniquement liée à une situation vécue par l'individu: le fait d'être géolocalisé à un endroit précis, par exemple, déclenchera l'envoi d'une publicité ou d'un contenu ciblé.

Des services personnalisés bien commodes

C'est indéniable, les avantages[3] offerts par les services personnalisés sont nombreux, à commencer par les recommandations ciblées qui s'appuient sur des points d'intérêt, des événements, des actualités, ou les offres promotionnelles pour un service de proximité ou un produit, ou des suggestions de films, de livres etc.

Autre exemple : les résultats retournés par les moteurs de recherche, pour la plupart, personnalisent les réponses en fonction du profil de l'internaute (et ce qu'il est au moment de sa requête : consommateur, parent, simple curieux etc.). Cette personnalisation est effectuée dès le mot clé tapé pour identifier la sémantique associée à la requête. Ainsi le mot « souris » peut signifier « rongeur » pour un vétérinaire, « souris d'agneau » pour un gastronome, ou « dispositif de pointage pour ordinateur » pour un internaute. Les applications mobiles (un mobile est personnel et souvent à portée de main) s'inscrivent aussi dans cette logique : celles relatives à la santé ou au bien-être (par exemple les nouveaux dispositifs FitBit / Vivosport) se révèlent parfois un outil précieux en proposant des astuces pour améliorer son hygiène de vie, suivre un traitement à distance, ou alerter l'utilisateur quant à d'éventuels problèmes de santé corrélés à une maladie.

Les risques de fuite de données personnelles, les abus observés sur l'exploitation de ces données, les dérives vers des environnements de surveillance et de notation sociale généralisés vont-ils nous faire regretter ces avantages, ou pouvons-nous corriger ces méfaits à temps ?

Une réaction individuelle et collective

Un nombre grandissant d'utilisateurs se sentent préoccupés[4] par les conséquences négatives des collectes massives de données, résultant du suivi à grande échelle de leurs vies quotidiennes. Ce suivi permanent n'est pas sans poser des questions en termes de respect des droits et libertés fondamentaux (droit au respect de la vie privée, à la non-discrimination mais aussi à la liberté de déplacement, à la liberté d'expression ou de rassemblement) et de valeurs sociétales partagées.

En 2014, un rapport intitulé *Big Data and Privacy*, publié par la Maison Blanche, a ainsi mis en lumière les défis posés par la collecte des données personnelles par des systèmes techniques omniprésents. En 2016, l'Union européenne (UE) a adopté[5] le Règlement Général sur la Protection des Données personnelles (RGPD), dont l'objectif est de renforcer la protection des citoyens. Le texte précise notamment les conditions dans lesquelles il est obligatoire d'obtenir le *consentement* de la personne concernée avant de collecter ses données personnelles, en particulier pour les données sensibles (origines raciales, orientations sexuelles, etc.) et les données relatives aux mineurs. Le RGPD introduit également une nouvelle obligation, l'obligation de *responsabilité*. Selon ce principe aussi appelé *accountability*, tout responsable de traitement qui traite des données personnelles doit pouvoir démontrer à tout moment qu'il respecte les obligations définies par le RGPD.

Les utilisateurs sont amenés à rechercher le bon **compromis** entre vie privée et services personnalisés. Notre panorama des technologies PETs vous aidera à faire ce choix.

Une étude de la Chaire Valeurs et Politiques des Informations Personnelles

Une enquête conduite en 2017 par la Chaire Valeurs et Politiques des Informations Personnelles avec Médiamétrie (voir page 187 et suivantes) a montré que certains utilisateurs-consommateurs adoptent des stratégies de protection, notamment en utilisant des outils logiciels qui les empêchent d'être tracés ou qui leur permettent de naviguer anonymement sur Internet. Cela suppose un investissement, notamment en terme de temps. Afin d'exercer une certaine maîtrise sur leur profil informationnel, ils optent selon leur objectif soit pour un service personnalisé, soit pour un service générique.

_Et si la technologie résolvait la difficile équation entre "services personnalisés" vs "vie privée" ?

Partant des constats issus de cette enquête, l'équipe de recherche de la Chaire a réalisé une étude scientifique portant sur les technologies protectrices de la vie privée. Cette étude, restituée dans ce document, recense les technologies les mieux à même de répondre au besoin de personnalisation de services, les détaille techniquement et les analyse comparativement.

Une étude qui propose une cartographie des technologies PETs, et qui ouvre de nouvelles pistes.

Quelques défis, et pistes à explorer

Si notre étude apporte quelques réponses pour choisir les PETs adaptées, des sujets de recherche restent ouverts :

- Améliorer les performances des PETs proposées afin de les adapter aux capacités réduites des mobiles et objets connectés,
- Utiliser les systèmes d'Intelligence Artificielle pour mieux intégrer le respect de la vie privée lors de la personnalisation de services,
- Rechercher le meilleur compromis économique entre respect de la vie privée, utilisation des données personnelles et expérience utilisateur,
- Déterminer les coûts supportés par les industriels en cas d'intégration des PETs dans leurs solutions, à la fois en terme de développement, de modèle d'affaires et d'ajustement de leur analyse d'impact sur la protection des données personnelles,
- Étudier les PETs sous l'angle d'un moyen de contourner ou de forcer l'application de la législation.

Ce qu'un panorama des PETs nous apprend

Ce document établit une taxonomie et un panorama des Technologies préservant la vie privée (*Privacy Enhancing Technologies*, PETs).

La protection de la vie privée et la personnalisation des services semblent deux objectifs antagonistes à première vue. Protéger la vie privée nécessite en effet de ne pas divulguer de données personnelles, données qui justement sont nécessaires pour personnaliser des services sur la base d'un profil utilisateur. Cependant, suite à des progrès dans les technologies traitant de cryptographie et de statistique, ces contradictions apparentes peuvent être résolues grâce à des approches PETs spécifiques.

Nous distinguons ***huit technologies préservant la vie privée***, qui peuvent être rattachées à trois groupes distincts (voir figure 1 plus loin) :

- les ***techniques orientées utilisateur***, qui impliquent que l'utilisateur gère lui-même la protection de son identité par l'installation de logiciels spécifiques, depuis la certification des attributs le décrivant jusqu'au contrôle de leur divulgation
- les ***techniques orientées serveur***, qui nécessitent des outils permettant d'anonymiser les données personnelles et d'effectuer des calculs sur les données chiffrées,
- les ***techniques orientées canal de communication***, qui mettent en exergue les caractéristiques de ce dernier, comme le chiffrement, ainsi que l'utilisation de serveurs intermédiaires.

Trois groupes de techniques pour préserver sa vie privée en ligne

- Le *premier groupe* de mécanismes s'adresse directement à l'utilisateur, pour l'aider à gérer la protection de son identité, via l'installation de logiciels spécifiques, à mieux contrôler la divulgation de ses données et en certifier la véracité.

- Le *deuxième groupe* réunit les solutions déployées sur les serveurs pour anonymiser les données personnelles, travailler sur des contenus chiffrés, et détruire les données de façon automatique.

- Le *troisième groupe* agit sur le canal de communication entre utilisateurs (et serveurs) : il assure que les communications sont sécurisées ou il déploie des approches de tiers de confiance.

La figure 1 et les pages suivantes affinent cette description.

Autres panoramas existants

La littérature scientifique a offert ces dernières années plusieurs panoramas de solutions techniques préservant la vie privée en ligne, particulièrement dans les contextes de *cloud computing*, de *fog computing*, de l'Internet des objets et plus généralement de tout type de technologies impliquant du calcul. D'autres panoramas abordent la question selon des domaines précis : les applications de e-santé, les *smart cities*, les systèmes *pervasifs*, les services de recommandations, la cyber-sécurité. Quelques articles seulement traitent de la question spécifique de la préservation de la vie privée dans le cadre des services personnalisés. La liste de ces articles est disponible en annexe.

En 2007, Alfred Kosba a mis en exergue la tension[6] opérant entre les besoins de préservation de la vie privée et les objectifs de personnalisation, considérant trois catégories de systèmes web personnalisés respectueux de la vie privée. Son article présente également des pratiques utiles en la matière, et discute des lois, ainsi que des régulations proposées par des acteurs privés, ce qui à l'époque avait suscité de nombreux débats.

En 2012, Toch *et al.* ont proposé un panorama[7] et une analyse détaillée des risques afférents à la vie privée. Les auteurs considèrent trois techniques de personnalisation, fondées sur les réseaux sociaux, les comportements et la localisation.

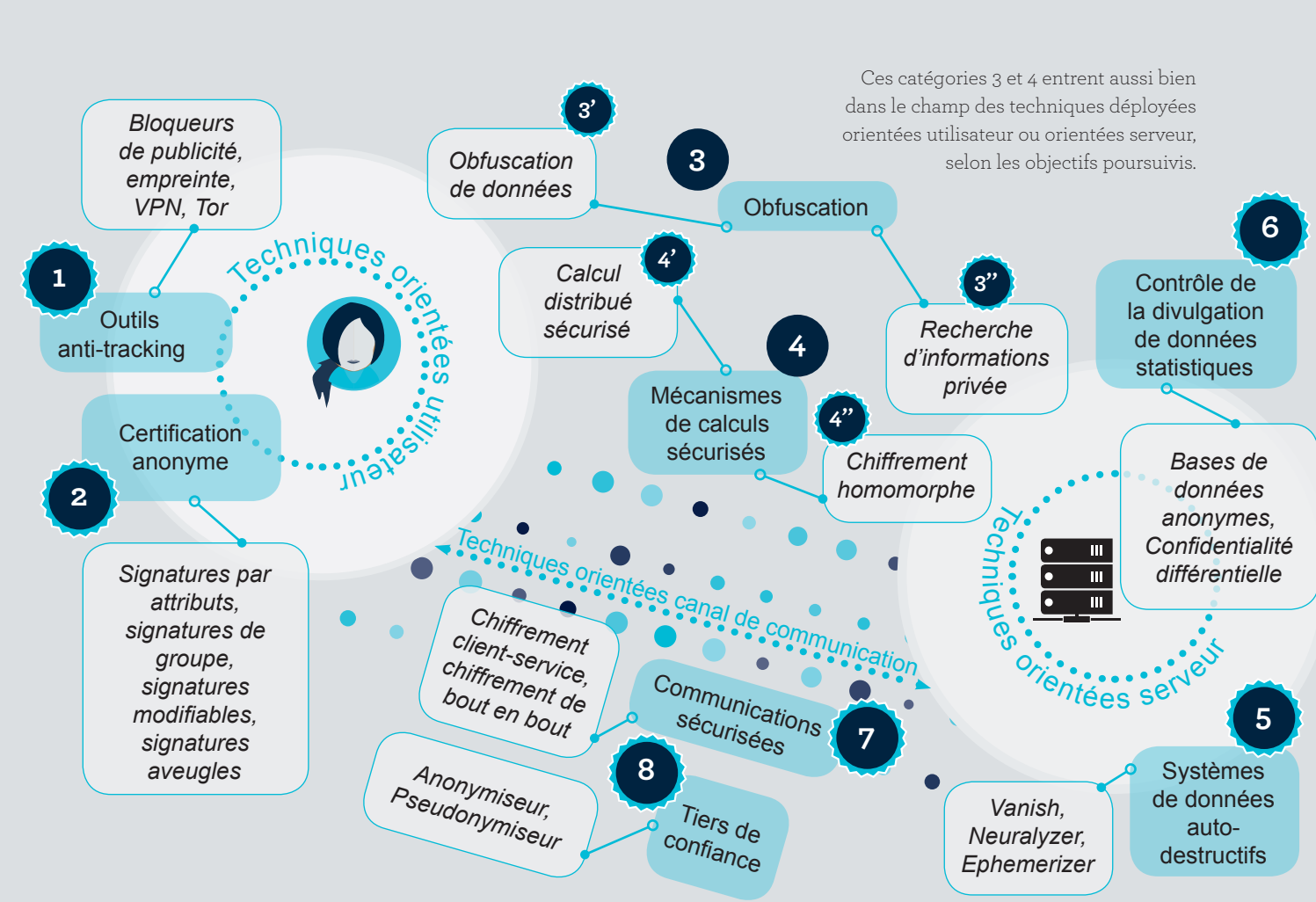


Fig. 1

une taxonomie des PETs

Toch *et al.* ont également fait une revue des solutions techniques permettant de préserver la vie privée dans le contexte des systèmes d'informations personnalisées, et ont proposé quelques directions de recherche.

En 2015, Parra-Arnau *et al.* ont présenté à leur tour une revue des technologies préservant la vie privée, ainsi que des métriques[8] associées. Les auteurs classent ces technologies en cinq groupes distincts :

- les outils anti-tracking,
- les techniques cryptographiques, avec un accent porté sur les recherches de contenus non intrusives,
- les systèmes de certification pseudonymes et anonymes,
- les approches privilégiant la coopération des utilisateurs,
- les techniques d'obfuscation de données.

Leur publication est l'occasion de lister des métriques concernant la vie privée, en particulier pour les techniques d'obfuscation de données. Les auteurs montrent ainsi qu'en général les mécanismes de préservation de la vie privée ont un impact sur l'utilité des données, et que savoir quantifier les paramètres de vie privée des utilisateurs mériterait d'être étudié plus avant.

Côté utilisateur, deux catégories principales, les outils anti-tracking (p. 47), et la certification anonyme (p. 61), ainsi que deux sous-catégories relevant de l'obfuscation (p. 99) et des mécanismes de calcul sécurisé (p. 103) sont à disposition des individus.

Côté serveur, l'obfuscation (p. 121) et les mécanismes de calcul sécurisé (p. 113) fournissent deux sous-catégories supplémentaires, complétées par les méthodes d'auto-destruction des données (p. 129), et de contrôle de la divulgation des données (p. 133).

Enfin, ***sur le canal de communication***, des communications sécurisées peuvent être mises en place (p. 153), et des intermédiaires de confiance peuvent créer de l'anonymat ou du pseudonymat (p. 161).

Applications de santé & protection de la vie privée

Les applications relatives au bien-être et au suivi des patients, dans le cadre d'un parcours de santé, de soins et de vie, sont particulièrement concernées par le large volume de données considérées comme sensibles, données qui sont collectées et stockées.

Il est indéniable que les dispositifs portatifs peuvent améliorer le quotidien des personnes, en facilitant la collecte de données relatives à leur santé (par exemple les bracelets FitBit / Vivosport qui enregistrent des informations de santé basiques). Ces données peuvent être utilisées pour repérer des maladies survenant conjointement et proposer de nouvelles options de traitement, sans oublier de surveiller à distance le bon déroulement des traitements. Cependant, ces bénéfices sont à mettre en balance avec les défis concernant la protection de la vie privée, l'identification précise des patients et la connaissance fine de leurs comportements.

De même, l'analyse des données massives provenant de larges cohortes de patients a permis de découvrir de nouveaux éléments pour étudier des pathologies et mieux comprendre le fonctionnement de certaines maladies. Si cela a été globalement utile pour les malades, cela a fourni également de nouvelles manières de détecter les soucis de santé de quelqu'un à son insu, via la corrélation de ses comportements avec les conditions, nouvellement découvertes, liées aux pathologies.

Considérations sur le respect de la vie privée

Dans un rapport[9] paru en 2015, l'Agence européenne chargée de la sécurité des réseaux et de l'information ordonne les risques sur la vie privée selon trois dimensions techniques :

1. les données traitées et collectées, considérées comme sensibles, personnelles ou facilitant l'identification,
2. les données qui peuvent être utilisées pour lever et / ou révoquer l'anonymat d'un utilisateur,
3. les *attaquants* possibles qui peuvent exploiter des données pendant leur transfert ou leur traitement, utiliser des connaissances externes ou contextuelles, ou agir de connivence avec d'autres acteurs.

La vie privée : une notion en mouvement

La notion de vie privée n'est pas conçue de façon identique selon les pays et les cultures, et elle n'a pas toujours recouvert les mêmes aspects. Nous avons vu (page 6) que le droit américain avait défini au départ le *Right to privacy* comme le droit à être laissé tranquille. Comment cette notion de tranquillité peut-elle perdurer dans un monde de plus en plus connecté, dans lequel les interactions avec les humains et les objets se multiplient sans cesse ?

Le ***droit au respect de la vie privée*** est consacré par l'article 12 de la Déclaration universelle des droits de l'Homme de 1948, l'article 8 de la Convention européenne des droits de l'Homme de 1950 et l'article 7 de la Charte des droits fondamentaux de l'Union européenne de 2000. De façon générale, il correspond à « *tout ce qui n'est pas la vie publique de l'individu* » pour reprendre les propos de Robert Badinter. Il s'agit du droit de ne pas révéler, ou voir révéler, des informations liées à son intimité (entendue comme la sphère physique mais aussi l'expression d'une relation avec autrui) et à son identité, afin de permettre à la personne de s'épanouir. Ce droit fondamental entend protéger l'opacité de l'individu, à la fois ce qui est lié à son ***intimité***, et ce qui relève du ***secret de ses communications***. C'est ce droit fondamental ainsi que le droit à la protection des données personnelles reconnu par l'article 8 de la Charte des droits fondamentaux de l'UE que le *Règlement* RGPD entend garantir en encadrant l'usage des données personnelles.

Il est possible de définir la vie privée par les exigences à respecter pour la préserver.

La notion de Data Protection by design

Data Protection by design[10], *security by design, privacy by design, data protection by default*… ces termes ont une histoire qu'il convient de connaître.

La démarche de *Privacy by design* est apparue à la fin des années 1990 sous l'impulsion d'Ann Cavoukian, la Commissaire à l'information et à la protection de la vie privée de l'Ontario (Canada). Partant du constat que le cadre légal n'intervient qu'a posteriori pour corriger les abus, la Commissaire proposait de traiter le problème à la source en intégrant le respect de la vie privée directement dans la conception et le fonctionnement des systèmes techniques, et ce, pendant toute la période d'utilisation des données personnelles jusqu'à leur obsolescence informationnelle. Le législateur européen intégrera cette notion dans le droit positif en 2016.

L'article 25-1 du RGPD introduit ainsi l'obligation de *Data protection by design*. Le responsable de traitements de données personnelles doit implémenter au cœur même de chaque système d'information des mesures à la fois techniques et organisationnelles pour mettre effectivement en œuvre les principes-clés définis par le règlement. Dans son article 25-2, le RGPD codifie ensuite le concept de protection des données par défaut (*Data protection by default*), avec comme objectif d'éviter une exploitation abusive par le responsable de traitement ou des tiers, ainsi qu'une réutilisation à d'autres fins que celles pour lesquelles les données ont été collectées.

Identité numérique, identification & authentification

L'identité numérique[11] est la « représentation informatique » d'une entité. Il s'agit de l'ensemble des moyens informatiques et technologiques qui permettent à cette entité de se projeter dans le monde du numérique, cette projection pouvant prendre plusieurs formes selon son contexte (administratif, professionnel, réseaux sociaux…).

Techniquement, cette représentation informatique est l'association d'une *entité* (personne physique ou morale, groupe de personnes… également désignée par souci de simplification : individu) à un ensemble de données numériques appelées *attributs*. Ces attributs peuvent être de natures très variées, statiques ou dynamiques : nom, prénom, adresse, centres d'intérêt, application utilisée et sa durée d'utilisation, géolocalisation…

L'action de ***s'identifier*** repose en général sur la simple déclaration par l'individu de son identité sous la forme d'un identifiant numérique de types numéro de sécurité sociale, nom, alias, pseudonyme, etc. ***S'authentifier*** suppose que l'individu déclare son identité ***et*** apporte la preuve de cette identité. Cette preuve, qui s'appelle en informatique un *crédentiel* (*credential*), peut provenir de l'individu lui-même ou bien d'un tiers. Elle peut prendre la forme d'un mot de passe, d'une *signature électronique* faisant en général référence à un *certificat électronique*, d'une empreinte biométrique… La distinction entre identification et authentification d'un individu réside dans le degré de *confiance* établi entre l'identité déclarée par l'individu et l'identité numérique dont il est détenteur.

À propos d'anonymat

Une solution technique pour protéger un individu consiste à couper le lien entre l'identité numérique et l'identité réelle de la personne. Deux propriétés répondent à cet objectif :

- le **pseudonymat** implique l'utilisation d'un identifiant, dit pseudonyme, qui doit pouvoir être aisément relié à une identité réelle par une entité légitime,
- l'**anonymat** implique que les identifiants utilisés doivent rendre très difficile – voire impossible et ce, de manière irréversible, pour les juristes – l'établissement de ce lien.

L'anonymisation des attributs permettant de couper le lien entre un ensemble d'attributs et les personnes concernées peut se faire par altération de la véracité des données décrites par ces attributs, ou par dilution dans des ensembles de valeurs plus grands.

Pour préserver la vie privée en ligne, un certain nombre d'exigences[12] ont été définies :

- ***l'anonymat de l'utilisateur***, qui signifie qu'il n'est pas identifiable parmi un ensemble de sujets (*anonymity set* en anglais). Plusieurs niveaux d'anonymat ont été définis dans la littérature, allant de l'*anonymat* complet (personne ne peut révéler l'identité de l'utilisateur) au *pseudo-anonymat* (l'identité n'est généralement pas connue mais peut être divulguée si nécessaire) et au *pseudonymat* (plusieurs identités virtuelles peuvent être créées et utilisées dans différents contextes).
- ***la minimisation des données***, composante fondamentale de la préservation de la vie privée, exigée par le RGPD. Selon ce principe, les fournisseurs de services doivent uniquement collecter et traiter les données strictement nécessaires pour atteindre les objectifs poursuivis par leur traitement, réduisant ainsi les risques de pistage et de profilage.
- ***l'inassociabilité***, un principe essentiel pour garantir l'anonymat, par lequel un attaquant se trouve incapable de relier deux ou plusieurs *élément d'intérêt* relatifs à un individu
  - l'inassociabilité dite «*issue-show*» garantit que toute information collectée afin de délivrer des crédentiels ne peut pas être utilisée ultérieurement pour lier le *jeton de présentation* au crédentiel
  - l'inassociabilité dite «*multi-show*» garantit que plusieurs jetons de présentation construits à partir du même crédentiel et transmis au cours de plusieurs sessions ne peuvent pas être reliés par le vérificateur

Exigences de protection de la vie privée

Un ensemble de propriétés est communément associé aux données afin qu'elles ne soient pas divulguées. Ces propriétés sont autant d'exigences permettant de juger des capacités des systèmes techniques PETs à atteindre leurs objectifs.

Il s'agit de la capacité à offrir et à conserver une *certaine forme* d'***anonymat de l'utilisateur***, le principe de ***minimisation des données*** collectées, et l'***inassociabilité***, l'impossibilité de relier des données, des transactions, des individus... entre eux.

Modèles de confiance et données personnelles

- *trusted model* : dans ce modèle, les utilisateurs confient leurs données personnelles à une entité externe, appelée *tiers de confiance*, qui se charge d'en assurer leur protection. Les anonymiseurs et les pseudonymiseurs sont les principaux exemples d'architectures suivant ce modèle. Si l'idée paraît simple, elle a pour principal inconvénient de nécessiter une infrastructure, et de supposer que les utilisateurs sont prêts à faire confiance à d'autres entités. En outre, de tels tiers peuvent éventuellement être amenés, par exemple par les autorités publiques, à révéler les informations sensibles collectées auxquelles ils ont accès.

- *untrusted model* : ici, les utilisateurs se méfient de tous les acteurs impliqués. Ne faisant confiance qu'à eux-mêmes, il leur incombe de protéger leur vie privée. Ils font appel par exemple à des mécanismes reposant sur l'obfuscation de données, opérés du côté de l'utilisateur. Dans ce modèle, la protection de la vie privée se fait au détriment de la fonctionnalité du système et de l'utilité des données.

- *semi-trusted model* : la confiance est distribuée sur plusieurs entités qui réalisent correctement les traitements attendus, excepté qu'elles peuvent tenter de collecter frauduleusement des données. Si ce modèle permet de se prémunir des entités curieuses isolées, cela n'empêche pas un sous-ensemble de ces entités de s'entendre secrètement et de compromettre la vie privée des autres utilisateurs.

Quels modèles de confiance ?

Les systèmes technologiques décrits dans ce document reposent sur des modèles de confiance, c'est-à-dire des descriptions des niveaux et formes de confiance entre entités.

Trois modèles de confiance sont ainsi considérés :

- le ***trusted model***, dans lequel des entités de confiance, externes, peuvent intervenir,
- le ***untrusted model***, inverse du précédent, décrivant un environnement où personne ne fait confiance aux autres entités,
- le ***semi-trusted model***, mixte, où la confiance ne repose pas sur une seule entité comme pour le *trusted model*, mais sur un ensemble d'entités.

Pour comprendre comment les techniques préservant la vie privée répondent à ce qu'on attend d'elles, il est nécessaire de définir auparavant quelques éléments de contexte : les types de services personnalisés, les données collectées, et les types d'attaquants et les menaces qu'ils font peser sur les utilisateurs.

Systèmes, applications et services personnalisés

_Services de recommandations

Cette classe de services permet aux utilisateurs de recevoir des recommandations personnalisées en fonction de leurs centres d'intérêt. Ceci est rendu possible grâce à l'analyse des données personnelles collectées au fur et à mesure de l'utilisation de ces services. Leur efficacité[13] est largement reconnue, puisque selon une étude parue en 2013, 35% des clients d'Amazon effectuaient un acte d'achat en suivant une recommandation suggérée, un chiffre qui montait à 75% chez Netflix.

Les services de recommandations s'appuient sur deux entités principales : l'utilisateur et le système de recommandation. Les données personnelles des utilisateurs sont en général stockées sur leur appareil (par exemple leur mobile), et collectées par le système de recommandation qui les analyse pour créer des ***profils***, des ***centres d'intérêt*** et des ***préférences***, éléments de base pour fournir des recommandations pertinentes. Ces dernières sont le plus souvent envoyées sous forme de messages, de notifications et de fenêtres pop-up.

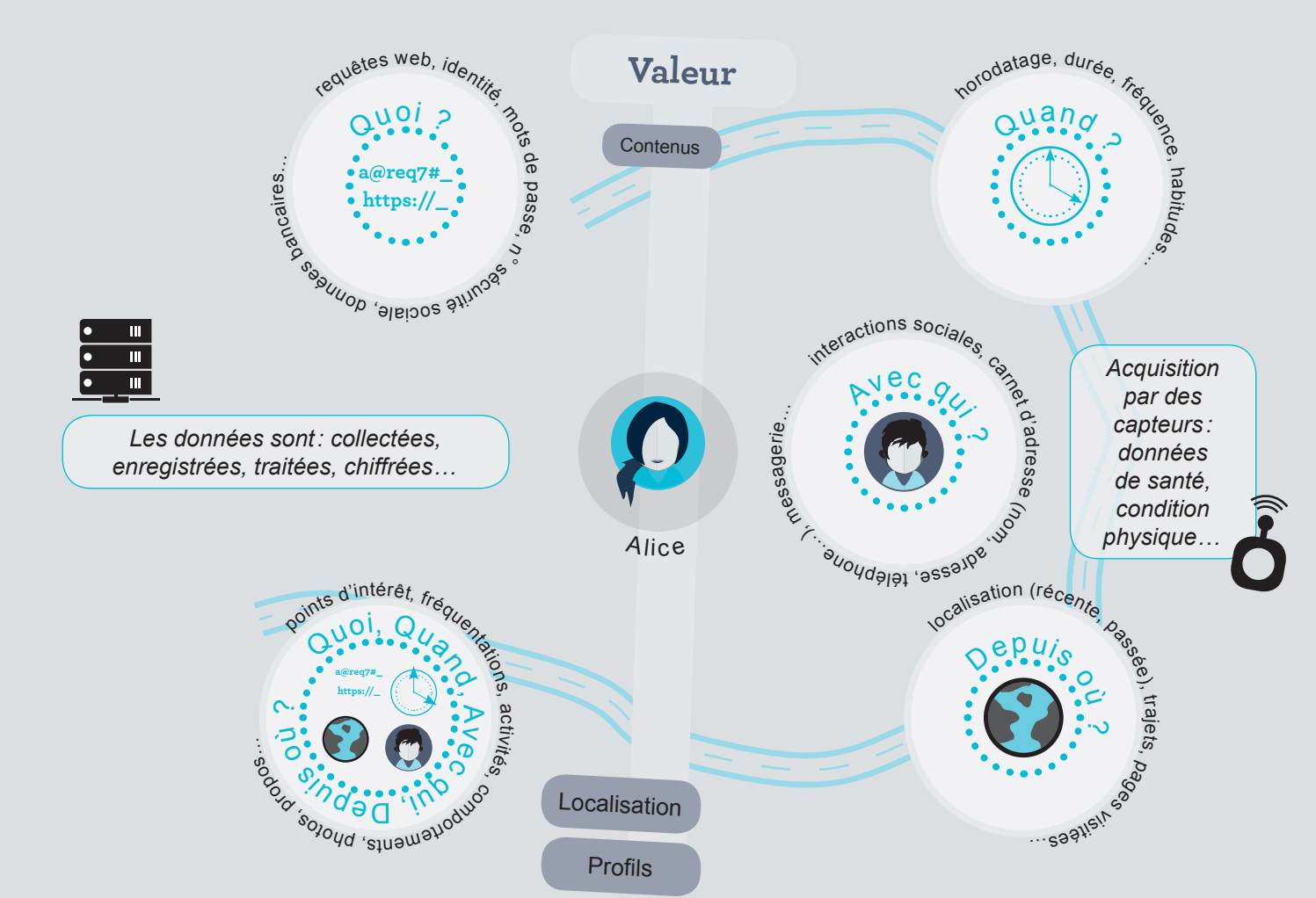


Fig. 2

quelles données sont collectées ?

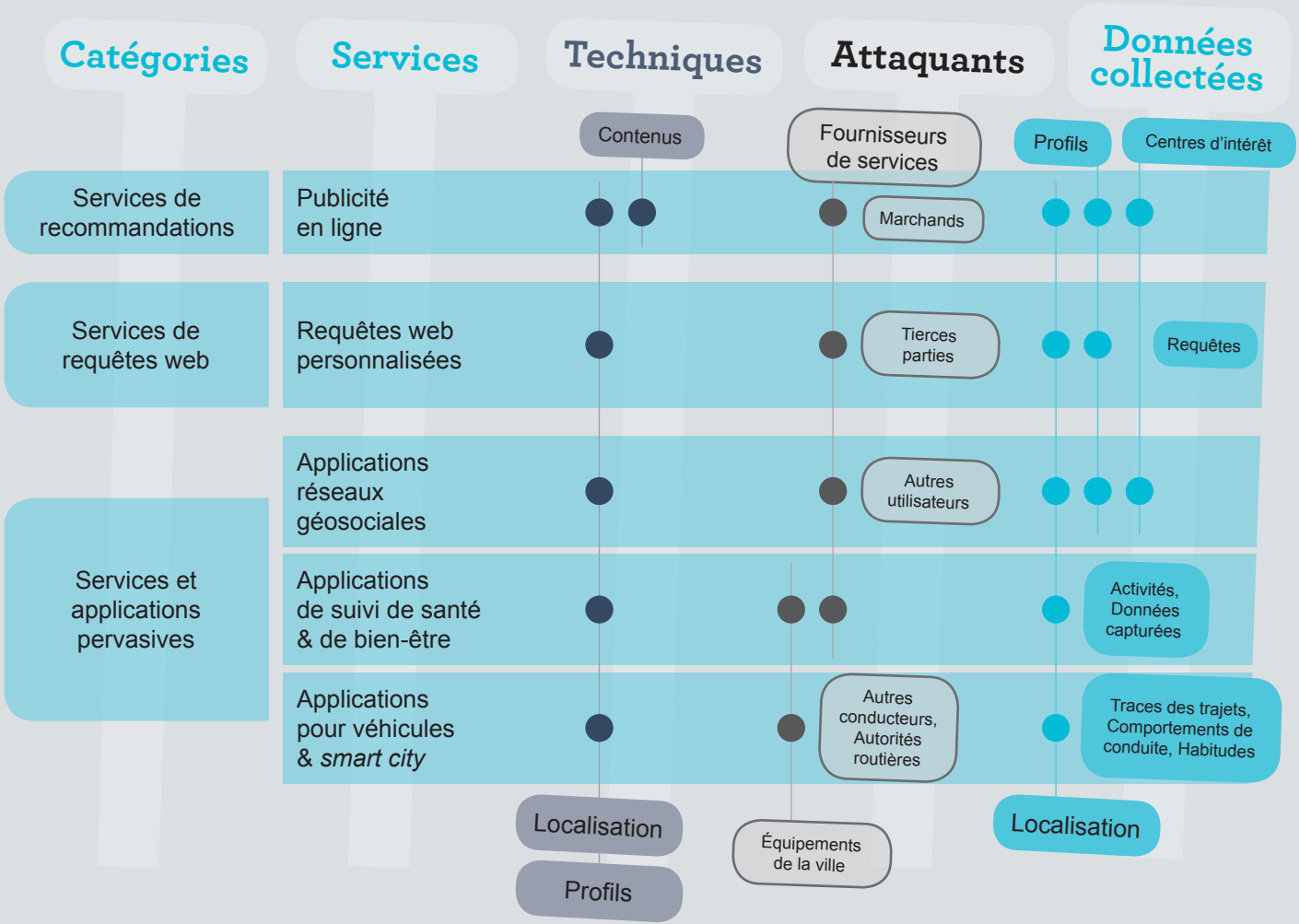
_Services de requêtes web

Les moteurs de recherche fonctionnant selon ce principe fournissent des résultats de requêtes web adaptés aux profils des requérants et à leurs centres d'intérêt tels qu'enregistrés lors d'utilisations précédentes, donnant l'impression de percevoir les intentions des utilisateurs et offrant ainsi une qualité d'expérience[14] accrue. Deux mécanismes principaux sont possibles : via l'historique des requêtes (modèle *click-log*) ou par l'identification du modèle des centres d'intérêt de l'utilisateur reposant sur des techniques de profilage.

_Services et applications pervasives

Sous ce vocable (*pervasive services*) est réunie une large classe de services allant des applications de réseaux géosociales aux applications véhiculaires, les services déployées dans les villes, et les applications de suivi de santé et de bien-être. Ces services sont personnalisés principalement en fonction de la localisation de l'utilisateur, soit via une prise en compte ponctuelle du lieu (mécanismes de *check-in*, usages de guides touristiques et suggestions de sortie, par exemple pour Foursquare et Facebook), soit en collectant en continu la localisation.

Avec « pervasif » et « géosociale », l'accent est mis sur le mode collaboratif, des objets communicants entre eux, dans le premier cas, et des usagers dans le second. Ces moments de collaboration, diffus, et enfouis dans les usages, peuvent finir par ne plus être remarqués, alors qu'ils nécessitent là encore l'échange de données personnelles.



Tab. 1

caractéristiques principales des services personnalisés

_Risques et menaces découlant de cette personnalisation

Concernant les ***services de recommandations***, les principales menaces découlent a) des données personnelles collectées transmises trop largement aux fournisseurs de recommandations qui b) peuvent les revendre sans consentement et c) peuvent être piratées, conduisant alors à la divulgation des ces données.

Les fuites de données collectées par les ***services de requêtes web*** ont pu miner la confiance des utilisateurs dans ces services (cf. affaire AOL[15] en 2006), et ont freiné l'enthousiasme de ces moteurs à fournir des services personnalisés, mettant ainsi en danger leur modèle économique. Autre rupture de confiance : quand ces fournisseurs de services revendent les données personnelles collectées, à l'insu des utilisateurs.

Enfin, dans le cadre des ***services et applications pervasives***, l'association d'un utilisateur avec des informations spatiales et temporelles spécifiques peut révéler ses affiliations, permet de retracer ses mouvements, et d'en déduire ses habitudes, ses centres d'intérêt, ses choix religieux ou ses problèmes de santé. Un fournisseur de services pourra ainsi a) connaître les lieux souvent fréquentés ou au contraire jamais visités, relier les personnes entre elles, ainsi que la fréquence et la durée de leurs interactions ; b) transmettre des données personnelles (obtenues grâce à des capteurs) à des tierces parties sans consentement, ce qui peut conduire à des discriminations (par ex. par les compagnies d'assurance[16]) ; c) collecter parfois plus de données que nécessaire, données qui deviennent des cibles de choix pour des attaquants.

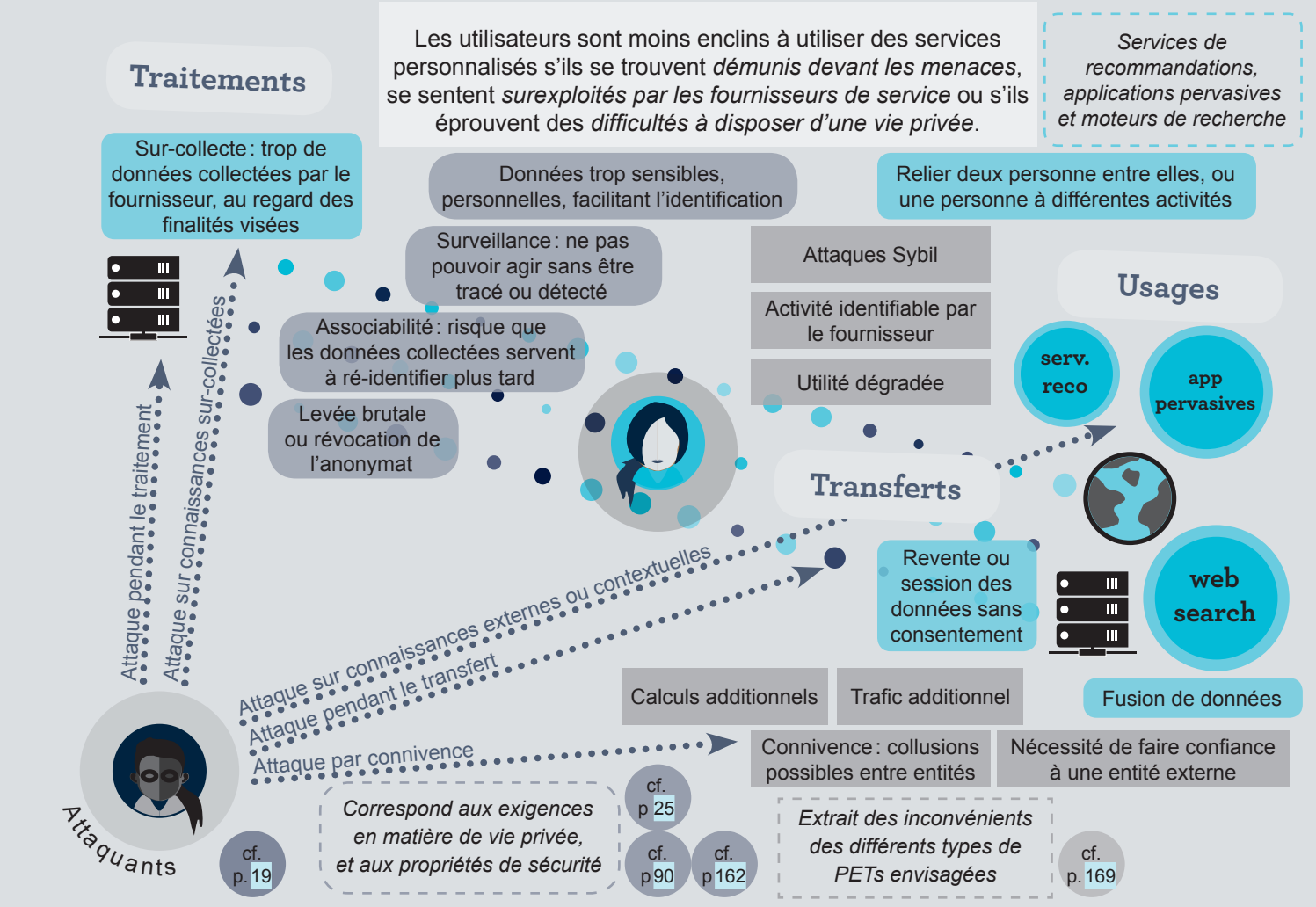


Fig. 3

risques et menaces sur les services personnalisés

Différentes approches pour personnaliser

Les techniques de personnalisation combinent des idées issues du profilage utilisateur, de la récupération d'informations, de l'Intelligence Artificielle et de la conception d'interfaces utilisateur, afin de fournir des services personnalisés aux utilisateurs finaux.

_Approches selon les profils utilisateur

En soumettant des requêtes à un moteur de recherche, en cliquant sur des liens d'articles de presse dans un contexte de système de recommandations d'actualités, en choisissant d'associer tel ou tel mot-clé à ses favoris ou ses photos, l'utilisateur donne des indices sur qui il est et ce qui l'intéresse, permettant ainsi aux fournisseurs de services d'élaborer un profil[17] sur des bases très fines. Ces profils sont généralement construits et modélisés sous formes d'histogrammes. Google News classe ainsi les actualités selon un ensemble prédéfini de thèmes, et les profils utilisateurs sont modélisés selon la distribution des clics parmi ces thèmes.

_Approches selon les informations spatio-temporelles

Deux niveaux d'approche sont à distinguer pour les services de personnalisation reposant sur la connaissance de la localisation[18] : a) les services de base qui rassemblent la navigation et la recherche de points d'intérêt ; b) les services dérivés qui rassemblent les services de pistage, le m-commerce, les jeux géolocalisés et les applications *solomo*.

Données personnelles : une réflexion complète que chacun doit mener…

Dans ce document, nous allons principalement discuter des technologies de préservation de la vie privée appliquées à certains types de données et d'échanges, mais nos usages et nos technologies évoluant plus rapidement encore, rien n'est faisable sans une attention personnelle sur nos propres comportements. En particulier, il ne suffit pas de protéger ses données personnelles, il faut apprendre à prendre soin de celles des autres. En le faisant, on adopte de bonnes pratiques qui nous protègent aussi.

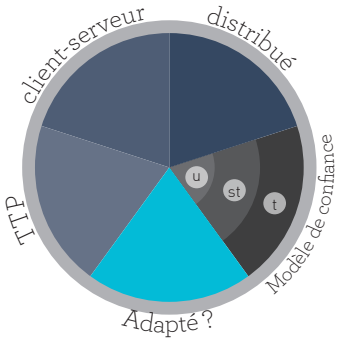
Prenons trois exemples. Les assistants vocaux[19] nous entourent et nous aident de plus en plus, qu'ils se trouvent dans des équipements particuliers à la maison, dans nos mobiles, et bientôt dans des objets du quotidien non spécifiques. Si vous les appréciez, qu'en est-il de votre entourage, vos enfants, vos hôtes ? Se poser la question de leur consentement à être écoutés est un début de réflexion utile sur ce que la voix possède de personnel. Dans la même veine, donner des échantillons de soi pour connaître son génome[20], c'est également donner accès aux informations d'autres personnes que soi. Enfin, des données perçues comme personnelles peuvent, au contact des autres utilisateurs, perdre cette qualité : la fonction proposée par la messagerie Telegram[21] début mars 2019 permettant d'effacer une conversation y compris sur le mobile de l'autre a ainsi entraîné de nombreux débats, compte-tenu des atteintes possibles aux données de l'autre.

Comprendre pour mieux choisir

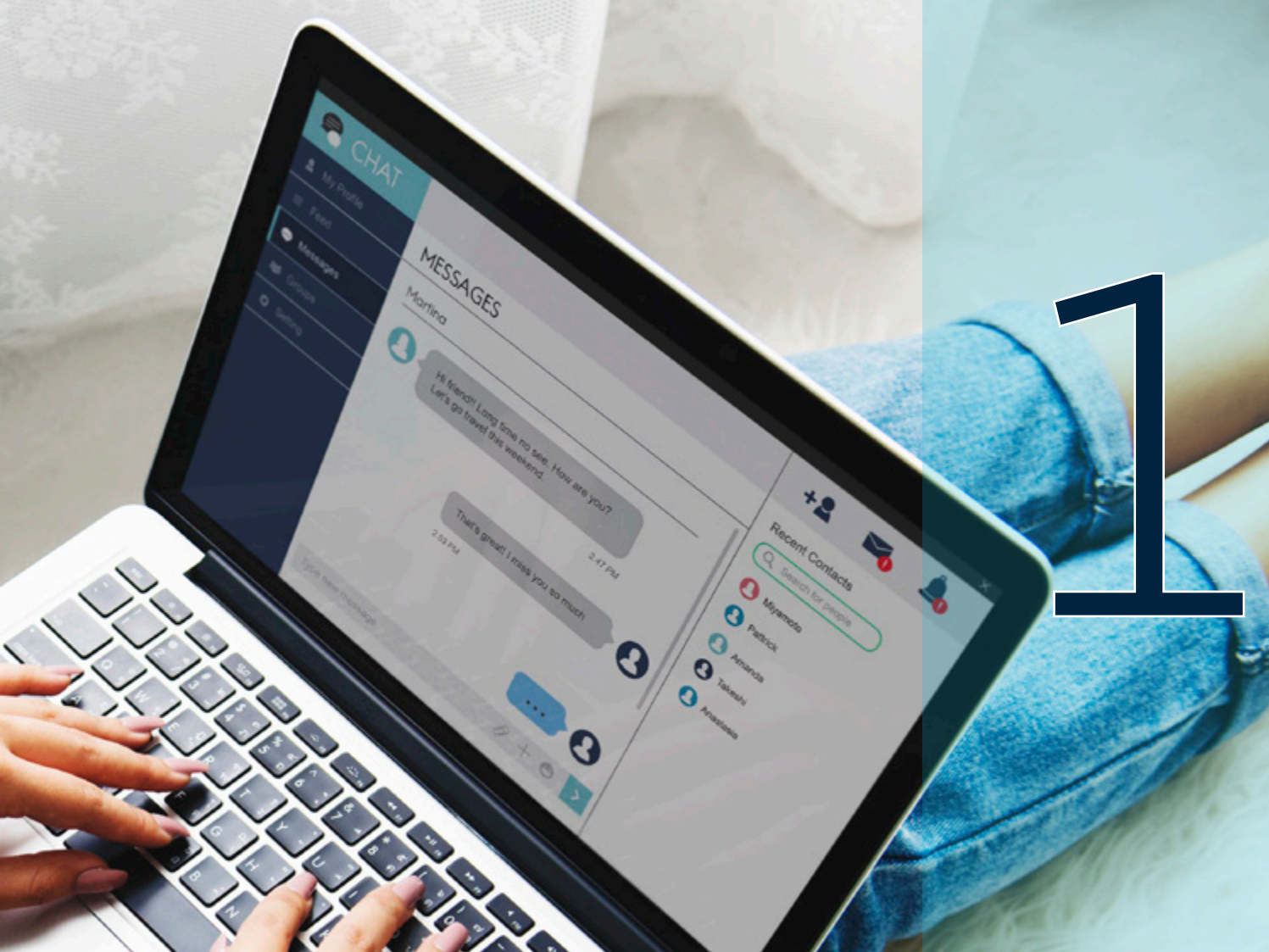
Par le titre de ce document, *« Personnalisation de services : quelles technologies pour la préservation de la vie privée ? »*, nous invitons les lecteurs à prendre connaissance des différents curseurs permettant de résoudre, selon les cas d'usage, le compromis entre l'utilisabilité de ses données pour bénéficier de services personnalisés, et la préservation de leur caractère privé.

Pour pouvoir faire ces choix, il faut comprendre le rôle et les intérêts des différentes entités impliquées. Il faut également garder en tête que les techniques sont amenées à évoluer, en fonction des modes de vie et des modèles de société.

Modèles de confiance sous-jacents et architectures techniques faisant les liens entre les entités impliquées sont deux dimensions importantes qui caractérisent les technologies PETs. Le radar ci-contre nous accompagnera tout du long pour commencer à se familiariser avec les différences entre PETs, avant un tableau comparatif complet page 169.



Les technologies effectivement adaptées pour permettre un compromis entre personnalisation et vie privée seront mises en exergue. Enfin, leurs avantages et inconvénients, ainsi que leurs domaines d'application les plus courants, régulièrement discutés, achèveront de brosser ce paysage.


CHAT
MESSAGES
Martina
Hi friend! Long time no see. How are you? Let's go travel this weekend.
2:47 PM
That's great! I miss you so much
2:53 PM
Recent Contacts
Search for people
Miyamoto
Patrick
Amanda
Takeshi
Anastasia

Les techniques orientées utilisateur

Ces techniques nécessitent l'implication active de l'utilisateur final pour protéger son identité et ses données personnelles.

Installer un outil *anti-tracking*, coopérer avec d'autres utilisateurs pour réaliser des calculs en toute intimité, adopter un système de certification anonyme ou obfusquer ses données sont les principales méthodes à disposition des individus soucieux de leur vie privée en ligne.

Commençons par les méthodes permettant de mieux gérer les cookies, éléments devenus très intrusifs au fil du temps.

Il existe des outils de pistage à état...

Il s'agit là principalement de techniques reposant sur des cookies stockant de l'information sur l'ordinateur de l'utilisateur. Dans ces fichiers très légers est ainsi conservée une information d'identification, qui sert au serveur à identifier le client pour lequel il a répertorié un certain nombre d'informations comme les dernières pages visitées…

Les cookies ont une date d'expiration au-delà de laquelle ils sont effacés du disque dur de l'utilisateur. Cette durée peut varier, et est en général assez longue. De plus, dès que l'utilisateur revient sur le site ayant initié un cookie, sa durée de vie est réinitialisée à sa valeur de départ, ce qui au bout du compte peut permettre à un site web de conserver beaucoup plus longtemps côté client l'état de sa navigation en son sein, qui aurait été perdue autrement.

On observe aujourd'hui de nombreux abus de cette qualité de stockage à état, et des avantages qu'offrent les cookies. Pour le comprendre, il faut se souvenir que les sites web sont construits à partir d'un grand nombre de ressources (textes, images…) qui chacune peut générer des cookies. Qui plus est, certaines de ces ressources ne sont pas nécessairement produites / possédées par le site web visité, et ces ressources *externes* sont également sources de cookies (voir figure ci-contre). Des parties tierces peuvent alors, par cet intermédiaire, collecter des informations d'usage, d'historique et d'habitudes de l'utilisateur.

Les cookies : un concept désormais familier

Si les *cookies* font aujourd'hui partie du paysage de la navigation sur le web, leur prolifération (aussi bien les *cookies directs* que les *cookies tiers*) et leur gestion au quotidien restent à maîtriser.

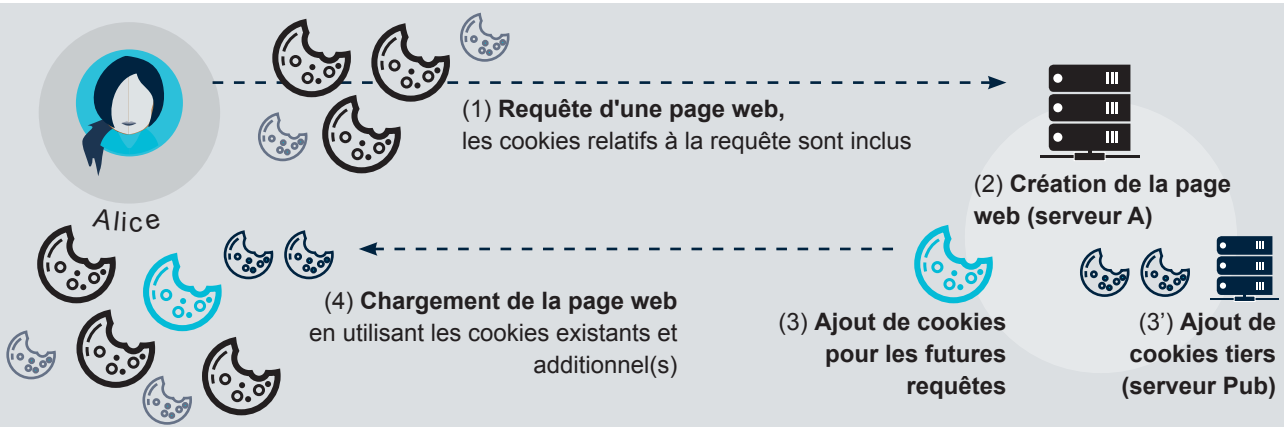


Fig. 4

cookies en cuisine

L'exemple classique est celui d'un site A qui inclut une ressource de type image provenant d'un site tiers Pub gérant des publicités. Lorsqu'un utilisateur visite A, un cookie-tiers provenant de Pub est stocké sur son ordinateur. Ce même utilisateur visite plus tard un site B qui s'avère faire appel également aux services de Pub, et notamment charger la même image. Pub, qui détecte la présence de son cookie antérieur, est alors en mesure de reconnaître l'utilisateur comme ayant déjà été vu, et d'en tirer des informations.

...et des outils de pistage sans état

Il est également possible de tracer des utilisateurs sans déposer de cookies sur leur ordinateur. Pour ce faire, le fournisseur de services extrait des informations des propriétés du navigateur web de l'utilisateur, et des mécanismes d'*empreinte numérique* sont exploités.

Ces techniques sont de plus en plus utilisées car elles permettent de passer outre les modes de *navigation privée* que les navigateurs proposent. Elles permettent également de contourner les réglementations européennes[22] en matière de cookies.

Outils de pistage et contre-mesures

Les outils de pistage sur le web sont des techniques qui permettent de mettre en œuvre la personnalisation des contenus, l'identification et le ciblage publicitaire ou de news. Leurs deux principaux objectifs sont d'améliorer l'expérience utilisateur et de maintenir une économie de l'Internet favorable aux publicitaires et éditeurs.

Quand ils deviennent trop intrusifs, trop présents, et que leurs actions ralentissent ou dégradent l'affichage des informations recherchées, les utilisateurs sont tentés d'installer des outils de contre-mesure pour reprendre la main et préserver leur vie privée.

S'additionnant aux systèmes de cookies-tiers, les techniques de suivi par empreinte numérique permettent aux entreprises publicitaires d'accumuler plus d'informations encore sur les utilisateurs. En partageant des empreintes collectées par des services différents, ces entreprises peuvent construire des profils utilisateurs pour les identifier plus facilement ou plus précisément.

Une autre donnée est à prendre en compte. Le langage interprété JavaScript étant aujourd'hui très répandu, et disposant de capacités étendues, il est devenu de plus en plus facile de construire (de calculer) des données de plus haut niveau (et donc porteuses de plus d'informations distinctives sur les utilisateurs) en accédant aux ressources rendues disponibles par le navigateur. Par exemple, la vitesse de mouvement de la souris lors d'une action de l'utilisateur, le temps moyen d'un double-clic, peuvent être observés et servir d'éléments, parmi d'autres, pour caractériser plus précisément tel ou tel utilisateur.

Les techniques à empreinte numérique ne reposent pas seulement sur le langage JavaScript pour obtenir des informations supplémentaires sur les navigateurs. Utiliser la balise `<canvas>`, les informations HTTP de type `User-Agent`, ou même la possibilité ou non de laisser des cookies sont autant de manières possibles d'établir une empreinte numérique d'un utilisateur.

Pister avec ou sans état

Deux catégories d'outils de pistage existent, selon le type de données qui sont demandées par le serveur web et stockées côté client. Les cookies, bien connus, font partie des technologies dites à état (*stateful*), l'état désignant ici la représentation en cours du profil de l'utilisateur, qui va être stockée sur sa propre machine. Cet état est ici l'essence même des échanges permettant le pistage. Mais il existe en informatique une autre manière de créer des échanges, sans faire référence à une succession d'états qui nécessiteraient d'être mémorisés (*stateless*). Exploiter les informations venant des requêtes des utilisateurs (adresse IP, type de navigateur…) relève de ces mécanismes.

Anti-tracking côté applications

Trois techniques utilisées par les outils de pistage peuvent être contrecarrées par les utilisateurs côté applications. Il s'agit des cookies en général, des codes JavaScript et Flash Script, et des techniques à base d'empreintes numériques.

_Cookies

Ces derniers peuvent être contrôlés directement dans le navigateur ou à l'aide d'outils de gestion adaptés. Le mode de navigation dit « privé » déconnecte la constitution de l'historique et réduit les traces localement. Dans certains navigateurs l'en-tête HTTP `Do Not Track` (DNT) peut être utilisé. L'insérer dans sa requête web signifie demander au serveur de ne pas tracer sa navigation, ou de ne pas le tracer de façon croisée quand il navigue sur d'autres sites (il s'agit là d'un point ambigu de la définition de cet en-tête). Cette technique n'est cependant pas standardisée, et rien ne dit que le serveur va appliquer honnêtement la directive DNT passée et cesser de pister. C'est pourquoi des techniques plus efficaces, disponibles dans des extensions des navigateurs, doivent être utilisées pour se protéger des différentes formes de cookies.

Privacy Badger, de l'Electronic Frontier Foundation (EFF), bloque les cookies tiers effectuant le pistage des utilisateurs à travers plusieurs sites, par exemple dès qu'un cookie tiers donné a été rencontré un certain (trop grand) nombre de fois.

Anti-tracking côté applications

Pour contrer le pistage, les utilisateurs peuvent agir pour commencer en paramétrant leurs applications et en y ajoutant des extensions particulières. Ces actions permettent de lutter contre les cookies, les codes JavaScript et FlashScript et les techniques d'empreintes numériques. Certaines méthodes, comme la demande `Do Not Track` associée à une requête, nécessitent le bon vouloir des fournisseurs de service. Les extensions reposent quant à elles sur des méthodes (gestions de listes noires, déclenchement de bannissement au-delà d'un seuil, refus de toute exécution de scripts…) qui ont leurs limites, et nécessitent des ajustements précis et éclairés.

AdBlock Plus et ***uBlock*** bloquent les publicités indésirables (sauf celles des annonceurs avec lesquels des accords ont été passés pour *AdBlock Plus*), ces publicités se présentant sous forme de popup, comme sur YouTube, dans Facebook, ou sous forme de bannières interactives. Les deux utilisent les *expressions régulières* pour réagir sur les URLs serveuses de publicités à bloquer. *AdBlock Plus* permet d'identifier également les publicités acceptables, et peut être configuré pour bloquer les domaines réputés diffuser des *malwares*. Tout comme pour les autres logiciels anti-tracking, AdBlock Plus et *uBlock* s'appuient sur des listes noires et des listes blanches et ils suppriment les *boutons médias sociaux* sur les pages.

Disconnect propose des extensions pour la navigation et l'utilisation des moteurs de recherche en mode privé, pour plusieurs navigateurs comme Firefox, Chrome, Opera… La fonction *Private Search* empêche les moteurs de pister les différentes requêtes, tandis que la fonction *Private Browsing* bloque les outils de pistage sur la base de listes noires. Seules les *requêtes http* nécessaires pour présenter des contenus utiles sont exécutées, chacune des requêtes étant classée dans une des quatre catégories suivantes : publicité, statistiques, sociale et contenus. Pour ce faire, *Disconnect* maintient une base de données des outils de pistage en parcourant lui-même le web à la recherche de requêtes tierces. Les utilisateurs ont par ailleurs toute latitude pour revalider un pisteur ou remettre en liste blanche un site.

Ghostery bloque les cookies pisteurs installés par défaut dans le navigateur. Reposant également sur des listes blanches et noire, il aide les utilisateurs à détecter et contrôler les balises JavaScript, bouts de code très répandus dans les sites web pour collecter des informations d'usages et d'habitudes via les cookies, voire même via des techniques sophistiquées telles que l'exploitation de la balise `<canvas>` comme nous l'avons vu plus tôt.

_JavaScript et Flash Script

Les techniques pour se protéger des scripts intrusifs sont rendues possibles via des extensions des navigateurs, comme ***NoScript*** ou ***ScriptSafe*** qui désactivent tous les types de scripts, s'ils n'ont pas été mis manuellement dans une liste blanche par l'utilisateur. ***Flashblock*** ne bloque que les contenus de type Flash.

_Empreintes numériques

Pour contrer le pistage sans état par empreinte numérique effectué côté serveur, il est possible de partager une même empreinte entre plusieurs utilisateurs pour se fondre dans la masse. Mais avant cela, il convient de noter que l'utilisation des contre-mesures présentées précédemment consiste en soi en une empreinte qui peut être utilisée côté serveur pour élaborer le profil de l'utilisateur.

Il est certain qu'utiliser un outil comme ***NoScript*** protège de toutes les techniques d'empreintes reposant sur JavaScript. La meilleure méthode consiste cependant à passer par ***Tor*** (voir page 54), où l'empreinte du navigateur est la même pour tous les utilisateurs et qui dispose de mécanismes visant à la préserver, comme le blocage des canvas.

Anti-tracking côté réseau

_Les réseaux privés virtuels

Appelés *Virtual Private Network* en anglais, leur acronyme VPN[23] est devenu courant en français. Il s'agit d'une technologie déployée sur le réseau pour créer un canal de communication sécurisé, c'est-à-dire chiffré, au travers d'un réseau d'accès public comme Internet, afin d'atteindre le réseau privé d'un fournisseur de service de VPN.

L'information transmise entre les deux sites, l'utilisateur d'une part et le réseau privé d'autre part, l'est à travers un tunnel chiffré et ne peut être lue par aucun tiers. L'ensemble repose sur plusieurs facteurs de sécurité.

Quatre solutions protocolaires principales utilisent les VPN : IPsec[24] (*Internet Protocol Security*), TLS (*Transport Layer Security*), PPTP (*Point-to-Point Tunelling Protocol*) et L2TP (*Layer Two Tunneling Protocol*).

En utilisant un VPN, l'utilisateur empêche son fournisseur d'accès Internet de voir ce qu'il fait et visualise, et empêche le fournisseur de services de savoir qui et où il est. Le VPN rend également possible d'outrepasser les restrictions d'accès géographiques imposées par certains services sur Internet. En revanche, le fournisseur de services VPN reste capable d'identifier l'utilisateur et d'en comprendre les comportements en ligne.

Anti-tracking côté réseau

Les serveurs web et les applications peuvent identifier la localisation des utilisateurs de manière approximative en se fondant par exemple sur leur adresse IP. La plupart d'entre elles sont en effet liées à des villes et des aires métropolitaines, et certaines à des emplacements encore plus spécifiques.

Alors que la majorité des outils de pistage sont déployés côté application, cacher son adresse IP de connexion au réseau est également nécessaire pour ces raisons. Deux techniques, les réseaux privés virtuels (VPN), et The Onion Router (Tor) sont utilisées à cet effet.

Construite sur la technologie VPN, ***Meddle***, un outil pour mobiles, a été conçu pour contrôler le trafic réseau en provenance des mobiles en le bloquant, le filtrant ou le modifiant. Avec cet outil, l'utilisateur peut observer toutes les connexions établies par son téléphone mobile, et également être prévenu de tout accès à ses données personnelles par des applications mobiles. Ceci permet à l'utilisateur de décider si telle ou telle information devrait être empêchée de circuler vers le réseau, ou bien si celle-ci devrait être modifiée.

CyberGhost est une plateforme VPN complète qui permet l'anonymat et le chiffrement des activités des utilisateurs tout en cachant leur adresse IP. Bénéficiant d'un déploiement important et stratégiquement distribué, cette solution est bien adaptée pour les PME, les indépendants et généralement les utilisateurs isolés désireux de sécuriser leurs transactions. *CyberGhost* dispose d'une interface conviviale et propose plusieurs fonctionnalités comme le camouflage d'adresse IP, le blocage des sites web mal intentionnés, le *kill switch* qui empêche toute information de fuiter en cas de coupure de VPN en bloquant l'accès à Internet…

OpenVPN est un logiciel libre sous licence GNU GPL qui permet à une entreprise ou des utilisateurs de mettre en œuvre un VPN qui repose sur le protocole TLS.

NordVPN est un fournisseur de service VPN personnel. Il propose une connectivité mondiale à très haut débit. Avec plus de 4800 serveurs répartis autour du globe, il offre le blocage des publicités et la plupart des services qu'on peut attendre d'un fournisseur VPN.

ExpressVPN a été présenté dans sa première mouture en 2009. Il s'agit d'un service de VPN mobile qui chiffre le trafic mobile des utilisateurs et cachent leur adresse IP. Fournissant une interface conviviale ainsi qu'un grand nombre de pays et de serveurs, c'est une solution très répandue. **ExpressVPN** propose un service de support abouti et un grand nombre de fonctionnalités de sécurité, comme par exemple la fonction *kill switch,* appelée *network lock* par ce fournisseur, afin de garantir que les activités des utilisateurs restent sûres et confidentielles.

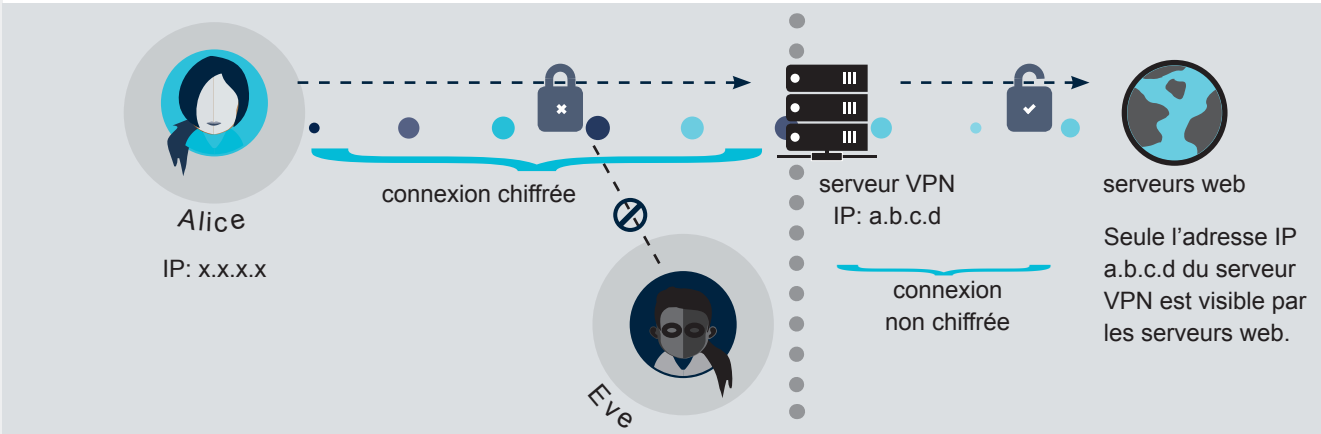


Fig. 5

_The Onion Router

Ce type de réseau est conçu pour anonymiser les communications sur Internet de manière à rendre ardue la possibilité d'établir une relation entre deux entités communicantes, comme par exemple un utilisateur et le serveur web qu'il visite. Pour mettre en place cette fonctionnalité, ces « réseaux d'anonymisation » reposent souvent sur des réseaux distribués en couches de chiffrement successives, et d'un routage dit « en oignon » dans lequel chaque routeur ne « pèle » qu'une couche de chiffrement pour connaître la destination suivante. Ces techniques sont utilisées dans le cadre d'applications construites sur TCP, comme la navigation sur le web.

Le réseau de ce style le plus connu est Tor, ***The Onion Router***, un réseau de tunnels virtuels conçu pour permettre la navigation web anonyme. Il se présente sous la forme d'un navigateur Tor.

Dans le routage en oignon, une chaîne de serveurs Tor est utilisée pour pousser les données d'un utilisateur vers leur destination, alors que le canal inter serveurs Tor est chiffré. Le premier serveur Tor de la chaîne doit sélectionner au moins trois nœuds Tor intermédiaires (routeurs oignon) vers la destination. Il chiffre le message plusieurs fois (chiffrement sur la base de clés symétriques), de telle façon que chaque nœud Tor doit enlever une couche de chiffrement pour que la destination puisse recevoir un message en clair. Chaque nœud intermédiaire ne connaît que l'adresse précédente et les adresses suivantes des nœuds impliqués dans le routage, ce qui permet de laisser dans le secret l'adresse de départ pour quiconque observe, sauf bien sûr lors du premier saut, et de masquer ce que l'utilisateur va voir, sauf lors du dernier saut.

L'anonymisation est l'aspect principal recherché par ce type de réseau. D'une part, le routage en oignon permet de ne pas exposer l'adresse IP. D'autre part, le navigateur Tor utilisé permet d'uniformiser d'un utilisateur à un autre tous les paramètres de navigation (résolution de l'écran…), ce qui contribue à masquer les activités de l'utilisateur.

À l'inverse, il convient de noter la différence fondamentale entre les VPN et les réseaux anonymisants. Si les VPN permettent bien de cacher l'adresse IP d'origine, ils ne fournissent pas l'anonymat puisque le fournisseur de service VPN reste capable d'identifier les utilisateurs de ses services. Il connaît en effet la véritable adresse IP du client avec lequel il a souvent un lien contractuel. Ils fournissent de fait une forme de pseudonymat du fait qu'un utilisateur de VPN est moins noyé dans la masse qu'un utilisateur Tor, et potentiellement (selon les VPNs), l'utilisateur peut garder la même adresse IP vis-à-vis des serveurs web, ce qui rend toujours possible le pistage. Pour Tor, l'adresse IP fournie aux serveurs est a contrario extrêmement changeante.

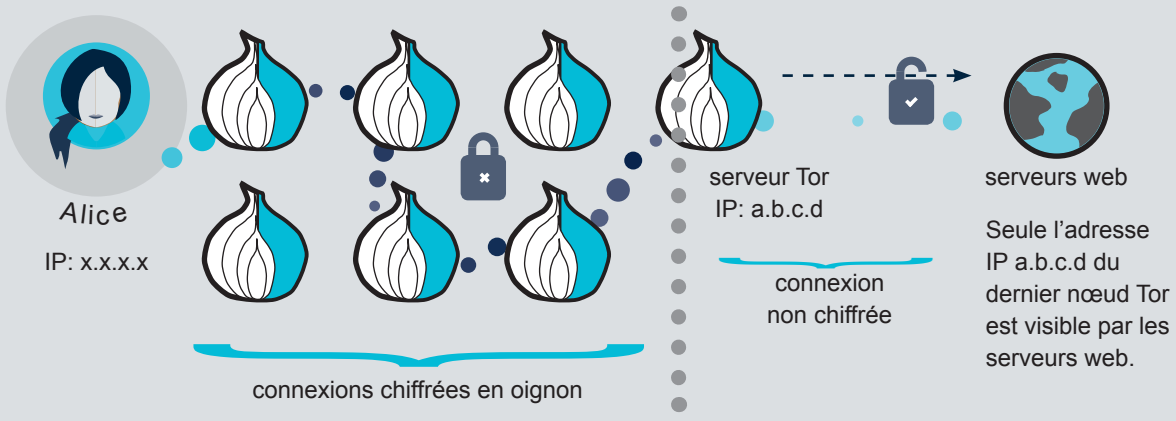


Fig. 6

Tor

Les bloqueurs de publicité : un comparatif

Deux critères permettent notamment de comparer les différents outils de blocage des publicités intempestives : leur *efficacité* et la *qualité d'expérience* offerte.

Dans une étude parue en 2017, Stefano Traverso ***et al.***[25] considèrent à la fois les exigences en matière de protection de la vie privée et de performance. Un certain nombre de plugins et de paramétrages sont utilisés, formant un banc de test pour naviguer sur des pages web tout en collectant des données de navigation. Chaque configuration contient un unique bloqueur de publicité, parmi ***AdBlock***, ***Privacy Badger***, ***Blur***, ***Disconnect***, ***uBlock*** et ***Ghostery***. Une configuration de référence (nommée *Plain*) est vierge de tout paramétrage anti-tracking. La campagne de mesures montre un panorama très diversifié, et aucun plugin ne se révèle capable de garantir une protection complète tout en améliorant les performances, comme promis.

En examinant les résultats de cette campagne de test, il est visible que la majorité des sites web considérés ne respectaient pas à cette époque la Directive européenne « *ePrivacy* ». Celle-ci interdit en effet aux sites web d'installer des cookies pour pister ou profiler les utilisateurs sans leur consentement explicite a priori.

Choisir ses outils de protection

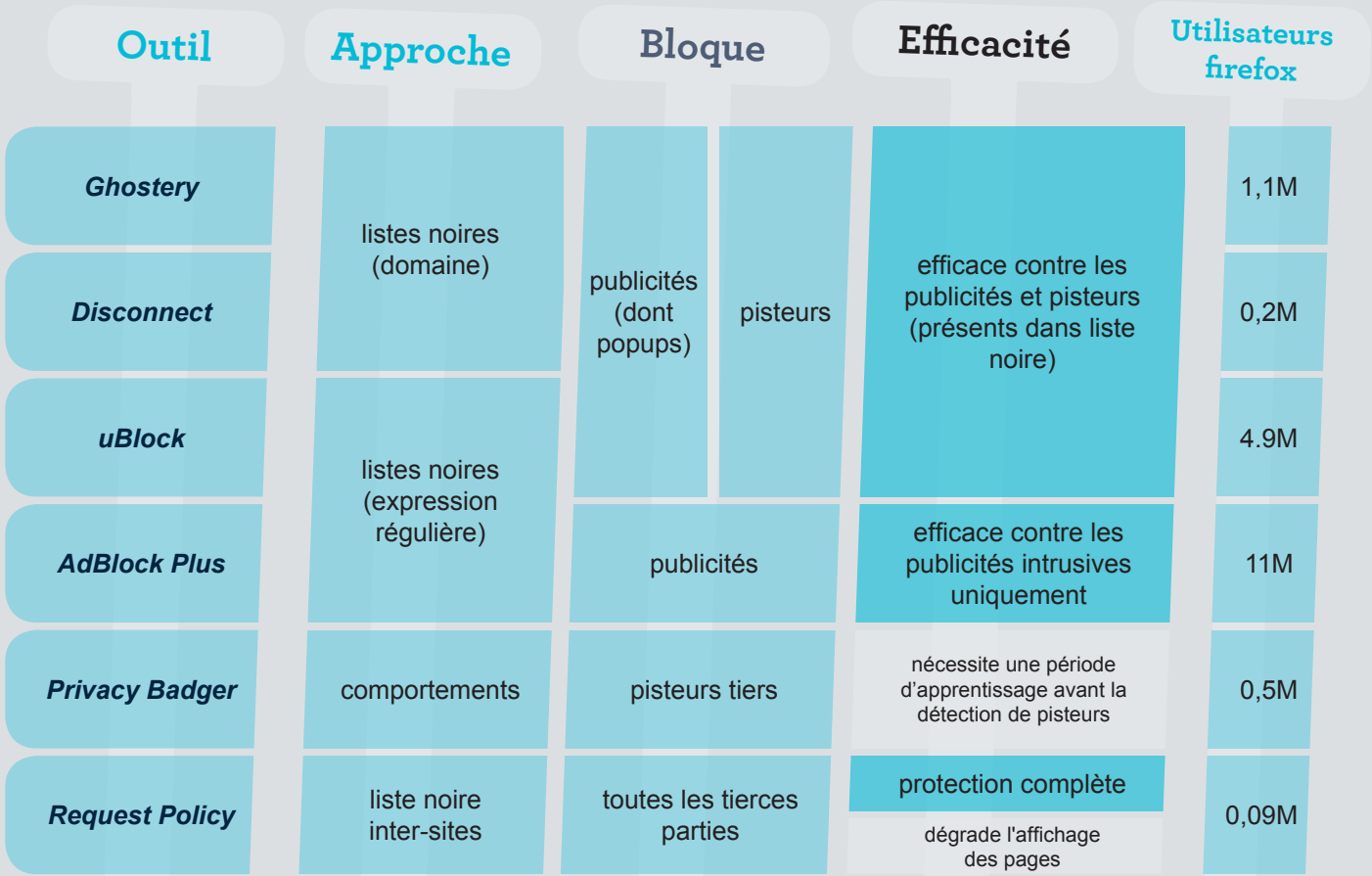
Les bloqueurs de publicité sont légion, et les procédés mis en place par les sites pour les contourner sont également imaginatifs. Pour faire un choix adapté à ses usages, et à ses souhaits en terme de niveau de protection, d'utilisabilité et de confort de navigation sur le web, il convient de comprendre les différentes manières de détecter et bloquer les pisteurs et les publicités qui en découlent, de savoir quels sont les éléments effectivement bloqués, et de jauger l'efficacité des outils de protection retenus.

Plusieurs comparatifs dans ce document sont ainsi proposés pour aider chacun à mieux gérer sa vie privée en ligne.

Notre tableau 2, élaboré à partir de cette étude 2017 de Stefano Traverso et.al, permet de comparer les six bloqueurs de publicité pré-cités selon des critères tant qualitatifs que quantitatifs :

- ***Approche :*** la manière utilisée pour détecter et bloquer les pisteurs
  - par leur nom de domaine d'origine
  - par expressions régulières
  - en se fondant sur leur comportement
  - en mettant sur liste noire les pisteurs vus sur plusieurs sites
- ***Bloque :*** ce qui est bloqué par chaque outil
  - pisteurs, publicités, cookies tiers…
- ***Efficacité :*** les avantages et inconvénients de chaque solution, en regard de leur capacité à préserver la vie privée et de leur *efficience*

Request Policy est remarquablement *efficiente*, en ce sens qu'elle bloque tous les contenus tiers, quels qu'ils soient. Ceci est au prix de l'*efficacité* globale, qui doit tenir compte de l'ensemble des objectifs, l'un d'entre eux étant la capacité à présenter du contenu web correctement. Ce n'est pas le cas ici, le rendu de la page web étant dégradé. ***Privacy Badger*** bloque plus généralement les sites qui ne respectent pas la demande `Do Not Track`. ***AdBlock Plus***, le plus utilisé est pourtant moins efficace que ***uBlock*** et ***Disconnect*** qui offrent une bonne protection contre les publicités et pisteurs connus des listes noires.



Tab. 2

Panorama des principaux outils anti-tracking

❗ Par souci de concision, nous désignons *AC*, ou *systèmes AC*, les systèmes reposant sur les preuves d'attributs anonymes (*Anonymous Credentials* ou *Anonymous Certification*). Deux composantes de ces systèmes, les *entités* impliquées et les *procédures*, doivent être bien comprises. Voir la figure «diagramme d'échanges entre entités AC», page 62.

Entités impliquées dans un système AC

Les systèmes AC[26] reposent sur des entités bien définies. Trois d'entre elles sont obligatoires : l'émetteur (*issuer*), le vérificateur et l'utilisateur (désigné également en sa qualité de signataire). Deux sont optionnelles : l'autorité de révocation et l'inspecteur.

Dans un système AC, un utilisateur est un entité décisive (*pivotal entity*) qui souhaite, tout en préservant sa vie privée, accéder à des services proposés par des fournisseurs de services qui vont jouer un rôle de vérificateur d'attributs. Chaque vérificateur impose une *politique de contrôle d'accès* à ses ressources et services, appelée *politique de présentation*, qui oblige les utilisateurs à se présenter avec un ensemble de preuves d'attributs (*set of credential*). L'utilisateur les obtient d'une autorité de confiance, l'émetteur. Il sélectionne un sous-ensemble de ces attributs certifiés, qu'il personnalise avant de les montrer au fournisseur de services, sous la forme d'un *jeton de présentation*. L'autorité de révocation est quant à elle responsable du maintien d'une liste de preuves d'attributs valides, et de leur révocation. Une fois révoqué, un crédentiel ne peut plus être utilisé pour construire un jeton de présentation. Enfin, l'inspecteur est une entité d'audit de confiance ayant aptitude à lever l'anonymat d'un utilisateur ou révoquer ses attributs, si nécessaire.

2

Certification anonyme

Deuxième série de techniques orientées utilisateur, les technologies de certification préservant la vie privée sont issues de concepts élaborés en 1982[27], également connus sous le nom de *Anonymous Credentials* ou *Anonymous Certification*. Concepts assez anciens, ils ont été complètement formalisés en 2001[28] par Camenisch & Lysyanskaya. Différentes implémentations ont été proposées depuis, considérées comme des composantes essentielles de tout système de gestion d'identités[29] qui préserve la vie privée. Un utilisateur *honnête* est ainsi capable de prouver à un fournisseur de services le lui demandant qu'il possède certains attributs (être majeur, posséder son permis…).

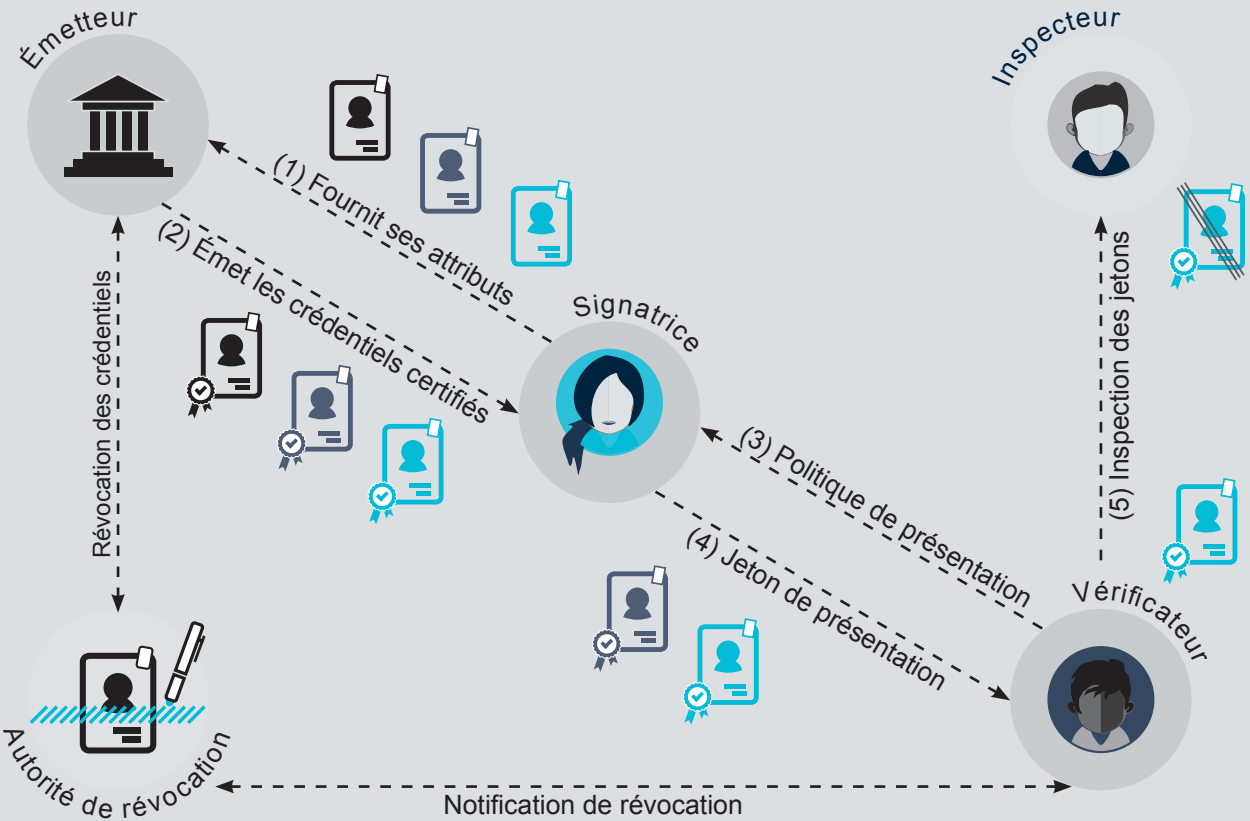


diagramme d'échanges entre entités AC

Fig. 7

Les attributs certifiés sont fournis par des *autorités* elles-même dignes de confiance. Ces techniques AC empêchent les fournisseurs de services de pister – ils le feraient en suivant les sessions successives d'échanges – les activités des utilisateurs. Ceux-ci émettent en effet une preuve d'attributs à chaque requête de telle manière qu'il n'est pas possible de les relier entre elles, ou à aucune autre de leurs données personnelles, même en cas de connivence entre fournisseurs de services (vérificateurs) et autorité de confiance (émetteurs). Fortement compatibles avec l'exigence de minimisation des données, un principe crucial des législations actuelles sur la protection des données (RGPD en Europe, *National Strategy for Trusted Identities in Cyberspace* aux États-Unis), ces techniques sont d'intérêt pour les fournisseurs de services qui veulent s'assurer de la qualité des données fournies par l'utilisateur.

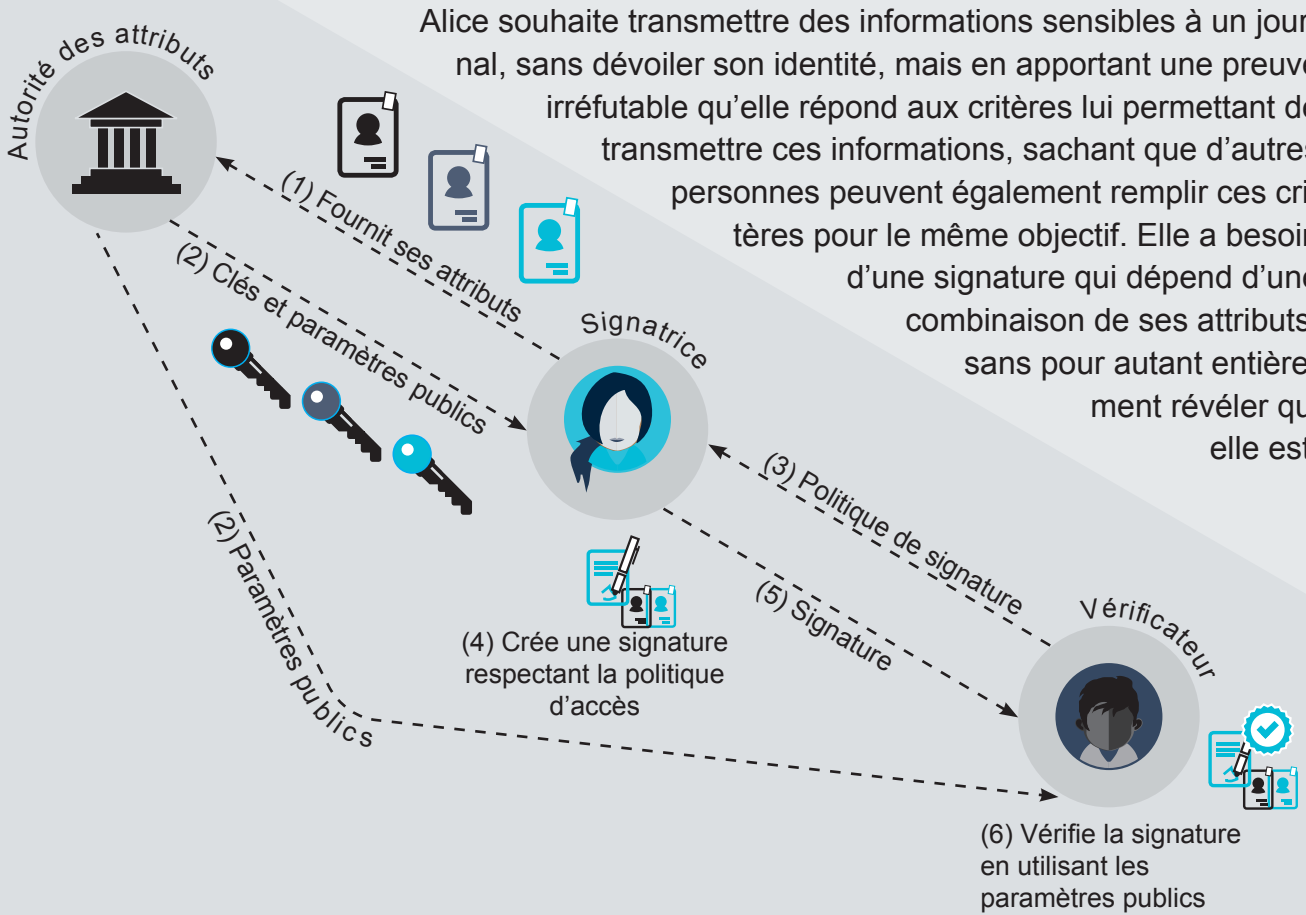
Procédures

Un système AC repose par ailleurs sur trois ***procédures*** principales et six $\mathrm{Algorithmes}$:

- $\textsc{Setup}$: algorithme qui, étant donné un niveau de sécurité souhaité ξ, crée les paramètres publics params et les paires de clés publiques-privées de l'émetteur $(\mathrm{pk}_{\mathrm{is}}, \mathrm{sk}_{\mathrm{is}})$.
- $\textsc{UserKG}$: algorithme de génération de la paire de clés $(\mathrm{pk}_{\mathrm{j}}, \mathrm{sk}_{\mathrm{j}})$ de l'utilisateur j. $\textsc{Setup}$ et $\textsc{UserKG}$ forment la ***procédure d'initialisation***.
- $\textsc{Issue} \leftrightarrow \textsc{Obtain}$: articulant deux algorithmes, ce protocole correspond à la ***procédure d'émission*** effectuée entre l'émetteur et l'utilisateur. L'algorithme $\textsc{Issue}$, exécuté par l'autorité émettrice, prend en entrée sa clé secrète $\mathrm{sk}_{\mathrm{is}}$, les params, la clé publique de l'utilisateur pk_{u}, et l'ensemble des attributs $\{a_{\mathrm{i}}\}_{\mathrm{i}=1..\mathrm{N}}$. Pour calculer son crédentiel C, l'utilisateur exécute $\textsc{Obtain}$, qui prend en entrée sa propre clé secrète sk_{u} et la clé publique de l'autorité émettrice $\mathrm{pk}_{\mathrm{is}}$.
- $\textsc{Show} \leftrightarrow \textsc{Verify}$: il s'agit de la ***procédure de présentation***, effectuée entre l'utilisateur et le vérificateur. L'algorithme $\textsc{Show}$ prend en entrée la clé publique de l'autorité émettrice $\mathrm{pk}_{\mathrm{is}}$, la clé secrète sk_{u} de l'utilisateur, l'ensemble des attributs $\{a_{\mathrm{i}}\}_{\mathrm{i}=1..\mathrm{N}}$, et le crédentiel C et produit, en tenant compte de la *politique de présentation*, un *jeton de présentation*. L'entité vérificatrice exécute l'algorithme $\textsc{Verify}$, prenant en entrée la clé publique de l'autorité émettrice $\mathrm{pk}_{\mathrm{is}}$, l'ensemble des attributs $\{a_{\mathrm{i}}\}_{\mathrm{i}=1..\mathrm{N}}$, et le jeton de présentation, qui produit un bit $\mathrm{b} \in \{0,1\}$ pour indiquer le succès ou l'échec de la vérification.

Deux solutions principales ont été proposées par les industriels du logiciel. IBM-Idemix[30], qui repose sur la signature Camenisch-Lysyanskaya, et qui est considérée comme une alternative pratique aux signatures de groupe (voir ce concept plus loin) ; Microsoft U-Prove[31], qui est construit sur une variante[32] des signatures aveugles.

Mais avant de décrire ces deux solutions (page 91), nous allons voir quelques-unes des *primitives cryptographiques* sur lesquelles les systèmes AC reposent, et leurs liens : les signatures de groupe (*group signatures*), les signatures modifiables (*sanitized signatures*), les signatures aveugles (*blind signatures*) et, pour commencer, les signatures reposant sur les attributs (*attribute based signatures*).



les signatures reposant sur les attributs

Fig. 8

Signatures reposant sur les attributs

Cette signature[33] est une primitive cryptographique polyvalente permettant à un utilisateur de signer électroniquement un message avec un contrôle fin sur les attributs certifiés révélés. En effet, chaque utilisateur, détenteur d'un ensemble d'attributs, doit obtenir une clé secrète pour chacun de ses attributs auprès d'une autorité émettrice de confiance.

Il est ainsi capable de signer un message prouvant qu'il satisfait un prédicat à l'aide de n'importe quel sous-ensemble de ses attributs.

Diverses implémentations techniques des signatures reposant sur les attributs

La littérature scientifique sur les systèmes cryptographiques fait le compte de plusieurs types d'implémentation des signatures reposant sur les attributs (*Attributed-based signature*, $\mathcal{ABS}$) :

- les attributs peuvent prendre des valeurs binaires, sur un bit, ou bien être une structure de données particulière,
- les *arbres d'accès* peuvent répondre à des besoins de politiques à seuil, monotones ou non-monotones,
- l'émission des clés secrètes des attributs peut être effectuée par une autorité unique, ou par un groupe d'autorités.

Les arbres d'accès (*access structure*) sont un concept issu de la théorie des ensembles, repris en cryptographie pour décrire les conditions à satisfaire pour qu'une opération cryptographique soit autorisée (par exemple une signature à base d'attributs prouve que le signataire dispose d'un ensemble d'attributs satisfaisant l'arbre d'accès). Ici l'*ensemble* considéré est celui des attributs, une *politique* désigne l'ensemble des règles qui définissent la sécurité d'accès à la ressource, une *politique à seuil* signifie qu'un nombre minimum d'attributs est nécessaire pour accéder à la ressource, et *monotone* / *non-monotone* désignent des relations d'ensemble dont les détails ne sont pas nécessaires ici.[34][35][36][37][38][39]

Quelques termes courants en cryptographie

Les signatures reposant sur les attributs, et les signatures en général, sont ce qu'on appelle une *primitive* cryptographique. Une telle primitive est un algorithme « de bas niveau » que l'on peut considérer comme une brique de base d'un système cryptographique. Cette brique est conçue pour effectuer une tâche précise, localement à une entité, de la manière la plus fiable possible. Concevoir un système cryptographique consiste notamment à combiner des primitives cryptographiques en un *protocole*. Intermédiaires entre les primitives et les protocoles, les *procédures* désignent des étapes principales effectuées pour appliquer le protocole, comme par exemple un échange d'informations entre deux entités.

Le terme *preuve* peut désigner des choses différentes. Il peut s'agit de *preuves* qu'un algorithme fait bien ce qu'il est censé faire. Il peut notamment s'agir de *preuves de sécurité*, démonstrations qui prouvent la sécurité d'un protocole. Il peut s'agir d'une *primitive* – par exemple la preuve à divulgation nulle de connaissance – utilisée par une entité pour apporter à une autre entité une preuve mathématique qu'il connaît un élément secret sans qu'il ait besoin de le divulguer. Enfin, ce terme peut désigner plus prosaïquement une *preuve d'identité*, ou crédentiel.

Enfin, les attributs sont combinables en utilisant les mathématiques de la *logique des prédicats*, *prédicat* désignant ici une propriété des objets du domaine considéré.

Modèle de sécurité des signatures reposant sur les attributs

Le premier modèle de sécurité des systèmes de type ABS a été proposé par Shahandashti *et al.*[40] en 2009. Les auteurs y présentaient les procédures principales, ainsi que les propriétés de sécurité telles que :

- ***la complétude (correctness),*** signifiant qu'un utilisateur honnête doit toujours pouvoir réussir à prouver la validité des preuves au vérificateur de manière anonyme,
- ***la non-forgeabilité (unforgeability),*** signifiant qu'un utilisateur ne disposant pas d'un crédentiel légitime ne doit pas être capable de générer un jeton de présentation valide,
- ***la confidentialité des attributs utilisés par le signataire (signer-attribute privacy),*** signifiant qu'il n'est pas possible de déduire de la signature fournie des éléments spécifiques sur les attributs du signataire.

Noter qu'il existe plusieurs niveaux de *signer-attribute privacy* et que le plus élevé garantit que même plusieurs signatures émises par un même signataire ne fournissent pas d'informations supplémentaires.

Par la suite, Maji *et al.*[41] en 2010, et El Kaafarani *et al.*[42] en 2014 introduisirent la propriété *perfect privacy*, qui stipule qu'une signature ne doit révéler ni l'identité de l'utilisateur signataire, ni l'ensemble des attributs utilisés pendant la procédure de signature.

Habituellement, un système construit sur les signatures $\mathcal{ABS}$ implique plusieurs entités : l'administrateur (*Signature Trustee*, ST), l'autorité des attributs (*Attribute Authority*, AA) et plusieurs utilisateurs et vérificateurs. ST est une entité globale générant les paramètres publics du système, tandis que AA émet les clés associées aux attributs que les utilisateurs possèdent.

Même avec la connaissance de ces attributs et des clés, il est impossible à une autorité AA trop curieuse d'identifier quels attributs ont été utilisés pour une signature valide donnée. Il lui est donc impossible d'associer une signature à un utilisateur donné, ou de détecter que plusieurs signatures proviennent d'un même utilisateur.

Cas d'usage en mode AC

Un médecin (l'utilisateur) obtient une preuve d'identité (par exemple sa carte professionnelle) du Ministère de la Santé (jouant ici le rôle d'entité émettrice), à partir d'un ensemble d'attributs :

$$\mathcal{S} = \{a_1 := \text{nom}; a_2 := \text{Martin}; a_3 := \text{ville}; a_4 := \text{Paris}; a_5 := \text{docteur}; a_6 := \text{chirurgie carcinologique}\}.$$

L'ensemble des attributs est associé à l'utilisateur propriétaire grâce à sa clé publique et est signé à l'aide de la clé secrète de l'émetteur, créant ainsi le crédentiel résultant, noté C.

Par la suite, le médecin peut prouver qu'il réside à Paris et qu'il est médecin, sans dévoiler ni son nom ni sa spécialité. Pour cela, le prédicat de signature suivant est défini :

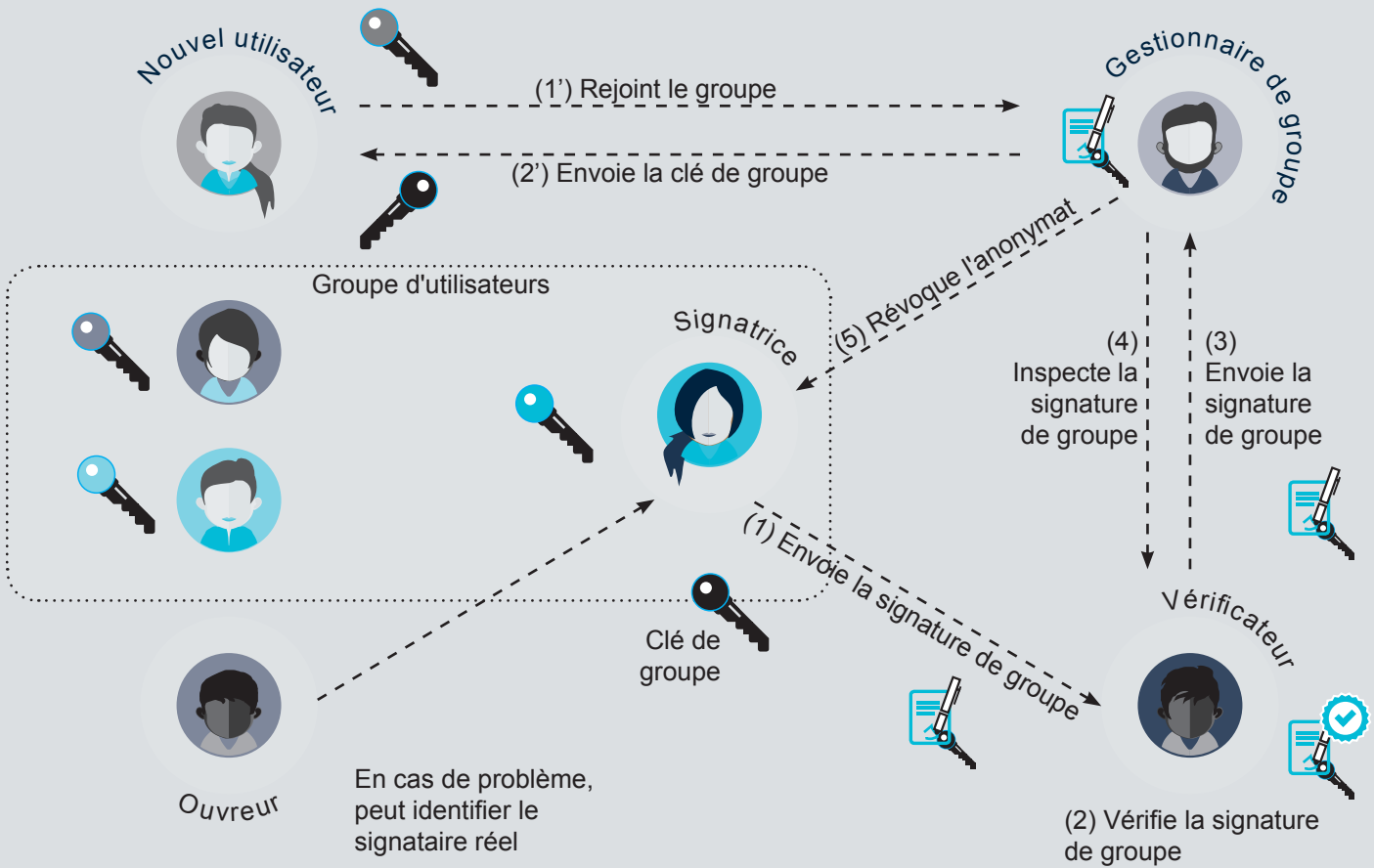
$$\Upsilon = (\text{médecin} \vee \text{soignant}) \wedge (\text{ville} \wedge (\text{NewYork} \vee \text{Paris} \vee \text{Tokyo})).$$

L'utilisateur dont les attributs *satisfont les arbres d'accès* est en mesure d'utiliser son crédentiel C pour extraire les valeurs associées aux attributs souhaités a_3 , a_4 et a_5. De cette manière, le médecin reste anonyme au sein du groupe de médecins résidant à Paris, tout en apportant une preuve solide à la requête émise, puisque la signature du Ministère de la Santé (jouant le rôle d'Autorité des attributs) est valide pour les attributs de ce médecin.

Des ABS aux systèmes AC

En 2016, Kaâniche & Laurent[43] ont proposé un nouveau schéma AC, reposant sur les signatures à base d'attributs, qui permet de rendre certaines propriétés d'un système AC (listées page 70).

Ce schéma permet la traçabilité des signatures via une nouvelle procédure d'inspection introduite pour supprimer l'anonymat et identifier l'utilisateur à l'origine d'une signature reposant sur les attributs ABS. L'inassociabilité entre émetteurs a également été rendue possible grâce à une nouvelle procédure d'émission.



les signatures de groupe en action

Fig. 9

Signatures de groupe ($\mathcal{GS}$)

Une signature de groupe est une primitive de clé publique qui permet aux membres d'un groupe de générer des signatures au nom du groupe auquel ils appartiennent. Un gestionnaire de groupe ajoute les membres au groupe, et les révoque. Ce gestionnaire peut être une entité en soi, ou une coalition d'entités (par exemple les membres du groupe), et peut aussi jouer le rôle –détaillé plus loin– de l'entité ouvreur.

Toute entité vérificatrice peut être certaine que la signature de groupe générée l'a bien été par un membre du groupe, sans en déduire son identité réelle ni être capable de l'identifier.

Applications des signatures de groupe

Les signatures de groupe sont utilisées dans plusieurs cas d'usage et avec différents paramétrages Une variante des signatures de groupe est ainsi utilisée comme brique de base des systèmes AC : la signature Camenisch-Lysyanskaya[44][45], utilisée dans le système Idemix –que nous détaillons plus loin. De manière générale, les signatures de groupe permettent le contrôle d'accès anonyme. Elles sont utilisées dans les systèmes de vote électronique (accès anonyme au bureau de vote), ou la monnaie électronique (accès anonyme au droit de payer).

Dans un contexte de système AC, il est possible de gérer plusieurs groupes. Un fournisseur de services peut souhaiter par exemple proposer des droits différenciés aux moins de 25 ans et au plus de 60. Une signature de groupe, par elle-même, ne permettrait pas de construire une telle offre. Au sein d'un système AC, cela devient possible. Le fournisseur n'a en effet pas besoin de connaître toutes les informations d'un utilisateur donné (nom, âge, adresse…) et doit juste s'assurer que certaines informations (âge, statut d'étudiant…) sont correctes et ont été certifiées par des autorités (les services de l'État, l'Université…). Grâce aux systèmes AC, l'utilisateur ne doit s'enregistrer qu'une seule fois pour appartenir à différents groupes (le groupe des personnes de moins de 25 ans, le groupe des personnes de nationalité française etc.)

Plusieurs mécanismes de $\mathcal{GS}$ supportant l'accès anonyme ont été proposés dans la littérature scientifique. Ainsi, une version appelée signatures de groupe liées (*linkable group signature*)

est une signature de groupe particulière pour laquelle une autorité est capable de révéler si deux signatures ont été générées par un même membre d'un groupe sans accéder à l'identité du signataire. Zheng *et al.*[46] ont proposé en 2018 un processus général de construction de ces signatures de groupe liées offrant des propriétés d'anonymat, et d'associabilité et la fonction d'audit permettant d'auditer les communications anonymes. Helback *et. al*[47] avaient présenté en 2008 une extension pour un système de vote électronique résistant aux attaques de type vol de vote, construite avec des signatures de groupes liées. L'idée était de permettre au votant de mettre son vote à jour autant de fois que souhaité (le tout dernier étant réputé celui souhaité) et de lier ces choix successifs, de sorte que l'attaquant ne puisse pas vérifier que le vote désiré est celui réellement retenu par le système.

Yan *et al.*[48] ont construit en 2016 un système de monnaie électronique autour d'une signature de groupe sans recours à des certificats électroniques. Ce mécanisme permet la non-forgeabilité, rend impossible la création de fausse monnaie ainsi que les *doubles paiements*. Cette solution de monnaie électronique supporte le système « *Fair off-line Multi-bank* » qui permet l'interaction de plusieurs banques même en mode déconnecté. Malina *et al.*[49] ont proposé en 2015 une combinaison du mécanisme de signature de groupe avec celui du chiffrement ElGamal, permettant de préserver l'anonymat des votants pendant le processus de vote et de décomptage. La propriété de responsabilité (voir pages suivantes) est assurée par la coopération du gestionnaire de groupe et de l'autorité responsable du vote pour révéler l'identité d'un des votants et le révoquer du système de vote si nécessaire.

Qu'est-ce que l'accountability ?

« *Accountability* », dans le cadre des données personnelles, a une signification technique – la propriété d'*accountability* – et juridique – le principe de responsabilité –.

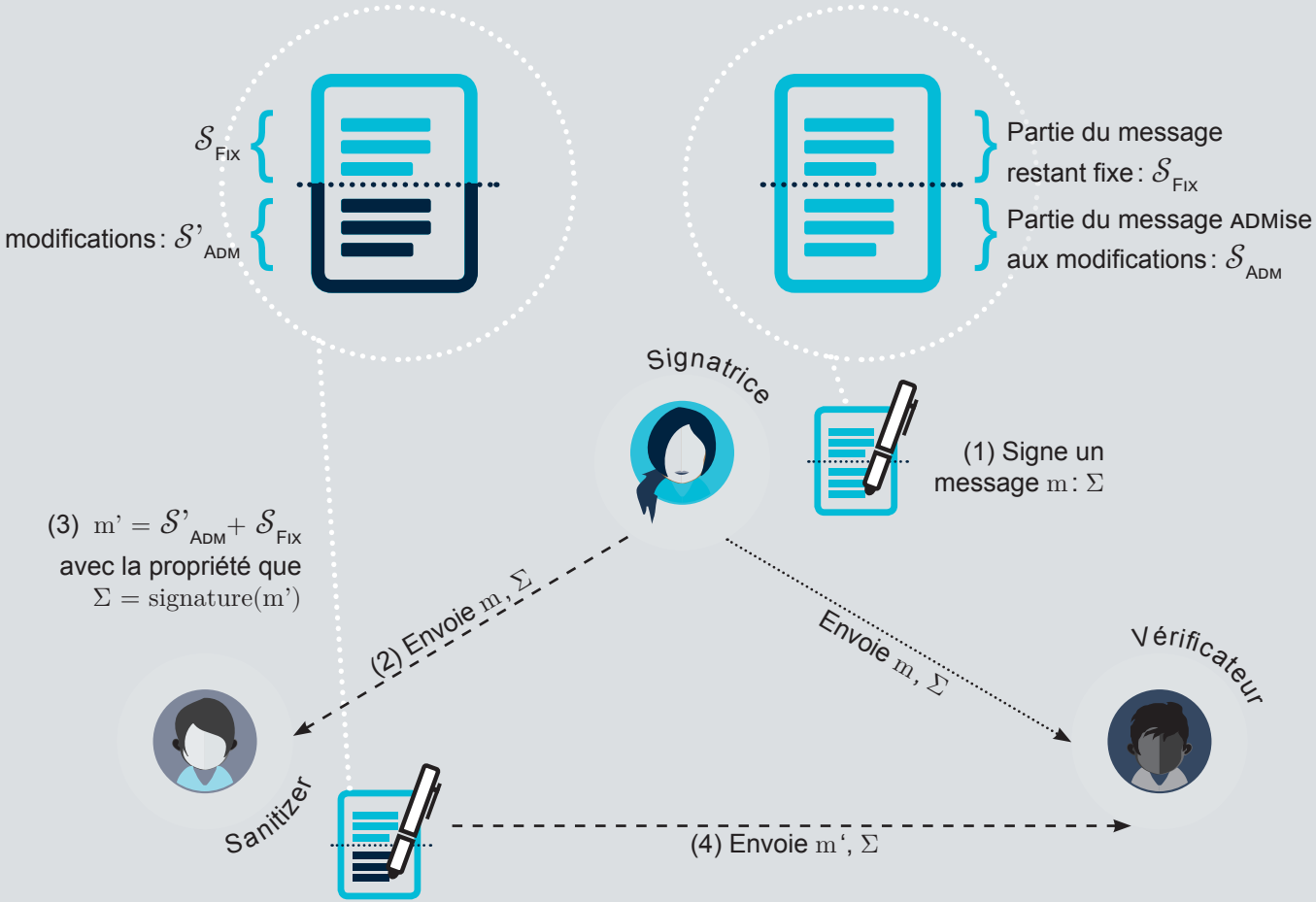
Du point de vue informatique, ***la propriété d'accountability*** fait référence à la responsabilité d'une entité (utilisateur, fournisseur de services…) à conserver des traces dans son système pour différentes actions futures. Dans le cadre des signatures de groupe, par exemple, un utilisateur peut ainsi engager le groupe en utilisant la signature de groupe, et en cas de problème, son identité doit pouvoir être révélée à, et par, l'entité qui en a le pouvoir.

Du point de vue juridique, ***le principe de responsabilité*** (*accountability*) correspond à l'obligation de rendre des comptes, et se situe plutôt à l'échelle des organisations. Ce principe a été reconnu explicitement dès 1980 dans les lignes directrices régissant la protection de la vie privée de l'Organisation de Coopération et de Développement Économiques (OCDE). Le point 14 y précisait ce « Principe de la responsabilité » : « *Tout maître de fichier devrait être responsable du respect des mesures donnant effet aux principes énoncés ci-dessus.* » Il est l'un des principaux concepts du cadre défini par l'Organisation Économique pour l'Asie-Pacifique (APEC) pour la protection de la vie privée (point 26 de l'*APEC Privacy Framework*). Il figure également dans la dernière version du projet de norme ISO 29100 établissant un cadre pour le respect de la vie privée.

Caractéristiques et usages des $\mathcal{GS}$

Par définition, les signatures de groupe sont anonymes et intraçables par quiconque, sauf pour une entité (*opener*) désignée pour ouvrir, si nécessaire, une signature et identifier le signataire effectif. De cette manière, deux propriétés principales sont assurées par les signatures de groupe : l'anonymat et la responsabilité (*accountability*).

De manière générale, les signatures de groupe permettent le contrôle d'accès anonyme. Elles sont utilisées dans les systèmes de vote électronique (accès anonyme au bureau de vote)[50][51][52][53], ou la monnaie électronique (accès anonyme au droit de payer)[54][55][56].



échanges lors d'une signature modifiable

Fig. 10

Sanitizable Signatures ($\mathcal{S}t\mathcal{S}$)

Ces ***signatures modifiables***, proposées en 2005 par Ateniese *et al* [57], ont été conçues pour autoriser une tierce partie, appelée *sanitizer*, à modifier certaines parties d'un message, sous le contrôle de l'utilisateur émetteur du message qui en a désigné à l'origine les parties modifiables. Le mécanisme de ces signatures modifiables doit respecter la propriété de complétude (*correctness*) : une signature honnêtement générée (*signing correctness*) et honnêtement modifiée (*sanitizing correctness*) doit être acceptée par le vérificateur, et des preuves honnêtement générées à partir de signatures valides (*proof correctness*) doivent être acceptées par l'entité vérificatrice.

_Construction des signatures modifiables

Pour permettre à une entité tierce de modifier des parties d'un message $\mathrm{m} \in \{0,1\}^*$, l'utilisateur le divise en N blocs $\mathrm{m}_1,\cdots,\mathrm{m}_\mathrm{N}$ et y associe un ensemble de blocs admis à être modifiés, $\mathcal{S}_{\text{ADM}} \subseteq \{1,\mathrm{N}\}$, et signe l'ensemble du message avec une clé associée au sanitizer. Ce dernier peut, grâce à cette clé, modifier les partie autorisées à la modification, de telle manière que la signature résultante reste valide avec la clé du signataire originel. Deux procédures sont définies :

- ADM fait correspondre pour chaque message m les blocs admissibles à modification. Elle rend le rang et la longueur de chaque bloc admissible.
- MOD fait correspondre tel bloc m_j de départ au bloc modifié $\mathrm{m'}_\mathrm{j}$ d'arrivée.

Une procédure optionnelle, FIX, est possible. Elle rend l'ensemble $\mathcal{S}_{\text{FIX}} \subseteq \{1,\mathrm{N}\}$ des parties devant rester fixes du message m.

Extension des signatures modifiables

Plusieurs extensions ont été proposées au concept initial. Canard et Jambert[58] ont ainsi montré en 2010 comment limiter l'ensemble des modifications possibles à un unique bloc, et comment forcer des modifications identiques sur différents blocs. En 2012, Canard *et al.*[59] ont étendu le concept pour des contextes avec plusieurs utilisateurs signataires et plusieurs sanitizers.

Motivations pour créer des signatures modifiables

Dans certaines organisations (domaine militaire, domaine de la santé…), il peut être nécessaire que des personnes puissent accéder à un même document, ou aux enregistrements d'une base de données, avec des niveaux d'accréditation différents. De même, certains documents doivent pouvoir être rendus accessibles à un plus large public, sans pour autant l'être entièrement. Les parties qui ne doivent pas être révélées sont dites caviardées, ce qui le plus souvent prend la forme d'un texte partiellement noirci [redacted].

Dans certains cas, une tierce partie autorisée peut souhaiter modifier le document, dans un cadre limité et contrôlé (modèle de confiance *semi-trusted*), et obtenir une signature valide pour le document modifié, sans faire appel au signataire originel (parce que sa clé a expiré, ou la signature était datée, ou il n'est pas disponible, ou le coût serait prohibitif). C'est avec ce type de cas d'usages en tête que Ateniese *et al.* ont construit les *sanitizable signatures*.

Extension des signatures $\mathcal{S}t\mathcal{S}$ vers ACS

Les signatures $\mathcal{S}t\mathcal{S}$ sont des primitives puissantes permettant de faire des traitements sur des données déjà authentifiées. Elles sont à ce titre très utilisées dans les systèmes ACS. Chow *et al.*[60] ont ainsi proposé un mécanisme de gestion des identités distribuées dans un contexte *cloud*. Par la combinaison de signatures de groupe et de signatures modifiables, il est possible de s'identifier chez différents fournisseurs de *cloud* en ne leur révélant pas des informations qui ne les concernent pas.

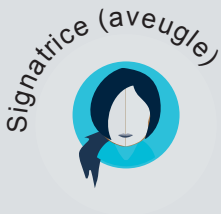
L'approche formalisée par Canard & Lecuyer en 2013 (voir ci-contre) étendent les $\mathcal{S}t\mathcal{S}$ avec les points suivants : la signature doit pouvoir être vérifiée sans la clé publique du *sanitizer*, de manière à préserver l'anonymat de l'utilisateur ; la traçabilité des signatures doit pouvoir être effectuée par une autorité indépendante quand il s'agit de retrouver le *sanitizer* associé à une paire donnée message-signature ; aucun algorithme de preuve n'est nécessaire ; les modifications sont seulement possibles dans des limites contrôlées ; le *sanitizer* doit modifier les parties qui peuvent l'être en utilisant une valeur aléatoire envoyée par le vérifieur à chaque session.

En 2018, Pàmies-Estrems *et al.*[61] ont proposé une solution dans le cadre des services construits autour des assistants vocaux comme Google Home ou Amazon Echo, via une combinaison de signatures modifiables et de techniques d'anonymisation (que l'on verra en page 135 et suivantes). Le mécanisme permet de ne pas révéler les informations d'identification lors du transfert de données entre l'assistant et les serveurs des fournisseurs.

Des $\mathcal{S}\mathrm{t}\mathcal{S}$ aux systèmes AC

Un utilisateur disposant d'un document contenant ses données personnelles, signé par une autorité, pourrait souhaiter utiliser ces crédentiels sans en révéler tout le contenu. Il jouerait ainsi lui-même le rôle de *sanitizer* sur son document, qui resterait toujours certifié par l'autorité.

Les signatures $\mathcal{S}\mathrm{t}\mathcal{S}$ ne le permettent cependant pas directement. C'est avec ce type de cas d'usage en vue qu'en 2013, Canard & Lecuyer ont formalisé[62] l'utilisation des $\mathcal{S}\mathrm{t}\mathcal{S}$ dans le cadre d'un système ACS, tout en fournissant une construction concrète.



possède (et garde pour elle) le secret x et rend publique sa clé de vérification $h = g^x$.

connaissent tous deux $h = g^x$

choisit (au hasard, et garde pour elle) w dans $\mathbb{Z}_q$
calcule (et communique au vérificateur) $a = g^w$

Phase d'engagement

Envoie a

$\mathbb{Z}_q$ est l'anneau des entiers relatifs modulo le nombre entier q, $\{1, 2, ... q\}$

Phase de défi

choisit α et β dans $\mathbb{Z}_q$
calcule $a' = a \cdot g^{\alpha} h^{\beta}$
calcule $c' = H(m||a')$
calcule $c = c' + \beta$

Envoie c

H est une fonction de hachage de $\{0,1\}^*$ vers $\mathbb{Z}_q$
Nota : m est le message, et $||$ représente la fonction de concaténation.

calcule $r = cx+w$

Phase de réponse

Envoie r

vérifie $a \stackrel{?}{=} g^r h^{-c}$
calcule $r' = r + \alpha$
signature (c,r) randomisée par le vérificateur pour présentation à la tierce-partie : (c',r')

Envoie $m,(c',r')$

vérifie $c' \stackrel{?}{=} H(m||g^{r'}h^{-c'})$

Paramètres publics : q, nombre premier, qui définit un groupe G d'ordre q engendré par : g.

protocole d'identification de Schnorr à l'aveugle

Fig. 11

Signatures aveugles

Les signatures aveugles, définies en 1982 par David Chaum[63], permettent d'obtenir une signature d'un signataire sans que ce dernier n'ait connaissance du contenu du message qu'il a signé. Elles sont utilisées quand le signataire et l'auteur du message signé sont deux personnes différentes, l'auteur masquant le message avant que le signataire n'intervienne.

Comme pour les signatures de groupe, les signatures aveugles sont utilisées notamment dans les systèmes de vote électronique, ou pour la monnaie électronique[64].

Protocole d'identification de Schnorr en situation aveugle

Le protocole de Schnorr[65] est un protocole classique d'identification à *divulgation nulle de connaissance*, proposé par le mathématicien Claus-Peter Schnorr en 1989. Ce protocole, dont les mathématiques (et la sécurité) reposent sur le problème complexe du logarithme discret, permet la création d'une signature numérique, en rendant la preuve non interactive. Il est constitué de trois phases : 1) une étape dite d'*engagement*, durant laquelle l'utilisateur qui doit prouver son identité effectue un certain calcul a sur la base d'un nombre tiré au hasard et d'un secret x et envoie ce calcul au vérificateur, 2) une phase de *défi* où le vérificateur envoie un nombre tiré aléatoirement c à l'utilisateur, et 3) une phase de *réponse* où l'utilisateur effectue un dernier calcul utilisant c, et dont le résultat est pour le vérificateur qui le reçoit une preuve que l'utilisateur connaît x.

Dans la version à l'aveugle de la signature Schnorr (voir figure 11 ci-dessus), l'utilisateur–signataire prouve sa connaissance du secret x sur la base de sa clé publique $\mathrm{h} = \mathrm{g}^{\mathrm{x}}$. Le vérificateur crée alors un défi de telle manière qu'une signature résultante est créée à partir de l'engagement et de la réponse du signataire. La signature résultante peut elle-même être vérifiée. L'entité vérificatrice peut alors faire suivre son message (dont le contenu n'a pas été révélé au signataire) et la signature à une tierce partie qui sera en mesure de vérifier que la signature a été générée par un utilisateur légitime.

Signatures aveugles dans un système ACS

Les signatures aveugles sont bien adaptées pour créer des jetons de présentation selon les exigences d'anonymat d'un système ACS. En effet, la validité de la signature, c'est-à-dire la complétude de la vérification, garantit que l'utilisateur qui a généré le jeton a les droits d'accès au service demandé. Le caractère aveugle de la signature garantit que le fournisseur de services n'est capable ni de reconnaître le signataire effectif, ni de déceler dans un jeton de présentation précédent de quoi le relier à un utilisateur particulier.

Ces mécanismes ont été appliqués sur différents cas d'usage avec différents paramétrages. Une variante des signatures aveugles est la composante principale des systèmes ACS, la signature Brand que l'on trouve dans la solution ***U-Prove***. Dans cette implémentation, l'émission d'un crédentiel se fait lors d'un protocole durant lequel l'émetteur signe à l'aveugle l'*engagement*. Les attributs utilisés sont connus à la fois de l'émetteur et de l'utilisateur, mais le crédentiel qui en résulte, ainsi que la clé secrète, ne sont connus que du seul utilisateur. Par conséquent, un émetteur et un vérificateur qui seraient de connivence, ne pourraient en aucun cas relier un crédentiel à un utilisateur particulier.

Une version dite « *self-blindable credentials* » a été proposée en 2001 par Verheul[66] sur la base de la signature Chaum-Pederson[67], dans laquelle chaque crédentiel aveuglement émis est différent, et donc intraçable. Ce qui n'est pas le cas avec U-Prove où le jeton de présentation est identique à chaque transaction et agit ainsi comme pseudonyme.

			Idemix	U-Prove
Propriétés	Anonymat		oui	oui
	Inassociabilité	*issue show*	oui	oui
		multi show	oui	
	Révocation		oui	oui
	Inspection		oui	oui
	Divulgation sélective		oui	oui
Performances	Émission d'un crédentiel (5 attributs)		2,6	5,5
	Divulgation sélective	preuve vide	1,5	0,9
		preuves de possession portant sur 3 attributs sur un total de 5	<1	0,6

performances exprimées en secondes, mesurées sur une carte à puces opérée sous MULTOS

comparaison des solutions Idemix et U-Prove

Tab. 3

Comparatif des solutions industrielles

Deux solutions industrielles principales reposant sur les ACS ont émergé : Idemix et U-Prove.

- Un crédentiel Idemix est une signature Camenisch-Lysyanskaya (CL) – une variante d'une signature de groupe, rappelons-le – générée par un émetteur à partir de la clé secrète d'un utilisateur et des valeurs d'attributs. Pour transformer un crédentiel en un jeton de présentation, l'utilisateur crée une *preuve à divulgation nulle de connaissance* montrant qu'il connaît une signature CL valide sur une valeur connue engagée.
- De son côté, U-Prove est construite à partir d'une variante des signatures aveugles, proposée par Brands[68] en 2000.

Idemix : https://www.zurich.ibm.com/identity mixer/

U-Prove : https://www.microsoft.com/en-us/research/project/u-prove/

Point communs entre Idemix et U-Prove

Tant Idemix que U-Prove préserve l'anonymat des utilisateurs durant le processus d'identification, grâce aux capacités des signatures CL et Brands, respectivement, et aux mécanismes de preuve à divulgation nulle de connaissance.

De plus, ces deux systèmes ACS héritent de la propriété de non-forgeabilité de leurs mécanismes de signature respectifs, rendant ainsi impossible à une entité n'appartenant pas à l'ensemble des utilisateurs autorisés d'exécuter le protocole d'identification avec succès auprès du fournisseur de services.

Différences entre Idemix et U-Prove

Contrairement à Idemix, U-Prove ne présente pas la propriété d'inassociabilité *multi-show*. Dit autrement, un crédentiel U-Prove est révélé à chaque fois qu'il est utilisé, ce qui fait qu'un utilisateur devra, pour empêcher d'être pisté, générer autant de crédentiels U-Prove différents que de sessions de présentations. Ceci découle du fait que le protocole d'émission d'Idemix repose sur les signatures CL dont le mécanisme de création permet d'obtenir plusieurs versions valides (*randomisable signature*).

Concernant les performances, les deux solutions sont consommatrices de ressources à des moments différents. Comme expliqué par Vullers *et al.* (2012[69], 2013[70]), les implémentations de U-Prove et de Idemix sur une carte à puce sous système d'exploitation MULTOS dans un contexte à cinq attributs montrent que la procédure d'émission de crédentiels U-Prove est beaucoup plus consommatrice de ressources (5,5 secondes pour s'exécuter) que celle d'Idemix (2,6 secondes). C'est la conséquence du nombre d'interactions nécessaires, qui conduit à des coûts en communication lourds. C'est le contraire pour la procédure de présentation, beaucoup plus consommatrice côté Idemix (1,5 secondes pour la présentation d'une preuve vide, contre 0,9 secondes avec U-Prove). Cette fois, il s'agit de calculs additionnels rendus nécessaires par les signatures CL qui sont construites sur des groupes *RSA* de grande taille.

	Propriétés ACS souhaitées	Signature Modifiable	Signature sur les Attributs	Signature de Groupe	Signature Aveugle
Propriétés de sécurité	Non-forgeabilité	non-forgeabilité	non-forgeabilité	non-forgeabilité	non-forgeabilité
		immutabilité ①			
	Anonymat	anonymat	anonymat	anonymat	anonymat
		vérifiable sans clé du signataire	vérifiable publiquement	vérifiable publiquement	
	Inassociabilité *multi-show*	oui	oui	oui	non
Propriétés fonctionnelles	Divulgation sélective	inhérent à cette signature	inhérent à cette signature		
	Procédure d'inspection	traçabilité	*accountability*	*accountability*	traçabilité

① *immutabilité : aucune modification n'est possible sur les parties fixées*

Propriété non atteinte

comparaison des mécanismes de signatures
en regard des exigences fonctionnelles et de sécurité

Tab. 4

Comparatif entre les mécanismes de certification selon les signatures utilisées

Idemix et U-Prove que nous venons d'évaluer sont des solutions industrielles complètes. Il est également intéressant de comparer les mécanismes de signatures dont les extensions, présentées plus haut, permettent de construire des systèmes AC. Parmi ces extensions, la traçabilité, ou le support de plusieurs signataires. Des différences plus importantes existent également, comme l'impossibilité intrinsèque des signatures aveugles à honorer la propriété d'inassociabilité *multi-show*, ainsi que les ressources nécessaires en communication et en temps de calcul.

Discussion sur les signatures

Dans le tableau 4 ci-dessus nous proposons une comparaison entre les propriétés – de sécurité et fonctionnelles – des mécanismes de signature examinés dans un contexte de certification anonyme.

Soulignons tout d'abord que tous les mécanismes de signature examinés sont considérés comme des primitives cryptographiques, tandis qu'un système AC fait référence à un système complet, impliquant plusieurs procédures et algorithmes. Par conséquent, bien que tous les schémas de signatures garantissent les propriétés de non forgeabilité et d'anonymat / confidentialité, il est normal – et légitime – que ces signatures ne prennent pas en charge toutes les propriétés de sécurité et fonctionnelles attendues d'un système AC. Rappelons également que, comme nous l'avons exposé dans les pages précédentes, toutes les signatures en question ont été étendues et leurs algorithmes adaptés pour supporter les propriétés de sécurité des systèmes AC.

En effet, les constructions [71] et [72], introduisant deux systèmes alternatifs différents basés respectivement sur $\mathcal{S}t\mathcal{S}$ et sur ABS, ont adapté ces primitives et fourni une nouvelle fonctionnalité : la traçabilité. Par exemple, [71] propose de concevoir un nouvel algorithme qui serait exécuté par une autorité distincte, l'inspecteur, afin de retrouver les modifications

d'une paire message-signature donnée. Une autre extension permettant le support de plusieurs signataires, c'est-à-dire les émetteurs d'un système AC, est également proposée, dans [73]. De plus, l'algorithme de modifications (*sanitizing*) est adapté pour limiter les modifications sur les valeurs / blocs admissibles, afin de garantir que ces modifications sont faites de manière contrôlée. [72] introduit également un algorithme exécuté par une autorité de confiance spécifique, à nouveau l'inspecteur, en charge de lever l'anonymat d'un utilisateur ayant employé une signature ABS. Et y ajoute une extension intéressante de l'algorithme d'émission des crédentiels, permettant à plusieurs émetteurs de les produire, à partir de la clé publique de l'utilisateur.

_Les différences de fond

Rappelons pour commencer que les signatures aveugles ne permettent pas, pour les raisons intrinsèques expliquées page 89, la propriété d'inassociabilité *multi-show*. L'autre différence importante réside dans les ressources nécessaires en communication et en calcul. La consommation en bande passante mesure ainsi les quantités de données échangées par les protocoles au regard du nombre d'attributs supportés par un crédentiel ou un jeton de présentation. Les coûts en calcul, pour l'utilisateur, l'émetteur, le vérificateur, sont relatives au nombre et à la complexité des calculs effectués lors des échanges protocolaires. Les critères de performance sont très importants lors du choix des mécanismes de signature, mais en tant que mécanismes purs, ces coûts sont plutôt intéressants, comparés aux solutions industrielles complètes Idemix et U-Prove.

_Obfuscation par ajout de données

Obfusquer les données transmises, une approche de l'anonymisation par l'introduction de bruit, c'est-à-dire d'informations erronées parmi les informations justes, est une technique utilisée dans le cadre des services nécessitant la géolocalisation. C'est ce que propose la solution ***DUMMY-Q***, introduite par Pingley *et al.*[74] en 2011. Un logiciel est installé côté client pour envoyer au service en question une série de requêtes avec des localisations fantaisistes, en plus de celle qui est réelle. Le service se trouve dans l'incapacité de déterminer quels attributs requis présentent un réel intérêt pour l'utilisateur.

En 2014, Niu *et al.*[75] ont proposé un algorithme permettant de cacher beaucoup plus sûrement la localisation réelle de l'utilisateur. En effet, le fournisseur de service ou un attaquant pourrait très bien détenir des informations sur sa localisation provenant d'autres sources d'informations. L'algorithme repose sur une mesure de l'*entropie* des informations en jeu.

L'obfuscation de données est également utilisée dans des contextes de navigation sur le web. Elovici *et al.*[76] ont proposé en 2016 ***PRAW***, un système de recherche sur le web préservant la vie privée. Ce système assure la confidentialité d'un groupe d'utilisateurs partageant la même connexion d'accès au web. Les auteurs proposent de cacher le profil réel des utilisateurs en générant par exemple de fausses requêtes web.

3’

Obfuscation des données

L'utilisateur peut empêcher un attaquant de le profiler précisément en obfusquant les informations qu’il divulgue explicitement ou implicitement. Il peut le faire en émettant des données fausses en même temps que les données authentiques, ou en supprimant certaines données.

Ces techniques sont utilisées dans le cadre des services utilisant la géolocalisation, ainsi qu'en navigation sur le web.

Cette *approche par obfuscation* de données se fait côté utilisateur. Celui-ci n'a donc besoin d'avoir ni recours ni confiance à une entité externe pour les réaliser.

À partir de *PRAW*, Ye *et al*[77]. proposent en 2009 une solution de navigation privée sur la base suivante : toute requête réellement souhaitée est envoyée avec la probabilité p, et une requête fantaisiste l'est alors avec la probabilité $1\text{-}p$, rendant impossible pour un attaquant de savoir quelle requête a de la valeur pour l'utilisateur.

En 2018, Masood *et al.*[78] introduisent à leur tour ***Incognito***, qui repose également sur des probabilités, mais cette fois celles permettant de prédire les risques associés à la divulgation de certaines données sur le web. Les données pour lesquelles le risque associé prédit est le plus élevé sont obfusquées avec des données sémantiques approchantes, d'une manière telle que les risques évalués se trouvent diminués.

_Obfuscation par suppression de données

Obfusquer des données ne se fait pas seulement par ajout de données fausses ou fantaisistes. Enlever des données est également utile, en particulier quand elles ne sont pas essentielles. C'est le cas de certaines URLs de requête web qui comportent une longue liste de paramètres pas forcément utiles. Parra-Arnau *et al.*[79] ont proposé en 2010 une technique d'élimination de ces paramètres, dans des contextes d'utilisation du web sémantique. Comme ces suppressions apportent des limites aux usages, les auteurs ont publié en 2012 une étude sur les compromis nécessaires dus à l'utilisation de cette méthode.[80]

Les deux techniques d'obfuscation, ajout et suppression de données, peuvent êtres combinées avec succès. Plusieurs études ont été conduites pour imaginer de tels systèmes hybrides. Chez Parra-Arnau *et al.*[81] en 2014, dans un contexte d'usage autour de recommandations, les utilisateurs soumettent de fausses évaluations qui ne reflètent pas leurs préférences et/ou s'interdisent de procéder à des évaluations. Chez Polatidis *et al.*[82] en 2017, dans un contexte de recommandation collaborative, différents niveaux de confidentialité sont définis, et les valeurs possibles pour les notations sont obfusquées dans la limite des ensembles de valeurs possibles. Avant chaque évaluation soumise au serveur, niveaux de confidentialité et ensembles de valeurs sont tirés au hasard en fonction du niveau de vie privée souhaité.

Alice, Bob et Carole disposent tous trois d'un secret, X_1, X_2 et X_3 respectivement.

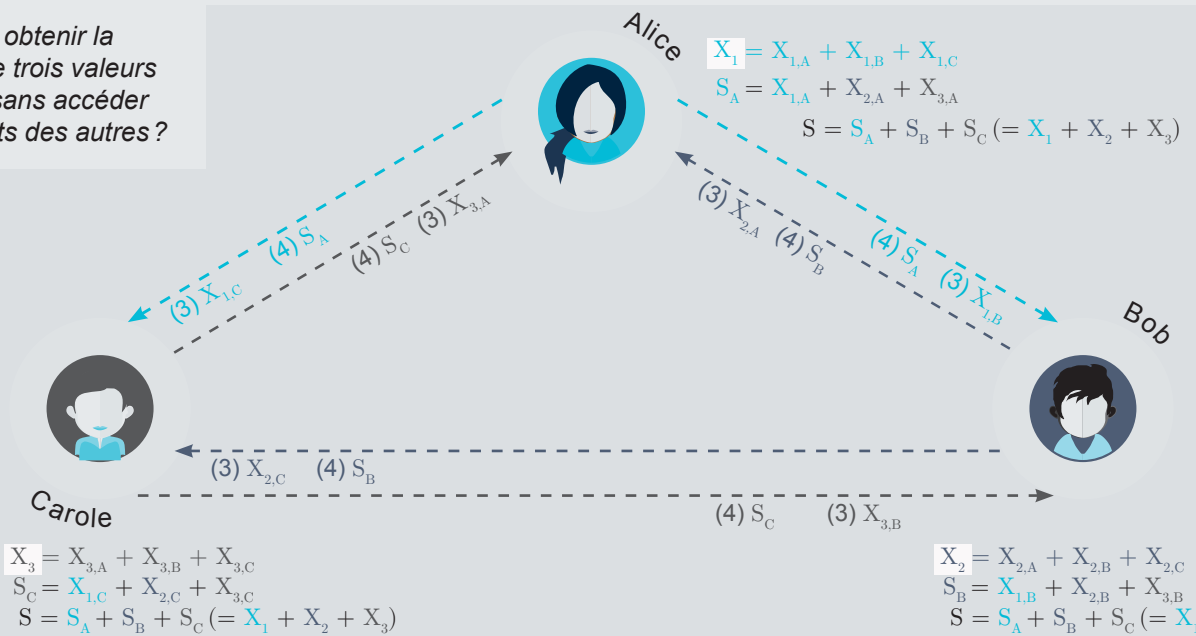
Tous trois partagent leur propre secret en trois morceaux, $X_{i,A}$, $X_{i,B}$ et $X_{i,C}$.

Alice, Bob et Carole envoient un de ces morceaux à leurs voisins.

Alice, Bob et Carole effectuent un premier calcul partiel $S_j = X_{1,j} + X_{2,j} + X_{3,j}$.
Puis envoient ce résultat aux mêmes voisins.

Il ne leur reste plus qu'à additionner S_A, S_B et S_C

Comment obtenir la somme de trois valeurs secrètes sans accéder aux secrets des autres ?



calcul multipartite sécurisé

Fig. 12

4’

Calculer sans tout montrer

Il est parfois nécessaire d'effectuer des calculs, de faire des requêtes dans des bases de données, ou de la fouille de données, sans s'exposer aux yeux des autres. Il en va de même pour certains types de vote, et pour des contextes de recherche de consensus ou d'évaluation en l'absence des valeurs exactes permettant ces calculs.

Le problème des millionnaires[83], proposé par Yao en 1982, illustre bien cela : comment deux entités peuvent-elles savoir laquelle des deux est la plus riche, sans s'échanger le montant exact de leurs avoirs ?

La généralisation du problème de Yao se fait ainsi : un groupe de participants souhaitent calculer la valeur d'une fonction f (par ex. $\mathrm{f} = \mathrm{add}()$) en utilisant leurs données d'entrée que seuls eux connaissent, sans les révéler aux autres. Ils doivent alors passer par un protocole interactif et itératif, de partage de calculs partiels, comme illustré figure 12 pour trois participants.

Plusieurs mécanismes de calcul multipartite sécurisé reposent sur un protocole d'évaluation de circuit[84]. Concrètement, les données d'entrée de chaque utilisateur sont représentées par un *circuit booléen*, ou un *circuit arithmétique*, et les échanges de communication entre les portes des circuits passent par le protocole *Oblivious Transfer*, en utilisant des variables aléatoires.

Le transfert inconscient (*Oblivious Transfer*) est une primitive cryptographique où l'émetteur envoie à un récepteur une information sélectionnée parmi plusieurs envois possibles, sans avoir la certitude que l'information arrive à destination, ni laquelle est utile au destinataire, et sans que ce dernier ne sache quelles autres informations il aurait pu acquérir.

Bien que ces mécanismes assurent le respect des données personnelles servant d'entrées aux calculs à effectuer, ils sont coûteux en calcul et en communication, ce qui peut être un obstacle dans certains équipements où les ressources sont une contrainte.

_Modèle de sécurité

Bien que ces mécanismes soient considérés comme relevant du modèle sans confiance, leur sécurité doit être élaborée comme si f était calculée par un tiers de confiance, ce qui correspond au modèle idéal. À aucun moment du protocole ci-dessus, les entités ne doivent en savoir plus sur les données des autres que ce qu'elles auraient su si f avait été exécutée par un tiers de confiance.

_Propriétés exigées

- ***complétude :*** une entité, ou un groupe d'entités, mal intentionnée ne doit pas pouvoir, en modifiant ses propres données d'entrée, influencer les résultats du calcul pour les parties honnêtes plus que ce qu'elle serait capable de faire dans le cas du modèle idéal.
- ***indépendance des données d'entrée :*** dans le modèle idéal, toutes les données d'entrée sont envoyées à une tierce partie de confiance pour effectuer les calculs, puis les données de sortie du calcul sont reçues. Les entités mal intentionnées ne savent rien des données d'entrée des parties honnêtes avant que celles-ci ne les envoient. Autrement dit, les données de chaque entité sont choisies de manière indépendante.
- ***équité :*** cette propriété assure que les parties honnêtes conservent un accès aux données de sorties correctes, même si un groupe, minoritaire, d'entités est mal intentionné. Cependant, cette propriété est inévitablement impossible à atteindre si ces entités mal intentionnées peuvent stopper le protocole avant la fin des calculs, empêchant les entités honnêtes d'accéder à leurs résultats. Plusieurs travaux ont été proposés pour surmonter cette difficulté en relâchant un peu les exigences de sécurité.

_Calcul multipartite pour applications pervasives

Plusieurs travaux ont été proposés afin de garantir la confidentialité des calculs collaboratifs pour les applications pervasives.

Reiter *et al.*[85] proposent ainsi en 1998 un nouveau protocole de calcul collaboratif : ***Crowds***. Celui-ci permet à un groupe d'utilisateurs de collaborer pour soumettre leurs messages à un serveur Web non fiable. Lorsqu'il envoie un message, l'utilisateur tire à pile ou face pour décider de le soumettre directement au fournisseur de services ou de l'envoyer à un autre utilisateur, qui répète ensuite cette décision aléatoire. Hélas ceci rend la communication très coûteuse, principalement en raison de ces sauts supplémentaires. Barba *et al.*[86] en 2013 ont cependant pu proposer, dans le cadre d'applications de transport, des travaux dérivés ayant des coûts de communication et de calcul plus acceptables.

En 2009, Pathak *et al.*[87] proposent un protocole évolutif pour du calcul collaboratif sur des données chiffrés de nature bancaire. Dans leur approche, des *tokens* sont créés en même temps que le processus de chiffrement est effectué. Les données ainsi mixées avec les *tokens* forment le texte chiffré qui est transmis, ce qui ne révèle aucune donnée personnelle originelle.

Implémentations

Trois approches différentes, ayant leurs avantages et leurs inconvénients, sont utilisées pour ce type de calcul :

- **par transfert inconscient :** très coûteux en calcul et en communication[88]
- **par partage de secret :** meilleure en coût de calcul, cette technique qui nécessite en revanche des canaux de communication sécurisés entre entités, est coûteuse en bande passante[89][90][91]
- **chiffrement homomorphe :** ici, aucun canal de communication sécurisé nécessaire, mais tout repose sur des calculs de grande complexité[92]

Erola *et al.*[93] appliquent en 2011 ces techniques sur le modèle de *Crowds* dans le domaine des requêtes web, les utilisateurs étant regroupés en fonction de leurs centres d'intérêt. En 2012, Rebollo *et al.*[94] s'inspirent également de Crowds, deux utilisateurs ou plus partageant une partie de leurs requêtes avant de les soumettre aux moteurs, cachant ainsi leurs centres d'intérêt à leur fournisseur de service.

_Calculs collaboratifs pour objets connectés

Dans le cadre des objets connectés, Jaydip Sen[95] discute dès 2010 des pistes de recherche pour le calcul réparti distribué sécurisé, car il observe que la plupart des constructions proposées jusqu'alors sont très liées à leur domaine d'application, et ne sont pas adaptées aux environnements pervasifs. Elles sont en effet très gourmandes en calcul ou en communication, voire les deux, ce qui pose des difficultés pour des objets ayant peu de ressources embarquées et peu d'énergie disponible.

Tonalyi *et al.*[96] présentent en 2017 un mécanisme d'agrégation de données préservant la confidentialité pour les applications embarquées sur les objets connectés, qui suit le protocole FHE (*Fully Homomorphic Encryption*). Chaque objet utilise un générateur de nombres pseudo-aléatoire pour calculer localement les mêmes éléments calculés par les autres objets et éviter de devoir les échanger.

Améliorations de ces techniques

Ces mécanismes sont coûteux en calcul et en communication, ce qui peut être un obstacle dans des cas d'usage – notamment en mobilité – où les équipements ont des limites en terme de ressources et d'énergie. Une partie des travaux de recherche récents consiste donc à améliorer ces algorithmes avec des coûts de calcul et de communication plus acceptables. Qui plus est, les solutions proposées ont souvent été fortement liées aux domaines d'application considérés, rendant peu propice une généralisation de ces méthodes. Proposer des algorithmes plus génériques, voire des composants électroniques adaptés favorisera le déploiement de ces techniques.


2

Les techniques orientées serveur

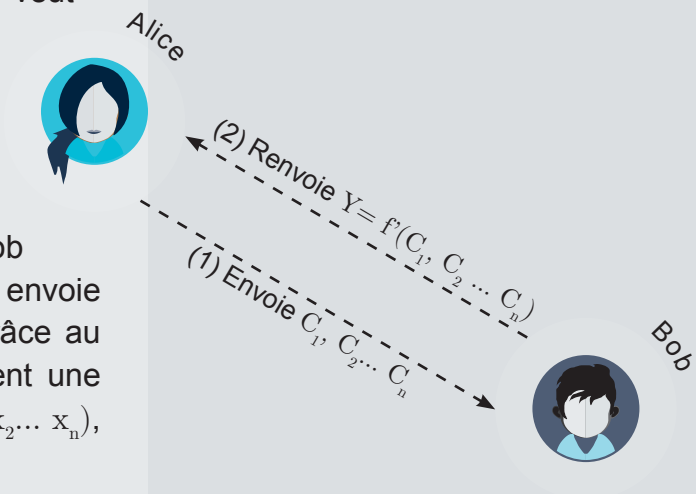
Ces techniques sont mises en œuvre côté serveur, notamment pour effectuer des traitements complémentaires nécessaires sur les données confiées par les utilisateurs, comme par exemple l'anonymisation des bases de données ou des calculs lourds sur des données chiffrées à la demande de l'utilisateur.

Parmi les PETs orientées serveur se trouvent ainsi les systèmes à données auto-destructibles et les mécanismes de contrôle de la divulgation de données statistiques. Poursuivons tout d'abord par deux sous catégories, le chiffrement homomorphe et la récupération d'information privée, qui appartiennent respectivement aux catégories : calculs sécurisés et obfuscation.

Alice souhaite effectuer un calcul, à savoir une fonction $f()$, sur des données $x_1, x_2 ... x_n$ lui appartenant, mais ne dispose pas de la puissance de calcul requise. Elle souhaite déléguer ce calcul à Bob, une entité qui dispose d'une capacité de calcul suffisante. Mais, Alice ne fait pas totalement confiance à Bob et ne veut pas qu'il ait connaissance de la valeur ou du contenu de ses données.

Alice va alors les chiffrer ($C_i = \text{ENC}(x_i)$) et les envoyer à Bob, qui récupère ainsi des données chiffrées $C_1, C_2 ... C_n$. Bob effectue le calcul $Y = f'(C_1, C_2 ... C_n)$, envoie ce résultat à Alice, qui le déchiffre. Grâce au caractère homomorphe de f, elle obtient une valeur $\text{DEC}(Y)$ qui n'est autre que $f(x_1, x_2 ... x_n)$, ce qu'elle cherchait à calculer.

Et jamais Bob n'a eu de quoi prendre connaissance des $x_1, x_2 ... x_n$ initiales.



Chiffrement homomorphe

Fig. 13

4”

Chiffrement homomorphe

Le chiffrement homomorphe se donne pour objectif de chiffrer les données de telle manière que des calculs puissent être effectués sur les données chiffrées. Autrement dit, nul besoin d'accéder aux valeurs en clair pour effectuer des calculs utiles.

Si la notion de *privacy homomorphism* date de 1978, les systèmes cryptographiques considérés comme partiellement homomorphes (comme RSA, ElGamal...), n’ont pas connu d’avancées notables pendant plus de 30 ans. Ce n’est que récemment que ces systèmes ont intégré *à la fois* les opérations de multiplication et d’addition.

Le chiffrement homomorphe : d'autres bases mathématiques

Les problèmes de chiffrements homomorphes sont un axe de recherche actif actuellement.

Les bases mathématiques en cryptographie sont de nature algébrique, et reposent sur la manipulation de concepts tels que les groupes, les anneaux, les corps, les espaces euclidiens, les espaces idéaux, les polynômes, les circuits. Ils relèvent également de la théorie de la complexité, avec les notions de fonctions calculables, de temps polynomiaux, de problèmes NP complets…Il s'agit souvent d'utiliser des fonctions f dont il est facile d'évaluer les sorties $\mathrm{f}(\mathrm{x})$, et très difficile de retrouver x en ayant seulement connaissance de $\mathrm{f}(\mathrm{x})$. Ainsi, les primitives cryptographiques utilisées dans le cadre du chiffrement RSA sont fondées sur la difficulté de factorisation des très grands nombres premiers. Celles du système cryptographique ElGamal, utilisé notamment par le logiciel PGP (*Pretty Good Privacy*) pour l'envoi sécurisé de courriers électroniques sont fondées sur le problème du logarithme discret. Ces deux systèmes ne sont que partiellement homomorphes (ne permettant qu'une seule opération sur les données chiffrées).

Les travaux de Craig Gentry reposent sur les réseaux euclidiens, mais les clés cryptographiques utilisées sont de tailles beaucoup plus importantes que les systèmes comme RSA et ElGamal, les informations chiffrées sont de tailles beaucoup plus conséquentes que les informations en clair, et les temps de calculs sur les informations chiffrées sont très grands. L'enjeu, sur ces fondements mathématiques, est d'obtenir des performances, en particulier des temps de calculs, acceptables pour une utilisation dans des solutions industrielles.

Chiffrement homomorphe complet

Boneh *et al.*[97] ont proposé en 2005 un mécanisme de chiffrement homomorphe plus élaboré, permettant d'effectuer sur les données chiffrées un nombre arbitraire d'additions, et une multiplication.

C'est en 2009 qu'un mécanisme réellement complet – dit idéal –, au nombre quelconque de calculs possibles sur les données chiffrées, est proposé par Craig Gentry [92] dans sa thèse. Ces travaux ont ouvert la porte à de nombreux autres, en passe de rendre les techniques de chiffrement homomorphe un des axes des PETs le plus en vue.

Bases d'un mécanisme de chiffrement homomorphe complet

Le mécanisme de chiffrement homomorphe inclut quatre algorithmes : KeyGen, Encrypt, Decrypt et Eval. L'algorithme KeyGen génère les paramètres publics pp et une paire de clés $(\mathrm{pk}, \mathrm{sk})$ (une publique, une secrète). L'algorithme Encrypt prend en entrée la clé publique pk et le message m et produit la sortie chiffrée CT, tandis que sa contrepartie Decrypt prend en entrée le texte chiffré CT et la clé secrète sk et produit le message m. L'algorithme Eval prend en entrée une clé d'évaluation evl (qui peut être considérée comme une partie de la clé publique pk), une fonction f de $\{0,1\}^{\mathrm{n}}$ vers $\{0,1\}$, l'ensemble des textes chiffrés $(\mathrm{CT}_1, \mathrm{CT}_2 ... \mathrm{CT}_n)$ et produit un texte chiffré. Bien représenter la fonction f est une des parties du problème et diffère selon les implémentations.

Le chiffrement homomorphe complet (*Fully Homomorphic Encryption*, FHE) présente deux catégories : le *ideal FHE*, au nombre quelconque de calculs possibles, y compris illimité, et le *leveled FHE*, pour un nombre limité de calculs, reposant sur des *circuits de taille polynomiale*. Les travaux de Gentry en 2009 d'un FHE pour un nombre arbitraire de multiplications et d'additions ont ouvert la voie à de nombreuses autres implémentations de FHE, certaines reposant sur le problème mathématique complexe des espaces idéaux et d'autres sur le problème tout aussi complexe du calcul approximatif du plus grand diviseur commun.

Le cas des services personnalisés

Pour être utiles dans un contexte de services personnalisés, les mécanismes de chiffrement homomorphe doivent avoir les deux propriétés suivantes :

- ***circuit privacy :*** propriété qui garantit que les entrées du processus exécuté côté serveur, en particulier la fonction f, restent protégées des yeux des clients. Les utilisateurs (et ceux qui seraient trop curieux) ne doivent pas avoir ou pouvoir établir de connaissances sur cette fonction,
- ***multi-hop homomorphism :*** propriété qui permet d'adapter la sortie d'un calcul homomorphe de manière à l'utiliser en entrée d'une autre fonction d'évaluation homomorphe.

Mise en œuvre concrète dans les services personnalisés

Le chiffrement homomorphe a été largement appliqué aux services de recommandation. En 2012, Erkin *et al.*[98] ont proposé un système où les données des utilisateurs (notes, éléments notés…) s'appuient sur un chiffrement homomorphe dit partiel (*Partially Homomorphic Encryption*, PHE). Le service de recommandation est ainsi capable de faire son travail sans accéder aux données des utilisateurs. La proposition d'Erkin *et al.* introduit par ailleurs une entité tierce, semi-fiable, pour effectuer des calculs supplémentaires sur les données chiffrées, qui peut apporter une aide supplémentaire au service de recommandation, ainsi qu'empêcher toute collusion entre des utilisateurs malveillants et le dit service.

En 2016, Badsha *et al.*[99] ont présenté un nouveau système de recommandation, respectueux de la vie privée, basé sur le chiffrement ElGamal[100], sans avoir recours à l'assistance d'une entité tierce. Toutes les entités impliquées doivent cependant collaborer avec le service de recommandation pour générer des recommandations privées. Les auteurs ont présenté par la suite une extension reposant sur le mécanisme de chiffrement homomorphe Boneh-Goh-Nissim[101]. Là, une partie tierce intervient pour permettre aux utilisateurs de déchiffrer les recommandations chiffrées envoyées par le fournisseur de service.

Ces techniques sont également utilisées pour les services de recherche de données sécurisée. L'utilisateur transmet sa requête chiffrée, et le service renvoie sa réponse chiffrée sans jamais avoir vu la requête en clair.

Applications

Les techniques de chiffrement homomorphe sont très utilisées dans le cadre des services de recommandation. Des données de différents types sont en effet échangées dans ces contexte, qui toutes peuvent offrir des informations de nature personnelle sur des individus différents : notes attribuées à des chauffeurs ou des loueurs de biens, ou centres d'intérêt des personnes cherchant des recommandations.
Ces techniques sont également utilisées pour la mise en place de serveurs de requêtes (bases de données) et de recherche sur le web, de manière à ce que les requêtes, et leurs résultats, soient effectuées tout en respectant la vie privée des entités impliquées.

Mécanismes de recherche d'informations privée

La recherche d'informations privée (*Private Information Retrieval*, PIR) permet à un utilisateur distant d'interroger une base de données sans révéler quelle donnée est effectivement d'intérêt pour lui. Bien noter que « privée » qualifie donc ici non pas le contenu des informations, mais l'action d'accéder à ces informations, dont la nature et le sens ne doivent être connus que de celle ou celui qui l'entreprend, et sans signification pour les autres entités impliquées ou cherchant à s'y intéresser.

Les mécanismes de PIR sont des solutions efficaces et adaptées aux grandes bases de données. Ils offrent des niveaux de vie privée élevés, et ne sont pas compatibles avec les objectifs de personnalisation de services. Ils ont fait l'objet de nombreux travaux.[102][104][105][106][107][108]

Ils sont par exemple largement utilisés dans les services de géolocalisation, quand il s'agit de trouver le point d'intérêt le plus proche de la position de l'utilisateur au moment de sa requête. Attallah et Frikken[109] ont proposé en 2004 une solution dans laquelle les données de localisation sont représentées côté serveur par un graphe acyclique orienté. Pour résoudre sa recherche de points d'intérêt, l'utilisateur doit envoyer un nombre de requêtes proportionnel à la profondeur du graphe.

3”

Recherche d'informations privée

Il s'agit là de la deuxième sous-catégorie des techniques d'***obfuscation***, relevant principalement du modèle semi-confiance, et dont la première sous-catégorie, l’obfuscation de données, a été présentée page 99. Les techniques de *Private Information Retrieval* (PIR)[110][111][112] consistent à rendre les communications (questions et réponses) inintelligibles pour cacher le sens exact des échanges entre les entités. On cherche à permettre l'accès à des données d'une base de données distante sans que soient révélées les données spécifiques auxquelles l'utilisateur porte un intérêt. Une solution triviale, et coûteuse en communications, est bien sûr de récupérer toute la base. Un juste milieu existe heureusement.

En 2011, Mittal *et al.*[103] ont proposé ***PIR-Tor***, une architecture pour le réseau Tor (voir page 54), dans laquelle les utilisateurs obtiennent de l'information de seulement quelques serveurs Tor en utilisant des techniques de type PIR, interdisant à d'éventuels observateurs de connaître les choix de relais effectués, et assurant ainsi anonymat et passage à l'échelle.

En 2014, Paulet *et al.*[113] ont proposé un nouveau cadre de services de requêtes dans des contextes de géolocalisation, assurant la confidentialité à la fois de l'utilisateur et du service de localisation. Un algorithme de transfert inconscient (voir page 104) est d'abord utilisé pour projeter la localisation de l'utilisateur dans une zone de la carte du service. Puis un algorithme PIR est employé pour obtenir tous les points d'intérêt présents dans cette zone.

Plus récemment en 2016, Ullah *et al.*[114] ont proposé ***Poshida***, un protocole PIR pour l'application du moteur de recherche web. L'idée est d'opacifier le profil utilisateur collecté et maintenu par le moteur de recherche. Plusieurs expériences ont été faites pour mesurer le niveau d'exposition de ces profils, compte-tenu des informations collectées par les moteurs de recherche. Les auteurs montrent que leur protocole est capable de cacher jusqu'à 85% du profil, et des attributs, d'un utilisateur vis-à-vis d'entités observatrices et trop curieuses.

Recherche d'informations privée

Les algorithmes PIR sont le plus souvent des mécanismes interactifs, que l'on peut classer selon leur niveau de résistance à la curiosité du fournisseur de services disposant de la base de données :

- mécanismes *Computational PIR* (cPIR) : pour lesquels le fournisseur de service est supposé disposer d'un nombre limité de ressources (puissance de calcul, temps raisonnable…) pour extraire des échanges les informations utiles pour lui.
- mécanismes *Information theoretic PIR* (itPIR) : pour lesquels le fournisseur de services est beaucoup plus puissant car doté de ressources en calcul théoriquement infinies

Un mécanisme cPIR est donc moins robuste qu'un mécanisme itPIR. Les mécanismes PIR peuvent également être classés selon le nombre de serveurs impliqués. Sont ainsi distingués les *single-server PIR* et les *multi-server PIR*.[115]

Mécanismes de recherche chiffrée

_Scénarios mono- et multi-propriétaires des données

Deux grandes catégories de recherche chiffrée (*Searchable Encryption*) doivent être distinguées : les scénarios avec plusieurs utilisateurs et un unique propriétaire (des contenus faisant l'objet de requêtes), et ceux avec plusieurs utilisateurs et plusieurs propriétaires. Dans le premier cas, le propriétaire des contenus chiffrés gère seul les capacités de recherche qui sont possibles à leur égard[116][117][118][119]. Dans le second cas, les multiples propriétaires de contenus chiffrés les destinent à des serveurs distants, sur lesquels les utilisateurs disposent de capacités de recherche[120][121][122].

Certains des scénarios à propriétaire unique nécessitent d'accorder des droits d'accès aux utilisateurs à travers une clé secrète partagée [117][118], tandis que d'autres fournissent ces accès sous la forme de portes dérobées. Plus récemment, le chiffrement à base d'attributs (*Attribute-based Encryption*, ABE) a été employé dans le cas multi-propriétaires pour construire des mécanismes de recherche sur des mots-clé autorisés. Pathak et al. [120] utilisent des politiques d'accès en clair, ce qui risque de révéler des informations sensibles sur la nature des accès et des types de recherche. Un serveur trop curieux qui effectue des recherches pourrait ainsi produire des statistiques d'accès et rompre la confidentialité recherchée.

Recherche d'informations privée

Les mécanismes SE (*Searcheable Encryption Schemes*) permettent quant à eux d'effectuer des recherches sur des données chiffrées distantes, sans jamais révéler de mots-clé en clair. Ils doivent remplir les obligations suivantes :

- ***keyword secrecy :*** aucune entité non autorisée ne doit avoir accès aux mots-clé des requêtes
- ***search pattern secrecy :*** l'information comme quoi tel document contient, ou ne contient pas, tel mot-clé ne doit pas être révélée
- ***access pattern secrecy :*** les résultats de la requête ne doivent pas être connus

_Comment déterminer la similarité entre données ?

Parmi les défis de la recherche d'informations privée se trouvent la recherche à mots-clé multiples, et la présentation des résultats selon un classement pertinent. Il s'agit notamment d'un calcul de similarité. Plusieurs méthodes sont disponibles dans la littérature scientifique. En 2014, Cao *et al.*[123] ont présenté une méthode de recherche multi-mots-clé avec résultats classés, sur des données chiffrées dans le *cloud*.

Pour déterminer la correspondance entre les données des documents cherchés et les mots-clé d'une requête, en prenant en compte les différentes sémantiques possibles (cf. page 8 l'exemple avec « souris »), une mesure de similarité appelée *coordinate matching* a été développée dans Witten *et al.*[124].

De leur côté, Cao *et al.* utilisent en particulier le produit interne de deux vecteurs pertinents pour ces calculs de similarité. Lors de la phase d'indexation des documents, chaque document est associé à un vecteur de bits, où chaque bit évalue quantitativement le nombre d'occurrences du mot clé dans le document. La requête des utilisateurs est elle-même représentée par un vecteurs de bits, les bits activés correspondant à la présence de tel mot-clé dans la requête. Ainsi, le produit interne de ces deux vecteurs permet de mesurer la similarité entre la requête et le document candidat.

Bien entendu, il n'est pas concevable que ces deux vecteurs circulent sans protection, l'un relevant de la confidentialité de l'indexation potentielle des documents, et l'autre de la confidentialité de la recherche effectuée par l'utilisateur. Cao *et al.* adoptent un calcul sécurisé de ce produit interne, adapté de techniques connues et sécurisées de calcul des k plus proches voisins (*k-Nearest Neighbor*, kNN). Voir également Wong *et al.*[125]

Mécanismes d'auto-destruction d'accès aux données

Ces mécanismes sont particulièrement adaptés dans des contextes d'applications web où les données sensibles des utilisateurs peuvent rester indéfiniment sur les serveurs. Ce type de technologie PET redonne du contrôle aux utilisateurs, par exemple sur leurs messages privés, leurs documents, leurs photographies…

Vanish est une solution de ce type construite sur la base de techniques de chiffrement, des infrastructures pair-à-pair et des *tables de hachage distribuées*. Vanish[126] chiffre les données des utilisateurs localement à l'aide d'une clé secrète générée au hasard. Cette clé est conservée juste le temps de générer des parts de secrets (*secret share*), puis détruite, tandis que les parts sont stockées sur des nœuds tirés au hasard dans la table de hachage. Ces derniers détruisent les parts après un certain temps, rendant les données associées inaccessibles de manière permanente.

Construit sur Vanish, une extension Firefox est disponible pour auto-effacer des mails de Gmail, tandis qu'une autre extension du navigateur est proposée pour intégrer des éléments auto-destructibles sur des sites web. Ces mécanismes ne résistent cependant pas aux attaques Sybil[127], qui compromettent le système en collectant les parts de secrets des tables de hachage avant leur péremption. Les versions plus modernes de Vanish (Vanish 0.1, Vanish 0.2) sont plus résistantes[128] à ce type d'attaques.

5

Protection contre la fuite de données

Parmi les nombreuses révélations de ces dernières années montrant que des données personnelles n'ont pas été assez protégées par les fournisseurs de service, celle de début 2018 sur les données collectées par FedEx a fait date. Les données de passeport, de permis de conduire, sous forme de documents scannés, ainsi que les adresses, codes postaux, numéros de téléphone de dizaines de milliers de clients étaient ainsi exposées sans protection dans le service de stockage Amazon S3 bucket. Datant de la période 2009-2012, ces données étaient visibles de tous depuis des années.

Ephemerizer est un protocole cryptographique proposé par Perlman[129] en 2005, qui est conçu pour assurer la destruction d'une donnée après une date d'expiration. L'*ephemerizer* en question est une entité responsable de la génération de clés temporaires et de la destruction des clés expirées. Les données sont chiffrées grâce aux clés temporaires et ne sont donc plus accessibles une fois ces dernières détruites.

Proposé en 2016, ***Neuralyzer*** introduit un mécanisme de destruction de données autour du concept de temps d'expiration flexible. Plutôt que d'être relié à un temps de validité fixe et pré-défini, la vie d'une donnée est reliée à un schéma de révocation distinguant plus de possibilités. Par exemple, une donnée peut être rendue inaccessible car elle a perdu tout intérêt, ou bien parce qu'elle a été utilisée un trop grand nombre de fois.

Pour réaliser ses objectifs, *Neuralyzer*[130] repose sur les mécanismes de cache du *DNS*. Les clés de déchiffrement des données sont divisées en plusieurs morceaux répartis sur différentes *entrées DNS*. Chaque clé est récupérable par toute entité ayant connaissance des entrées DNS hébergeant les morceaux. Chaque accès aux données étend automatiquement leur date de péremption, par définition du fonctionnement du cache DNS. En revanche, si les caches ne sont pas visités à temps, ils sont supprimés, et les clés de déchiffrement qu'ils abritaient avec.

Rendre les données « auto-destructibles »

Pour limiter l'impact de la fuite et de la publication de ses données personnelles, une solution consiste à rendre ces données auto-destructibles au bout d'un certain temps.

En réalité, c'est la méthode d'accès à ces données qui est auto-destructible, les données chiffrées n'étant pas touchées. Devenues inutilisables, ces données perdent leur intérêt pour le fournisseur de service, qui n'a plus qu'à les détruire.

Que cherche-t-on à anonymiser ?

Les techniques d'anonymisation sont de plus en plus pertinentes pour les industriels car ce sont les seules techniques permettant de faire sortir des données du champs d'application du RGPD. Par « techniques d'anonymisation », il est le plus souvent fait référence à la préservation de la confidentialité dans des contextes de bases de données. Cette confidentialité peut prendre deux formes :

- ***privacy of respondents :*** dans le cas général, les bases de données contenant souvent des informations statistiques sur des personnes, la confidentialité associée à ces données qui sont liées à des personnes est essentielle
- ***data owner privacy :*** dans un cadre collaboratif entre organisations, il est important que chaque responsable de traitement garde confidentielles ses propres bases de données.

Pour les applications pervasives, par exemple celles traitant des données de santé ou de bien-être, les deux qualités précédentes sont requises : les patients veulent que leur vie privée soit respectée, et les enregistrement médicaux doivent être protégés.

L'anonymisation n'est pas toujours suffisante. Il est parfois possible d'identifier des données personnelles en croisant plusieurs bases de données, qui pourtant, prises séparément, sont inoffensives et correctement anonymisées.

Gérer proprement les données statistiques

Les techniques de contrôle de la divulgation de données statistiques (*Statistical Disclosure Control, SDC*) servent à protéger les données dans les bases de données utilisées à des fins d'analyses statistiques (ou bases de données statistiques). Elles résolvent le compromis entre l'utilité de ces données et la préservation de leur caractère privé. Elles permettent la publication et l'accès à ces données sans qu'il soit possible d'identifier l'utilisateur qui en est à l'origine.

Ces techniques regroupent les mécanismes d'***anonymisation des bases de données***, comprenant également les mécanismes dits de ***confidentialité différentielle***.

La popularité grandissante des services utilisant la géolocalisation – sur mobile comme sur base fixe – augmente les risques de divulgation ou de fuite de données personnelles. Ceci concerne à la fois les individus utilisant ces services, ceux qui les accompagnent ou ceux qu'ils croisent. Ces services vont des outils classiques de cartographie à la recherche de points d'intérêt, de connaissances communes dans les alentours, et celles de bons plans commerciaux.

Les fournisseurs de tels services ne se contentent pas de prendre en compte la localisation au moment d'une requête. Ils peuvent également enregistrer les déplacements, les trajets, les horaires, et en déduire les modes de transport, les habitudes et des informations encore plus personnelles dès que les bâtiments et lieux fréquentés sont identifiés par ailleurs. Ces données collectées sont utilisées pour améliorer les services rendus en général – service marketing basé sur la localisation, publicité ciblée, service d'authentification, profilage comportemental, réseaux de transports – grâce aux informations statistiques collectées sur une masse d'individus ou à l'échelle d'un individu. Elles peuvent également être revendues à des tierces parties, sans que les utilisateurs ne le sachent.

Toutes ces données peuvent permettre à un attaquant de connaître précisément la localisation d'un individu, et donc l'espionner ou porter préjudice à ses biens en son absence. Dans certains cas, les traces peuvent devenir des quasi-identifiants, permettant de remonter jusqu'à l'utilisateur précis, en particulier si les traces ont seulement été pseudonymisées. C'est pourquoi les techniques d'anonymisation, à commencer par le k-anonymat exposé à présent, sont de première importance en contexte de géolocalisation.

k-anonymat, l-diversité et t-proximité

Les principales techniques[131] pour anonymiser les bases de données en mode non interactif sont le k-anonymat, la l-diversité et la t-proximité (*k-anonymity*, *l-diversity*, *t-closeness*). Noter que ces techniques, développées à l'origine pour des bases de données statiques, ont été étendues depuis pour des contextes dynamiques.

Souffrant de plusieurs limitations, et notamment des attaques divulguant les attributs, le k-anonymat a été amélioré en 2006 par la notion de l-diversité, et cette dernière par la t-proximité.

Pour définir la notion de k-anonymat, il faut tout d'abord distinguer trois types d'attributs pour les données d'un ensemble $\mathcal{S}$:

- ***identifiants :*** attributs qui permettent d'identifier très exactement un individu, comme par exemple un numéro de sécurité sociale, ou un identifiant fiscal. Ce type d'attribut est généralement laissé de côté.
- ***attributs clés :*** attributs qui peuvent, une fois combinés avec des informations externes à $\mathcal{S}$, permettre de ré-identifier les individus. L'âge ou le code postal en sont des exemples. Contrairement aux précédents, il est nécessaire de conserver un certain nombre de ces attributs clés dans $\mathcal{S}$.
- ***attributs sensibles :*** attributs dont les valeurs peuvent être hautement intéressantes (religion, salaire…) pour des entités externes attaquantes, et qui doivent donc être hautement protégées.

k-anonymat

Proposé en 1998 par Samarati et Sweeney, le k-anonymat est une méthode d'anonymisation qui repose sur une *approche par généralisation et suppression*.

Afin d'éviter la réidentification d'un individu porteur d'un ensemble d'attributs ayant des valeurs trop précises (par exemple une adresse postale), soit on supprime certains éléments trop identifiants, soit on crée une dilution regroupant au moins k individus (en passant dans cet exemple à l'échelle de la commune), l'objectif étant de rendre les k individus indiscernables.

_Des limites du k-anonymat

Le k-anonymat est un des modèles possibles d'anonymisation de données. Il a cependant plusieurs limites[132], et notamment son manque de résistance aux attaques de divulgation des attributs.

Conçues initialement pour des bases de données statiques, les techniques k-anonymat ont été étendues à des systèmes plus dynamiques et dépendant d'un contexte. C'est le cas des services de géolocalisation, où un attaquant peut identifier un utilisateur en examinant ses paramètres spatio-temporels. Élargir le périmètre correspondant à la localisation précise du demandeur permet de le fondre dans la masse, mais rend beaucoup moins pertinentes les réponses du fournisseur de service en dégradant ainsi la qualité du service. Shin *et al.* proposent ainsi de choisir parmi plusieurs profils d'utilisateur, correspondant à des niveaux de confidentialité prédéfinis différents.

D'autres services ne reposent pas sur des informations de localisation ou d'autres données contextualisées, comme les activités, les habitudes, les centres d'intérêt. Cependant, comme l'ont démontré Riboni *et al.*[133] en 2008, un attaquant peut identifier un utilisateur protégé par du k-anonymat, en observant le comportement d'utilisateurs potentiels en regard des réponses du service. Les auteurs proposent la mise en place d'une entité intermédiaire de confiance, chargée de calculer les niveaux de violation possible de la confidentialité (sur la base de métriques données) et d'avertir l'utilisateur avant qu'il ne se connecte au service.

De manière intuitive, assurer l'anonymat dans un ensemble de données nécessite d'identifier les attributs clés, qui sont des quasi identifiants. Un enregistrement présent dans l'ensemble k-anonymisé $\mathcal{S}_k$ ne peut pas être relié à son original dans $\mathcal{S}$ puisque au moins k-1 autres enregistrements ont la même valeur dans $\mathcal{S}_k$.

Le k-anonymat offre un premier compromis entre la perte d'informations intéressantes pour l'analyse de données et le risque de divulgation d'identité. Malheureusement, il n'est pas résistant à l'attaque par divulgation d'attributs.

_Des limites de la l-diversité

Si la l-diversité est plus résistante que le k-anonymat aux risques de révélation des attributs, elle reste vulnérable à ce type d'attaques dans les cas où les valeurs des attributs sensibles rendus l-diversifiés sont toujours sémantiquement similaires.

Par exemple, en considérant l'exemple où des données de patients ont été rendues 3-diversifiées sur l'attribut sensible `Maladie` en le diluant dans le 3-uplet `{cancer du sein, cancer du poumon, cancer du foie}`, un attaquant qui aurait la connaissance qu'un individu donné appartient à ce groupe pourrait en déduire qu'il a un cancer.

De même, si l'attribut est une valeur numérique, et que la l-diversification se fait sur un ensemble de valeurs trop proches, l'attaquant peut toujours faire une bonne estimation de la valeur sensible d'origine.

La t-proximité, introduite plus loin, apporte des solutions à ces limitations.

l-diversité

En situation de k-anonymat, il est possible de déduire qu'un individu possède un attribut sensible particulier si tous les individus d'une même classe possèdent un même attribut sensible (cette classe est appelée *classe d'équivalence*). Pour pallier ce problème, une première idée est d'ajouter une contrainte : le champ sensible de la classe d'équivalence doit prendre au moins l valeurs distinctes.

Bien qu'améliorant le procédé d'anonymisation des données et le niveau de protection des individus, il reste des cas où cette protection n'est pas suffisante, notamment quand les l valeurs distinctes restent dans un champ sémantique trop resserré.

_l-diversité en contexte de géolocalisation

Les techniques de l-diversité ont été utilisées dans de nombreux contextes en géolocalisation. En 2009 Liu *et al.*[134] ont proposé que le contenu des requêtes de localisation soit suffisamment différent au sein d'une même zone regroupant plusieurs utilisateurs (zone dite masquée, *cloaked zone*, construite dynamiquement avant toute requête). Ce faisant, il est impossible de remonter d'une requête à son utilisateur source, avec une probabilité inférieure à un seuil prédéfini.

De leur côté, Bamba *et al.*[135] présentent ***PrivacyGrid***, une solution pour applications mobiles. Dans ce cadre, des profils de préférences de confidentialité de la localisation sont définis. Ils permettent aux utilisateurs mobiles de choisir explicitement leurs paramétrages, que ce soit pour les mesures de localisation elles-même (via du k-anonymat ou de la l-diversité), ou pour les mesures de qualité de service. Pour ce faire, les utilisateurs mobiles communiquent avec un certain nombre de *serveurs LBS* à travers un serveur proxy procédant à l'anonymisation des données de localisation. Ce dernier intercepte toutes les données de localisation émises par les utilisateurs, et les l-diversifie en utilisant des données géolocalisées publiques disponibles par ailleurs.

Toujours en 2009, Xue *et al.*[136] ont formalisé le concept de localisation diversifiée. Il s'agit toujours d'améliorer la mise en place d'un k-anonymat en associant à chaque requête au moins l lieux à la sémantique éloignée : parking, commerce, université… Les auteurs proposent un algorithme permettant de construire de tels groupes de l- localisations variées sémantiquement.

Plus récemment, en 2015, Wang *et al.*[137] ont proposé un nouveau schéma de protection des attributs sensibles de localisation. La solution permet de générer une zone couvrante sur une aire continue de LBS, avec des serveurs d'anonymisation qui effectuent leur travail à l'aide de techniques de k-anonymat et de l-diversité.

En 2017, Ye *et al.*[138] définisent un nouvel algorithme de l-diversité, utilisant la connaissance des réseaux routiers pour améliorer la confidentialité des trajets des utilisateurs. L'algorithme propose ainsi un ensemble de trajets alternatifs possibles pour cacher le trajet effectif. Chaque localisation renvoyée à un serveur LBS correspond à une zone couvrante qui regroupe au moins l autres trajets, ces derniers ayant été générés à l'avance.

Le k-anonymat est habituellement utilisé en conjonction avec la t-proximité[139] pour améliorer la confidentialité des requêtes. Riboni *et al.* (auteurs cités plus haut pour leur publication de 2008 sur le k-anonymat) ont ainsi proposé dans une publication de 2009 une technique reposant sur la généralisation des informations spatio-temporelles, dans le cadre d'un service de requêtes sur des localisations. Le constat initial était que les utilisateurs avaient tendance à effectuer des requêtes simultanément, ou avaient tendance à répéter les mêmes requêtes, rendant des attaques pour lever leur confidentialité plus faciles avec ces particularités en tête. En utilisant ces deux techniques d'anonymisation des données pour les rendre moins parlantes, les auteurs limitent l'effet de ce type d'attaques.

_Des limites de la t-proximité

Bien que la t-proximité améliore véritablement la résistance aux attaques sur la divulgation des attributs, ce pour quoi elle a été conçue, sa mise en place diminue particulièrement l'utilité des données concernées, qui se trouvent trop obfusquées ou trop généralisées pour permettre des traitements ayant du sens.

Par ailleurs, et comme le relèvent Sondeck, Laurent et Frey[140] en 2017, ce n'est pas tant la valeur elle-même du fameux t qui est importante, c'est la classification des valeurs des attributs qui nécessite le plus de soin, et il s'agit d'une tâche qui fait appel à la subjectivité de celui qui l'entreprend.

t-proximité

Introduite en 2007 par Li *et al.*, la t-proximité pousse encore plus loin que le k-anonymat ou la l-diversité les protections contre les attaques consistant à déduire des valeurs d'attributs pour un individu donné. Le facteur t représente cette fois une mesure de l'écart entre la distribution d'un attribut sensible dans la classe d'équivalence et la distribution de l'attribut dans la base de données originale.

Un ensemble de données satisfait la t-proximité si, pour chaque classe d'équivalence, la distribution d'un attribut sensible n'est pas éloignée d'un seuil t de la distribution de l'attribut dans la base de données originale.

Mathématiques de la confidentialité différentielle

Une fonction probabiliste $\mathcal{F}$ assure la confidentialité différentielle de paramètre ε (ε-différentiellement confidentielle) si pour tout jeu de donnée $\mathcal{D}_1$ et $\mathcal{D}_2$ qui diffèrent d'au plus un élément, et pour tout sous-ensemble $\mathcal{S}$ de l'image de $\mathcal{F}$,

$$\Pr[\mathcal{F}(\mathcal{D}_1) \in \mathcal{S}] \leq exp(\varepsilon) \times \Pr[\mathcal{F}(\mathcal{D}_2) \in \mathcal{S}]$$

Le paramètre public ε n'est pas fixé, et doit être déterminé par l'entité qui réalise l'anonymisation en fonction de ses besoins de protection et du contexte d'utilisation. En effet, plus ε est petit, plus l'algorithme est anonymisant, mais plus la perte d'information est grande.

Cette construction mathématique est indépendante de toute connaissance auxiliaire qu'un attaquant, ou un utilisateur, pourrait posséder par ailleurs sur la base de données considérée. Elle s'étend également à la confidentialité de groupes d'utilisateurs, ainsi que pour le cas où un même utilisateur contribue plusieurs fois au contenu de la base de données. Cela est particulièrement intéressant quand on considère qu'un groupe d'utilisateurs (et l'équivalent pour un utilisateur aux multiples contributions) peut estimer que leurs données groupées sont susceptibles de créer des fuites d'informations, alors que chaque donnée isolée ne le fait pas.

Confidentialité différentielle

Les méthodes précédentes, de k-anonymat à t-proximité, ne donnent pas de garanties formelles (preuves mathématiques) de leur efficacité à limiter les informations que l'on peut apprendre sur un utilisateur. Une technique plus récente, dite de confidentialité différentielle, est très en vogue depuis 2006 (travaux de Dwork *et al.*).

Elle consiste à nouveau à introduire de l'aléa dans les données sous la forme d'un bruit de Laplace plus ou moins important en fonction du niveau de confidentialité attendu et du niveau d'utilité attendu de ces données.

La littérature scientifique relate plusieurs exemples de cas d'usage de la confidentialité différentielle. En 2014, Riboni *et al.*[141] proposent un tel mécanisme dans le cadre de la mise à disposition de données de localisation à une entité de recommandation non fiable. Un intermédiaire de confiance est chargé de collecter les données de localisation et de leur appliquer la confidentialité différentielle avant de les porter à la connaissance de l'entité non fiable. Cette approche permet à la fois de se protéger de l'entité non fiable et des actions malveillantes d'un éventuel tiers forgeant de fausses requêtes de manière à déduire les endroits visitées par l'usager d'origine. Dans la même veine, Chen *et al.*[142] considèrent le problème de la publication de données utiles de la structure d'un réseau social, sans qu'il soit possible à un attaquant de déduire des liens particuliers entre individus, même dans le cas de données fortement corrélées.

En 2017, To *et al.*[143] appliquent ces idées à la protection des localisations d'individus, connues de leurs opérateurs télécom (par exemple par triangulation dans le cas des mobiles), en les bruitant avant de les mettre à disposition de services comme des applications de *crowdsourcing* spatiales. Asghar *et al.*[144] appliquent de leur côté ces techniques dans des contextes de publication des usages d'un réseau de transport. Ce faisant, ils conservent la confidentialité des trajets, et habitudes de trajets, des individus participant à l'étude des usages, tout en permettant l'extraction et l'utilisation de données permettant d'imaginer de futurs services de transport à haute valeur ajoutée.

Confidentialité différentielle locale

Quand la confidentialité différentielle est appliquée directement par l'utilisateur source des données, elle prend le nom de confidentialité différentielle locale[145] (*Local Differential Privacy*, LDP). Dans ce cas, c'est l'utilisateur qui obfusque lui-même ses données. Ce type d'approche a fait l'objet d'applications utilisées largement par des acteurs majeurs. ***RAPPOR***, par Google, permet d'identifier les sites web populaires sans rien révéler de leurs visiteurs. Apple a indiqué lors de sa conférence annuelle des développeurs en 2016 employer ces techniques et ne pas construire de profils utilisateurs. Microsoft déploie également cette approche dans son outil ***Telemetry***.


3

Les techniques orientées canal

Les deux dernières familles de technologies sont celles qui sont mises à disposition pour protéger la confidentialité des données circulant sur un canal de communication – entre utilisateurs ou entre utilisateurs et serveurs.

Elles peuvent nécessiter l'intervention d'une entité médiatrice, ou le chiffrement des communications.

Commençons par la sécurisation des communications et terminons par l'intervention des tiers de confiance.

Où passent les données, au fond ?

Plus de 430 câbles sous-marin parcourent la planète pour véhiculer Internet d'un continent à l'autre, d'un rivage à l'autre. Des coupures involontaires ont montré par le passé que des pays pouvaient se retrouver isolés plusieurs jours. Quand une telle liaison est rompue, c'est l'accès aux données personnelles hébergées sur les serveurs *cloud* qui n'est plus possible, rendant inopérant de nombreux services. C'est pourquoi certains pays ont adopté une politique de localisation de données sur leur propre territoire. Les grands acteurs de l'Internet investissent également dans la pose des câbles, et toute cette infrastructure échappe peu à peu aux États souverains.

7

Des communications sécurisées

Les liens physiques de communication ne permettent pas une garantie de confidentialité suffisante, et il est impossible de les surveiller physiquement de manière permanente. Non seulement le transfert de données privé entre utilisateurs (et via des serveurs) doit être protégé sur cette partie de la communication, mais même l'accès à des ressources publiques doit se faire de manière protégée pour empêcher un observateur de capturer des identifiants ou de construire un profil de l'utilisateur, ce qui pourrait être utilisé par la suite pour pister les utilisateurs.
Deux techniques sont proposées aux utilisateurs : le chiffrement client-service et le chiffrement de bout en bout.

_TLS, SSH, IPsec, des technologies du quotidien

TLS très répandu auprès du grand public et ses versions antérieurs connues sous l'acronyme SSL, SSH surtout connu des administrateurs de réseaux informatiques et IPsec (dans l'un de ses modes de fonctionnement) reposent sur des technologies de cryptographie à clé publique. Les utilisateurs et les serveurs impliqués peuvent donc créer le canal sécurisé sans avoir à partager de secrets.

TLS permet de sécuriser la grande majorité des échanges sur Internet. Il est construit sur une infrastructure de certificats à clés publiques, permettant d'authentifier le serveur. En cas de compromission de l'autorité génératrice de ces certificats, c'est l'ensemble des canaux de communication bénéficiant de sa protection qui sont compromis.

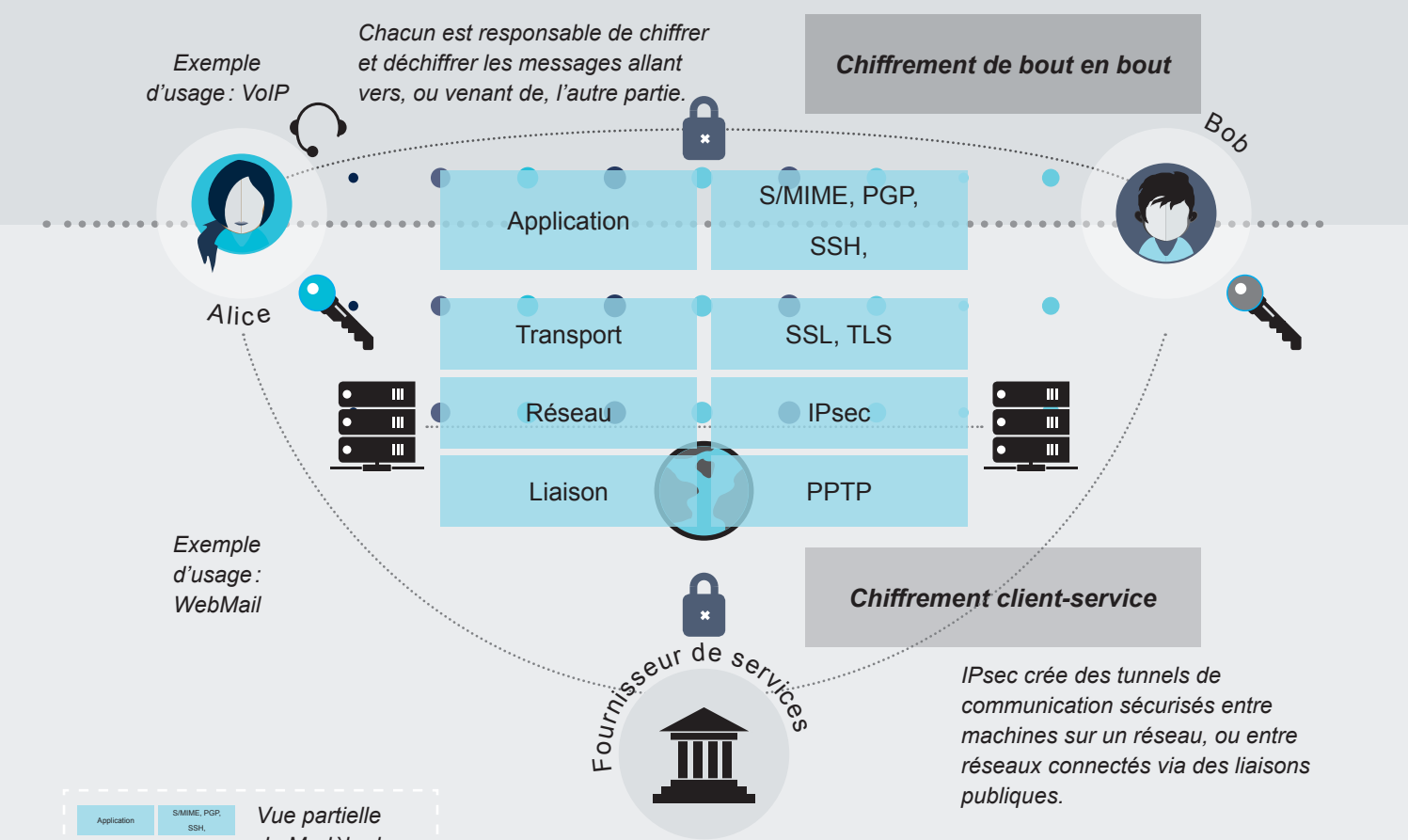
Le protocole IPsec (*Internet Protocol Security*)[146] permet de créer des communication sécurisées entre machines sur un réseau, ou des tunnels de communications sécurisées entre réseaux connectés via des liaisons publiques, plus couramment appelés VPN (voir page 50).

Notez également que Tor (voir page 54) assure un chiffrement en oignon avec des communications chiffrées entre le client et un serveur, en plus de garantir une bonne protection contre le pistage par des fournisseurs de services.

Chiffrement client-service basique

Afin de préserver les communications de la surveillance permanente, des fournisseurs de services proposent des canaux de communication chiffrés. Pour assurer un niveau de sécurité suffisant, il faut bien sûr que ces canaux soient implémentés et configurés correctement.

Deux technologies de cryptographie à clé publique, ***Transport Layer Security*** (TLS)[147][148] et ***Secure Shell*** (SSH)[149] fournissent les niveaux de confidentialité et d'authentification suffisants entre un utilisateur et un serveur. TLS, ou équivalent, est même considéré comme recommandé pour toute interaction sur le réseau.



Application	S/MIME, PGP, SSH,
Transport	SSL, TLS
Réseau	IPsec
Liaison	PPTP

Vue partielle du Modèle de protocoles en couches OSI

IPsec crée des tunnels de communication sécurisés entre machines sur un réseau, ou entre réseaux connectés via des liaisons publiques.

TLS et SSH reposent tous deux sur des technologies de cryptographie à clé publique pour l'établissement de la session.

des communications sécurisées

Fig. 14

Chiffrement de bout en bout

Il s'agit ici de chiffrer l'ensemble de la communication entre deux utilisateurs finaux, chacun d'entre eux étant responsable de chiffrer et déchiffrer les messages allant vers, ou venant de, l'autre partie. Aucun intermédiaire, fournisseur de service ou tierce partie, ne doit être en mesure de déchiffrer les messages échangés.

Les services concernés, qui émergent ces dernières années, vont de la voix sur IP (*Voice-over-Internet-Protocol*, VoIP) à la messagerie électronique, en passant par la messagerie instantanée et les réseaux sociaux.

PreVeil, une technologie de courrier électronique chiffré de bout en bout a été présentée par Popa *et al.*[150] et a fait l'objet d'un brevet en 2018. Elle repose sur la cryptographie à clé publique. Chaque utilisateur dispose d'une paire de clés, l'une publique et l'autre privée. L'utilisateur émetteur chiffre son message avec la clé publique de l'utilisateur récepteur, qui sera ainsi le seul à pouvoir le déchiffrer. PreVeil permet aux utilisateurs d'accéder à leurs mails à partir de plusieurs appareils, en assurant le transfert sécurisé de leur clé privée vers ces appareils, et offre également un service de recouvrement sécurisé de clé privée grâce à des groupes d'utilisateurs de confiance (*approval group*). Dans le même esprit, ***CryptoPhone*** et ***Signal*** ont été proposés pour fournir des communications chiffrées de bout en bout.

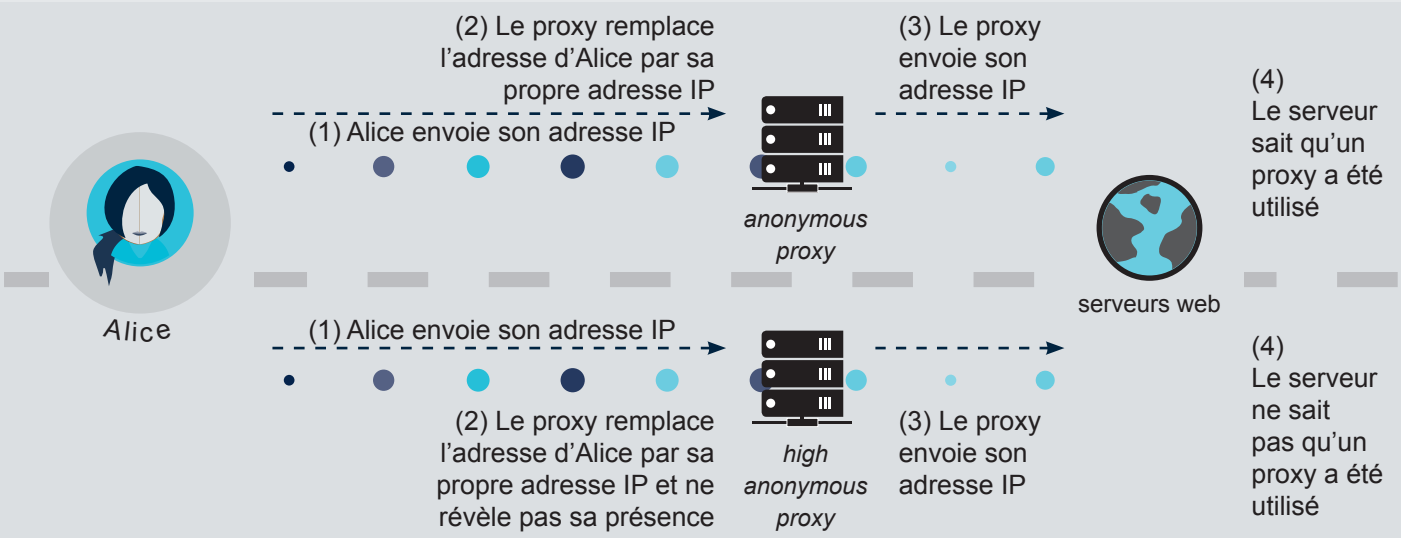
Ermoshina *et al.*[151] ont publié en 2016 un panorama de 30 protocoles de messagerie chiffrée de bout en bout, sous l'angle des garanties en matière de vie privée et de sécurité. Les auteurs observent que, malgré les promesses, ces services collectent et conservent de nombreuses données personnelles, le justifiant a posteriori au nom d'une meilleure expérience utilisateur rendue possible grâce à des données à usage statistique. Un autre paradoxe est celui du récit des mauvais usages (par des terroristes, par exemple) de ces messageries, récit qui percute celui de l'émancipation et de la protection des droits et libertés individuelles. Il s'agit également là d'un débat sur les différences entre les solutions décentralisées et pair-à-pair et ces solutions nouvelles qui offrent certes de l'*empowerment* mais parfois hors d'un cadre légal, le tout dans un contexte plus large de gouvernance par les infrastructures.

Les fournisseurs de service aident les utilisateurs à s'authentifier pour établir ce canal chiffré de bout en bout. Pour autant, les clés permettant d'assurer ensuite la confidentialité et l'intégrité des données ne doivent pas leur être confiées, et doivent être utilisées sur l'appareil de l'utilisateur. Cependant, un minimum d'informations doit tout de même leur être accessible, pour qu'ils puissent router correctement les données ou offrir des services à valeur ajoutée et améliorer ainsi l'expérience utilisateur.
Si ***Pretty Good Privacy*** (PGP) et S/MIME[152] sont devenues des solutions standard pour des échanges protégés de messagerie de bout en bout, d'autres solutions ont émergé plus récemment, avec un discours axé également sur la protection des libertés individuelles.

Les serveurs proxy, utilisés comme anonymiseurs, sont des techniques de type TTP très connues. Il s'agit d'entités intermédiaires placées entre deux entités communicant entre elles. Deux types de proxy sont à considérer :

- les *anonymous proxy,* qui cachent l'adresse IP de l'utilisateur en la remplaçant par la leur. Cependant, les autres entités peuvent identifier qu'un proxy est dans la boucle.
- les *high anonymous proxy*, qui fournissent le même service, tout en ne révélant pas qu'ils sont là.

Noter qu'aucun service personnalisé ne peut être offert puisque les serveurs proxy relaient toutes les informations qui pourraient être utiles à cela comme si elles venaient d'un unique utilisateur. Le classique problème du pistage de l'utilisateur (voir page 40) est déplacé des serveurs web au proxy qui est en effet le seul à pouvoir relier un pseudonyme, des requêtes et les réponses.



les différents serveurs proxy

Fig. 15

8

Techniques à tiers de confiance

L'utilisation d'identifiants uniques à travers plusieurs systèmes pose un problème critique : ils facilitent la collecte de données, leur corrélation et leur combinaison pour en déduire de nouvelles informations, sans nécessiter le consentement des utilisateurs les produisant. Toute brèche de sécurité révèle alors des données personnelles parfaitement identifiables et qui peuvent être reliées à leur propriétaire.

Pour contrer ce problème, plusieurs mécanismes construits sur l'anonymat et le pseudonymat ont émergé. Relevant du modèle de confiance *trusted*, il s'agit des techniques dites *Trusted Third Party* (TTP).

Systèmes à pseudonymes

Les systèmes à pseudonymes s'appuient sur des pseudonymes dépendant du contexte et sans possibilité de les relier entre eux. Cependant, il est toujours possible, au cas par cas[153], avec l'intervention de l'entreprise mettant en œuvre le service de proxy, de retrouver un utilisateur particulier associé à un ou plusieurs pseudonymes, et ayant émis un ensemble de jeux de données.

Un système à pseudonymes permet à la fois l'*auditabilité* et l'assurance que la confidentialité sur les données est préservée. Chaque utilisateur peut créer différents pseudonymes, à partir d'une clé privée et d'une chaîne de caractères arbitraire dépendant du contexte, ce qui lui permet de prouver qu'il connaît le secret ayant permis de créer son pseudonyme.

Le système doit remplir trois exigences de sécurité principales :

- ***extractibilité des clés:*** par laquelle les fournisseurs de service ont la garantie qu'un utilisateur honnête connaît effectivement la clé secrète utilisée pour créer le pseudonyme,
- ***résistance aux collisions :*** par laquelle il est garanti que la probabilité que deux utilisateurs se retrouvent avec le même pseudonyme dans un contexte donné est extrêmement faible,
- ***inassociabilité :*** par laquelle il est garanti qu'il n'est pas possible de lier des utilisateurs ou de les identifier par le biais de leurs pseudonymes, et ce quelque soit le ou les contextes considéré(s).

Dans le cas des techniques à base de pseudonymes, les fournisseurs de service ne peuvent pas déduire l'identité de l'utilisateur, mais seulement celle de la tierce partie impliquée dans la communication. Mais cette entité de confiance, centrale dans le dispositif, peut facilement associer les communications aux utilisateurs, et déduire leurs activités de l'étude de ces échanges.

En 2015[154], un mécanisme permettant une conversion aveugle entre utilisateur et pseudonyme a été proposé, assurant ainsi l'anonymat et l'inassociabilité. Ce système a été amélioré en 2017[155] avec un convertisseur dit inconscient (*oblivious*) qui offre la propriété d'audit, un point essentiel dans un contexte juridique ou commercial.

_Chiffrement polymorphe

Verheul *et al.* ont proposé en 2016 d'utiliser une approche dite *Polymorphic Encryption and Pseudonymisation* (PEP) dans un contexte d'applications de santé. Les utilisateurs envoient leurs données au serveur avec un *chiffrement polymorphe*, et, quand cela est nécessaire, peut permettre à telle ou telle entité le déchiffrement des seules données partielles utiles, gardant ainsi le contrôle du qui, où et pourquoi. Le compromis entre personnalisation et vie privée est résolu par le système PEP qui permet aux utilisateurs de choisir automatiquement leur pseudonyme selon le contexte. Ces pseudonymes ne peuvent être levés que par des entités (par exemple, les médecins) qui connaissent de toute façon les identités réelles. Plusieurs expérimentations sont conduites actuellement au sein d'un projet de recherche avec des patients réels à Nijmegen aux Pays-Bas.

	conversion aveugle	multi-convertor	audit utilisateur	audit public
Wouters et al., **Secure and privacy friendly logging for egovernment services.** 2008	non	non	oui	oui
Pulls et al., **Distributed privacy-preserving transparency logging.** 2013	non	non	oui	oui
Camenisch et al., **(un) linkable pseudonyms for governmental databases.** 2015	oui	non	non	non
Camesnisch et al., **Privacy-preserving user-auditable pseudonym systems.** 2017	oui	non	oui	non
Kaâniche & Laurent, **Blockchain-based data usage auditing.** 2018	oui	oui	oui	oui

comparaison entre systèmes pseudonymes

Tab. 5

_Blockchain-based Data Usage Auditing (BDUA)

Kaâniche et Laurent[156] ont proposé en 2018 d'adjoindre une *blockchain* pour assurer la préservation des données personnelles lors de la conduite des audits par les entités autorisées.

Dans le système qu'elles proposent, le convertisseur qui aide les utilisateurs à dériver un pseudonyme par serveur ne connaît pas le pseudonyme dérivé par l'utilisateur, mais reste tout de même la seule entité capable de les relier, sans savoir à quel utilisateur ils correspondent. Il permet également de conduire, à partir des transactions blockchain enregistrées, un audit respectueux des données personnelles.

Comparaison des techniques PETs

À présent que l'ensemble des technologies préservant la vie privée ont été détaillées et illustrées, il est temps d'établir le panorama comparatif de synthèse. Qu'elles soient orientées utilisateur ou orientées serveur, ou bien encore activées sur le canal de communication, elles ne relèvent pas toutes du même modèle de confiance et n'ont pas la même architecture de déploiement (et donc des entités impliquées différentes). Leurs inconvénients sont également de natures différentes et leurs cas et contextes d'usage varient d'autant. Enfin, certaines ne sont pas adaptées pour le déploiement de services personnalisés, tant elles protègent les données personnelles.

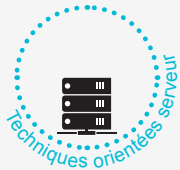
Technologies PETs adaptées pour la personnalisation de services

/

Technologies PETs ***non*** adaptées pour la personnalisation de services

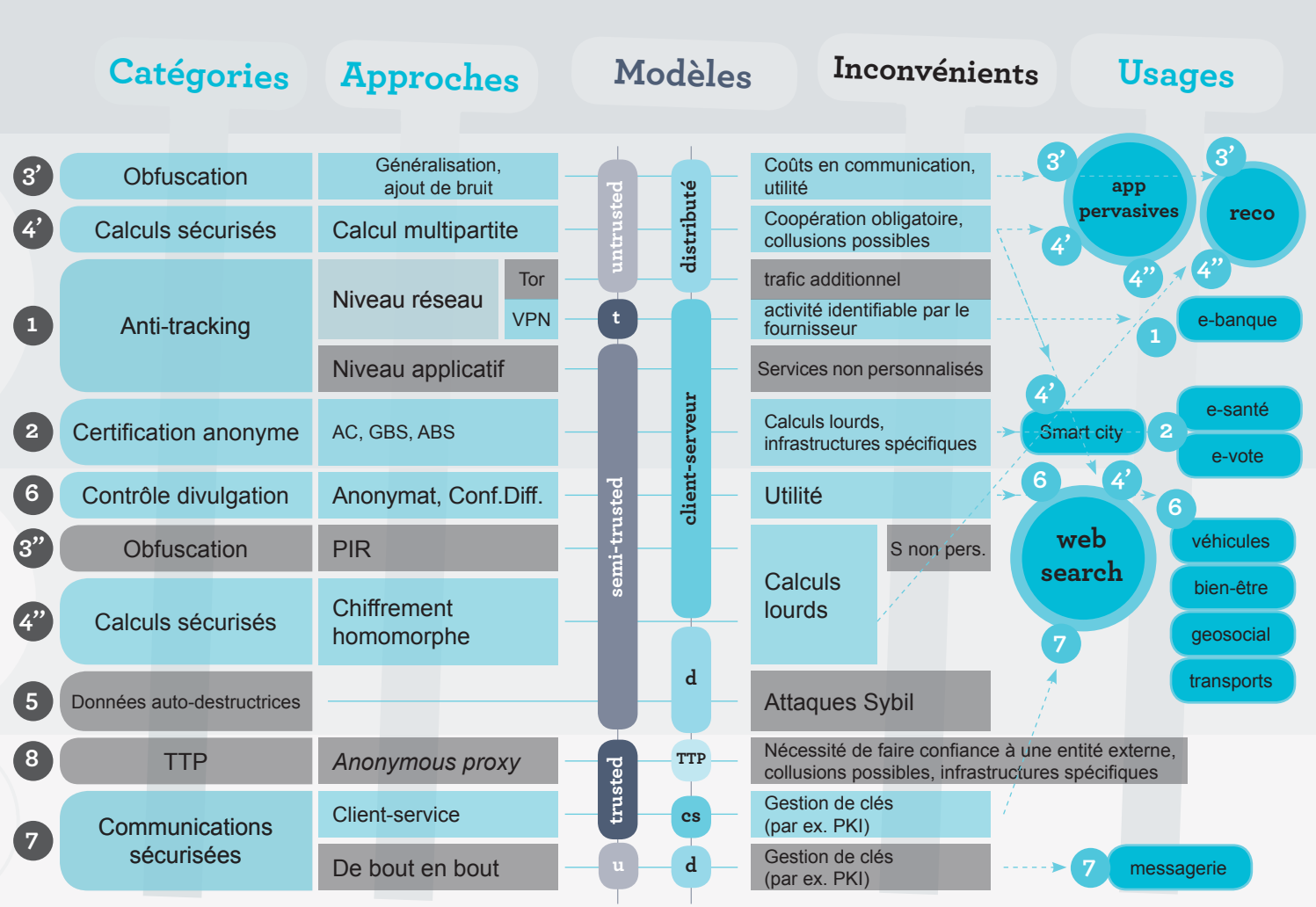
techniques orientées utilisateur, qui s'adressent directement à l'utilisateur, pour l'aider à gérer la protection de son identité, via l'installation de logiciels spécifiques, à mieux contrôler la divulgation de ses données et en certifier la véracité.

techniques orientées serveur, déployées sur les serveurs pour anonymiser les données personnelles, travailler sur des contenus chiffrés, et détruire les données de façon automatique.



techniques orientées canal de communication, qui assurent que les communications sont sécurisées entre utilisateurs (et serveurs) ou qui déploient des approches de tiers de confiance.

Comparaison entre PETs



Tab. 6

Comparaison entre PETs

Discussion selon les modèles de confiance

_Modèle trusted

Trois techniques relevant du modèle de confiance *trusted*, les TTP, les communications sécurisées client-service, et les VPN en tant qu'outil d'anti-tracking de niveau réseau.

Les communications par VPN peuvent également relever des communications client-service sécurisées car les deux reposent sur un canal sécurisé capable d'atténuer les velléités d'un adversaire cherchant à obtenir des informations à partir des échanges entre le client et le fournisseur de VPN. Une différence existe, cependant. Les communications via VPN aboutissent chez le fournisseur VPN, tandis que les communications sécurisées client-service aboutissent chez le fournisseur de service. Dans le premier cas, l'utilisateur fait confiance au fournisseur de VPN pour gérer le canal sécurisé et sa confidentialité. Dans le second, il doit faire confiance à l'ensemble des fournisseurs avec lesquels il interagit, et doit établir un canal sécurisé reposant sur des infrastructures à clé publique (*Public Key Infrastructure*, *PKI*) dans lesquelles il doit également avoir confiance.

Comme les solutions VPN, les approches de type TTP mettent en œuvre un intermédiaire de confiance qui, contrairement au VPN, anonymise ou pseudonymise les communications.

Les techniques TTP d'anonymiseurs ne peuvent pas être utilisées pour assurer des services personnalisés, puisque les serveurs intermédiaires mélangent des données de plusieurs utilisateurs avant de les transmettre aux fournisseurs de service.

Pour résoudre le problème du compromis entre personnalisation et vie privée, des systèmes (comme BDUA, vu page 165) facilitant la conversion d'identités vers des pseudonymes peuvent être déployés, dans des contextes d'applications pervasives et notamment dans les scénarios de transfert de données.

À l'inverse, les communications sécurisées client-serveur sont adaptées à la personnalisation de services, principalement pour la recherche d'informations sur le web et un certain nombre d'applications pervasives comme le e-commerce et le e-banking.

Construites sur les relations de confiance entre les clients et les fournisseurs de service, elles permettent des offres de service très adaptées aux usages des utilisateurs et à leurs préférences. En revanche, elles n'offrent aucune protection contre des fournisseurs de service qui auraient tendance à être trop curieux.

_Modèle semi-trusted

Sous ce modèle se retrouvent six techniques, déployées aussi bien côté serveur que côté client. Auprès de l'utilisateur on trouve ainsi les outils d'anti-tracking de niveau applicatif, la certification anonyme, tandis que les solutions de contrôle de divulgation des données statistiques (SDC), les systèmes de recherche d'informations privée (PIR), les mécanismes de données auto-destructibles et le chiffrement homomorphe sont déployés côté serveur. Toutes ces techniques supposent que les fournisseurs de services effectuent honnêtement les traitements attendus.

Les PETs relevant de ce modèle correspondent le plus souvent à des architectures client-serveur. Ainsi, chaque client, à son niveau, met en place des techniques qui révèlent le minimum d'informations aux fournisseurs de service, les empêchant ainsi d'élaborer des profils précis des utilisateurs. En conséquence, ces mécanismes sont généralement peu adaptés à la mise en place de services personnalisés, comme les services de recherche sur le web. Cependant des solutions prometteuses existent pour certains applications pervasives au sein des véhicules et de type géosociales.[157]

_Modèle untrusted

Quatre techniques sont rattachées à ce modèle : les outils d'anti-tracking déployés au niveau réseau sur le modèle Tor, le calcul sécurisé multipartite, les communications sécurisées de bout en bout et les techniques d'obfuscation de données. Toutes sont construites sur des architectures distribuées.

Fournissant une protection de la vie privée particulièrement aboutie, elles ne sont pas très adaptées aux services personnalisés, puisqu'elles cachent efficacement les informations qui seraient utiles aux fournisseurs pour construire les profils des utilisateurs. De plus, elles engendrent des coûts additionnels en calcul et en communication, provenant du nombre de participants (équipements ou individus) qui doivent coopérer. Ces techniques ne sont par conséquent pas adaptées aux applications et services pervasifs dont l'accès est réalisé depuis des équipements aux ressources limitées.

Les techniques de calcul sécurisé multipartite sont adaptées aux applications de la *smart city* dans le cas où les équipements impliqués ne sont pas contraints en ressources. C'est le cas par exemple des services de collecte de la consommation[158] d'énergie dans des habitations individuelles, qui pourront ainsi ne pas faire fuiter d'informations privées sur les habitants. De leur côté, les communications sécurisées[159] de bout en bout sont largement utilisées dans les applications de messagerie pour assurer la confidentialité des échanges entre participants.

Discussion selon les cas d'usage

La question ici est simple : les techniques de préservation de la vie privée étudiées sont-elles adaptées, ou non, à la création de services personnalisés ? Les réponses permettent de jauger le compromis possible entre vie privée et services personnalisés.

_Les technologies PETs non adaptées à la personnalisation de services

Sont regroupées dans cette catégorie les mécanismes de données auto-destructibles, les solutions PIR, les communications sécurisées de bout en bout et, parmi les outils d'anti-tracking, Tor et les outils anti-cookies. Toutes ces technologies sont déployées dans le cas où les utilisateurs sont préoccupés par leur vie privée, la confidentialité de leur données et de leurs échanges, et ne font pas confiance aux fournisseurs de service. Ils font donc le choix de gagner sur le tableau de la vie privée, et de perdre sur celui de la personnalisation des services.

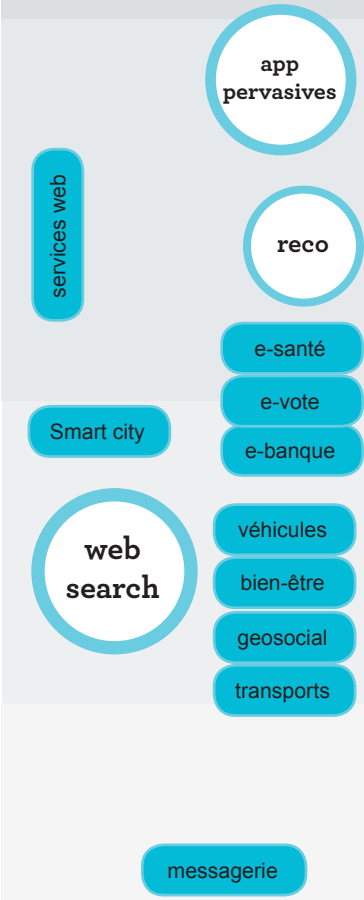
Bien que dans leur cas, les techniques comme les mécanismes de données auto-destructibles et les solutions PIR permettent aux données de résider un temps chez les fournisseurs de services, elles attendent de ces derniers un traitement honnête et un minimum de collaboration pour garantir la vie privée, et ne sont donc pas considérées comme éligibles aux services personnalisés.

_Les technologies PETs adaptées à la personnalisation de services

Dans cette famille de technologies respectueuses de la vie privée et qui permettent l'élaboration de services personnalisés, distinguons trois sous-familles importantes : celles pour les applications de recherche sur le web, celles pour les applications pervasives, et celles pour les systèmes de recommandation.

web search

Les technologies adaptées à la recherche sur le web regroupent les mécanismes de calcul sécurisé multipartite, les communications sécurisées client-service sécurisées et le contrôle de la divulgation des données statistiques (SDC). Bien que les deux premières reposent sur des architectures et des modèles de confiance différents, elles garantissent toutes deux la confidentialité des recherches, la première protégeant des fournisseurs trop curieux, et la seconde des attaques externes.



Un exemple est celui d'une recherche collaborative où des utilisateurs vont coopérer en agrégeant leurs requêtes avant de les envoyer au moteur de recherche. Le principal problème de ce type d'approche est d'une part le coût supplémentaire en communications et en calculs et d'autre part la pertinence des résultats renvoyés, qui dépendront de l'homogénéité des profils des membres du groupe d'utilisateurs.

Noter que les outils d'anti-tracking de niveau applicatif (anti-cookies) sont également largement installés par les utilisateurs. Cependant, ces outils ne sont pas nécessairement fournis en gardant en tête les besoins de vie privée. Ils le sont surtout pour bloquer des publicités et autres popups indésirables.

Les approches SDC sont déployées par les moteurs de recherche pour assurer la vie privée des utilisateurs en anonymisant l'enregistrement de leurs requêtes. Ces bases de données de requêtes sont en effet essentielles aux moteurs pour suivre les usages statistiques et pour alimenter leur modèle économique. Proposant un bon compromis entre confidentialité et personnalisation, elles sont les technologies les plus utilisées côté serveur.

Ces approches sont en effet très appréciées des industriels, car elles permettent d'obtenir des données anonymisées et d'en faire libre usage du fait que ces données sortent du cadre réglementaire du RGPD.

app pervasives

Les approches adaptées aux applications pervasives recouvrent le calcul sécurisé – calcul sécurisé multipartite et chiffrement homomorphe –, les solutions d'obfuscation de données et les mécanismes de certification anonyme. Dans ce type d'environnement, le choix des PETs les plus efficaces dépend à la fois des modèles de confiance et des modèles d'architecture. Par exemple, dans les applications pour la *smart city* s'appuyant sur des *réseaux ad hoc*, le calcul sécurisé collaboratif est mis en œuvre, tandis que dans l'e-commerce répondant à une architecture client-serveur, la certification anonyme et les communications client-service sécurisées sont préférées.

reco

Trois techniques sont particulièrement adaptées pour les services de recommandation : l'obfuscation de données et le calcul sécurisé – calcul sécurisé multipartite et chiffrement homomorphe –. Ces services permettent aux utilisateurs de recevoir des recommandations en fonction de leurs centres d'intérêt, proposées par une ou plusieurs entités recommandatrices. L'obfuscation de données – par injection d'avis factices ou la suppression d'avis redondants ou insincères – est ici une technique qui préserve la vie privée avec un coût acceptable en communications et calculs supplémentaires. Cependant, le compromis vie privée / personnalisation n'est pas toujours optimal, et quelques travaux ont été proposés pour analyser l'impact de l'adoption de ces méthodes (voir Parra-Arnau *et al.*[81]).

Rester vigilant sur les futures menaces

Un panorama des techniques tel qu'il vient d'être fait ne permet pas pour autant de prédire quelles technologies risquent de créer de nouvelles brèches dans la vie privée. De nouvelles métriques doivent être envisagées pour mieux saisir la dynamique du compromis vie privée / personnalisation. De plus, les capacités grandissantes de traitement de données, et de croisement avec d'autres ensembles de données et de profils construits par les grands acteurs de l'Internet, conduisent ces derniers à proposer de plus en plus de services personnalisés… rendant d'autant plus critiques et préoccupantes les futures fuites de données qui risquent de se produire.

Une offre qui creuse son sillon

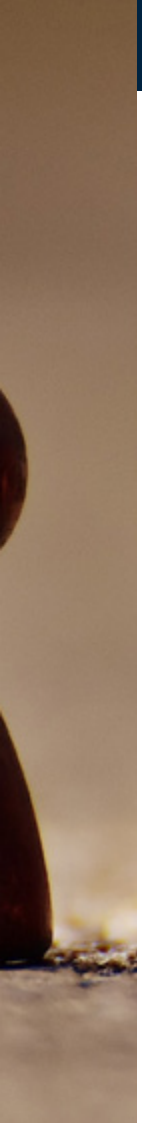
Les PETs sont des technologies prometteuses qui peuvent protéger la vie privée requise par les utilisateurs, particulièrement dans un contexte où les régulations se renforcent, tout en permettant de construire des offres personnalisées avec une qualité de service et une expérience utilisateur optimale.

Alors que chacun cherche à équilibrer selon ses besoins le compromis vie privée / personnalisation, elles sont peu à peu adoptées par les professionnels. Pour autant, des recherches supplémentaires sont nécessaires pour l'emploi des PETs dans les applications courantes.

Défis & perspectives

Nous sommes début 2019, et il n'est pas une semaine sans que de nouvelles affaires impliquant des atteintes aux données personnelles soient médiatisées, augmentant la prise de conscience du public.

Cependant, les travaux de recherche en sont encore à leurs prémices, et les solutions proposées sont souvent loin d'exister sous forme d'implémentations opérationnelles. Les défis qui sont devant nous ne sont du reste pas tous techniques. La nécessité d'offrir une expérience utilisateur de qualité, ainsi que les aspects juridiques et économiques, sont également en jeu.



Défis techniques

Trois principaux défis techniques sont à considérer : l'implémentation des outils d'audit, l'irruption du *Machine Learning* dans la personnalisation, les contraintes des environnements pervasifs. Chacun apporte son lot de questions ouvertes, et des liens existent entre eux.

_Les outils d'audit préservant la vie privée

Trop peu de travaux ont abordé pour l'instant la question de la transparence et celle de l'audit, mis en avant par les récentes réglementations, en particulier le Règlement Général sur la Protection des Données. Deux aspects sont en jeu ici. D'une part les utilisateurs ont besoin de se voir garantir un accès – et des capacités de traitement et de partage – ***transparent*** et contrôlé à leurs données, de telle manière que des utilisateurs non autorisés, ou des serveurs avec lesquels aucune relation de confiance n'a été établie, ne puissent accéder à ces données. D'autre part, la nécessité d'une validation externe de la conformité au RGPD des services offerts par un fournisseur, sous la forme d'un ***audit*** ou d'une certification effectuée par une autorité compétente, implique l'accès aux données des utilisateurs sous une forme qui en préserve absolument le caractère privé et personnel.

Transparence et capacité d'audit : deux exigences mises en avant par les récentes réglementations.

Dans leurs travaux de 2017, Jan Camenisch et Anja Lehmann ont proposé un mécanisme d'échange de données distribuées permettant un contrôle respectueux de la confidentialité des données. Chaque propriétaire de données peut en effet conduire un audit précis à partir d'un fichier de *journalisation* public.

Aujourd'hui, la technologie blockchain[160][161][162] peut apporter à ces questions certaines réponses. Le principe général derrière la blockchain est en effet que *toutes les transactions sont partagées par tous les nœuds d'un réseau, mises à jour par les actions de minage sur des nœuds particuliers, visibles par tout le monde, et sous le contrôle et la propriété de personne en particulier*.

C'est fort de ces caractéristiques que Nesrine Kaâniche et Maryline Laurent ont proposé en 2018 un mécanisme d'audit préservant la vie privée construit sur les transactions au cœur de la blockchain, audit qui peut être conduit aussi bien par les utilisateurs eux-même (le souhait de transparence) que par les autorités (le besoin d'audit de conformité).

> *La blockchain peut fournir le mécanisme d'audit préservant la confidentialité des données.*

Dans leur proposition, les transactions sont enregistrées par chacune des parties concernées et chaque entité peut accéder à différents niveaux d'informations dans la blockchain sous couvert d'un chiffrement multi-niveaux, ce qui lui permet de relier ou pas plusieurs évènements inscrits dans la blockchain et de mener à bien un audit.

Cependant, la blockchain et les *smart contracts*, dans leur fondamentaux techniques actuels, ne permettent pas d'assurer une confidentialité correcte des transactions.

Toute la séquence d'actions effectuées lors d'un *smart contract* est en effet propagée dans l'ensemble du réseau des nœuds blockchain et/ou enregistrée sur la blockchain elle-même, devenant ainsi visible aux yeux de tous. Même si les utilisateurs effectuant les transactions ont la possibilité de créer des pseudonymes à la demande, la valeur de chaque transaction, et les nouveaux états des comptes qui en résultent sont publiquement visibles. Il a du reste été montré dès 2013[163], dans le cas de la blockchain Bitcoin, qu'il était possible de ré-identifier des individus, par analyse des structures de graphes transactionnels[164].

La confidentialité des transactions reste cependant à créer.

Cette absence de confidentialité sur les transactions est un obstacle de taille vers une adoption plus large de la blockchain. Dans le secteur bancaire, notamment, les transactions financières ne peuvent pas souffrir d'un tel manque. Le secret des transactions en est même à notre époque la pierre angulaire. Plusieurs évolutions à ce sujet, portant soit sur les crypto-monnaies et leurs transactions, soit sur les *smart contracts*, sont apparues depuis 2016. Pour ces derniers, des pistes à base de preuve à divulgation nulle de connaissance sans interaction sont envisagées pour éviter de rendre publiques les valeurs d'entrées des utilisateurs du contrat.

_L'apprentissage machine au cœur des services personnalisés

Parmi les techniques les plus souvent convoquées pour construire des services personnalisés, le *machine learning*, un des modes d'apprentissage les plus en vogue dans le champ de l'Intelligence Artificielle, apporte toute sa puissance et son efficacité[165][166][167][168]. Cette dernière est tout à fait dans son élément pour traiter des données non structurées disponibles en masse, avec une forte dimension spatio-temporelle. Avec ces techniques, il est possible d'identifier sans a priori des modèles sur les ensembles de données étudiés, et donc ici de construire des modèles représentatifs des utilisateurs à partir de leurs habitudes, de leurs déplacements, de leurs centres d'intérêt. Après avoir collecté des données, il est possible d'utiliser les techniques de *machine learning* pour générer les services personnalisés[169][170]. Les moteurs de recherche en général, les services où la recommandation, y compris entre utilisateurs, est une brique importante, et un bon nombre d'applications pervasives sont de grands consommateurs de *machine learning*.

Pour autant, ces solutions mettent souvent l'accent sur les performances, au détriment d'une prise en compte sérieuse des questions de préservation de la vie privée. C'est tout un équilibre, entre vie privée, efficacité, expérience utilisateur et performances qu'il faut atteindre. Et il n'existe que peu de travaux de recherche présentant des résultats offrant de bons compromis. Quelques avancées récentes sur les systèmes cryptographiques peuvent cependant être exploitées, tout en prenant en compte à la fois l'efficacité et l'efficience.[171][172][173][174]

_Des contraintes supplémentaires en environnement pervasif

Les services mobiles personnalisés suivent la même croissance que celle du parc des téléphones eux-même. Sur ce type d'usage, les considérations sur le respect de la vie privée sont également de mise. Elles souffrent tout autant des compromis nécessaires à faire entre les différentes propriétés souhaitées des services proposés. Une caractéristique supplémentaire est cependant fondamentale : quelles que soient leurs raisons d'être, les techniques déployées sur un téléphone mobile doivent prendre en compte les ressources limitées à disposition. Un axe de recherche consiste donc à imaginer et faire le design de solutions légères de préservation de la vie privée, au sens où elles sont économes des capacités de calcul, de mémoire, et de trafic de communication. La consommation d'énergie est également une préoccupation au cœur de l'expérience utilisateur. Des puces spécifiques, comme les ***Intel SGX*** (Intel Software Guard Extensions) font ainsi partie des solutions matérielles développées pour effectuer des opérations sécurisées avec un coût de calcul supplémentaire minimal.

> *La gestion des ressources est la contrainte principale en mode pervasif.*

Lors de la recherche des compromis techniques, il est bon de faire des choix en prédisant l'état de connaissance, ou de compétence, de l'attaquant. Nul besoin peut-être de sortir une grosse artillerie au détriment de la qualité de service. Depuis les travaux de Canetti en 2001[175], prolongés par d'autres auteurs en 2018[176] pour les systèmes à réputation, on sait remplacer un protocole de sécurité idéal par un autre protocole, tout en ayant la preuve mathématique de l'équivalence. C'est sans doute une piste à explorer.

Stratégies des consommateurs et citoyens français

La **Chaire Valeurs et Politiques des Informations Personnelles** a conduit en 2017 une étude auprès de 2000 répondants, utilisateurs français du web, pour comprendre et analyser l'impact de la collecte abusive de données sur les comportement des utilisateurs. Il a été montré que les utilisateurs du web devenaient plus vigilants, prenant leurs décisions de partager certains types de données en fonction des catégories de services utilisés (banque, réseaux sociaux, achats en ligne, téléphonie mobile…). 10 % d'entre eux indiquent préférer ne pas partager de données si le choix leur en est laissé.

	Réseau social	Banque	Site gouvernemental	Site e-commerce	Opérateur télécom
Nom	36%	80%	79%	63%	73%
Contact	10%	75%	74%	52%	66%
Documents d'id	2%	56%	65%	6%	20%
Infos bancaires	2%	72%	48%	24%	32%
Infos de santé	3%	7%	39%	2%	3%
Centres d'intérêt	42%	9%	9%	28%	13%
Localisation	16%	14%	17%	11%	18%
Habitudes web	11%	5%	7%	14%	16%
Achats internet	6%	16%	5%	36%	8%
Liste d'amis/contacts	30%	4%	4%	4%	6%
Messages/images…	38%	4%	5%	6%	7%
Aucune information	38%	14%	14%	24%	19%
Au moins 1 info	62%	86%	86%	76%	81%

- *Question : Si vous pouvez choisir, à laquelle de ces entités êtes-vous prêts à donner cette information ?*
- *Base : utilisateurs du web âgés de 15 ans et plus (n = 2051)*

Fig. 16

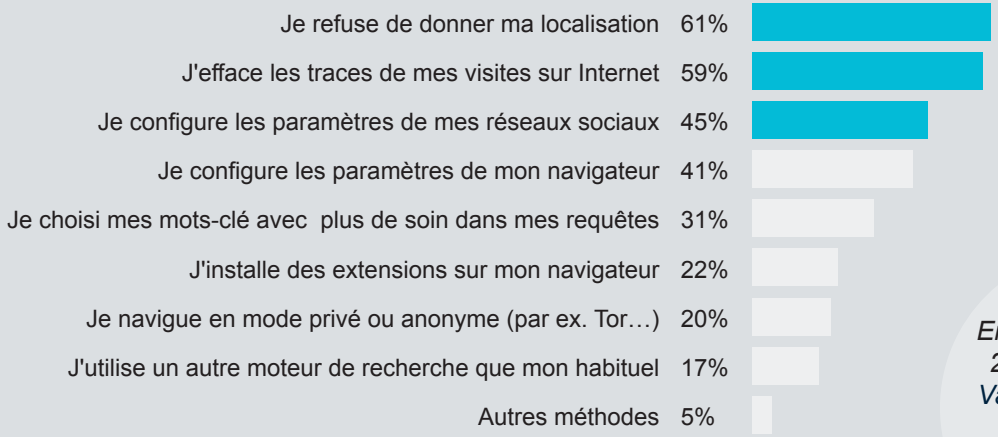
Réussir l'expérience utilisateur

Les défis techniques listés ci-dessus ont plusieurs fois fait référence à l'*expérience utilisateur*. Cette dernière définit à la fois le périmètre des données personnelles, et l'utilité qui est prête à être perdue par l'adoption de PETs. Sur le premier point, des retours d'usages de nouveaux équipements comme des appareils photo, des lunettes de réalité augmentée ou des téléviseurs connectés ont montré à quels points ces objets pouvaient être intrusifs et avides de données personnelles. L'idée qu'ils puissent constituer une menace à terme, avec des conséquences sur la vie des utilisateurs, se fraye un chemin dans les esprits. Cependant, la nature des données qui ne devraient pas être ainsi collectées ou communiquées, et si elles le doivent, à qui, diffère selon les individus, en fonction de leur statut social, leur contexte culturel, ou du contexte en général.

> *L'expérience utilisateur définit à la fois le périmètre des données personnelles et l'utilisabilité des PETs.*

_Vigilance accrue

54% des internautes étaient plus vigilants sur Internet en 2017 par rapport aux années précédentes, et parmi ces 54%, à la question "*En quoi êtes-vous plus vigilants ?*", ils précisaient :



- *Question : En quoi êtes-vous plus vigilants ?*
- *Base : utilisateurs du web âgés de 15 ans et plus (n = 2051)*

Enquête conduite en 2017 par la Chaire Valeurs et Politiques des Informations Personnelles avec Médiamétrie

Fig. 17

_Quelques règles pour une expérience utilisateur réussie

Au vu de ces résultats, un premier chantier consiste à imaginer et concevoir des solutions techniques qui offrent un panel de protections flexibles, adaptables à chaque utilisateur selon son contexte, ses besoins, ses préférences.

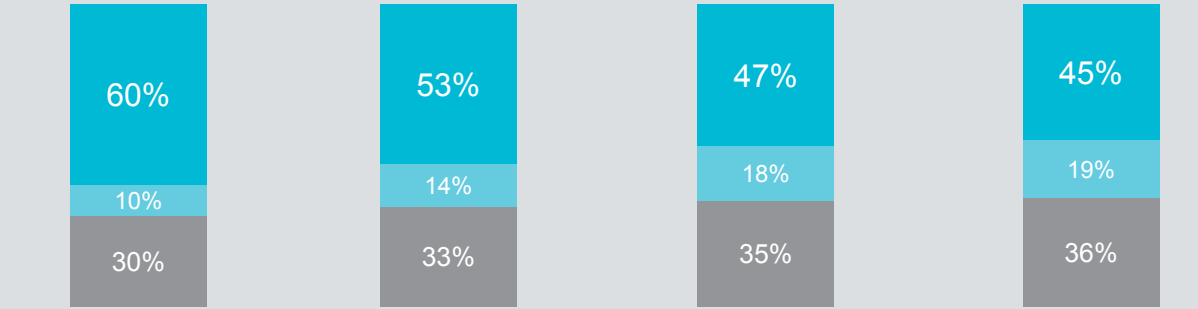
Le second chantier est celui, bien sûr, de la recherche de l'équilibre et des compromis entre la protection de l'utilisateur, l'efficience de l'expérience utilisateur offerte et l'utilité des données qui circulent. Massaguer *et al.*[172] ont ainsi modélisé la notion de vie privée en fonction de l'*utilité négative* associée à chaque information diffusée. En découle une construction mathématique qui permet de maximiser l'utilité des données partagées par les utilisateurs tout en restant à un certain niveau d'exigence de confidentialité.

Un autre aspect clé d'une meilleure expérience utilisateur consiste à fournir des applications ergonomiques et conviviales, de manière à ce que l'utilisateur ne passe pas par mille réglages fastidieux pour paramétrer les technologies selon ses besoins propres.

Enfin, fournir des outils de monitoring est hautement souhaitable, pour que les utilisateurs soient en mesure de vérifier, de manière totalement transparente, comment leurs données sont effectivement traitées, et si elles le sont bien telles qu'ils l'avaient souhaité en paramétrant leurs choix. Il en va de la construction de la confiance des utilisateurs envers ces technologies.

_Paramétrages

L'objectif principal des internautes paramétrant leurs préférences pour leur comptes Internet est de réduire le nombre de données collectées, et non pas d'améliorer la qualité des recommandations personnalisées.



▸ *Question : Pour quelle raison paramétrez-vous les préférences en matière de vie privée de vos comptes Internet ?*
▸ *Base : utilisateurs du web âgés de 15 ans et plus (n = 2051)*

Fig. 18

Défis juridiques

Le Règlement Général sur la Protection des Données (RGPD), dont il a souvent été question, n'est pas le seul texte juridique à prendre en compte.

Il convient en effet de prendre en compte notamment le règlement (UE) 2018/1807 établissant un cadre applicable au libre flux des données à caractère non personnel et d'anticiper l'application du futur règlement e-Privacy.

Le RGPD n'est pas le seul texte juridique à prendre en compte.

À consulter :

Règlement (UE) 2018/1807 du Parlement européen et du Conseil du 14 novembre 2018 établissant un cadre applicable au libre flux des données à caractère non personnel dans l'Union européenne

e-Privacy : Proposition de règlement du Parlement européen et du Conseil, concernant le respect de la vie privée et la protection des données à caractère personnel dans les communications électroniques et abrogeant la directive 2002/58/CE (règlement «vie privée et communications électroniques»), 10 janvier 2017, COM(2017) 10 final

Le règlement (UE) 2018/1807 établissant un cadre applicable au libre flux des données à caractère non personnel précise la place du curseur entre les informations qui peuvent être qualifiées de données personnelles au sens du RGPD et celles qui peuvent être considérées comme des données non personnelles. Lorsque les données personnelles et les données à caractère non personnel d'un ensemble sont inextricablement liées, il précise que le RGPD s'applique.

L'objectif du règlement e-Privacy proposé par la Commission européenne le 10 janvier 2017 est d'actualiser et d'unifier les règles fixées par l'actuelle directive 2002/58/EC. Il s'agit de renforcer le respect de la vie privée dans les communications électroniques en incluant dans le champ du règlement l'ensemble des fournisseurs, y compris les services « *Over-The-Top* » comme WhatsApp, Facebook Messenger, Skype, Gmail. Le texte entend également réglementer l'usage des *métadonnées* ; il renforce la place du consentement en ce qui concerne les cookies et les communications de prospection directe (spam).

Dans un avenir qui reste encore indéterminé début 2019, compte tenu des négociations en cours, le Règlement e-Privacy serait une *lex specialis* complétant le RGPD en vigueur depuis le 25 mai 2018. Cependant, l'intense lobbying des parties prenantes et les élections des représentants au Parlement européen en mai 2019 rendent cette trajectoire incertaine.

Dans la mesure où les services personnalisées impliquent de collecter et de traiter des données personnelles, le RGPD s'applique. Il est va ainsi de l'obfuscation (en particu-

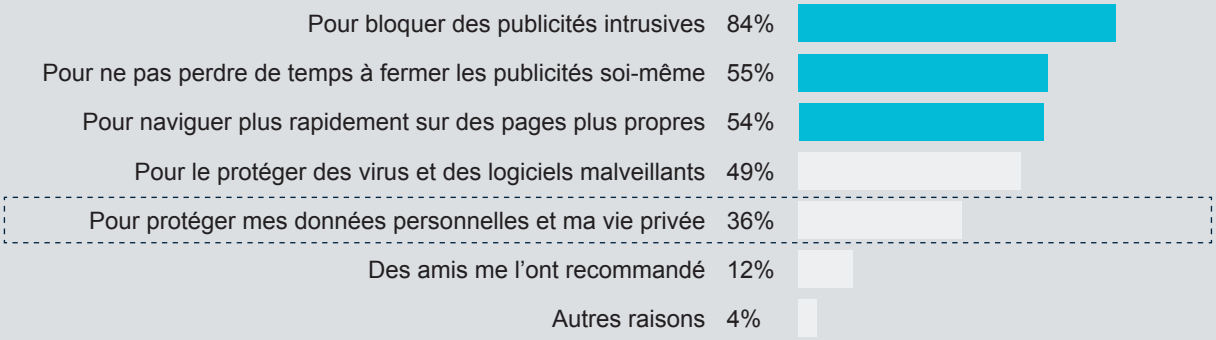
lier l'obfuscation de données), de la pseudonymisation et de l'anonymisation mises en place par les fournisseurs de services. Il convient également de prendre en compte les lignes directrices définies par le Comité européen de protection des données, qui entendent clarifier et illustrer d'exemples concrets le nouveau cadre juridique issu du RGPD : https://www.cnil.fr/fr/reglement-europeen/lignes-directrices et plus complet mais en anglais : https://edpb.europa.eu/our-work-tools/general-guidance/gdpr-guidelines-recommendations-best-practices_en.

Les activités d'audit dans un contexte de données personnelles constituent un défi technique déjà souligné. Un point important est que ces outils d'audit soient utiles, et utilisables. Il est essentiel de fournir aux utilisateurs des moyens conviviaux leur permettant de s'assurer, et de leur garantir, que leur consentement (ou non) est respecté lors des phases de collecte, de traitement, d'analyse et d'enregistrement de leurs données.

De plus, et selon les cas d'usage considérés, certaines procédures techniques sont nécessaires pour pouvoir identifier un utilisateur malveillant, ce qui par ricochet soulève un certain nombre de difficultés, compte-tenu de l'ensemble des parties à auditer (personnes concernées par les données, autorités pertinentes…). Dans ces cas, en effet, les exigences du RGPD comme l'obligation d'intégrité et de confidentialité, les principes de protection des données dès la conception et par défaut, le droit à l'oubli doivent être respectées.

_Bloqueurs de publicités

55% des internautes indiquaient dans l'étude avoir déjà utilisé un bloqueur de publicités sur au moins un de leurs écrans.



- *Question : Pour quelle(s) raison(s) utilisez-vous un bloqueur de publicités ?*
- *Base : utilisateurs du web âgés de 15 ans et plus, utilisant un bloqueur de publicités (n = 1126)*

Fig. 19

Défis économiques

L'étude conduite en 2017 par la Chaire Valeurs et Politiques des Informations Personnelles avec Médiamétrie, citée plus haut, a mis en lumière également plusieurs défis relevant du champ de l'économie.

Pour commencer, le nombre important (54% début 2017, rappelons-le) d'utilisateurs de plus en plus réticents à partager leurs données personnelles représente un risque qui ne doit pas être sous-estimé par les acteurs de la nouvelle économie qui utilisent ces données dans leur modèle d'affaires.

Cependant, il ne faut pas en déduire que l'utilisation de stratégies individuelles de protection agirait comme un frein sur les échanges économiques. Par exemple, l'usage de plus en plus répandu des bloqueurs de publicité n'est pas une menace, car l'étude a montré que les utilisateurs concernés effectuent plus d'achats en ligne. Disposer d'un moyen de protection renforcerait le sentiment, subjectif, d'être maître de ses choix.

Proposer des solutions aux utilisateurs renforçant leur sentiment d'encapacitation (*empowerment*) a cependant un coût, qu'il s'agisse de la mise en œuvre d'une véritable stratégie de crédibilité de la marque définie par l'entreprise, ou de la mise en conformité aux obligations imposées par le RGPD. Ce type de démarche impose de revisiter, voire parfois de repenser, le design des systèmes d'information et des services. Elle implique de mettre en place une politique de communication à destination des personnes concernées pour réduire l'asymétrie d'information. Cette politique peut alors prendre la forme d'un label ou d'une certification. En outre, ce type d'approche doit être évalué régulièrement, en fonction des évolutions économiques et technologiques en général, et des nouvelles PETs proposées en particulier.

L'implémentation de PETs constitue donc un investissement à la fois monétaire et temporel important, notamment pour les PME. Son résultat direct, recherché, est la réduction du nombre de données personnelles collectées, à tout le moins de la force du lien entre une information et une personne.

Si certains prestataires nous ont fait part de leurs craintes – notamment que l'obtention de données pseudonymisées est susceptible de compliquer la collecte de données pertinentes et le recueil du consentement de la personne concernée – la démarche PETs constitue un investissement qui sur le long terme permettra de renforcer l'indispensable confiance des utilisateurs.

Mots-clé

Technologies de préservation de la vie privée, données personnelles, services de recommandations, moteurs de recherche web, applications pervasives, services géolocalisés, profilage comportemental, tendances en cryptographie, communications sécurisées, certification anonyme, recherche d'informations privée, calcul sécurisé multipartite, chiffrement homomorphe, modèles de confiance.

Keywords

Privacy Enhancing Technologies, recommendation services, web-search engines, pervasive applications, location-based services, profile-based services, cryptographic trends, secure communications, anonymous certification, private information retrieval, secure multi-party computation, homomorphic encryption, trust models.

Glossaire

Nous réunissons dans cette section quelques-uns des termes et expressions qui n'ont pas été complètement développés dans l'ouvrage et d'autres qui ont bénéficié d'explications plus détaillées, l'ensemble pouvant servir tout à la fois de références et de clés d'entrée lorsque l'on parle des technologies de préservation de la vie privée.

Anonymat : l'impossibilité de remonter à une personne, et ce, de façon irréversible.

Arbre d'accès : (*access structure*) concept issu de la théorie des ensembles, repris en cryptographie pour décrire les conditions à satisfaire pour qu'une opération cryptographique soit autorisée (par exemple une *signature* à base d'attributs prouve que le signataire dispose d'un ensemble d'*attributs* satisfaisant l'arbre d'accès).

Attaquants : (*adversaries*) *entités* œuvrant pour extraire des informations ou pour obtenir des droits d'accès de façon illégale.

Attributs : un utilisateur dispose habituellement d'un jeu d'attributs le décrivant : prénom, nom, sexe, âge, adresse, situation professionnelle…

Auditabilité : capacité d'un système à pouvoir être audité.

Authentification : le mécanisme prouvant que l'utilisateur est bien celui qu'il prétend être, le *crédentiel* fourni servant de preuve.

Autorité : entité dans laquelle les utilisateurs placent une certaine confiance et/ou qui possède des pouvoirs particuliers. Ce peut-être par exemple une autorité de certification, de délivrance, de révocation… Voir à ce sujet : *tiers de confiance*.

Blockchain : registre distribué et sécurisé, enregistrant un ensemble de transactions, dont le fonctionnement repose sur la coopération d'un ensemble d'entités. Cette organisation décentralisée, à la fois dans le stockage du registre, la validation des transactions et l'exécution de codes (smart

contracts) évite le problème classique de devoir faire confiance à une tierce partie. Après validation consensuelle et collective, les transactions sont groupées par blocs chaînés les uns aux autres jusqu'au tout premier, le tout faisant appel à des techniques cryptographiques de pointe.

Boutons médias sociaux : boutons d'action aux couleurs des réseaux sociaux, invitant à partager les contenus associés. Le code JavaScript sous-jacent de ces boutons permet souvent le pistage des utilisateurs qui les croisent.

Clé cryptographique : paramètre utilisé en entrée d'une opération cryptographique (opérations qui ne se limitent pas au chiffrement et au déchiffrement, et se rapportent également au scellement de documents, à l'établissement de *signatures électroniques*…). Dans le cas du chiffrement, une clé est dite symétrique quand elle sert à chiffrer *et* à déchiffrer. En cryptographie asymétrique, l'utilisateur possède une clé publique et une clé privée (qui ne peut pas se déduire de sa clé publique), la clé publique permettant par exemple à un correspondant de chiffrer un message que seul l'utilisateur pourra déchiffrer avec sa clé privée.

Certificat électronique : un document électronique permettant de rendre public le lien entre une *entité* et sa *clé cryptographique* publique, la *confiance* dans ce lien reposant sur la crédibilité de l'*autorité* émettrice du certificat.

Circuit arithmétique, booléen, de taille polynomiale : le terme « circuit » ici fait référence à l'assemblage d'opérations décrivant un processus qui doit être suivi, et au graphe sous-jacent qui le représente. Une variable donnée peut ainsi subir une séquence d'opérations arithmétiques, ou une variable booléenne (variable à deux états, *vrai* ou *faux* généralement) participer à diverses opérations logiques (l'algorithme), avant de produire le résultat recherché. « De taille polynomiale » fait référence à un certain type de complexité de l'algorithme (nombre d'instructions, place mémoire…), complexité bornée par une fonction polynomiale de la valeur des variables traitées.

Chiffrement homomorphe : chiffrement dont les caractéristiques algébriques sont telles que certaines opérations sur les données chiffrées donnent, après déchiffrement, le même résultat qui aurait été obtenu sur les données d'origine, ouvrant ainsi la voie à des systèmes où des entités externes peuvent faire des calculs sur des données sans connaître ni le sens des données d'origine, ni le résultat de leurs calculs.

Cloud computing : puissance de calcul ou de stockage disponible à distance via un réseau (Internet, le plus souvent) et louée à la demande.

Confiance : voir **Signes de confiance – l'impact des labels sur la gestion des données personnelles**, Paris. Chaire Valeurs et Politiques des Informations Personnelles, coordonné par Claire Levallois-Barth, 2018.

Consentement : désigne plus particulièrement le recueil du consentement des personnes à autoriser le traitement de leurs données. Le RGPD impose que le consentement soit libre, spécifique, éclairé et univoque.

Cookies : petites quantités d'information (souvent sous forme de petits fichiers) envoyées à l'origine par un serveur web (puis circulant) et contenant des informations pouvant ainsi persister sur plusieurs sessions. L'objectif affiché est d'améliorer la qualité d'expérience utilisateur en lui offrant un service personnalisé, d'éviter qu'il ne se réauthentifie au cours de ses interactions. On parle de cookies directs, cookie interne, cookie d'origine, cookie de domaine – *first-party cookie* en anglais. Le serveur peut également inclure des cookies tiers ou *third-party cookies* en anglais.

Crédentiel : *(credential)* un élément d'information servant à prouver l'identité d'une personne déclarée et prenant la forme d'un mot de passe statique ou à usage unique, d'une signature électronique etc. Dans cet ouvrage, ce terme a été choisi plutôt qu'une traduction plus classique, mais plus limitée, comme « preuve » ou « garantie ».

Directive : acte normatif pris par les institutions de l'Union européenne, donnant des objectifs à atteindre aux pays membres, avec un délai. À ne pas confondre avec *Règlement*.

Divulgation nulle de connaissance : voir : *preuve*

Divulgation sélective des attributs : mécanisme qui permet à l'utilisateur de divulguer le minimum d'attributs nécessaires pour réaliser une transaction électronique.

DNS : *Domain Name System*, service informatique distribué permettant de traduire les noms de domaine Internet en leur adresse IP (ainsi que d'autres enregistrements qui peuvent être associés).

Donnée à caractère personnel : « *toute information se rapportant à une personne physique identifiée ou identifiable (ci-après dénommée «personne concernée»); est réputée être une «personne*

physique identifiable» une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale. » (art. 4 du RGPD)

Doubles paiements : fait référence à une faille dans un système transactionnel qui permettrait de voir se réaliser deux fois l'effet d'une transaction monétaire, le temps que la transaction soit prise en compte.

Efficacité : (*effectiveness*) propension à atteindre un objectif.

Efficience : (*efficiency*) capacité à produire le maximum de résultats avec le minimum d'efforts. Doit bien être distingué de *efficacité*.

Éléments d'intérêt : (*items of interest*) tout élément visant à enrichir la base de connaissance d'un attaquant en vue de mener une attaque. Il peut s'agir des messages, de l'identité des utilisateurs, de leurs actions… Ne pas confondre avec « centres d'intérêt ».

Empreinte (numérique) : 1) ensemble d'informations émises par un terminal et ses applications (ex : type de terminal, version de l'OS, langue, liste des applications, résolution de l'écran…), utilisées par des entités (ex : fournisseurs de services) pour tracer un utilisateur au cours de ses différentes activités sur Internet ; 2) résultat d'une fonction de *hachage*.

Entité : (*entity*) désigne tour à tout un individu, un serveur, une *autorité* de confiance, un fournisseur de services, un *attaquant*… qui sont impliqués dans un échange ou dans le fonctionnement d'un service.

Entrée DNS : ou enregistrement DNS (dans la base de données DNS), comme une adresse IP, un serveur de messagerie associé à un nom de domaine… L'un d'entre eux, appelé `TXT`, permet d'enregistrer des informations sous forme textuelle.

Entropie : mesure mathématique du bruit contenu dans un signal émis par une source, ce qui revient également à mesurer la quantité d'information contenue dans ce signal.

Expérience utilisateur : (*User eXperience, UX*) fait référence à la qualité de l'expérience vécue par les utilisateurs dans leurs interactions avec un service.

Expression régulières : (*regex*) chaîne de caractères qui décrit, selon une syntaxe précise, un ensemble

plus large de chaînes de caractères possibles. Par ex. les entrées en bleu de ce glossaire sont décrites avec `^[^:]*:`. Bien manier les *regex* facilite par exemple le filtrage par mots-clé.

Fog computing : informatique géodistribuée. Désigne l'idée d'effectuer des calculs localement aux objets connectés pour optimiser le volume nécessaire de communications vers les services de traitement distants.

Hachage : (*hash*) fonction unidirectionnelle qui calcule, à partir d'une information en entrée (par exemple un message, un fichier), une *empreinte* de cette information, de taille fixe (taille dépendant de la fonction) dont le résultat est très fortement lié au contenu de l'information en entrée. Le résultat d'un hachage s'appelle, selon le contexte, et selon les propriétés souhaitées, la somme de contrôle, l'empreinte numérique, le condensat, le condensé, l'empreinte cryptographique.

Identité : voir **Identités numériques**, Paris. Chaire Valeurs et Politiques des Informations Personnelles, coordonné par Claire Levallois-Barth, 2016.

Inconscient : (*oblivious*) qualifie dans cet ouvrage des *primitives cryptographiques* (comme le transfert inconscient) ou des techniques (comme le convertisseur inconscient) dans lesquelles une entité effectue son travail sans en connaître tous les tenants, comme par exemple les données parmi celles transmises qui sont véritablement utiles.

Jeton de présentation : (*presentation token*) éléments d'information qui sont transmis sous forme d'une structure de données par un utilisateur à un fournisseur de services pour prouver la validité d'un ou plusieurs de ses attributs dans le cadre de la certification anonyme.

Journalisation : (*log file*) historique daté et classé d'événements, enregistré à des fins d'analyse par exemple.

Lex specialis : loi spéciale qui suit la loi générale *lex generalis* et peut la supplanter sur certains de ses aspects. Provient du latin « *Lex specialis derogat legi generali* ».

Machine Learning : apprentissage machine, une des techniques de l'Intelligence Artificielle qui provoque un regain d'intérêt ces dernières années car, les approches statistiques utilisées confèrent aux ordinateurs une certaine capacité à apprendre à partir de données.

Malware : logiciel malveillant.

Métadonnées : donnée servant à caractériser une donnée ou ensemble de données (ex : numéro de version, date de dernière mise à jour…).

Navigation privée : mode de navigation offert dans plusieurs navigateurs web faisant en sorte que l'historique de navigation et certaines traces comme les *cookies* ne sont pas stockés ou sont automatiquement effacés à la fermeture du navigateur.

Over-The-Top : (*offre hors du fournisseur d'accès à Internet*) désigne les offres des fournisseurs de services contournant le fournisseur d'accès Internet au sens où ce dernier ne participe ni au contrôle ni à la distribution du contenu.

Pervasif : un environnement pervasif, ou ubiquitaire, désigne à la fois l'environnement constitué par les objets communicants, et l'idée que l'informatique s'est diffusée, insinuée, propagée (du latin *pervadere*) dans ces objets, fournissant aux utilisateurs autant de moyens d'accès à la technologie.

PKI : (*Public Key Infrastructure*) infrastructure à *clés* publiques, ensemble d'entités physiques (équipements…), de logiciels et de procédures humaines destinées à gérer les clés publiques des utilisateurs d'un système.

Politique de contrôle d'accès : (*access control policy*) ensemble de règles définies par une organisation pour permettre à des *entités* d'accéder sous certaines conditions à un ensemble de ressources.

Politique de présentation : (*presentation policy*) forme particulière de politique de contrôle d'accès employée dans le cas de la certification anonyme.

Prédicat : en logique mathématique, désigne une propriété d'un objet du domaine considéré.

Preuve : selon le contexte, désigne une preuve d'*identité* (besoin de s'assurer que l'utilisateur qui se présente sous un pseudonyme ou de façon anonyme est bien une personne physique réelle), une preuve de *consentement* (le fournisseur de services en aura besoin pour prouver par la suite que l'utilisateur était d'accord pour la collecte et le traitement ses données) ou encore une preuve d'*attributs* (preuve permettant de s'assurer que l'utilisateur a telle ou telle caractéristique, comme un âge minimum, ou un permis de conduire…). Une *preuve à divulgation nulle de connaissance* est souvent un *protocole* sécurisé dans lequel une *entité* prouve à une autre entité (vérificateur) qu'un *prédicat* est vrai sans révéler aucune autre information.

Primitive cryptographique : algorithme « de bas niveau », brique de base d'un système cryptographique.

Privacy by Design : respect de la vie privée dès la conception. Voir en détail page 21.

Privacy Impact Assessment : (Étude d'Impact sur la Vie Privée) procédure visant, pour une organisation, à identifier et minimiser les risques pesant sur la vie privée et la protection des données personnelles qu'elle traite. Opération à réaliser avant le déploiement d'un système technique et tout au long de son cycle de vie.

Procédure : (dans un système cryptographique) étapes principales des traitements réalisés en soutien à un protocole cryptographique. Voir page 64.

Protocole : ensemble de règles communément admises, souvent sous forme de standards, de sorte que les systèmes aient un comportement normatif et soient interopérables. HTTP est un protocole de communication.

Pseudo-anonymat : mécanisme d'anonymat dans lequel l'*identité* n'est (donc) pas connue, mais peut être révélée si nécessaire.

Pseudonymat : relatif aux pseudonymes, qui se présentent sous la forme d'alias ou de noms d'emprunt, pour dissimuler l'identité de la personne, le pseudonymat pouvant éventuellement être levé par certaines *entités*.

Qualité d'expérience : (*Quality of Experience, QoE*) ensemble des caractéristiques, objectives et subjectives, susceptibles de satisfaire (et donner *confiance*, fidéliser…) une personne utilisant un service ou un produit.

Règlement : acte juridique de l'Union européenne, directement applicable sans aucune transcription nationale, de manière simultanée et uniforme à l'ensemble des États membres de l'Union.

Requête HTTP : requête transmise par un client à destination d'un serveur, et respectant le standard HyperText Transfer Protocol, protocole développé pour le Web.

Réseaux ad hoc : réseau d'objets ne s'appuyant pas sur une infrastructure de réseau, chaque objet devenant un nœud du réseau participant de manière dynamique au routage des données y circulant.

Responsabilité (obligation de) : (*accountability*) principe introduit par le RGPD selon lequel tout responsable de traitement qui traite des données

personnelles doit pouvoir démontrer à tout moment qu'il respecte les obligations définies par le RGPD.

RSA : relatif au chiffrement RSA, algorithme cryptographique asymétrique du nom de ses inventeurs, Ronald Rivest, Adi Shamir et Leonard Adleman.

Serveurs / services LBS : LBS pour *Location-based service*.

Serveurs proxy : désigne aussi bien le matériel que le logiciel placé comme intermédiaire entre deux entités pour offrir un certain type de services, comme par exemple l'accélération de la navigation (grâce à des proxy qui hébergent des caches DNS), ou encore une navigation anonyme avec l'intervention d'un proxy anonymiseur ou bien d'un ensemble de proxys Tor (qui mettent en œuvre un routage en oignon anonymisé).

Signature électronique : un élément d'information servant à prouver l'authenticité d'un document ou d'un message transmis. La signature est générée sur ce document ou message à l'aide de la *clé cryptographique* privée de l'*entité* émettrice. La vérification de la signature nécessite la connaissance de la clé publique complémentaire qui est généralement publiée dans un *certificat électronique*. Notons que dans le cas de signatures basées sur les *attributs*, une signature est générée à l'aide de plusieurs clés privées (une clé par attribut) et est vérifiée à l'aide de l'arbre d'accès, ce qui prouve que le signataire satisfait bien certaines conditions de possession d'attributs.

Smart city : (*ville intelligente, ville sensible*) zone urbaine utilisant les données numériques produites en son sein (habitants, visiteurs, services publics, capteurs…) ou externes pour permettre une gestion et une vie quotidienne améliorées.

Solomo : acronyme et concept marketing pour « social, local, mobile ».

Table de hachage distribuée : (*distributed hash table, DHT*) permet d'identifier et de retrouver une information dans un système réparti sur la base d'une structure de données, appelée table de hachage distribuée, de type clé-valeur.

Tiers de confiance : entité en laquelle il est nécessaire d'avoir confiance pour qu'un système fonctionne correctement avec des garanties quant à la sécurité ou au respect de la vie privée. Il peut s'agir d'autorités émettrices de certificats électroniques, ou des anonymiseurs sur lesquels les utilisateurs reportent leur confiance pour assurer l'anonymat des échanges avec un serveur (voir page 61).

Bibliographie

Cette section reprend les références bibliographiques (publications scientifiques et autres textes) citées dans cet ouvrage. La lecture des deux premiers livres de la Chaire Valeurs et Politiques des Informations Personnelles est également recommandée. Quelques publications scientifiques majeures, qui ont ouvert la voie à de nouvelles pistes de recherche, sont indiquées par un ⓘ.

_Textes introductifs

[1] **How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read.** Bernard Marr, Forbes, 21 mai 2018

[2] Claudio Bettini and Daniele Riboni. **Privacy protection in pervasive systems: State of the art and technical challenges.** Pervasive and Mobile Computing, 17:159–174, 2015.

[3] Sur le principe de commodité, voir : Du sujet de droit au sujet libidinal, L'emprise du numérique sur nos sociétés, Mark Hunyadi, Esprit.

[4] Johann Čas. **Ubiquitous computing, privacy and data protection: Options and limitations to reconcile the unprecedented contradictions.** In Computers, privacy and data protection: An element of choice, pages 139–169. Springer, 2011

[5] Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (Règlement général sur la protection des données), JOUE L 119/1 du 4 mai 2016. (accès PDF 88 pages)

[6] Alfred Kobsa. **Privacy-enhanced personalization.** Communications of the ACM, 50(8):24–33, 200

[7] Eran Toch, Yang Wang, and Lorrie Faith Cranor. **Personalization and privacy: a survey of privacy risks and remedies in personalization-based**

systems. User Modeling and User-Adapted Interaction, 22(1-2):203–220, 2012

[8] Javier Parra-Arnau, David Rebollo-Monedero, and Jordi Forné **Privacy-enhancing technologies and metrics in personalized information systems.** In Advanced Research in Data Privacy, pages 423–442. Springer, 2015.

[9] Rapport de 2015 de l'Agence européenne chargée de la sécurité des réseaux et de l'information (European Union Agency for Network and Information Security, ENISA)

[10] George Danezis, Josep Domingo-Ferrer, Marit Hansen, Jaap-Henk Hoepman, Daniel Le Metayer, Rodica Tirtea, and Stefan Schiffner. **Privacy and data protection by design-from policy to engineering.** arXiv preprint arXiv:1501.03726, 2015.

[11] Fiche n°3 **Réflexions sur l'évolution de l'identité numérique en sciences informatiques**, du Cahier n°1, **Identités numériques**, de la Chaire Valeurs et Politiques des Informations Personnelles

[12] Andreas Pfitzmann and Marit Hansen. **A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management.** 2010

[13] Ian MacKenzie, Chris Meyer, and Steve Noble. **How retailers can keep up with consumers.** McKinsey & Company, 2013.

[14] Andreas Krause and Eric Horvitz. **A utility-theoretic approach to privacy in online services.** Journal of Artificial Intelligence Research,. 39:633–662, 2010.

[15] Katie Hafner. **Researchers yearn to use aol logs, but they hesitate.** New York Times, 23, 2006

[16] Wencheng Sun, Zhiping Cai, Yangyang Li, Fang Liu, Shengqun Fang, and Guoyan Wang. **Security and privacy in the medical internet of things: A review.** Security and Communication Networks, 2018

[17] Javier Parra-Arnau, David Rebollo-Monedero, and Jordi Forné. **Measuring the privacy of user profiles in personalized information systems.** Future Generation Computer Systems, 33:53–63, 2014

[18] Shengling Wang, Qin Hu, Yunchuan Sun, and Jianhui Huang. **Privacy preservation in location-based services.** IEEE Communications Magazine, 56(3):134–140, 2018

[19] Why companies want to mine the secrets in your voice. Voices are highly personal, hard to fake, and contain surprising information about our mental

health and behaviors. Angela Chen, The Verge, Mar 14, 2019

[20] Pourquoi faire analyser son ADN pour connaître ses origines est une très mauvaise idée. Cécilia Leger, Numérama, 4 novembre 2018

[21] Taking Back Our Right to Privacy. Blog Telegram, Mar 24, 2019

_Les techniques orientées utilisateur

[22] Règlement Général sur la Protection des Données (Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (Règlement général sur la protection des données), JOUE L 119/1 du 4 mai 2016

[23] RFC 2547, BGP/MPLS VPNs, E. Rosen, Y. Rekhter, 1999.

[24] Naganand Doraswamy and Dan Harkins. **IPSec: the new security standard for the Internet, intranets, and virtual private networks.** Prentice Hall Professional, 2003

[25] Stefano Traverso, Martino Trevisan, Leonardo Giannantoni, Marco Mellia, and Hassan Metwalley. **Benchmark and comparison of trackerblockers: Should you trust them?** In Network Traffic Measurement and Analysis Conference (TMA), 2017, pages 1–9. IEEE, 2017

[26] Jan Camenisch, Stephan Krenn, Anja Lehmann, Gert L. Mikkelsen, Gregory Neven, and Michael O. Pederson. **Scientific comparison of abc protocols: Part i – formal treatment of privacy enhancing credential systems**, 2014.

[27] David Chaum. **Blind signatures for untraceable payment.** In Advances in Cryptology: Proceedings of Crypto'82, 1982.

[28] Jan Camenisch and Anna Lysyanskaya. **An efficient system for nontransferable anonymous credentials with optional anonymity revocation.** In Proceedings of the International Conference on the Theory and Application of Cryptographic Techniques: Advances in Cryptology, EUROCRYPT'01, London, UK, UK, 2001. Springer-Verlag

[29] **« La gestion des identités numériques »**, (Ed. M. Laurent, S. Bouzefrane), collection ISTE, ISBN: 978-1-78405-056-6 (papier), ISBN : 978-1-78406-056-5 (ebook), 2015.

[30] Jan Camenisch and Els Van Herreweghen. **Design and implementation of the idemix anonymous credential system.** In Proceedings of the 9th ACM Conference on Computer and Communications Security, CCS'02, New York, NY, USA, 2002 ACM

[31] Microsoft. **U-prove community technology.** 2013

[32] Stefan A. Brands. **Rethinking Public Key Infrastructures and Digital Certificates: Building in Privacy.** MIT Press, Cambridge, MA, USA, 2000.

[33] Hemanta K. Maji, Manoj Prabhakaran, and Mike Rosulek. **Attribute based signatures.** Cryptology ePrint Archive, Report 2010/595, 2010

[34] Hemanta K. Maji, Manoj Prabhakaran, and Mike Rosulek. **Attribute based signatures.** Cryptology ePrint Archive, Report 2010/595, 2010

[35] Tatsuaki Okamoto and Katsuyuki Takashima. **Efficient attribute-based signatures for non-monotone predicates in the standard model.** PKC'11, 2011

[36] Jin Li, Man Ho Au, Willy Susilo, Dongqing Xie, and Kui Ren. **Attribute-based signature and its applications.** ASIACCS '10, 2010

[37] Siamak F. Shahandashti and Reihaneh Safavi-Naini. **Threshold attribute-based signatures and their application to anonymous credential systems.** AFRICACRYPT '09, 2009

[38] Javier Herranz, Fabien Laguillaumie, Benoit Libert, and Carla Rafols. **Short attribute-based signatures for threshold predicates.** In Topics in Cryptology – CT-RSA 2012. 2012

[39] Yan Zhang and Dengguo Feng. **Efficient attribute proofs in anonymous credential using attribute-based cryptography.** In Proceedings of the 14th International Conference on Information and Communications Security, ICICS'12, 2012

[40] Siamak F. Shahandashti and Reihaneh Safavi-Naini. **Threshold attribute-based signatures and their application to anonymous credential systems.** AFRICACRYPT '09, 200

[41] Hemanta K. Maji, Manoj Prabhakaran, and Mike Rosulek. **Attribute-based signatures.** Cryptology ePrint Archive, Report 2010/595, 2010

[42] Ali El Kaafarani, Essam Ghadafi, and Dalia Khader. **Decentralized traceable attribute-based signatures.** In Topics in Cryptology CT-RSA 2014

[43] Nesrine Kaâniche and Maryline Laurent. **Attribute-based signatures for supporting anonymous certification.** In European

Symposium on Research in Computer Security, pages 279–300. Springer, 2016

[44] Jan Camenisch and Anna Lysyanskaya. **An efficient system for nontransferable anonymous credentials with optional anonymity revocation.** In Proceedings of the International Conference on the Theory and Application of Cryptographic Techniques: Advances in Cryptology, EUROCRYPT '01, London, UK, UK, 2001. Springer-Verlag.

[45] Jan Camenisch and Anna Lysyanskaya. **A signature scheme with efficient protocols.** In SCN, volume 2576 of Lecture Notes in Computer Science. Springer, 2002

[46] Haibin Zheng, Qianhong Wu, Bo Qin, Lin Zhong, Shuangyu He, and Jianwei Liu. **Linkable group signature for auditing anonymous communication.** In Australasian Conference on Information Security and Privacy, pages 304–321. Springer, 2018

[47] Jörg Helbach, Jörg Schwenk, Sven Schäge, and Bregenz EVOTE08. **Code voting with linkable group signatures.** Electronic Voting, 50, 2008

[48] Liang Yan, Zhang Xiao, and Zheng Zhiming. **An electronic cash system based on certificateless group signature.** International Journal of Security and Its Applications, 10(2):287–300, 2016

[49] L. Malina, J. Smrz, J. Hajny, and K. Vrba. **Secure electronic voting based on group signatures.** In Telecommunications and Signal Processing (TSP), 2015 38th International Conference on, pages 6–10. IEEE, 2015

[50] Nadia El Mrabet and Marc Joye. **Guide to Pairing-Based Cryptography.** CRC Press, 2017

[51] Alexander Schneider, Christian Meter, and Philipp Hagemeister. **Survey on remote electronic voting.** arXiv preprint arXiv:1702.02798, 2017

[52] Lukas Malina, Jan Smrz, Jan Hajny, and Kamil Vrba. **Secure electronic voting based on group signatures.** In Telecommunications and Signal Processing (TSP), 2015 38th International Conference on, pages 6–10. IEEE, 2015

[53] Giuseppe Ateniese and Gene Tsudik. **Some open issues and new directions in group signatures.** In International Conference on Financial Cryptography, pages 196–211. Springer, 1999

[54] Anna Lysyanskaya and Zulfikar Ramzan. **Group blind digital signatures: A scalable solution to electronic cash.** In International Conference on Financial Cryptography, pages 184–197. Springer, 199

[55] Liang Yan, Zhang Xiao, and Zheng Zhi-ming. **An electronic cash system based on certificateless group signature.** International Journal of Security and Its Applications, 10(2):287–300, 2016

[56] Greg Maitland and Colin Boyd. **Fair electronic cash based on a group signature scheme.** In International Conference on Information and Communications Security, pages 461–465. Springer, 2001

[57] Giuseppe Ateniese, Daniel H. Chou, Breno de Medeiros, and Gene Tsudik. **Sanitizable signatures.** In Proceedings of the 10th European Conference on Research in Computer Security, ESORICS'05, pages 159–177, Berlin, Heidelberg, 2005. Springer-Verlag

[58] Sébastien Canard and Amandine Jambert. **On extended sanitizable signature schemes.** In Proceedings of the 2010 International Conference on Topics in Cryptology, CT-RSA'10, Berlin, Heidelberg, 2010. Springer-Verlag.

[59] Sébastien Canard, Amandine Jambert, and Roch Lescuyer. **Sanitizable signatures with several signers and sanitizers.** In Proceedings of the 5th International Conference on Cryptology in Africa, AFRICACRYPT'12, Berlin, Heidelberg, 2012. Springer-Verlag

[60] S. S. Chow, Y.-J. He, L. C. Hui, and S. M. Yiu. **Spice-simple privacy-preserving identity-management for cloud environment.** In International Conference on Ap- plied Cryptography and Network Security, pages 526-543. Springer, 2012

[61] D. Pàmies-Estrems, N. Kaaniche, M. Laurent, J. Castella-Roca, and J. Garcia- Alfaro. **Lifelogging protection scheme for internet-based personal assistants.** In Data Privacy Management, Cryptocurrencies and Blockchain Technology, pages 431-440. Springer, 2018.

[62] Sébastien Canard and Amandine Jambert. **On extended sanitizable signature schemes.** In Proceedings of the 2010 International Conference on Topics in Cryptology, CT-RSA'10, Berlin, Heidelberg, 2010. Springer-Verlag

[63] David Chaum **Blind signatures for untraceable payment.** In Advances in Cryptology: Proceedings of Crypto'82, 1982

[64] A. Lysyanskaya and Z. Ramzan. **Group blind digital signatures: A scalable solution to electronic cash.** In International Conference on Financial Cryptography, pages 184-197. Springer, 1998.

[65] C.-P. Schnorr. **Efficient identication and signatures for smart cards.** In Conference on the Theory and Application of Cryptology, pages 239-252. Springer, 1989

[66] E. R. Verheul. **Self-blindable credential certicates from the weil pairing.** In C. Boyd, editor, ASIACRYPT, volume 2248 of Lecture Notes in Computer Science. Springer, 2001

[67] D. Chaum and T. P. Pedersen. **Wallet databases with observers.** In Proceedings of the 12th Annual International Cryptology Conference on Advances in Cryptology, CRYPTO '92, pages 89-105, London, UK, UK, 1993. Springer-Verlag

[68] Stefan A. Brands. **Rethinking Public Key Infrastructures and Digital Certificates: Building in Privacy.** MIT Press, Cambridge, MA, USA, 2000.

[69] W. Mostowski and P. Vullers. **Efficient U-Prove implementation for anonymous credentials on smart cards.** 2012

[70] P. Vullers and G. Alpar. **Efficient selective disclosure on smart cards using Idemix.** 2013

[71] S. Canard and R. Lescuyer. **Protecting privacy by sanitizing personal data: A new approach to anonymous credentials.** In Proceedings of the 8th ACM SIGSAC Symposium on Information, Computer and Communications Security, ASIA CCS '13, New York, NY, USA, 2013. ACM

[72] N. Kaaniche and M. Laurent. **Attribute-based signatures for supporting anonymous certication.** In European Symposium on Research in Computer Security, pages 279-300. Springer, 2016.

[73] S. Canard, A. Jambert, and R. Lescuyer. **Sanitizable signatures with several signers and sanitizers.** In Proceedings of the 5th International Conference on Cryptology in Africa, AFRICACRYPT'12, Berlin, Heidelberg, 2012. Springer-Verlag

[74] Aniket Pingley, Nan Zhang, Xinwen Fu, Hyeong-Ah Choi, Suresh Subramaniam, and Wei Zhao. **Protection of query privacy for continuous location based services.** In Infocom, 2011 Proceedings IEEE, pages 1710–1718. IEEE, 2011

[75] Ben Niu, Qinghua Li, Xiaoyan Zhu, Guohong Cao, and Hui Li. **Achieving k-anonymity in privacy-aware location-based services.** In INFOCOM, 2014 Proceedings IEEE, pages 754–762.IEEE, 2014

[76] Yuval Elovici, Bracha Shapira, and Adlay Meshiach. **Cluster-analysis attack against a private web solution (PRAW).** Online Information Review, 30(6):624–643, 2006

[77] Shaozhi Ye, Felix Wu, Raju Pandey, and Hao Chen. **Noise injection for search privacy protection.** In Computational Science and Engineering, 2009. CSE'09. International Conference on, volume 3, pages 1–8. IEEE, 2009

[78] Rahat Masood, Dinusha Vatsalan, Muhammad Ikram, and Mohamed Ali Kaafar. **Incognito: A method for obfuscating web data.** In Proceedings of the 2018 World Wide Web Conference on World Wide Web, pages 267–276. International World Wide Web Conferences Steering Committee, 2018

[79] Javier Parra-Arnau, David Rebollo-Monedero, and Jordi Forné. **A privacy-preserving architecture for the semantic web based on tag suppression.** In International Conference on Trust, Privacy and Security in Digital Business, pages 58–68. Springer, 2010.

[80] Javier Parra-Arnau, David Rebollo-Monedero, Jordi Forné Jose L Muñoz, and Oscar Esparza. **Optimal tag suppression for privacy protection in the semantic web.** Data & Knowledge Engineering, 81:46–66, 2012

[81] Javier Parra-Arnau, David Rebollo-Monedero, and Jordi Forné. **Optimal forgery and suppression of ratings for privacy enhancement in recommendation systems.** Entropy, 16(3):1586–1631, 2014

[82] Nikolaos Polatidis, Christos K Georgiadis, Elias Pimenidis, and Haralambos Mouratidis. **Privacy-preserving collaborative recommendations based on random perturbations.** Expert Systems with Applications, 71:18–25, 2017.

[83] Andrew C. Yao. **Protocols for secure computations.** In Proceedings of the 23rd Annual Symposium on Foundations of Computer Science, SFCS '82, pages 160–164, Washington, DC, USA, 1982. IEEE Computer Society

[84] Wenliang Du and Mikhail J Atallah. **Privacy-preserving cooperative statistical analysis**. In Computer Security Applications Conference, 2001. ACSAC 2001. Proceedings 17th Annual, pages 102–110. IEEE, 2001

[85] Michael K Reiter and Aviel D Rubin. **Crowds: Anonymity for web transactions.** ACM transactions on information and system security (TISSEC), 1(1):66–92, 1998

[86] Carolina Tripp Barba, Luis Urquiza Aguiar, Mónica Aguilar Igartua, Javier Parra-Arnau, David Rebollo-Monedero, Jordi Forné, and Esteve Pallarès. **A collaborative protocol for anonymous reporting in vehicular ad hoc networks.** Computer Standards & Interfaces, 36(1):188–197, 2013

[87] Rohit Pathak and Satyadhar Joshi. **Smc protocol for privacy preserving in banking computations along with security analysis.** In Internet, 2009. AH-ICI 2009. First Asian Himalayas International Conference on, pages 1–5. IEEE, 2009

[88] M. J. Fischer, S. Micali, and C. Rackoff. A **secure protocol for oblivious transfer (extended abstract).** In Journal of Cryptology, pages 191–195. Springer-Verlag, 1996.

[89] Ronald Cramer, Ivan Bjerre Damgård, et al. **Secure multiparty computation.** Cambridge University Press, 2015.

[90] Yehuda Lindell and Benny Pinkas. **Secure multiparty computation for privacy-preserving data mining.** Journal of Privacy and Confidentiality, 1(1):5, 2009.

[91] Ronald Cramer, Ivan Damgård, and Ueli Maurer. **General secure multiparty computation from any linear secret-sharing scheme.** In Advances in CryptologyEUROCRYPT 2000, pages 316–334. Springer, 2000.

[92] C. Gentry. ⓘ **A fully homomorphic encryption scheme.** PhD thesis, Stanford, CA, USA, 2009.

[93] Arnau Erola, Jordi Castellà-Roca, Alexandre Viejo, and Josep M Mateo-Sanz. **Exploiting social networks to provide privacy in personalized web search.** Journal of Systems and Software, 84(10):1734–1745, 2011.

[94] David Rebollo-Monedero, Jordi Forne, and Josep Domingo-Ferrer. **Query profile obfuscation by means of optimal query exchange between users.** IEEE Transactions on Dependable and Secure Computing, 9(5):641–654, 2012

[95] Jaydip Sen. **Privacy preservation technologies in internet of things.** In Proceedings of the International Conference on

Emerging Trends in Mathematics, Technology and Management, pages 496–504, 2010.

[96] Samet Tonyali, Kemal Akkaya, Nico Saputro, A Selcuk Uluagac, and Mehrdad Nojoumian. **Privacy preserving protocols for secure and reliable data aggregation in IoT-enabled smart metering systems.** Future Generation Computer Systems, 2017.

_Les techniques orientées serveur

[97] D. Boneh, E. Goh, and K. Nissim. ⓘ **Evaluating 2-dnf formulas on ciphertexts.** In Proceedings of the Second international conference on Theory of Cryptography, TCC'05, Berlin, Heidelberg, 2005. Springer-Verlag.

[98] Zekeriya Erkin, Thijs Veugen, Tomas Toft, and Reginald L Lagendijk. **Generating private recommendations efficiently using homomorphic encryption and data packing.** IEEE transactions on information forensics and security, 7(3):1053–1066, 2012.

[99] Shahriar Badsha, Xun Yi, and Ibrahim Khalil. **A practical privacypreserving recommender system.** Data Science and Engineering, 1(3):161–177, 2016.

[100] T. Elgamal. ⓘ **A public key cryptosystem and a signature scheme based on discrete logarithms.** volume 31, page 469472, 1985.

[101] Shahriar Badsha, Xun Yi, Ibrahim Khalil, and Elisa Bertino. **Privacy preserving user-based recommender system.** In Distributed Computing Systems (ICDCS), 2017 IEEE 37th International Conference on, pages 1074–1083. IEEE, 2017.

[102] Carlos Aguilar-Melchor, Joris Barrier, Laurent Fousse, and Marc-Olivier Killijian. **Xpir: Private information retrieval for everyone.** Proceedings on Privacy Enhancing Technologies, 2016(2):155–174, 2016

[103] Prateek Mittal, Femi G Olumofin, Carmela Troncoso, Nikita Borisov, and Ian Goldberg. **Pir-Tor: Scalable anonymous communication using private information retrieval.** In USENIX Security Symposium, page 31, 2011

[104] Ning Cao, Cong Wang, Ming Li, Kui Ren, and Wenjing Lou. **Privacy preserving multi-keyword ranked search over encrypted cloud data.** IEEE Transactions on parallel and distributed systems, 25(1):222–233, 2014

[105] Ibrahim Lazrig, Tarik Moataz, Indrajit Ray, Indrakshi Ray, Toan Ong, Michael Kahn, Frédéric

Cuppens, and Nora Cuppens. **Privacy preserving record matching using automated semi-trusted broker.** In IFIP Annual Conference on Data and Applications Security and Privacy, pages 103–118. Springer, 2015

[106] Foteini Baldimtsi and Olga Ohrimenko. **Sorting and searching behind the curtain.** In International Conference on Financial Cryptography and Data Security, pages 127–146. Springer, 2015

[107] Chi Chen, Xiaojie Zhu, Peisong Shen, Jiankun Hu, Song Guo, Zahir Tari, and Albert Y Zomaya. **An efficient privacy-preserving ranked keyword search method.** IEEE Transactions on Parallel and Distributed Systems, 27(4):951–963, 2016

[108] Payal Chaudhari and Manik Lal Das. **Privacy-preserving attribute based searchable encryption.** IACR Cryptology ePrint Archive, 2015:899, 2015

[109] Mikhail J Atallah and Keith B Frikken. **Privacy-preserving location-dependent query processing.** In Pervasive Services, 2004. ICPS 2004. Proceedings. The IEEE/ACS International Conference on, pages 9–17. IEEE, 2004

[110] William Gasarch. **A survey on private information retrieval.** The Bulletin of the EATCS, 82(72-107):1, 2004

[111] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. **Private information retrieval.** volume 45, nov 1998

[112] Nesrine Kaâniche and Maryline Laurent. **Data security and privacy preservation in cloud storage environments based on cryptographic mechanisms.** Computer Communications, 111:120–141, 2017

[113] Russell Paulet, Md Golam Kaosar, Xun Yi, and Elisa Bertino. **Privacy-preserving and content-protecting location based queries.** IEEE transactions on knowledge and data engineering, 26(5):1200–1210, 2014.

[114] M. Ullah, R. Khan, and M. A. Islam. **Poshida, a protocol for private information retrieval.** In Innovative Computing Technology (INTECH), 2016 Sixth International Conference on, pages 464–470. IEEE, 2016

[115] R. Ostrovsky and William E. Skeith, III. **A survey of single-database private information retrieval: Techniques and applications.** In Proceedings of the 10th International Conference

on Practice and Theory in Public-key Cryptography, PKC'07, Berlin, Heidelberg, 2007.Springer-Verlag

[116] Zhicheng Dou, Ruihua Song, and Ji-Rong Wen. **A large-scale evaluation and analysis of personalized search strategies.** In Proceedings of the 16th international conference on World Wide Web, pages 581–590. ACM, 2007

[117] Wenliang Du and Mikhail J Atallah. **Privacy-preserving cooperative statistical analysis.** In Computer Security Applications Conference, 2001. ACSAC 2001. Proceedings 17th Annual, pages 102–110. IEEE, 2001.

[118] Ahmed Kosba, Andrew Miller, Elaine Shi, Zikai Wen, and Charalampos Papamanthou. Hawk: **The blockchain model of cryptography and privacy-preserving smart contracts.** In 2016 IEEE symposium on security and privacy (SP), pages 839–858. IEEE, 2016

[119] Benedikt Bünz, Jonathan Bootle, Dan Boneh, Andrew Poelstra, Pieter Wuille, and Greg Maxwell. **Bulletproofs: Short proofs for confidential transactions and more.** In Bulletproofs: Short

[120] Rohit Pathak and Satyadhar Joshi. **SMC protocol for privacy preserving in banking computations along with security analysis.** In Internet, 2009. AH-ICI 2009. First Asian Himalayas International Conference on, pages 1–5. IEEE, 2009

[121] Zoubin Ghahramani. **Probabilistic machine learning and artificial intelligence.** Nature, 521(7553):452, 2015

[122] Gulden Uchyigit and Matthew Y Ma. **Personalization techniques and recommender systems**, volume 70. World Scientific, 2008

[123] Ning Cao, Cong Wang, Ming Li, Kui Ren, and Wenjing Lou. **Privacypreserving multi-keyword ranked search over encrypted cloud data.** IEEE Transactions on parallel and distributed systems, 25(1):222–233, 2014.

[124] I. H. Witten, A. Moat, and T. C. Bell. **Managing gigabytes: compressing and indexing documents and images.** Morgan Kaufmann, 1999.

[125] W. K.Wong, D. W.-l. Cheung, B. Kao, and N. Mamoulis. **Secure knn computation on encrypted databases.** In Proceedings of the 2009 ACM SIGMOD International Conference on Management of data, pages 139{152. ACM, 2009.

[126] Roxana Geambasu, Tadayoshi Kohno, Amit A Levy, and Henry M Levy. **Vanish: Increasing data**

privacy with self-destructing data. In USENIX Security Symposium, volume 316, 2009

[127] Scott Wolchok, Owen S Hofmann, Nadia Heninger, Edward W Felten, J Alex Halderman, Christopher J Rossbach, Brent Waters, and Emmett Witchel. **Defeating vanish with low-cost sybil attacks against large dhts.** In NDSS, 2010.

[128] Lingfang Zeng, Zhan Shi, Shengjie Xu, and Dan Feng. **Safevanish: An improved data self-destruction for protecting data privacy.** In 2nd IEEE International Conference on Cloud Computing Technology and Science, pages 521–528. IEEE, 2010.

[129] Radia Perlman. **The ephemerizer: Making data disappear.** 2005

[130] Apostolis Zarras, Katharina Kohls, Markus D¨urmuth, and Christina Pöpper. **Neuralyzer: flexible expiration times for the revocation of online data.** In Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy, pages 14–25. ACM, 2016.

[131] Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer, and Muthuramakrishnan Venkitasubramaniam. **l-diversity: Privacy beyond k-anonymity.** In Data Engineering, 2006. ICDE'06. Proceedings of the 22nd International Conference on, pages 24–24. IEEE, 2006.

[132] Josep Domingo-Ferrer and Vicenc¸ Torra. **A critique of k-anonymity and some of its enhancements.** In Availability, Reliability and Security, 2008. ARES 08. Third International Conference on, pages 990–993. IEEE, 2008

[133] Daniele Riboni, Linda Pareschi, and Claudio Bettini. **Shadow attacks on users anonymity in pervasive computing environments.** Pervasive and Mobile Computing, 4(6):819–835, 2008.

[134] Fuyu Liu, Kien A Hua, and Ying Cai. **Query l-diversity in location based services**. In 2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware, pages 436–442. IEEE, 2009

[135] Bhuvan Bamba, Ling Liu, Peter Pesti, and Ting Wang. **Supporting anonymous location queries in mobile environments with privacy grid.** In Proceedings of the 17th international conference on World Wide Web, pages 237–246. ACM, 2008.

[136] Mingqiang Xue, Panos Kalnis, and Hung Keng Pung. **Location diversity: Enhanced privacy protection in location based services.** In

International Symposium on Location-and Context-Awareness, pages 70–87. Springer, 2009.

[137] Yong Wang, Yun Xia, Jie Hou, Shi-meng Gao, Xiao Nie, and Qi Wang. **A fast privacy-preserving framework for continuous location-based queries in road networks.** Journal of Network and Computer Applications, 53:57–73, 2015.

[138] Ayong Ye, Yacheng Li, Li Xu, Qing Li, and Hui Lin. **A trajectory privacy-preserving algorithm based on road networks in continuous location-based services.** In Trustcom/BigDataSE/ICESS, 2017 IEEE, pages 510–516. IEEE, 2017

[139] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. **t-closeness: Privacy beyond k-anonymity and l-diversity.** In Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on, pages 106–115. IEEE, 2007

[140] Louis-Philippe Sondeck, Maryline Laurent, and Vincent Frey. **The semantic discrimination rate metric for privacy measurements which questions the benefit of t-closeness over l-diversity.** In SECRYPT 2017: 14th International Conference on Security and Cryptography, volume 6, pages 285–294. Scitepress, 2017

[141] Daniele Riboni and Claudio Bettini. **Differentially-private release of check-in data for venue recommendation.** In Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on, pages 190–198. IEEE, 2014

[142] Rui Chen, Benjamin CM Fung, S Yu Philip, and Bipin C Desai. **Correlated network data publication via differential privacy.** The VLDB Journal, 23(4):653–676, 2014

[143] Hien To, Gabriel Ghinita, Liyue Fan, and Cyrus Shahabi. **Differentially private location protection for worker datasets in spatial crowdsourcing.** IEEE Transactions on Mobile Computing, 16(4):934–949, 2017.

[144] Hassan Jameel Asghar, Paul Tyler, and Mohamed Ali Kaafar. **Differentially private release of public transport data: The opal use case.** arXiv preprint arXiv:1705.05957, 2017.

[145] Graham Cormode, Somesh Jha, Tejas Kulkarni, Ninghui Li, Divesh Srivastava, and Tianhao Wang. **Privacy at scale: Local differential privacy in practice.** In Proceedings of the 2018 International Conference on Management of Data, pages 1655–1658. ACM, 2018.

_Les techniques orientées canal

[146] Naganand Doraswamy and Dan Harkins. **IPsec: the new security standard for the Internet, intranets, and virtual private networks.** Prentice Hall Professional, 2003.

[147] T. Dierks and E. Rescorla. **RFC 5246 - the transport layer security (TLS) protocol version 1.2.** Technical report, aug

[148] Olivier Levillain. **A study of the TLS ecosystem.** PhD thesis, Institut National des Télécommunications, 2016

[149] Tatu Ylonen and Chris Lonvick. **The secure shell (ssh) protocol architecture.** 2006

[150] Raluca Ada Popa, Nickolai Zeldovich, Sanjeev Verma, Randall Steven Battat, and Aaron Delano Burrow. **Secure sharing, April 24 2018. US Patent 9,954,684.**

[151] Ksenia Ermoshina, Francesca Musiani, and Harry Halpin. **End-to-end encrypted messaging protocols: An overview.** In International Conference on Internet Science, pages 244–254. Springer, 2016.

[152] Blake Ramsdell. **Secure/multipurpose internet mail extensions (s/mime) version 3.1 message specification.** 2004.

[153] Jan Camenisch, Stephan Krenn, Anja Lehmann, Gert Læssøe Mikkelsen, Gregory Neven, and Michael Østergaard Pedersen. **Formal treatment of privacy-enhancing credential systems.** In International Conference on Selected Areas in Cryptography, pages 3–24. Springer, 2015.

[154] Jan Camenisch and Anja Lehmann. **(un) linkable pseudonyms for governmental databases.** In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, pages 1467– 1479. ACM, 2015

[155] Jan Camenisch and Anja Lehmann. **Privacy-preserving user-auditable pseudonym systems.** In IEEE European Symposium on Security and Privacy (EuroS&P), pages 269–284. IEEE, 2017

[156] Nesrine Kaâniche and Maryline Laurent. **Blockchain-based data usage auditing.** In IEEE International Conference on Cloud Computing 2018. IEEE, 2018.

_

[157] L. Wang, D. Yang, X. Han, T. Wang, D. Zhang, and X. Ma. **Location privacy-preserving task allocation for mobile crowdsensing with dierential geoobfuscation.** In Proceedings of the 26th International Conference on World Wide Web, pages 627-636. International World Wide Web Conferences Steering Committee, 2017

[158] M. A. Rahman, M. H. Manshaei, E. Al-Shaer, and M. Shehab. **Secure and private data aggregation for energy consumption scheduling in smart grids.** IEEE Transactions on Dependable and Secure Computing, 14(2):221-234, 2017.

[159] K. Ermoshina, F. Musiani, and H. Halpin. **End-to-end encrypted messaging protocols: An overview.** In International Conference on Internet Science, pages 244-254. Springer, 2016

[160] **Signes de confiance – l'impact des labels sur la gestion des données personnelles**, Paris, Chapitre 11. Chaire Valeurs et Politiques des Informations Personnelles, coordonné par Claire Levallois-Barth, 2018

[161] Melanie Swan. **Blockchain: Blueprint for a new economy.** O'Reilly Media, Inc., 2015.

[162] Michael Crosby, Pradan Pattanayak, Sanjeev Verma, and Vignesh Kalyanaraman. **Blockchain technology: Beyond bitcoin.** Applied Innovation, 2:6–10, 2016.

[163] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M Voelker, and Stefan Savage. **A fistful of bitcoins: characterizing payments among men with no names.** In Proceedings of the 2013 conference on Internet measurement conference, pages 127–140. ACM, 2013.

[164] Dorit Ron and Adi Shamir. **Quantitative analysis of the full bitcoin transaction graph.** In International Conference on Financial Cryptography and Data Security, pages 6–24. Springer, 2013

[165] Zoubin Ghahramani. **Probabilistic machine learning and artificial intelligence.** Nature, 521(7553):452, 2015.

[166] Gulden Uchyigit and Matthew Y Ma. **Personalization techniques and recommender systems,** volume 70. World Scientific, 2008

[167] Pavel Hamet and Johanne Tremblay. **Artificial intelligence in medicine.** Metabolism, 69:S36–S40, 2017

[168] Peter Brusilovski, Alfred Kobsa, and Wolfgang Nejdl. **The adaptive web: methods and strategies of web personalization,** volume 4321. Springer Science & Business Media, 2007.

[169] Valeria Nikolaenko, Stratis Ioannidis, Udi Weinsberg, Marc Joye, Nina Taft, and Dan Boneh. **Privacy-preserving matrix factorization.** In Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security, pages 801–812. ACM, 2013

[170] Valeria Nikolaenko, Udi Weinsberg, Stratis Ioannidis, Marc Joye, Dan Boneh, and Nina Taft. **Privacy-preserving ridge regression on hundreds of millions of records.** In Security and Privacy (SP), 2013 IEEE Symposium on, pages 334–348. IEEE, 2013.

[171] Sheng Gao, Jianfeng Ma, Cong Sun, and Xinghua Li. **Balancing trajectory privacy and data utility using a personalized anonymization model.** Journal of Network and Computer Applications, 38:125–134, 2014.

[172] Daniel Massaguer, Bijit Hore, Mamadou H Diallo, Sharad Mehrotra, and Nalini Venkatasubramanian. **Middleware for pervasive spaces: Balancing privacy and utility.** In ACM/IFIP/USENIX International Conference on Distributed Systems Platforms and Open Distributed Processing, pages 247–267. Springer, 2009

[173] Ye Wang, Yuksel Ozan Basciftci, and Prakash Ishwar. **Privacy utility tradeoffs under constrained data release mechanisms.** arXiv preprint arXiv:1710.09295, 2017.

[174] Sana Belguith, Nesrine Kaaniche, Maryline Laurent, Abderrazak Jemai, and Rabah Attia. Phoabe: **Securely outsourcing multi-authority attribute based encryption with policy hidden for cloud assisted IoT.** Computer Networks, 133:141–156, 2018.

[175] Ran Canetti. **Universally composable security: A new paradigm for cryptographic protocols.** In Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on, pages 136–145. IEEE, 2001.

[176] Johannes Blömer, Fabian Eidens, and Jakob Juhnke. **Practical, anonymous, and publicly linkable universally-composable reputation systems.** In Cryptographers Track at the RSA Conference, pages 470–490. Springer, 2018.

Lectures web complémentaires

Des services personnalisés de plus en plus invasifs : http://www.bbc.com/news/world-us-canada-23123964 – http://fortune.com/2018/04/10/facebook-cambridge-analytica-whathappened/ – https://www.numerama.com/sciences/436626-pourquoi-faire-analyser-son-adn-pour-connaitre-ses-origines-est-une-tres-mauvaise-idee.html – https://www.theverge.com/2019/3/14/18264458/voice-technology-speech-analysis-mental-health-risk-privacy – https://telegram.org/blog/unsend-privacy-emoji – https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/ – *Des services personnalisés bien commodes : Sur le principe de commodité, et une approche critique, voir* https://esprit.presse.fr/article/mark-hunyadi/du-sujet-de-droit-au-sujet-libidinal-41976 – *Data protection by design :* https://www.gendarmerie.interieur.gouv.fr/crgn/content/download/1033/16080/version/2/file/Revue%20N%C2%B0263.pdf – *La vie privée en images à travers les siècles :* https://medium.com/the-ferenstein-wire/the-birth-and-death-of-privacy-3-000-years-of-history-in-50-images-614c26059e – *À propos d'anonymat :* http://www-public.imtbs-tsp.eu/~lauren_m/articles/2015-Bigdata-anonymisation-DP.pdf – *Outils de pistage :* https://amiunique.org/ – https://panopticlick.eff.org/ – *Anti-tracking :* https://adblockplus.org/ – https://www.ublock.org/ – https://disconnect.me/ – https://www.ghostery.com/ – https://noscript.net/ – https://flashblock.en.softonic.com/ – https://www.theverge.com/2016/9/13/12890050/adblock-plus-now-sells-ads – https://www.wired.com/2016/03/heres-how-that-adblocker-youre-using-makes-money/ – https://adblockplus.org/acceptable-ads-agreements – https://help.getadblock.com/support/solutions/articles/6000092027-why-does-adblock-allow-non-intrusive-ads – https://exodus-privacy.eu.org/fr/ – https://iris.polito.it/retrieve/handle/11583/2679579/159720/tracker-blockers.pdf – https://donottrack-doc.com/fr/ – https://www.eff.org/privacybadger – *VPN :* https://www.openvpn.net/ – https://meddle.mobi/ – *Tor :* https://metrics.torproject.org/ – https://tails.boum.org/ – https://www.torproject.org/projects/torbrowser.html.en – *U-Prove :* https://www.microsoft.com/en-us/research/project/u-prove/ – *Idemix :* https://www.zurich.ibm.com/identity mixer/ *Obfuscation :* https://cfeditions.com/obfuscation/ – *Questions juridiques : Règlement (UE) 2016/679* https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32016R0679&from=FR – https://edpb.europa.eu/our-work-tools/general-guidance/gdpr-guidelines-recommendations-best-practices_en – https://www.cnil.fr/fr/reglement-europeen/lignes-directrices – https://www.cnil.fr/fr/RGPD-analyse-impact-protection-des-donnees-pia – https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=uriserv%3AOJ.L_.2018.303.01.0059.01.FRA – https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:52017PC0010&from=EN

Remerciements

Les auteures tiennent à remercier les membres et les mécènes de la Chaire Valeurs et Politiques des Informations Personnelles, en particulier Antoine Dubus, Jonathan Keller, Claire Levallois-Barth et Martin Quinn, pour la richesse des échanges pluridisciplinaires et pour leurs remarques constructives.

Les auteures tiennent également à remercier Stéphane Bortzmeyer, Mathieu Goessens, et Louis-Philippe Sondeck pour leurs relectures attentives du document.

La Chaire de recherche Valeurs et Politiques des Informations Personnelles

Créée par l'Institut Mines-Télécom en mars 2013, la Chaire regroupe une équipe pluridisciplinaire de chercheurs travaillant à la fois sur les aspects juridiques de régulation et de conformité, techniques de sécurité des systèmes et des données, économiques de partage des informations personnelles et philosophiques de responsabilisation et d'anticipation des conséquences sociétales.

Elle bénéficie du soutien de huit partenaires : IN Groupe (ex Imprimerie Nationale), BNP Paribas, Orange, QWANT, SOPRA STERIA (mécènes fondateurs), Dassault Systèmes (mécène associé), de la collaboration de la Commission nationale de l'informatique et des libertés (CNIL) et de la Direction interministérielle du numérique et du système d'information et de communication de l'État (DINSIC), et du support de la Fondation Mines-Télécom.

La Chaire est coordonnée par Claire Levallois-Barth, maître de conférences en droit à Télécom ParisTech, et a été cofondée avec Ivan Meseguer, Affaires Européennes, Direction Recherche et Innovation de l'IMT, Maryline Laurent, professeure en sciences de l'informatique à Télécom SudParis, Patrick Waelbroeck, professeur en sciences économiques à Télécom ParisTech et Pierre-Antoine Chardel, professeur en philosophie à l'Institut Mines-Télécom Business School.

La Chaire Valeurs et Politiques des Informations Personnelles se propose d'aider les entreprises, les citoyens et les pouvoirs publics dans leurs réflexions sur la collecte, l'utilisation et le partage des informations personnelles, à savoir les informations concernant les individus (leurs vies privées, leurs activités professionnelles, leurs identités numériques, leurs contributions sur les réseaux sociaux, etc.), incluant celles collectées par les objets communicants qui les entourent. Ces informations fournies par les personnes, ou traces de leurs activités et interactions, posent en effet de nombreuses questions en termes de valeur sociale, valeur économique, politique de contrôle et politique de régulation.

Les travaux de la Chaire sont conduits selon cinq axes de recherche transdisciplinaires :

- les identités numériques ;
- la gestion des informations personnelles ;
- les contributions et traces ;
- les informations personnelles dans l'internet des objets ;
- les politiques des informations personnelles.

En plus de la publication d'articles de recherche et la participation aux colloques et conférences, la Chaire organise régulièrement des évènements ouverts à tous, pour sensibiliser le grand public sur ces enjeux majeurs du monde numérique.

Institut Mines-Télécom

Les partenaires de la Chaire Valeurs et Politiques des Informations Personnelles (CVPIP)

MÉCÈNES FONDATEURS

IN GROUPE

orange

BNP PARIBAS

sopra steria

Qwant

MÉCÈNE ASSOCIÉ

DASSAULT SYSTEMES

PARTENAIRES QUALIFIÉS

CNIL. COMMISSION NATIONALE INFORMATIQUE & LIBERTÉS

DINSIC*

**Direction Interministérielle du Numérique et du Système d'Information et de Communication de l'État*

9 782955 730874

EN SAVOIR PLUS

www.informations-personnelles.org
youtube.informations-personnelles.org
@CVPIP

CONTACTS

CLAIRE LEVALLOIS-BARTH
Coordinatrice de la Chaire
claire.levallois@imt.fr

ANNE-CATHERINE AYE
Assistante de la Chaire
cvpip@imt.fr
+33 1 45 81 72 53

Télécom ParisTech - IMT
46 rue Barrault | F-75634 Paris Cedex 13