Sujet de Recherche - Thèse CIFRE

Partage de documents sécurisé dans le Cloud Personnel

Paul Tran-Van

1. Le contexte

1.1. Economie et protection des données personnelles

Nous sommes témoins d'une accumulation exponentielle de données personnelles sur les serveurs : données stockées par les administrations, les hôpitaux, les assurances; données captées automatiquement par les sites web visités, données issues de capteurs (quantified-self, smart meters et plus généralement la multitude d'objets intelligents formant l'Internet des objets), ceci sans oublier les données numériques possédées ou créées par les individus eux-mêmes (e.g. photos, agendas, factures, etc.). Toutes ces données aboutissent aujourd'hui sur les serveurs des majors de l'internet par obligation ou sur le *Cloud* par choix pratique des individus (disponibilité 24h/24, tolérance aux pannes). Les données personnelles sont ainsi devenues le *nouveau pétrole* de l'internet et sont monopolisées par les services en ligne. Toutefois, comme l'a démontré l'affaire PRISM, nous ne sommes pas à l'abri d'une crise profonde : tous ces services nous rapprochent des dystopies décrites dans les romans d'anticipation. En effet, nombreuses sont les violations de la vie privée, dues à de la négligence, à un usage abusif, ou tout simplement à des attaques[1]. Face à cette situation, le World Economic Forum a formulé le besoin de voir émerger des plate-formes de gestion de données personnelles permettant à chaque individu de collecter, gérer et partager ses données dans différents contextes avec de réelles garanties de protection de la vie privée [WEF13].

Malheureusement, aucune solution disponible aujourd'hui ne permet de réaliser de telles plate-formes. Les solutions centralisées, incluant la majorité des coffre-forts personnels de données[2], privilégient la définition de services innovants au détriment de la sécurité. Au mieux, ces approches expriment de bonnes pratiques en terme de préservation de la vie privée et offrent aux individus la possibilités d'exprimer des politiques personnalisées de contrôle d'accès et d'usage (e.g., Bases de données Hippocratiques [AKS+02]) mais aucune n'a la capacité de garantir une mise en œuvre incontournable de ces politiques. Même les approches intégrant aux serveurs une protection cryptographique des données telles que CryptDB [PRZ+12] ou complémentant ces serveurs avec du matériel sécurisé telles que TrustedDB [BaS11] ne peuvent répondre aux deux problèmes intrinsèques des approches centralisées. Premièrement, les individus se trouvent toujours exposés à un changement éventuels des chartes de confidentialité (ex: changement de pratiques commerciales,

[1] http://www.datalossdb.org/

[2] Par exemple Personal (http://www.personal.com), My personal vault (http://www.mypersonalvault.com), or Mydex (http://www.mydex.org).

fusion/acquisition de sociétés) ou au non respect de ces dernières (ex: usage abusif de l'hébergeur, demande gouvernementale). Ensuite, la centralisation des données de tous les individus en un point unique attise les convoitises et multiplie le risque d'attaques en maximisant le ratio bénéfice/coût de ces attaques.

Pour répondre à ces critiques, de multiples solutions décentralisées émergent (voir notamment une discussion intégrant de nombreux exemples dans [NTB+12]). Cependant, ces solutions décentralisées sacrifient généralement la fonctionnalité et les perspectives de services innovants sur l'autel de la protection de la confidentialité. Soit les données sont gérées en silos isolés empêchant tout rapprochement, croisement et analyse de données, soit le partage de données est réalisé sur la seule base de la confiance entre individus, sans garantie tangible de sécurité[3].

Face à cette situation, la coopération naissante entre la société Cozy Cloud et l'équipe de recherche SMIS (équipe mixte INRIA-CNRS-UVSQ) ouvre une nouvelle voie dans la protection et le partage des données personnelles. Les sous-sections suivantes résument les apports respectifs de Cozy Cloud et de SMIS à cette problématique et montrent la complémentarité de ces apports. Cette complémentarité est à la base de travaux communs déjà engagés qui serviront de cadre technique et scientifique à cette thèse.

1.2. Cozy Cloud : interopérabilité des données personnelles 'by design'

Cozy Cloud est une startup française qui propose un "Personal Cloud" révolutionnant la gestion des données personnelles. Le Personal Cloud, tel que le propose Cozy, focalisé sur l'empowerment de l'utilisateur, **apporte de nouveaux modèles économiques pour la valorisation des données personnelles, dans le respect de la confidentialité**. En tant qu'individus, nous voudrions bien nous affranchir d'une quantité toujours croissante de mots de passe, être capable de rechercher dans toutes nos données par une simple requête, que les services disponibles sur nos données puissent collaborer, que les objets intelligents que nous utilisons de plus en plus fréquemment soient compatibles entre eux. Bref, nous voudrions être plus proactifs en ce qui concerne notre santé, nos finances, notre consommation d'énergie, notre temps … tout en ne faisant aucune concession sur la préservation de notre vie privée.

Or l'organisation en silos du web est telle qu'actuellement aucun service ne peut accéder à l'ensemble de nos données pour faire ce type de croisements. La solution, les geeks la connaissent bien : il suffit d'avoir son propre serveur sur lequel on héberge toutes ses données et ses propres web services. Mais cette solution n'est pas abordable par le commun des mortels. C'est le challenge de Cozy Cloud, faire du serveur un nouveau device grand public, que tout un chacun peut administrer aussi facilement que son smartphone et sur lequel il va réunir toutes ses données et tous ses services web.

Concrètement votre Cozy qui est votre serveur personnel, votre « **Personal Cloud** », tourne sur votre propre équipement, que ce soit une machine dans votre salon, la box de votre accès internet, ou une machine que vous louez chez un hébergeur. Quand vous vous connectez à votre Cozy, vous accédez à votre Home où vous avez les icones et widgets des services que vous avez installés sur votre serveur. Depuis le **market place** de votre Cozy, vous avez accès

[3] voir par exemple l'initiative FreedomBox (http://freedomboxfoundation.org/).

aux services développés par des tiers et d'un click vous les installez après avoir validé les droits d'accès demandés par l'application. Cozy est **Open Source**, la communauté peut à la fois vérifier son intégrité et contribuer en proposant des apps, correctifs et améliorations.

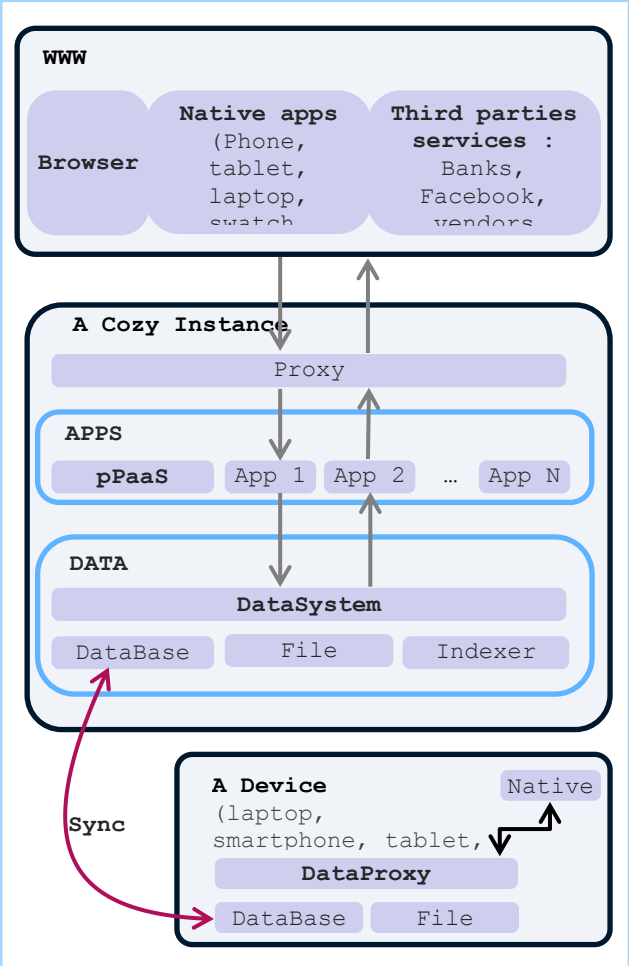
Sur son Cozy nous retrouvons les classiques mails, contacts, agenda, synchronisation de fichiers, mais aussi les robots qui se connectent à nos fournisseurs pour récupérer relevés bancaires, factures, historiques de consommation, données produites par vos capteurs issus de vos terminaux mobiles, de votre maison connectée, de votre internet des objets… Une fois que l'on possède son cloud personnel, on s'affranchit des contraintes qu'imposent les GAFA (Google, Apple, Facebook, Amazon…) et un nouvel écosystème d'usages plus riches et véritablement au service de l'utilisateur peut se développer.

Cozy est une plateforme qui doit masquer à l'utilisateur la complexité technique tout en offrant aux "apps" des services transverses permettant des usages avec les données personnelles impossibles à réaliser aujourd'hui. Pour cela, Cozy redéfinit ce qu'est un "serveur" et abandonne l'approche classique. **Pour Cozy, un serveur est une collection de services http qui se connaissent et sont habilitées à travailler ensemble**.

L'architecture de Cozy est **interopérable "by design"**. Elle repose sur 2 briques essentielles :

Le **pPaaS** : **un personal PaaS** qui déploie les apps à la demande de l'utilisateur. Chaque app est un service http à part entière, le développeur travaille avec la technologie et le framework de son choix, il n'a pas de SDK à apprendre.

Le **DataSystem** : service de persistance des données, permettant leur partage entre les "apps" (contacts, agenda, données bancaires, historique de poids, réglage de la chaudière…) et permettant également aux apps de réagir lorsqu'une donnée évolue.. **Cozy est une architecture "user centric", donc "data centric".** Le DataSystem est le cœur de Cozy. La base utilisée, CouchDB, est une base no-sql orientée documents qui se synchronise avec des bases installées sur les terminaux de l'utilisateur rendant possible la "continuité du contexte numérique" de l'utilisateur de terminal en terminal.



Cozy, en devenant notre "**domicile numérique**" où sont fédérées toutes nos données, fluidifie nos usages :

- **Single Sign On** : en une seule connexion j'accède à MON "walled garden".
- **Recherche globale** : je peux rechercher dans TOUTES mes données – contacts, notes, mails, photo…- en une seule requête.

- **Mes Apps sont intégrées** les unes aux autres, c'est-à-dire que les modifications faites par les unes sont prises en compte par les autres, alors même que leurs développeurs ne se sont pas concertés.
- **Mes données sont unifiées**, je n'ai par exemple qu'une seule base de contacts. Ceux-ci ne sont plus morcelés et dupliqués entre Gmail, Facebook, Linkedin…
- **Mes apps coordonnent mes objets connectés**, en récupèrent les données et les actionnent.
- Le **personal analytics devient accessible :** Le Big Data est appliqué aux entreprises, à la société et à la science, des domaines qui conjuguent des enjeux et volumes de données importants. Or, avec la numérisation de tous les aspects de la vie, de tous les gestes du quotidien, l'individu est en passe de disposer d'une masse de données personnelles considérable. Et lui aussi a des enjeux en termes de gestion de ses finances, de sa santé, de sa consommation d'énergie, de ses achats, de ses informations. Cozy lui donne la possibilité de faire du Big Data sur ses données.

1.3. SMIS : Protection de la vie privée 'by design'

SMIS (Secured and Mobile Information Systems)[4] est une équipe de recherche INRIA-CNRS-UVSQ qui développe l'essentiel de son activité autour de la protection de la vie privée. Dans [AAB+10], SMIS a décrit la vision d'un *Personal Data Server* (PDS) dont l'objectif est de construire une alternative crédible à la centralisation systématique des données personnelles sur des serveurs. L'idée maîtresse de PDS est d'embarquer dans des composants matériels sécurisés (ex: cartes à puce à grande capacité de stockage) des briques logicielles capables d'acquérir, de stocker et de gérer des formes variées de données personnelles (ex: feuilles de salaire, factures, relevés bancaires, données médicales, traces de géolocalisation, etc) en fonction des applications cibles. Ces briques logicielles ont vocation à constituer un véritable Serveur Personnel de Données capable d'interopérer avec des serveurs/services externes tout en garantissant un contrôle total du porteur sur ses données personnelles.

Un PDS tire sa sécurité du composant matériel sécurisé dans lequel il est embarqué. S'il existe une grande variété de tels composants sécurisés (mass storage SIM card, clé USB sécurisée, secure token, smart dongle, etc), tous peuvent être abstraits par (1) un Environnement d'Exécution de Confiance et (2) un grand espace de stockage (potentiellement sans confiance, et donc devant contenir uniquement des données chiffrées). L'Environnement d'Exécution de Confiance est habituellement composé d'un microcontrôleur sécurisé qualifié de *tamper-resistant* (résistant aux attaques physiques), tandis que l'espace de stockage est habituellement constitué d'une mémoire Flash externe (puce de Flash ou carte micro-SD). Le moteur PDS s'exécute dans l'Environnement d'Exécution de Confiance et les données de l'individu sont hébergées dans la Flash externe et protégées cryptographiquement. Le moteur PDS est un véritable SGBD (Système de Gestion de Bases de Données) embarqué dans un composant matériel sécurisé. Il est capable de stocker les données sous forme de tables, de les indexer, de les interroger via des requêtes SQL, de garantir leur intégrité logique (contraintes d'intégrité) et physique (atomicité transactionnelle) et surtout de les protéger par une politique de contrôle d'accès assertionnelle (i.e., tel utilisateur peut accéder aux données satisfaisant tel prédicat SQL). La

[4] Equipe de recherche commune à l'INRIA, au CNRS (laboratoire PRiSM) et à l'Université de Versailles Saint-Quentin en Yvelines (http://www-smis.inria.fr/).

mise en œuvre d'un tel SGBD embarqué pose de multiples verrous scientifiques liées aux contraintes combinées du microcontrôleur (e.g., très faible RAM) et de la Flash NAND (e.g., coût des réécritures aléatoires, Block-erase-before-page-rewrite, wear leveling, etc). Attaquer ces verrous nécessite une profonde redéfinition des principes classiques de gestion de bases de données (stockage, indexation, requêtes, transactions) [ABP+10].

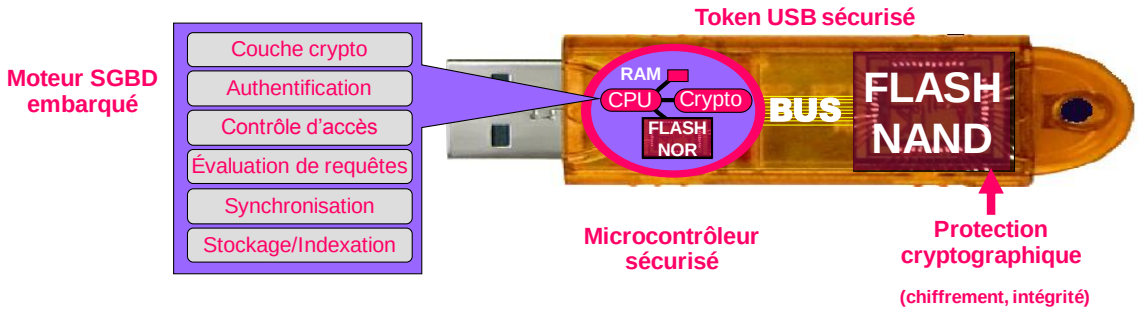


Figure 2: architecture logicielle et matérielle d'un Personal Data Server

La confiance qu'il est possible de porter à un PDS est due à une combinaison de facteurs: (1) le fait qu'un PDS porte les données d'un seul individu, qu'aucun code externe ne puisse être installé dessus et qu'il faille être en sa possession physique pour l'attaquer, (2) le microcontrôleur tamper-resistant, qui rend les attaques physiques ou par canaux cachés très difficiles, (2) l'auto-administration du PDS, qui évite les attaques internes par l'administrateur, et (4) le fait que le propriétaire du PDS lui-même ne peut pas accéder à toutes les données stockées, puisqu'il doit s'authentifier et accède aux données selon ses privilèges[5]. La combinaison des points (1) à (3) diminue significativement le ratio bénéfice/coût d'une attaque et minimise ainsi l'intérêt d'une telle attaque. Le point (4) est également d'une grande importance dès lors que l'on envisage des traitements distribués entre plusieurs individus. En effet, ce point garantit qu'un PDS peut exporter des données vers un autre PDS sans risque de perte de confidentialité. Dit autrement, un individu peut exporter des données vers un partenaire en y associant des règles de contrôle d'usage qui seront assurées par le PDS de ce partenaire. On parle dans ce cas de *sticky policies* (politiques de contrôles d'accès et d'usage collées aux données et indissociables de ces dernières). Cette caractéristique permet d'envisager des traitements distribués qualifiés de 'privacy by design', comme par exemple l'anonymisation de données décentralisée [ANP11, ANP13] ou la gestion de requêtes distribuées sur un ensemble de PDS [QBP14].

Un prototype opérationnel de PDS a été développé et déposé à l'APP [ABP+09]. Ce prototype est embarqué dans un token USB sécurisé, d'où son nom PlugDB du fait qu'il est connectable/déconnectable à l'envie sur tout terminal. PlugDB est utilisé sur le terrain pour une expérimentation de dossiers médico-sociaux portables et sécurisés facilitant la coordination des soins au domicile de personnes dépendantes sur le territoire des Yvelines [ABB+08]. Depuis, la vision Personal Data Server a évoluée vers une vision plus large appelée Trusted Cells et englobant les données personnelles issues de l'Internet des objets [ABB+13].

[5] Au même titre qu'un porteur de carte bancaire par exemple n'a pas accès aux secrets cryptographiques embarqués dans sa propre carte à puce.

1.4. Une convergence des approches

Sans conteste, les approches développées parallèlement par Cozy Cloud et SMIS sont très complémentaires. Les deux approches sont centrées sur la création d'un serveur personnel de données permettant de stocker, analyser, croiser des données personnelles entre de multiples applications, le tout de façon décentralisée et sous le contrôle de l'utilisateur.

Cozy Cloud met l'accent sur la structuration de cet espace personnel, sur l'administration de la plate-forme et sur le déploiement et l'interopérabilité des applications. D'où la qualification d'une approche *interopérable par design*. Si la sécurité est une composante essentielle d'un serveur personnel de données, elle repose aujourd'hui dans Cozy Cloud sur des outils classiques : identification, authentification, contrôle d'accès, chiffrement des données sensibles. Ceci avec les limites connues de ces outils classiques : problèmes de sécurisation des clés de chiffrement et de leur restauration en cas de perte, faible résistance aux attaques y compris logicielles, difficulté de déléguer la confiance à un autre serveur personnel et donc d'implémenter des scénarios collaboratifs d'échange de données personnelles avec un réel contrôle d'usage.

Inversement, SMIS met l'accent sur la sécurité des données avec une garantie tangible de résistance aux attaques apportée par le moteur PlugDB embarqué sur un composant sécurisé matériellement. PlugDB apporte donc des réponses aux problèmes de séquestre sécurisé de clés de chiffrement, à la tolérance aux attaques logicielles et matérielles et à la capacité de déléguer la confiance à tout serveur personnel reposant sur PlugDB. D'où la qualification d'une approche *Privacy by design*. Par contre, PlugDB n'apporte pas de réponse à la problématique de structuration et d'administration de l'espace personnel de l'utilisateur, ni au déploiement des applications, points forts de Cozy Cloud.

Une collaboration naturelle s'est donc établie et va conduire à un transfert technologique permettant de connecter les deux outils pour profiter du meilleur des deux mondes. L'idée est simple. Il s'agit d'utiliser PlugDB comme un co-DataSystem sécurisé pour Cozy Cloud. Ainsi, Cozy Cloud garde une architecture conforme à l'existant tout en délégant à PlugDB les parties critiques de la sécurité, à savoir le contrôle d'accès et la gestion des clés de chiffrement. Le modèle initial simple suivant est envisagé : (1) le granule de contrôle d'accès est le document, (2) des méta-données sont associées aux documents et peuvent être utilisées pour définir des règles de contrôle d'accès (ex: autoriser l'accès de tel document à tel utilisateur ou rôle en fonction du sujet, de l'auteur, du type de document, etc), (3) les documents sensibles sont stockés chiffrés dans le DataSystem de Cozy Cloud (i.e., CouchDB), (4) les clés de chiffrement et les règles de contrôle d'accès sont stockées et évaluées dans PlugDB, (5) quand une demande de document est envoyée par une App à Cozy Cloud, celui-ci route la requête vers PlugDB qui évalue la ou les règles de contrôle d'accès associées et renvoie la clé permettant de déchiffrer le document dans le cas positif, (6) Cozy Cloud peut alors déchiffrer le document stocké dans son DataSystem et le renvoyer à l'App. Si le débit du composant sécurisé le permet, il sera envisagé de faire déchiffrer le document directement par PlugDB afin qu'aucune clé de chiffrement ne soit jamais exposée (i.e., ne sorte de l'environnement confiné de PlugDB).

C'est cette architecture, que Cozy Cloud et SMIS commencent à mettre en place ensemble, qui servira de cadre à cette thèse.

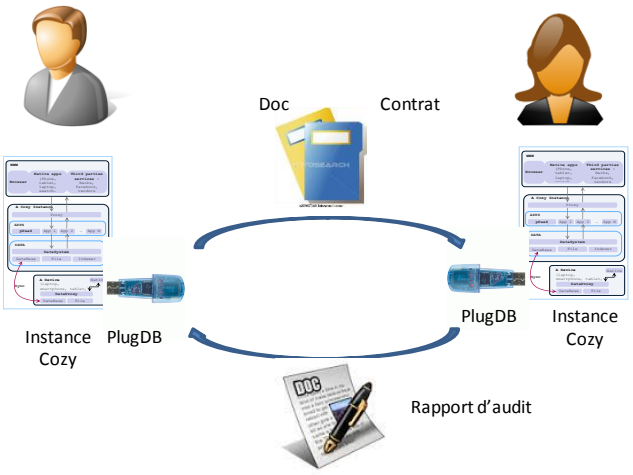
2. Objectif de la thèse

2.1. Objectif général

Dire que l'architecture Cozy-SMIS servira de cadre à cette thèse sous-entend que la thèse prend pour acquis l'existence de cette architecture. Cette dernière sera en effet développée dans le cadre d'un accord de transfert technologique mais ce développement n'offre pas en soi de nouveau sujet de recherche propre à réaliser une thèse de doctorat. En fait l'architecture Cozy-SMIS offre une garantie tangible de sécurité à l'utilisateur face aux attaques menées contre son propre serveur personnel. Par contre, cette architecture n'intègre pas de mécanisme permettant d'échanger avec cette même garantie des données personnelles entre plusieurs individus. C'est justement l'objectif de cette thèse que de :

1. Proposer un modèle de partage de documents entre individus permettant d'exprimer des règles d'usage sous forme d'un contrat.
2. Concevoir un protocole sécurisé donnant une garantie tangible à l'émetteur du document que le récepteur suivra scrupuleusement les termes de ce contrat.
3. Implémenter l'ensemble dans l'architecture Cozy-SMIS pour en faire un outil opérationnel.

L'objectif pour le point 1 est de s'inspirer de travaux existants et si possible de standards existants afin d'assurer le succès industriel de la solution. Le point 2 est de fait le verrou scientifique et technologique adressé dans cette thèse. A notre connaissance, peu de travaux de l'état de l'art portent sur cette problématique et l'équipe de recherche SMIS a une réelle expertise sur ce sujet. Le point 3 est évidemment l'aboutissement espéré et la valorisation effective du travail de cette thèse.



2.2. Bref état de l'art des modèles de partage

Dans le monde des bases de données, le partage d'information entre utilisateurs est traditionnellement régulé par des mécanismes de contrôle d'accès. Trois familles de modèles de contrôle d'accès sont particulièrement utilisées. Le modèle MAC (Mandatory Access Control) associe des labels aux objets et aux sujets (e.g. Unclassified, Confidential, Secret, Top Secret) et régule les accès par deux règles : no-read-up (impossibilité de lire des données classifiée à un niveau supérieur à celui du lecteur) et no write-down (impossibilité de déclassifier une donnée). C'est un modèle très sécuritaire mais très rigide et mal adapté au contexte qui nous intéresse ici. Le modèle DAC (Discretionary Access Control) est un modèle décentralisé ou chaque créateur d'un objet a la capacité de définir les règles d'autorisation qu'il souhaite. Le modèle RBAC (Role-Based Access Control) assigne des rôles aux utilisateurs et des permissions à ces rôles, ceci afin de réduire la complexité d'administration de politiques de sécurité impliquant un grand nombre d'acteurs [FKC03]. Sur l'initiative

RBAC, de nombreux modèles ont été étudiés dans la littérature, l'étendant avec des contraintes temporelles ou géographiques ou organisant les politiques par le biais de concepts additionnels (e.g., équipes, tâches, organisations) comme par exemple dans les modèles TBAC ou OrBAC. Cependant, ces modèles ont été conçus prioritairement pour réguler les accès à une base de données traditionnelle multi-utilisateurs et capturent mal l'expression d'un contrôle fin sur l'usage pouvant être fait des données, un pré-requis dès lors que l'on parle de protection de la vie privée.

Le modèle UCON (Usage CONtrol model) [ZPS+05] pose les fondements d'une nouvelle génération de modèles de contrôle d'accès. Dans ce modèle, la décision d'autoriser une action est déterminée par la combinaison d'autorisations classiques (ex: le sujet S1 est autorisé à exécuter l'action A2 sur le document D3) avec des obligations (ex: journaliser cette action dans un objectif d'audit ultérieur) et des conditions contextuelles (ex: uniquement pendant les heures de bureau). Dans cet esprit [YBZ07] a proposé un modèle de contrôle d'accès basé *'objectif'* (Purpose-Based Acces Control) et [NTB+07] a intégré ce concept *d'objectifs* dans RBAC (P-RBAC).

Du point de vue des standards, les modèles MAC, DAC et RBAC ont tous été standardisés au travers de SQL notamment. Des modèles de contrôle d'usage ont également été standardisés, comme par exemple XrML (eXtensible Rights Markup Language) par ContentGuard dans le contexte des droits d'accès digitaux (DRM). XACML (eXtensible Access Control Markup Language) est une spécification promue par OASIS qui définit un langage pour le contrôle d'accès, la circulation des règles et l'administration de la politique de sécurité des systèmes d'information dans les architectures SOA. XDI (XRI Data Interchange) est un format d'échange de données sémantique et un protocole d'échanges à base de contrats en cours de développement par OASIS.

Ainsi, de nombreux travaux et standards portent sur les modèles de contrôle d'accès et d'usage, sur la syntaxe des langages permettant d'exprimer des politiques associées mais peu de travaux s'intéressent à la sécurisation effective de ces politiques.

3. Organisation de la thèse

L'organisation de la thèse suivra les 3 objectifs énoncés précédemment:

1. **Proposer un modèle de partage de documents :** Les travaux évoqués dans la section précédente ouvrent des pistes intéressantes et devront être explorés plus avant pendant les premiers mois de la thèse. L'enrichissement de l'état de l'art doit cependant s'entendre comme un processus continu tout au long de la thèse. Si le choix du modèle n'est pas fait à l'heure actuelle, il est probable qu'il s'oriente vers un modèle à base de contrats qui englobe les points suivants : (1) quelles sont les parties du contrat et comment prouvent-elles leur identité, (2) quels sont leurs privilèges respectifs, quels usages sont autorisés, sous quelles conditions et avec quelles obligations, (3) que se passe-t-il en fin de contrat et (4) comment vérifie-t-on que les termes du contrat ont bien été respectées. La contribution scientifique de la thèse n'étant pas focalisée sur ce point, une reprise de travaux et de standards existants (ou fait par d'autres en parallèle) est escomptée.

2. **Concevoir un protocole sécurisé implémentant ce modèle :** l'objectif est de donner une garantie tangible à l'émetteur d'un document que le récepteur suivra scrupuleusement les termes de ce contrat. Si de nombreux travaux portent sur la sécurisation du contrôle d'accès dans des architectures centralisées, très peu de travaux portent sur la sécurisation du contrôle d'usage dans des environnements décentralisés (à l'exception d'outils DRM très spécifiques). C'est donc sur ce point que portera la contribution scientifique de la thèse. Elle s'appuiera sur l'expérience de SMIS dans ce domaine et la capacité d'exploiter des composants logiciels embarqués dans des architectures sécurisées matériellement. Nous sommes confiants sur le fait que de tels travaux puissent aboutir à des publications internationales de haut niveau. Le choix du modèle de partage (point 1) pouvant prendre un temps important, il est possible de commencer à raisonner sur le point 2 rapidement, sur la base d'un modèle abstrait de contrats.
3. **Implémenter l'ensemble dans l'architecture Cozy-SMIS :** c'est bien entendu l'aboutissement du travail de thèse dont on peut espérer une valorisation importante pour Cozy Cloud comme pour SMIS. En effet, les approches Cozy Cloud et SMIS sont toutes les deux reconnues comme très innovantes dans leur domaine et par leur communauté respective. On peut donc attendre de leur mariage une vraie rupture du point de vue technologique, économique et développement des usages. L'intégration de PlugDB dans l'architecture Cozy étant une option proposée dans l'offre commerciale de Cozy Cloud, il est essentiel que l'implémentation réalisée puisse s'adapter à la présence ou l'absence d'un composant PlugDB, avec un impact sur les garanties sécuritaires obtenues mais sans impact sur les fonctionnalités.

L'étudiant partagera son temps de façon équitable entre l'entreprise et l'INRIA mais pas obligatoirement de façon linéaire. Il sera plus fréquemment à l'INRIA pour les points 1 et 2 et plus fréquemment chez Cozy Cloud pour le point 3 mais sera présent au moins un jour par semaine chez l'un des deux partenaires chaque semaine.

L'étudiant sera encadré par Benjamin André côté Cozy Cloud et Philippe Pucheral côté SMIS.

L'étudiant sera inscrit à l'Université de Versailles Saint-Quentin en Yvelines (UVSQ) sous la direction de Philippe Pucheral, professeur UVSQ. Il sera membre de l'école doctorale STV lors de la première année de thèse. Il sera ensuite associé à l'école doctorale STIC de la future Université Paris-Saclay tout en restant inscrit administrativement à l'UVSQ[6].

[6] L'UVSQ est membre de la FCS Paris-Saclay et la mise en place des nouvelles écoles doctorales est prévue courant 2015.

4. Compétences des acteurs

Cozy Cloud - Benjamin André

Ingénieur ENSIMAG, CEO de Cozy Cloud, Benjamin ANDRE porte la vision du projet. Il a déjà fondé une startup et a travaillé en tant que directeur de projets pour la DSI de grands groupes (Accenture, Alcatel, Total). Mais Cozy Cloud, c'est également Frank Rousseau, CTO de la startup, qui a travaillé en tant que consultant puis en startup pendant 7 ans. C'est aussi une équipe actuelle de 8 ingénieurs (ENSIMAG & UTC) portant l'effectif total de la société à 10 personnes.

Bien que créée récemment, Cozy Cloud, ce sont déjà :

Des références

OVH, 3ième hébergeur mondial, fort de leur succès sur les serveurs personnels, sont convaincus qu'avec Cozy ils seraient en mesure de louer des serveurs personnels à une audience bien plus large que celle qu'ils adressent aujourd'hui. Alban Schmutz, Vice-président, nous a demandé de lui faire une offre commerciale dès que nous serons prêts et nous a adressé une lettre d'intention officialisant l'intérêt d'OVH pour le Personal Cloud.

Cozy a été sélectionné pour faire partie du groupe de travail sur le "plan filière cloud", aux côté de 12 autres sociétés (OVH, Atos, Orange, Cegid, Numergy, Talend…) pour remettre au gouvernement des préconisations pour le développement de la filière cloud.
Cozy y représente la composante "Personal Cloud".

LeWeb London 2013

Accéléré par Mozilla à San Francisco

Presse internationale comme Wired

Jeune Entreprise innovante du pôle

Une belle traction

Partenaires commerciaux: des expérimentations en cours

Cozy a été retenu par l'expérimentation MesInfos, coordonné par la FING (Daniel Kaplan) et financé par un consortium de grands groupes ainsi que par la DGCIS. MesInfos est l'équivalent des projets MiData au UK et Smart Disclosure aux USA.

Le changement de paradigme apporté par Cozy a séduit La Poste qui y voit le futur de son service de gestion de données personnelles (Digiposte). Cozy développe donc pour La Poste une preuve de concept.

Contacts en cours : Télécommunication (Orange, Bouygues Telecom), Banques (BNP), Energie (GDF)

Autres partenaires

Opérationnels : Les startups Privowny , Docker, Respect Network, soutien du Pôle de compétitivité PICOM.

Scientifiques : les équipes de recherche de Serge Abiteboul (INRIA/ENS Cachan) et de Philippe Pucheral (INRIA/CNRS/UVSQ).

SMIS - Philippe Pucheral

INRIA est l'Institut National de Recherche en Informatique et Automatique. Il est dédié aux recherches fondamentales et appliquées dans le domaine des STIC. Plus de 3000 chercheurs travaillent dans 175 équipes de recherche.

SMIS (Secured and Mobile Information Systems) est une de ces équipes de recherche, commune avec l'Université de Versailles Saint-Quentin en Yvelines (UVSQ) et le CNRS (laboratoire PRiSM). L'équipe SMIS (http://www-smis.inria.fr), constituée d'une quinzaine de personnes, travaille sur la gestion de données dédiée à des plate-formes matérielles spécialisées (e.g., puces, capteurs, mémoire Flash) et sur la protection de la confidentialité des données (e.g., modèles de contrôle d'accès et d'usage et sécurisation à base de hardware sécurisé, chiffrement de bases de données, anonymisation de données). SMIS est un des pionniers internationaux dans l'usage de hardware sécurisé pour protéger les données personnelles dans des architectures décentralisées. SMIS a réalisé plusieurs prototypes primés (e-gate '04, SIMagine '05) et un de ses prototypes est utilisé dans une expérimentation terrain de dossiers médicaux sécurisés et mobiles (DMSP: Dossier Médico-social Sécurisé Partagé).

Philippe Pucheral est Professeur à l'UVSQ, responsable du master recherche COSY et dirige l'équipe SMIS à l'INRIA Rocquencourt. Son domaine de recherche couvre la gestion de données embarquées et mobiles ainsi que la sécurisation des données, avec des applications directes à la protection des données à caractère personnel. Il a (co-)écrit plus de 80 articles de journaux et de conférences, 4 livres, 4 brevets et a reçu 5 awards internationaux (EDBT '92, VLDB '00, e-gate '04, SIMagine '05, PST'11). Il a (co-)encadré une vingtaine de thèses dont une récompensée par l'Accessit du prix de thèse ASTI 2007.

Paul Tran-Van

Paul Tran-Van est un ancien étudiant de l'école d'ingénieur ENSIIE d'Evry, diplômé en novembre 2013. Il a effectué sa dernière année d'étude en bi-cursus dans le master de recherche COSY à l'UVSQ, dirigé par Philippe Pucheral.

Paul a ensuite réalisé son stage de fin d'étude dans l'équipe SMIS au sein de l'INRIA, où il a pu découvrir le monde de la recherche dans un environnement extrêmement compétent et motivant. Suite à cette expérience enrichissante, il a décidé de poursuivre dans cette voie en souhaitant s'engager dans une thèse. C'est dans cette optique qu'il a été recruté par l'équipe SMIS dans un contrat dit de 'Relais thèse', afin de préparer au mieux la transition entre le stage de fin d'étude et le début de la thèse.

Il dispose d'un double profil, technique et théorique. Son profil technique a été façonné d'une part via ses études, en DUT Informatique puis en école d'ingénieur, et également par ses premières expériences professionnelles dans l'industrie, notamment dans l'entreprise Exa Informatique où il a travaillé 1 an en alternance et effectué un stage de 3 mois. Son profil théorique s'est quant à lui essentiellement constitué par le master de recherche COSY et son travail dans l'équipe SMIS depuis 1 an.

Cette double facette est un atout pour une thèse CIFRE, qui demande à la fois une approche théorique de recherche ainsi qu'une compétence technique indispensable pour la réalisation en entreprise.

5. Apports de la thèse

Apport pour Cozy Cloud

En l'état, la plate-forme Cozy Cloud souffre de 2 problèmes importants :

- la gestion des clés de chiffrement et des règles de contrôle d'accès correspondantes
- l'utilisateur a retrouvé le contrôle de ses données, mais perd la facilité de partage. Il possède son silo personnel, mais a perdu les commodités sociales des silos collectifs que sont les réseaux sociaux.

Ces problèmes sont inhérents à l'auto hébergement de son serveur personnel et l'ensemble des acteurs qui s'intéressent à cette approche butent de la même manière dessus. Parvenir à trouver une solution robuste techniquement et qui n'introduise pas de complexité pour l'utilisateur sera un différenciant fort. Sans compter que Cozy a un business model en B2B2C, c'est-à-dire que ses clients (Orange, OVH et La Poste) ou leurs clients potentiels (banques, énergéticiens, administrations publiques, opérateurs de télécommunication) sont des entreprises très sensibles à ces sujets à la fois de sécurité et de partage. L'enjeux pour Cozy est véritablement essentiel.

L'approche de plugDB est très pertinente et complémentaire car elle allie :

- Un faible coût : un token sécurisé qui coûte une dizaine d'euros.
- Un business model : ce token devient partie intégrante d'un business model potentiel pour certains clients de Cozy qui peuvent proposer ce token à la vente auprès de leurs utilisateurs (i.e., voici les clés de votre coffre-fort).
- Une sécurité technique forte
- Un usage simple : je branche ou débranche ma clé, que je peux facilement regénérer en cas de perte
- Apporte une perception physique, intuitive, de la sécurité apportée (un token que possède l'utilisateur)

Réussir à intégrer dans PlugDB à la fois le stockage des clés de chiffrement et l'organisation du partage des données entre utilisateurs est une solution quasi parfaite qui donnera une longueur d'avance à Cozy par rapport aux projets concurrents mais surtout permettra de clairement se positionner en concurrent avéré des actuels réseaux sociaux.

Cette thèse sera encadrée par Benjamin ANDRE, ingénieur ENSIMAG, co fondateur de Cozy Cloud et se déroulera en interaction avec les autres membres de l'équipes, tous également ingénieurs en informatique & mathématique appliquée (3 ENSIMAG et 2 UTC).

A noter qu'Orange Labs est très intéressé par les modèles de partages de documents pour envisager des nouvelles modalités de socialisation en ligne. L'équipe de François-Gaël Ottogalli qui en en charge de la R&D sur les clouds personnels chez Orange Labs suivra de près nos travaux et nous regarderons à monter des démonstrateurs ensemble ou même à enrichir mutuellement nos travaux en fonction de nos avancées respectives.

Apport pour SMIS

SMIS a une longue tradition de développement de prototypes. Certains sont simplement destinés à valider des travaux de recherche et sont démontrés dans de grandes conférences internationales telles que VLDB, Sigmod ou EDBT ou sont présentés à des concours logiciels. D'autres atteignent une robustesse quasi industrielle et sont utilisés dans des expérimentations terrain. C'est notamment le cas de PlugDB utilisé dans l'expérimentation DMSP (cf Section 4) et présenté à l'Assemblée Nationale lors de l'audition publique 'Dossier Médical Personnel' organisée par l'Office Parlementaire des Choix Scientifiques et Technologiques (OPECST) en avril 2009. L'objectif poursuivi va ici au-delà de la validation de nos recherches, il est d'avoir un véritable impact sociétal. Nous considérons que la protection des données personnelles dans un monde du tout numérique est un défi majeur posé à la communauté scientifique et aux industriels, l'affaire PRISM ne faisant qu'exacerber cette conviction[7]. Notre autre conviction est que le mariage des technologies Cozy-PlugDB peut apporter une réelle rupture dans les usages. Cette thèse participe à cette dynamique et la renforce, tout en abordant un sujet de recherche très porteur et sur lequel SMIS souhaite absolument s'investir.

Apport pour Paul Tran-Van

Paul Tran-Van a été sensiblisé à la protection des données personnelles lors de ses travaux dans l'équipe SMIS. Il est convaincu que la vision portée par Cozy Cloud s'inscrit parfaitement dans cette problématique et représente l'avenir pour la gestion des données personnelles.

Pour lui, cette thèse CIFRE est l'opportunité de travailler sur des problématiques de recherche très intéressantes et dont l'application concrète sera mise en œuvre industriellement.

Ceci cadre parfaitement avec sa volonté de travailler dans l'innovation, en gardant un pied dans la recherche et l'autre dans l'entreprise, lui conférant ainsi une bi-culturalité aux bénéfices non négligeables sur le marché de l'emploi.

De plus, l'encadrement à la fois par l'équipe SMIS et Cozy Cloud amènera inévitablement une montée en compétence véritablement profitable.

[7] voir le tutoriel organisé par des membres de SMIS [ANS14], slides en ligne.

6. Bibliographie

[AAB+10] T. Allard, N. Anciaux, L. Bouganim, Y. Guo, L. Le Folgoc, B. Nguyen, P. Pucheral, Ij. Ray, Ik. Ray, S. Yin, 'Secure Personal Data Servers: a Vision Paper', *Proc. of the 36th International Conference on Very Large Data Bases (VLDB)*, Singapore, PVLDB 3(1): 25-35, September 2010.

[ABB+13] N. Anciaux, P. Bonnet, L. Bouganim, B. Nguyen, I. Sandu Popa, P. Pucheral. 'Trusted Cells: A Sea Change for Personal Data Services'. *Proc. of the 6th Biennal International Conference on Innovative Data Systems Research (CIDR)*, Asilomar, USA, January 2013.

[ABB+08] N. Anciaux, M. Berthelot, L. Braconnier, L. Bouganim, M. De la Blache, G. Gardarin, P. Kesmarszky, S. Lartigue, J-F. Navarre, P. Pucheral, J-J. Vandewalle, K. Zeitouni, 'A Tamper-Resistant and Portable Healthcare Folder', International Journal of Telemedicine and Applications (IJTA), Vol. 2008, 2008.

[ABP+09] N. Anciaux, L. Bouganim, P. Pucheral, S. Yin, M. Benzine, K. Jacquemin, D. Shasha, C. Salperwyck, M. E. Kholy, Logiciel PlugDB-engine, enregistrement APP n° IDDN.FR.001.280004.000.S.C.2008.0000.10000, 27 avril 2009.

[ANP11] T. Allard, B. Nguyen, P. Pucheral. 'Safe Realization of the Generalization Privacy Mechanism'. *Proc. of the 9th International Conference on Privacy, Security and Trust (PST)*, Montreal, Canada, July 2011. Best paper award.

[ANS14] Nicolas Anciaux, Benjamin Nguyen, Iulian Sandu-Popa, *Managing Personal Data with Strong Privacy Guarantees*, in *17th International Conference on Extending Database Technology (EDBT)*, 2014. Slides en ligne (http://www.prism.uvsq.fr/~beng/wiki/index.php/Publications#2014)

[ANP13] T. Allard, B. Nguyen, P. Pucheral. " METAP : Revisiting Privacy-Preserving Data Publishing using Secure Devices", Distributed and Parallel Database Journal (DAPD), to appear. (http://link.springer.com/article/10.1007%2Fs10619-013-7122-x)

[ABP+10] N. Anciaux, L. Bouganim, P. Pucheral, Y. Guo, L. Le Folgoc, S. Yin, "MILo-DB: a Personal, Secure and Portable Database Machine", Distributed and Parallel Database Journal (DAPD), Special Issue on Secure and Privacy-aware Data Management, 2013.

[AKS+02] R. Agrawal, J. Kiernan, R. Srikant, Y. Xu: Hippocratic Databases. VLDB, 2002.

[BaS11] S. Bajaj, R. Sion: TrustedDB: a trusted hardware based database with privacy and data confidentiality. SIGMOD Conference, 2011.

[FKC03] D. Ferraiolo, D.R. Kuhn, and R. Chandramouli, *Role-Based Access Control*, Artech House, 2003.

[NTB+07] Q. Ni, A. Trombetta, E. Bertino, and J. Lobo, "Privacy-aware role based access control," *Proceedings of the 12th ACM symposium on Access control models and technologies SACMAT 07*, vol. 16(8), 2007.

[QBP14] Quoc-Cuong To, Benjamin Nguyen, Philippe Pucheral, Privacy-Preserving Query Execution using a Decentralized Architecture and Tamper Resistant Hardware,

in 17th International Conference on Extending Database Technology (EDBT), 2014.

[NTB+12] A. Narayanan, V. Toubiana, S. Barocas, H. Nissenbaum, D. Boneh: A Critical Look at Decentralized Personal Data Architectures CoRR abs/1202.4503: (2012).

[PRZ+12] R.A. Popa, C. Redfield, N. Zeldovich, H. Balakrisnan, CryptDB: processing queries on an encrypted database, Communications of the ACM, Volume 55 Issue 9, September 2012.

[WEF13] The World Economic Forum. Rethinking Personal Data: Strengthening Trust. May 2012.

[YBZ07] N. Yang, H. Barringer, and N. Zhang, "A Purpose-Based Access Control Model," *Third International Symposium on Information Assurance and Security*, 2007.

[ZPS+05] X. Zhang, F. Parisi-Presicce, R. Sandhu, and J. Park, "Formal model and policy specification of usage control," *ACM Transactions on Information and System Security*, vol. 8(4), 2005.