RESEARCH CENTRE

Saclay - Île-de-France

IN PARTNERSHIP WITH:

Université Versailles Saint-Quentin

2020 ACTIVITY REPORT

Project-Team PETRUS

PErsonal & TRUSted cloud

DOMAIN

Perception, Cognition and Interaction

THEME

Data and Knowledge Representation and Processing

Contents

Project-Team PETRUS

Creation of the Team: 2016 December 01, updated into Project-Team: 2017 July 01

Keywords

Computer sciences and digital sciences

A1.1.8. – Security of architectures

A1.1.9. – Fault tolerant systems

A1.3. – Distributed Systems

A3.1.2. – Data management, quering and storage

A3.1.3. – Distributed data

A3.1.5. – Control access, privacy

A3.1.6. – Query optimization

A3.1.9. – Database

A3.1.11. – Structured data

A4.5. – Formal methods for security

A4.7. – Access control

A4.8. – Privacy-enhancing technologies

Other research topics and application domains

B2.5.3. – Assistance for elderly

B6.4. – Internet of things

B6.6. – Embedded systems

B9.10. – Privacy

1 Team members, visitors, external collaborators

Research Scientists

- Nicolas Anciaux [Team leader, Inria, Senior Researcher, HDR]
- Luc Bouganim [Inria, Senior Researcher, HDR]

Faculty Members

- Philippe Pucheral [Univ de Versailles Saint-Quentin-en-Yvelines, Professor, HDR]
- Iulian Sandu Popa [Univ de Versailles Saint-Quentin-en-Yvelines, Associate Professor]
- Guillaume Scerri [Univ de Versailles Saint-Quentin-en-Yvelines, Associate Professor]

Post-Doctoral Fellows

- Mariem Habibi [Inria, from Feb 2020]
- Julien Loudet [Inria, until Aug 2020]

PhD Students

- Robin Carpentier [Univ de Versailles Saint-Quentin-en-Yvelines]
- Ludovic Javet [Inria]
- Riad Ladjel [Univ de Versailles Saint-Quentin-en-Yvelines]
- Julien Mirval [Inria]
- Dimitrios Tsolovos [Inria, until Mar 2020]

Technical Staff

- Julien Mirval [Inria, Engineer, from Sep 2020 until Oct 2020]
- Laurent Schneider [Inria, Engineer, until Jul 2020]
- Floris Thiant [Inria, Engineer, from Feb 2020]

Administrative Assistants

- Katia Evrat [Inria, from Oct 2020]
- Emmanuelle Perrot [Inria, until Aug 2020]

External Collaborator

- Benjamin Nguyen [INSA CENTRE VDL, HDR]

2 Overall objectives

We are witnessing an exponential accumulation of personal data on central servers: data automatically gathered by administrations and companies but also data produced by individuals themselves (e.g., photos, agendas, data produced by smart appliances and quantified-self devices) and deliberately stored in the cloud for convenience. The net effect is, on the one hand, an unprecedented threat on data privacy due to abusive usage and attacks and, on the other hand, difficulties in providing powerful user-centric services (e.g. personal big data) which require crossing data stored today in isolated silos. The Personal Cloud paradigm holds the promise of a Privacy-by-Design storage and computing platform, where each individual can gather her complete digital environment in one place and share it with applications and users, while preserving her control. However, this paradigm leaves the privacy and security issues in user's hands, which leads to a paradox if we consider the weaknesses of individuals' autonomy in terms of computer security, ability and willingness to administer sharing policies. The challenge is however paramount in a society where emerging economic models are all based - directly or indirectly - on exploiting personal data.

While many research works tackle the organization of the user's workspace, the semantic unification of personal information, the personal data analytics problems, the objective of the PETRUS project-team is to tackle the privacy and security challenges from an architectural point of view. More precisely, our objective is to help providing a technical solution to the personal cloud paradox. More precisely, our goals are (i) to propose new architectures (encompassing both software and hardware aspects) for secure personal cloud data management and formally prove important bricks of the architecture, (ii) propose new data administration models reaching the main requirements of a personal cloud (decentralized access and usage control models, data sharing, data collection and retention models, etc.) and study the enforcement of the resulting privacy policies based on secure hardware and formally proven architectural components, (iii) propose new secure distributed database indexing models, privacy preserving query processing strategies and data anonymization techniques for the personal cloud.

3 Research program

To tackle the challenge introduced above, we identify three main lines of research:

- (Axis 1) Personal cloud server architectures. Based on the intuition that user control, security and privacy are key properties in the definition of trusted personal cloud solutions, our objective is to propose new architectures (encompassing both software and hardware aspects) for secure personal cloud data management and formally prove important bricks of the architecture. We also focus in this axis on administration models and their enforcement in relation to the architecture of the system, so that the exclusive control of a non expert individual can be ensured.
- (Axis 2) Global query evaluation. The goal of this line of research is to provide capabilities for crossing data belonging to multiple individuals (e.g., performing statistical queries over personal data, computing queries on social graphs or organizing participatory data collection) in a fully decentralized setting while providing strong and personalized privacy guarantees. This means proposing new secure distributed database indexing models and query processing strategies. In addition, we concentrate on locally ensuring to each participant the good behaviour of the processing, such that no collective results can be produced if privacy conditions are not respected by other participants.
- (Axis 3) Economic, legal and societal issues. This research axis is more transverse and entails multidisciplinary research, addressing the links between economic, legal, societal and technological aspects. We will follow here a multi-disciplinary approach based on a 3-step methodology: i) identifying important common issues related to privacy and to the exploitation of personal data; ii) characterizing their dimensions in all relevant disciplines and jointly study their entanglement; iii) validating the proposed analysis, models and trade-offs thanks to in vivo experiments.

These contributions will also rely on tools (algorithms, protocols, proofs, etc.) from other communities, namely security (cryptography, secure multiparty computations, formal methods, differential privacy,

etc.) and distributed systems (distributed hash tables, gossip protocols, etc.). Beyond the research actions, we structure our software activity around a single common platform (rather than isolated demonstrators), integrating our main research contributions, called PlugDB. This platform is the cornerstone to help validating our research results through accurate performance measurements on a real platform, a common practice in the DB community, and target the best conferences. It is also a strong vector to federate the team, simplify the bootstrapping of new PhD or master students, conduct multi-disciplinary research and open the way to industrial collaborations and technological transfers.

4 Application domains

4.1 Personal cloud, home care, IoT, sensing, surveys

As stated in the software section, the Petrus research strategy aims at materializing its scientific contributions in an advanced hardware/software platform with the expectation to produce a real societal impact. Hence, our software activity is structured around a common Secure Personal Cloud platform rather than several isolated demonstrators. This platform will serve as the foundation to develop a few emblematic applications. Several privacy-preserving applications can actually be targeted by a Personal Cloud platform, like: (i) smart disclosure applications allowing the individual to recover her personal data from external sources (e.g., bank, online shopping activity, insurance, etc.), integrate them and cross them to perform personal big data tasks (e.g., to improve her budget management) ; (ii) management of personal medical records for care coordination and well-being improvement; (iii) privacy-aware data management for the IoT (e.g., in sensors, quantified-self devices, smart meters); (iv) community-based sensing and community data sharing; (v) privacy-preserving studies (e.g., cohorts, public surveys, privacy-preserving data publishing). Such applications overlap with all the research axes described above but each of them also presents its own specificities. For instance, the smart disclosure applications will focus primarily on sharing models and enforcement, the IoT applications require to look with priority at the embedded data management and sustainability issues, while community-based sensing and privacy-preserving studies demand to study secure and efficient global query processing. Among these applications domains, one is already receiving a particular attention from our team. Indeed, we gained a strong expertise in the management and protection of healthcare data through our past DMSP (Dossier Medico-Social Partagé) experiment in the field. This expertise is being exploited to develop a dedicated healthcare and well-being personal cloud platform. We are currently deploying 10000 boxes equipped with PlugDB in the context of the DomYcile project. In this context, we are currently setting up an Inria Innovation Lab with the Hippocad company to industrialize this platform and deploy it at large scale (see Section the bilateral contract OwnCare II-Lab).

5 Highlights of the year

- During the lockdown, the Petrus team co-led and contributed to the PréLiFa project with collegues from the Inria M3DISIM team and the Hippocad company (our partner in OwnCare II-Lab, see section 8.1) and many other collegues from Inria and AP-HP, to design and implement an operational solution to keep the link between patients in intensive care units and their relatives, while the medical staff were overwhelmed and the families completely cut off from their relatives due to the covid situation. We deployed the PréLiFa platform on three AP-HP sites: Lariboisière, Saint-Louis and Beaujon Hospitals, before AP-HP, satisfied with the solution, organized its own broader deployment (for more information, see press release or Inria website).
- Our paper on consent-driven data reuse for participatory systems [13] was selected as a best paper candidate at the PerCom conference. This study is the result of a collaboration with the Inria Mimove team.
- The deployment in the field of a secure personal medical folder based on our PlugDB PDMS solution has started in the French Yvelines district. So far, 4.000 hardware devices integrating our solution have been manufactured and delivered to the Yvelines district (10.000 elderly patients are targeted in the near term in this district). see section 8.1 for more details.

6 New software and platforms

6.1 New software

6.1.1 PlugDB

Keywords: Databases, Personal information, Privacy, Hardware and Software Platform

Functional Description: PlugDB is a complete platform dedicated to a secure and ubiquitous management of personal data. It aims at providing an alternative to a systematic centralization of personal data. The PlugDB engine is a personal database server capable of storing data (tuples and documents) in tables and BLOBs, indexing them, querying them in SQL, sharing them through assertional access control policies and enforcing transactional properties (atomicity, integrity, durability).

The PlugDB engine is embedded in a tamper-resistant hardware device combining the security of smartcard with the storage capacity of NAND Flash. The personal database is hosted encrypted in NAND Flash and the PlugDB engine code runs in the microcontroller. Complementary modules allow to pre-compile SQL queries for the applications, communicate with the DBMS from a remote Java program, synchronize local data with remote servers (typically used for recovering the database in the case of a broken or lost devices) and participate in distributed computation (e.g., global queries). PlugDB runs both on secure devices provided by Gemalto and on specific secure devices designed by PETRUS and assembled by electronic SMEs. Mastering the hardware platform opens up new research and experiment opportunities (e.g., support for wireless communication, secure authentication, sensing capabilities, battery powered ...). PlugDB engine has been registered first at APP (Agence de Protection des Programmes) in 2009 - a new version being registered every two years - and the hardware datasheets in 2015.

PlugDB has been experimented in the field, notably in the healthcare domain. We also recently set up an educational platform on top of PlugDB, named SIPD (Système d'Information Privacy-by-Design) and used at ENSIIE, INSA CVL and UVSQ through the Versailles Sciences Lab fablab, to raise students awareness of privacy protection problems and embedded programming. As a conclusion, PlugDB combines several research contributions from the team, at the crossroads of flash data management, embedded data processing and secure distributed computations. It then strongly federates all members of our team (permanent members, PhD students and engineers). It is also a vector of visibility, technological transfer and dissemination and gives us the opportunity to collaborate with researchers from other disciplines around a concrete privacy-enhancing platform.

PlugDB is now being industrialized in the context of the OwnCare Inria Innovation Lab (II-Lab). In OwnCare, PlugDB acts as a secure personal cloud to manage medical/social data for people receiving care at home. It should be deployed over 10.000 patient in the Yvelines district. The industrialization process covers the development of a complete testing environment, the writing of a detailed documentation and the development of additional features (e.g., embedded ODBC driver, TPM support, flexible access control model and embedded code upgrade notably). It has also required the design of a new hardware platform equipped with a battery power supply, introducing new energy consumption issues for the embedded software..

URL: https://project.inria.fr/plugdb/

Authors: Nicolas Anciaux, Luc Bouganim, Philippe Pucheral, Aydogan Ersoz, Laurent Schneider, Quentin Lefebvre

Contacts: Aydogan Ersoz, Laurent Schneider, Philippe Pucheral, Nicolas Anciaux, Luc Bouganim

6.2 New platforms

Participants Floris Thiant *(correspondent)*, Nicolas Anciaux, Luc Bouganim, Iulian Sandu Popa, Guillaume Scerri.

Personal Data Management Systems (PDMS) arrive at a rapid pace boosted by smart disclosure initiatives and new regulations such as GDPR. However, our recent survey [1] indicates that the existing PDMS solutions cover partially the PDMS data life-cycle and, more importantly, focus on specific privacy threats depending on the employed architecture. To address this issue, we proposed in [1] a logical reference architecture for an extensive (i.e., covering all the major functionalities) and secure (i.e., circumventing all the threats specific to the PDMS context) PDMS. We also discussed several possible physical instances fo the architecture and showed that TEEs (Trusted Execution Environments) are a prime option for building a trustworthy PDMS platform [2].

Hence, based on our previous studies, we currently develop an extensive and secure PDMS (ES-PDMS) platform using the state-of-the-art TEE technology available today, i.e., Intel Software Guard eXtension (SGX). Our ES-PDMS software stack can be deployed on any SGX-enabled machine (i.e., any relatively recent computer having an Intel CPU). We have acquired a server with 6 Intel Xeon E-2276G CPU cores allowing a collaborative development of the ES-PDMS protoype by the Petrus team as well as the related Personal Cloud applications. We note that the Intel Xeon CPU series offer access to the Intel Attestation Service[1] which is required for the remote enclave attestation process. This allows us to implement (and not only to simulate) remote attestations required by several use-cases in the PDMS context (e.g., distributed computations between a network of PDMSs or attesting results of local computations on the PDMS to a third party).

7 New results

7.1 Security properties of a PDMS (Axis 1)

Participants Guillaume Scerri *(correspondent)*.

The architecture proposed by the Petrus team for Extensible and Secure Personal Data Management Systems (ES-PDMS) [1] is based on three layers : a *core* engine –small trusted code base (TCB)–, *data tasks* processing personal data –large code modules in which trust is limited– and *applications* with authorized access to the computation results –on which no security assumption can be made–. The different components of the architecture are ruled by different security properties and must be *composed* at runtime to provide the expected data oriented (rich) functionality while meeting the desired security objectives [2]. Performing a proof of such a large andcomplex system directly would be too complex to do without decomposing it in smaller subprotocols and goals. As a first step towards a formal evaluation of this proposal, in collaboration with Hubert Comon and Charlie Jacomme (LSV, ENS Paris-Saclay), we proposed at CCS [14] a novel composition framework together with a variety of composition theorems allowing to split the security proof of an unbounded number of sessions of a compound protocol into simpler goals. The application of the proposed techniques (and more generally, formal methods) to the case of PDMS is part of the future work.

7.2 PDMS architecture for Intel SGX (Axis 1)

Participants Iulian Sandu Popa *(correspondent)*, Nicolas Anciaux, Luc Bouganim, Robin Carpentier, Floris Thiant.

To design a plateform for ES-PDMS [1], our approach is to use trusted execution environments. The challenge is to orchestrate the different data-related tasks using secure hardware enclaves, in order to provide rich data-oriented functionality while meeting the desired security objectives. We have begun the implementation of a new PDMS plateform based on Intel SGX (see section 6.2). By decomposing the computation into data tasks, we aim to guarantee the *bilateral security* property, on the one hand by

[1] https://software.intel.com/content/www/us/en/develop/topics/software-guard-extensions/attestation-services.html

protecting the data confidentiality and privacy of the PDMS owner, and on the other hand by guaranteeing the integrity of the computed results shared with third parties. This work is part of Robin Carpentier's PhD. A demonstration of our proposal has been designed and is being implemented for submission to the next VLDB conference.

7.3 Consent-driven data reuse (Axis 1)

Participants Mariem Brahem *(correspondent)*, Nicolas Anciaux, Guillaume Scerri.

Participatory sensing allows individuals to contribute data across time and space in order to feed general interest computation tasks. However, there are major obstacles including the preservation of the privacy of contributors. This consideration has led to two main approaches, sometimes combined, which are, respectively, to trade privacy for rewards, and to take advantage of privacy-enhancing technologies anonymizing the collected data. Although relevant, we claim that these approaches do not sufficiently take into account the consent of the participants to the use of the personal data provided and may even lead to *defects in consent* even in the presence of a trusted system. To address this issue, we introduce the ℓ-completeness property, which ensures that the data provided can be reused for all the tasks to which their contributors consent as long as they are analyzed with a same set of $\ell - 1$ other sources. We propose a clustering strategy sensitive to the data distribution, which is shown to optimize data reuse and utility. This study [13], conducted in collaboration with Valerie Issarny (Inria Mimove team), appears at PerCom. The design of an participatory sensing architecture leveraging on trusted execution environements to support stronger attackers models is ongoing.

7.4 DISPERS (Axis 2)

Participants Luc Bouganim *(correspondent)*, Julien Loudet, Julien Mirval, Iulian Sandu Popa.

Personal Data Management System (PDMS) solutions are flourishing, boosted by smart disclosure initiatives and new regulations. PDMSs allow users to easily store and manage data directly generated by their devices or resulting from their interactions. Users can then leverage the power of their PDMS to benefit from their personal data, for their own good and even in the interest of the community. The PDMS paradigm brings thus exciting perspectives by unlocking novel usages, but also raises security issues. An effective approach, considered in several recent works, is to let the user data distributed on personal platforms, secured locally using different hardware and/or software security mechanisms. This study goes beyond the local security issues and addresses the important question of querying securely this massively distributed personal data. To this end, we proposed DISPERS, a fully-distributed PDMS peer-to-peer architecture. DISPERS allows users to securely and efficiently share and query their personal data, even in the presence of malicious nodes. We considered three increasingly powerful threat models and derived for each a security requirement that must be incrementally fulfilled to minimize data disclosure: (i) *hidden communications*, (ii) *random dispersion of data*, and (iii) *collaborative proofs* to resist respectively to potentially spied, leaking or corrupted nodes. We showed that the expected security level can be guaranteed with very high probability and validated experimentally the efficiency of the proposed protocols, allowing for tunable trade-off between the security level and its cost. This work led to a long journal paper, based on our previous works [6, 1] and an extension of Julien's Phd results. This paper was recently submitted to the VLDB Journal. The follow-up of this work is the topic of Julien Mirval's PhD (CIFRE) thesis (see section 8.2) which has recently started (Nov 2020) and focuses on distributed and secure machine learning protocols with PDMSs.

7.5 Secure distributed queries over large sets of PDMS home boxes (Axis 2)

Participants Nicolas Anciaux *(correspondent)*, Riad Ladjel, Luc Bouganim, Ludovic Javet, Philippe Pucheral, Guillaume Scerri.

Smart disclosure initiatives and new regulations such as GDPR allow individuals to get the control back on their data by gathering their entire digital life in a Personal Data Management Systems (PDMS). Multiple PDMS architectures exist and differ on their ability to preserve data privacy and to perform collective computations crossing data of multiple individuals (e.g., epidemiological or social studies) but none of them satisfy both objectives. The emergence of Trusted Execution Environments (TEE) changes the game. In a journal paper at TLDKS [12], we propose a solution called Trusted PDMS, combining the TEE and PDMS properties to manage the data of each individual, and a complete framework to execute collective computation on top of them, with strong privacy and fault tolerance guarantees. We demonstrate the practicality of the solution through a simulation based on a real case-study being conducted over 10.000 patients in the healthcare field. In this same context, Ludovic Javet started a PhD thesis in order to further study the processing of distributed queries in a delay-tolerant environment, typical of PDMS home boxes.

7.6 Hidding communications patterns in distributed queries (Axis 2)

Participants Riad Ladjel *(correspondent)*, Nicolas Anciaux, Guillaume Scerri.

The execution of distributed queries on populations of PDMS involves communication patterns between computing nodes (see [4]), which may depend on the values of the personal data being processed (a computing node may aggregate personal data corresponding to a given range of sensitive values). An attacker observing communications, even encrypted, can potentially infer personal information about participants. The use of traditional solutions to conceal data-dependent communications at runtime results in either significant performance penalties or privacy gains that are difficult to quantify. Chapter 4 of Riad Ladel's PhD manuscript [16] formulates as an ϵ-differential privacy problem the issue of concealing communication patterns in distributed queries. In section 5 of [12], we propose a simple solution to the specific case of home box PDMS. In collaboration with Aurélien Bellet (Inria Magnet team), we are currently studying the proposed algorithms using the differential confidentiality model (ϵ,δ)-differential privacy, allowing a finer analysis of the trade-offs between privacy, performance and utility in practice.

7.7 Empowerment and Big Data on Personal Data (Axis 3)

Participants Nicolas Anciaux *(correspondent)*, Riad Ladjel.

The place of individuals and the control of their data have emerged as central issues in the European data protection regulation. The *empowerment* of the individual has notably resulted in the recognition of a new prerogative for the individual: the right to the portability of personal data. The corollary of this new right is the design and deployment of technical platforms, commonly known as Personal Data Mangement Servers (PDMS) or PIMS, allowing the individual to consolidate all his or her data in a single system managed under his or her control. On the strength of these technical and legal innovations, several questions arise: what forms of empowerment are targeted in practice? What are the appropriate conditions to guarantee the objective pursued? At the crossroads of these questions, one dimension appears to be insufficiently exploited: that of *agency*. During this period, in collaboration with law researchers Célia Zolynski (IRJS, Sorbonne Univ.) and Sébastien Chaudat (DANTE, Univ. of Versailles), we have transposed this notion from the social sciences to the management of personal data and provided a new reading of the empowerment measures of Big Data functionalities on personal data. This study led to two journal publications in the Dalloz IP/IT [15] and in the Global Privacy Law Review [11].

7.8 OwnCare Inria Innovation Lab (Axis 3)

Participants Philippe Pucheral *(correspondent)*, Nicolas Anciaux, Luc Bouganim, Laurent Schneider.

The OwnCare IILab was created in January 2018 (see section: Bilateral Contracts with Industry) and involves the Hippocad SME and the PETRUS team around the management of medical-social data at patient's home. The objective is to build a fully decentralized and highly secure personal medical-social folder based on PlugDB, and deploy it at large scale. Besides this industrial objective, the goal is also to leverage and validate the PETRUS research contributions related to secured Personal Cloud architectures. Before the creation of the OwnCare IILab initiative, PlugDB was an advanced research prototype. It is now evolving towards a transferable industrial product. To reach this state, a considerable effort has been made in terms of development, testing platform, validation procedures and documentation. PlugDB engine is regularly registered at APP (Agence de Protection des Programmes), for both the PlugDB hardware datasheets and the code of the PlugDB-engine. The next PlugDB code registration will cover all functionalities added since the beginning of the IILab, notably: dynamic upgrade of the embedded code, TPM-based secure boot, ad-hoc embedded stored procedures, RBAC-style access control model, aggregate computation, SSL certificate management, event/error logging mechanism. Some of these developments are highly challenging considering the embedded context and the energy consumption constraints we have to face (the current device hosting PlugDB is based on two microcontrollers – MCU – powered by small batteries). Typically, we had to implement the first coupling between a TPM and a STM MCU, a lightweight version of SSL that accommodates MCU resources and energy-saving synchronization protocols between 2 MCU.

8 Bilateral contracts and grants with industry

8.1 Bilateral contracts with industry

OwnCare II-Lab (Jul 2017 - Dec 2020) Partners: PETRUS, Hippocad
End 2016, the Yvelines district lauched a public call for tender to deploy a personal medical-social folder aiming at covering the whole distinct (10.000 patients). The Hippocad company, in partnership with Inria, won this call for tender with a solution called DomYcile and the project was launched in July 2017. DomYcile is based on a home box combining the PlugDB hardware/software technology developed by the Petus team and a communication layer based on SigFox. Hippocad and Petrus then decided to launch a joint II-Lab (Inria Innovation Lab) named OwnCare in January 2018. The objective is threefold: (1) build an industrial solution based on PlugDB and deploy it in the Yvelines district in the short-term, (2) use this Yvelines testbed to improve the solution and try to deploy it at the national/international level in the medium-term and (3) design flexible/secure/mobile personal medical folder solutions addressing new usages (IoT, machine learning models, distributed statistics, etc.) in the long-term. In 2020, important progress has been made on Point (1) with the manufacture of 4.000 home boxes being deployed in the Yvelines district, R&D efforts have been put on Point (2) to improve the robustness and ease of use of the platform and new research works have been initiated on Point (3) notably regarding the design of decentralized privacy-preserving distributed computations (see section 7.5).

8.2 Bilateral grants with industry

Cozy Cloud CIFRE - Mirval contract (Nov 2020 - Oct 2023) Partners: Cozy Cloud, PETRUS
A third CIFRE PhD thesis has been started between Cozy Cloud and Julien Mirval from PETRUS. Cozy Cloud is a French startup providing a personal Cloud platform. The Cozy product is a software stack that anyone can deploy to run his personal server in order to host his personal data and web services. The objective of this thesis is to propose appropriate solutions to effectively train an AI model (e.g., a deep neural network) in a fully distributed system while providing strong security guarantees to the participating nodes. The results, in the form of protocols and distributed and secure execution algorithms, will be applied to practical cases provided by the Cozy Cloud company.

9 Partnerships and cooperations

9.1 National initiatives

9.1.1 ANR PerSoCloud (Jan 2017 - Mar 2021)

Partners: Orange Labs (coordinator), PETRUS (Inria-UVSQ), Cozy Cloud, U. of Versailles.
The objective of PerSoCloud is to design, implement and validate a full-fledged Privacy-by-Design Personal Cloud Sharing Platform. One of the major difficulties linked to the concept of personal cloud lies in organizing and enforcing the security of the data sharing while the data is no longer under the control of a central server. We identify three dimensions to this problem. Devices-sharing: assuming that the primary copy of user U1's personal data is hosted in a secure place, how to share and synchronize it with U1's multiple (mobile) devices without compromising security? Peers-sharing: how user U1 could exchange a subset of his-her data with an identified user U2 while providing to U1 tangible guarantees about the usage made by U2 of this data? Community-sharing: how user U1 could exchange a subset of his-her data with a large community of users and contribute to personal big data analytics while providing to U1 tangible guarantees about the preservation of his-her anonymity? In addition to tackling these three scientific and technical issues, a legal analysis will guarantee compliance of this platform with the security and privacy French and UE regulation, which firmly promotes the Privacy by Design principle, including the current reforms of personal data regulation.

9.1.2 GDP-ERE, DATAIA project (Sept. 2018 - Jan. 2022)

Partners: DANTE (U. of Versailles), PETRUS (Inria-UVSQ).
The role of individuals and the control of their data is a central issue in the new European regulation (GDPR) enforced on 25th May 2018. Data portability is a new right provided under those regulations. It allows citizens to retrieve their personal data from the companies and governmental agencies that collected them, in an interoperable digital format. The goals are to enable the individual to get out of a captive ecosystem, and to favor the development of innovative personal data services beyond the existing monopolistic positions. The consequence of this new right is the design and deployment of technical platforms, commonly known as Personal Cloud. But personal cloud architectures are very diverse, ranging from cloud based solutions where millions of personal cloud are managed centrally, to self-hosting solutions. These diversity is not neutral both in terms of security and from the point of view of the chain of liabilities. The GDP-ERE project tends to study those issues in an interdisciplinary approach by the involvement of jurists and computers scientists. The two main objectives are (i) to analyze the effects of the personal cloud architectures on legal liabilities, enlightened by the analysis of the rules provided under the GDPR and (ii) to propose legal and technological evolutions to highlight the share of liability between each relevant party and create adapted tools to endorse those liabilities.
http://dataia.eu/actualites/linstitut-dataia-vous-presente-le-projet-gdp-ere-rgpd-et-cloud-personnel-de-lempowerment

9.2 Regional initiatives

9.2.1 Postdoc DIM RFSI, Ile-de-France Region (2019 - 2020)

Partners: Inria (PETRUS).
This project is a continuation of Julien Loudet's Phd thesis. Julien finalized a CIFRE thesis defended in October 2019. This thesis is the result of a solid collaboration (another CIFRE thesis was defended in 2018) between the PETRUS team and the startup Cozy Cloud, which is also working on the personal cloud issue. The project finances 8 months of postdoc for Julien. This postdoc allowed to finalize an extended journal paper on the thesis results (DISPERS protocol) and to collaborate with Julien Mirval, a PhD candidate of a new thesis in collaboration with Cozy Cloud exploring decentralized automatic learning techniques in the personal cloud context (see section 8.2).

9.2.2 Prevadom, Paris-Saclay grant (2020-2021)

Partners: Inria (PETRUS).
The objective of the Prevadom project is to integrate IoT facilities and machine learning algorithms into the DomYcile PlugDB-enabled boxes to detect and ideally prevent loneliness and despair situations. Such situations are peculiarly alarming during periods where patients are confined at home (e.g., during a pandemy). To this end, the PlugDB-enabled boxes will be augmented with IoT communication facilities in order to collect data from sensors (e.g., light, temperature, hygrometry, motion) and quantified-self devices (e.g., thermometer, oxymeter, blood pressure). Such raw data are highly intrusive. The objective is then to store them and analyse them with machine learning models locally (i.e., inside the box) and only externalize aggregated data or launch alarms when required.

10 Dissemination

10.1 Promoting scientific activities

10.1.1 Scientific events: organisation

Member of the organizing committees

- Iulian Sandu Popa: Organizing Chair, IEEE International Conference on Mobile Data Management (MDM 2020), June 30-July 3, 2020
- Luc Bouganim: Co-organizer "École thématique BDA Masses de Données Distribuées", Bastia (cancelled in 2020, trying in 2021)

10.1.2 Scientific events: selection

Chair of conference program committees

- Niocolas Anciaux: vice-president of the jury of the 5th edition of CNIL-Inria Privacy Award.

Member of the conference program committees

- Nicolas Anciaux: Int. Conf. on Extending Database Technology (EDBT 2021), IEEE International Conference on Smart Computing (SmartComp 2021), Atelier sur la Protection de la Vie Privée (APVP 2020, finally cancelled)
- Luc Bouganim: Int. Conf. on Extending Database Technology (EDBT 2021), Int. Conf. on Extending Database Technology (EDBT 2020), 36ème Conférence sur la Gestion de Données – Principes, Technologies et Applications (BDA 2020).
- Iulian Sandu Popa: IEEE Int. Conf. on Data Engineering (ICDE 2020), Int. Conf. on Scientific and Statistical Database Management (SSDBM 2020), Int. Conf. on Data Science, Technology and Applications (DATA 2020), IEEE Int. Conf. on Mobile Cloud Computing, Services, and Engineering (MobileCloud 2020)

10.1.3 Journal

Member of the editorial boards

- Nicolas Anciaux: Associate Editor at the VLDB Journal

10.1.4 Scientific expertise

- Nicolas Anciaux: Member of the program committee of the DATAIA Paris-Saclay Institute

10.1.5 Research administration

- Nicolas Anciaux: responsible of the "Formation par la Recherche (FpR)" and represent of the "Mission Jeunes Chercehurs (MJC)" at Inria-Saclay
- Nicolas Anciaux: member of the "COnseil de la POLitique doctorale (COPOL)" at University Paris-Saclay
- Nicolas Anciaux: correspondent for the STIC doctoral school at Inria-Saclay
- Nicolas Anciaux: member of the "Bureau du Comité des Projets (BCEP)" at Inria-Saclay
- Nicolas Anciaux: member of the "Bureau du Laboratoire" at DAVID Lab.
- Luc Bouganim: Member of the Scientific Commission (CS) of Inria Saclay-IDF (Cordi-S, Post-Doc, Delegation)
- Luc Bouganim: PhD thesis referent for the Doctoral School of University Paris-Saclay
- Philippe Pucheral: member of the 'Commission recherche' and of the 'Conseil Académique CaC' of University of Versailles
- Philippe Pucheral: Member of the steering committee of the ED STIC doctoral school of University Paris-Saclay, 'Data, Knowledge and Interactions' committee
- Philippe Pucheral: Member of the 'Habilitation à Diriger des Recherches' committee of University of Versailles - UFR des Sciences
- Philippe Pucheral: member of the "Bureau du Laboratoire" at DAVID Lab.
- Iulian Sandu Popa: PhD thesis referent for the Doctoral School of University Paris-Saclay
- Guillaume Scerri: PhD thesis referent for the Doctoral School of University Paris-Saclay

10.2 Teaching - supervision - juries and selection committees

10.2.1 Teaching

- Licence : Iulian Sandu Popa, Bases de données (niveau L3), 96, UVSQ, France. Guillaume Scerri, Fondements de l'informatique (niveau L1), 36, UVSQ, France. Guillaume Scerri, Théorie des Langages (niveau L2), 51, UVSQ, France.
- Master : Iulian Sandu Popa, Bases de données relationnelles (niveau M1), Gestion des données spatiotemporelles (niveau M2), Sécurité des bases de données (niveau M2), 96, UVSQ, France. Philippe Pucheral, responsible of the DataScale master, courses in M1 and M2 in databases and in security, introductory courses for jurists,UVSQ, France. Guillaume Scerri, Sécurité et bases de données pour juristes, 6, UVSQ, France. Guillaume Scerri, Sécurité, 3, UVSQ, France.
- Engineers school : Nicolas Anciaux, courses on Databases (module IN206, niveau M1), 21, and Advanced databases (module ASI13, niveau M2), 24. Philippe Pucheral, Gestion de Données Avancées (niveau M1), 30, ENSIIE Evry, France. Luc Bouganim, Bases de données relationnelles (niveau M1), 32, ENSTA, France.
- MOOC : "Villes intelligentes : défis technologiques et sociétaux", organisé par Valérie Issarny et Nathalie Mitton. Co-Auteurs: Nicolas Anciaux, Stéphane Grumbach, Valérie Issarny, Nathalie Mitton, Christine Morin, Animesh Pathak et Hervé Rivano. Sessions sur la plateforme FUN, avec 13125 inscrits depuis mai 2019.

10.2.2 Supervision

- PhD: Riad Ladjel: Calculs Distribués et Sécurisés pour le Cloud Personnel, Paris-Saclay, Dec 8, 2020, Nicolas Anciaux, Philippe Pucheral and Guillaume Scerri
- PhD in progress: Robin Carpentier, Secure and efficient data processing in trusted execution environments for the personal cloud, UVSQ, October 2018, Nicolas Anciaux, Iulian Sandu Popa and Guillaume Scerri
- PhD in progress: Julien Mirval, DISSEC-ML : DIStributed and SECure Machine Learning on Personal Clouds, UVSQ, November 2020, Luc Bouganim and Iulian Sandu Popa
- PhD in progress: Ludovic Javet, Requêtes distribuées respectueuses de la vie privée dans un environnement partiellement connecté, Inria, Januray 2020, Luc Bouganim, Nicolas Anciaux and Philippe Pucheral

10.2.3 Juries and selection committees

- Nicolas Anciaux : Member of the selection committee (COS) for the position of Associate Professor No. 503 of the ENSIIE in system and networks (recruitment campaign 2020).
- Nicolas Anciaux : Member of the PhD jury of Riad Ladjel (Paris-Saclay University, 8/12/2020).
- Luc Bouganim: Reviewer of the HDR of Patricia Serrano Alvarado (Nantes University, 16/06/2020)
- Luc Bouganim: Reviewer of the PhD of N. Andriamilanto (IRISA, IRT b<>com, 17/12/2020)
- Philippe Pucheral : President of the selection committee (COS) for the position of Associate Professor No. 4219 of University of Versailles (recruitment campaign 2020)
- Philippe Pucheral : Member of the selection committee (COS) for the position of Professor No. 2201 of Paris-Saclay University (recruitment campaign 2020)
- Philippe Pucheral : Member of the PhD jury of Benjamin Moreau (Nantes University, 06/11/2020)
- Guillaume Scerri : Invited member of the PhD jury of Riad Ladjel (Paris-Saclay University, 8/12/2020).

10.3 Popularization

10.3.1 Internal or external Inria responsibilities

- Iulian Sandu Popa: Member of the technological development comission (CDT - "Commission de Développement Technologique") for INRIA Saclay-IdF (since 2019)

10.3.2 Interventions

- "Démonstration de PlugDB - projet DomYcile", Nicolas Anciaux, Ambition Inria 2023, espace des démonstrations, Gaîté lyrique, Feb 2020.
- Table ronde "Ethics and AI", Nicolas Gogin, Jérôme Perrin, Nicolas Anciaux, Laurence Devillers and Michèle Sebag, DATAIA Day, 1st Data Science, Intelligence & Society Conference, Mar 2020.
- Présentation "Research Projects Overview: GDP-ERE", Célia Zolynski and Nicolas Anciaux, DATAIA Day, 1st Data Science, Intelligence & Society Conference, Mar 2020.
- "L'agentivité, c'est le pouvoir d'utiliser collectivement nos données personnelles, sous forme de Big Data citoyen", Laboratoire d'Innovation Numérique de la CNIL (LINC), par Félicien Vallet, 18 mai 2020, suite à une présentation à la CNIL des travaux de l'équipe Petrus par Nicolas Anciaux. (lien)
- Interview de Nicolas Anciaux pour un article de la *Revue des deux mondes*, "Vous reprendrez bien un cookie ?", par Kyrill Nikitin, Nov 2020.

11 Scientific production

11.1 Major publications

[1] N. Anciaux, P. Bonnet, L. Bouganim, B. Nguyen, P. Pucheral, I. Sandu-Popa and G. Scerri. 'Personal Data Management Systems: The security and functionality standpoint'. In: *Information Systems* 80 (2019), pp. 13–35. DOI: 10.1016/j.is.2018.09.002. URL: https://hal.archives-ouvertes.fr/hal-01898705.

[2] N. Anciaux, L. Bouganim, P. Pucheral, I. S. Popa and G. Scerri. 'Personal Database Security and Trusted Execution Environments: A Tutorial at the Crossroads'. In: *Proceedings of the VLDB Endowment (PVLDB)* (Aug. 2019). DOI: 10.14778/3352063.3352118. URL: https://hal.inria.fr/hal-02269292.

[3] C. Berthet, C. Zolynski, N. Anciaux and P. Pucheral. '" Contenus numériques et récupération des données : un nouvel outil d' 'empouvoirement' du consommateur ? "' In: *Dalloz IP/IT* IP IT / 10 (Jan. 2017). URL: https://hal.inria.fr/hal-01429939.

[4] R. Ladjel, N. Anciaux, P. Pucheral and G. Scerri. 'Trustworthy Distributed Computations on Personal Data Using Trusted Execution Environments'. In: *TrustCom 2019 - The 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications / BigDataSE 2019 - 13th IEEE International Conference on Big Data Science and Engineering*. Rotorua, New Zealand, Aug. 2019. DOI: 10.1109/TrustCom/BigDataSE.2019.00058. URL: https://hal.archives-ouvertes.fr/hal-02269207.

[5] S. Lallali, N. Anciaux, I. Sandu-Popa and P. Pucheral. 'Supporting secure keyword search in the personal cloud'. In: *Information Systems* 72 (Dec. 2017), pp. 1–26. DOI: 10.1016/j.is.2017.09.003. URL: https://hal.inria.fr/hal-01660599.

[6] J. Loudet, I. Sandu-Popa and L. Bouganim. 'SEP2P: Secure and Efficient P2P Personal Data Processing'. In: *EDBT 2019 - 22nd International Conference on Extending Database Technology*. Lisbon, Portugal, Mar. 2019. URL: https://hal.inria.fr/hal-01949641.

[7] S. J. Pan, I. Sandu-Popa and C. Borcea. 'DIVERT: A Distributed Vehicular Traffic Re-Routing System for Congestion Avoidance'. In: *IEEE Transactions on Mobile Computing* 16.1 (Jan. 2017), pp. 58–72. DOI: 10.1109/TMC.2016.2538226. URL: https://hal.inria.fr/hal-01426424.

[8] G. Scerri, B. Warinschi, M. Barbosa and B. Portela. 'Foundations of Hardware-Based Attested Computation and Application to SGX'. In: *IEEE European Symposium on Security and Privacy, EuroS&P 2016*. IEEE European Symposium on Security and Privacy, EuroS&P 2016, Saarbrücken, Germany, March 21-24, 2016. Saarbrücken, Germany: IEEE, Mar. 2016, pp. 245–260. DOI: 10.1109/EuroSP.2016.28. URL: https://hal.inria.fr/hal-01417137.

[9] C. Q. To, B. Nguyen and P. Pucheral. 'Private and Scalable Execution of SQL Aggregates on a Secure Decentralized Architecture'. In: *ACM Transactions on Database Systems* 41.3 (2016), p. 46. DOI: 10.1145/2894750. URL: https://hal.archives-ouvertes.fr/hal-01296432.

[10] P. Tran-Van, N. Anciaux and P. Pucheral. 'SWYSWYK: A Privacy-by-Design Paradigm for Personal Information Management Systems'. In: *International Conference on Information Systems Development (ISD)*. Cyprus, Cyprus, Sept. 2017. URL: https://hal.inria.fr/hal-01675090.

11.2 Publications of the year

International journals

[11] N. Anciaux, C. Zolynski, S. Chaudat and R. Ladjel. 'Empowerment and Big Personal Data: from Portability to Personal Agency'. In: *Global Privacy Law Review* (2021). URL: https://hal.inria.fr/hal-03140409.

[12] R. Ladjel, N. Anciaux, P. Pucheral and G. Scerri. 'Secure distributed queries over large sets of personal home boxes'. In: *Transactions on Large-Scale Data- and Knowledge-Centered Systems* (10th Sept. 2020). URL: https://hal.archives-ouvertes.fr/hal-02941076.

International peer-reviewed conferences

[13] M. Brahem, G. Scerri, N. Anciaux and V. Issarny. 'Consent-driven data use in crowdsensing platforms: When data reuse meets privacy-preservation'. In: PerCom 2021 - IEEE International Conference on Pervasive Computing and Communications. Kassel / Virtual, Germany: http://www.percom.org/, 22nd Mar. 2021. URL: https://hal.inria.fr/hal-03097047.

[14] H. Comon, C. Jacomme and G. Scerri. 'Oracle simulation: a technique for protocol composition with long term shared secrets'. In: ACM CCS 2020. CCS '20: Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. Orlando, United States, 9th Nov. 2020, pp. 1427–1444. URL: https://hal.inria.fr/hal-02913866.

Scientific book chapters

[15] N. Anciaux and C. Zolynski. 'Empowerment et Big Data sur données personnelles : de la portabilité à l'agentivité'. In: *Le Big Data et le Droit*. Thèmes et Commentaire. 2020. URL: https://hal.inria.fr/hal-02349274.

Doctoral dissertations and habilitation theses

[16] R. Ladjel. 'Secure Distributed Computations for the Personal Cloud'. Paris-Saclay, 8th Dec. 2020. URL: https://hal.inria.fr/tel-03141197.