IN PARTNERSHIP WITH:
CNRS

Université Versailles Saint-Quentin

Activity Report 2014

Project-Team SMIS

Secured and Mobile Information Systems

IN COLLABORATION WITH: Parallelisme, réseaux, systèmes, modélisation (PRISM)

RESEARCH CENTER
Paris - Rocquencourt

THEME
Data and Knowledge Representation and Processing

Table of contents

Project-Team SMIS

Keywords: Databases, Privacy, Ubiquitous Computing, Distributed System, Information Indexing And Retrieval

Creation of the Project-Team: 2004 September 01.

1. Members

Research Scientists
Nicolas Anciaux [Inria, Researcher, HdR]
Luc Bouganim [Inria, Senior Researcher, HdR]

Faculty Members
Philippe Pucheral [Team leader, Univ. Versailles, Professor, HdR]
Benjamin Nguyen [Univ. Versailles, Associate Professor, until Aug 2014, now Professor at INSA CVL, HdR]
Iulian Sandu Popa [Univ. Versailles, Associate Professor]

Engineers
Aydogan Ersoz [Inria, from Apr 2014]
Quentin Lefebvre [Inria]

PhD Students
Matias Bjørling [Univ. of Copenhagen, co-supervision]
Niv Dayan [Univ. of Copenhagen, co-supervision]
Athanasia Katsouraki [Inria, CORDI-S]
Saliha Lallali [Inria, granted by ANR KISS project]
Quoc-Cuong To [Inria, CORDI-S]
Paul Tran Van [CIFRE Cozy]

Visiting Scientist
Philippe Bonnet [Marie-Curie grant, Inria, until Jul 2014]

Administrative Assistant
Laurence Bourcier [Inria]

2. Overall Objectives

2.1. Overall Objectives

The research work within the project-team is devoted to the design and analysis of core database techniques dedicated to the definition of secured and mobile information systems.

Ubiquitous computing and ambient intelligence entail embedding data in increasingly light and specialized devices (chips, sensors and electronic appliances for smart buildings, telephony, transportation, health, etc.). These devices exhibit severe hardware constraints to match size, security, power consumption and also production costs requirements. At the same time, they could highly benefit from embedded database functionalities to store data, analyze it, query it and protect it. This raises a first question "Q_1: *How to make powerful data management techniques compatible with highly constrained hardware platforms?*". To tackle this question, SMIS contributes to the design and validation of new storage and indexing models, query execution and optimization techniques, and transaction protocols. The relevance of this research goes beyond embedded databases and may have potential applications for database servers running on advanced hardware.

By making information more accessible and by multiplying –often transparently– the means of acquiring it, ubiquitous computing involves new threats for data privacy. The second question addressed by the project-team is then "Q_2: *How to make smart objects less intrusive?*". New access and usage control models have to be devised to help individuals keep a better control on the acquisition and sharing conditions of their data. This means integrating privacy principles like user's consent, limited collection and limited retention in the access and usage control policy definition. This also means designing appropriate mechanisms to enforce this control and provide accountability with strong security guarantees.

In parallel, thanks to a high degree of decentralization and to the emergence of low cost tamper-resistant hardware, ubiquitous computing contains the seeds for new ways of managing personal/sensitive data. The third question driving the research of the project-team is therefore "Q_3: *How to build privacy-by-design architectures based on trusted smart objects?*". The objective is to capitalize on embedded data management techniques, privacy-preserving mechanisms, trusted devices and cryptographic protocols to define an integrated framework dedicated to the secure management of personal/sensitive data. The expectation is showing that credible alternatives to a systematic centralization of personal/sensitive data on servers can be devised and validating the approach through real case experiments.

3. Research Program

3.1. Embedded Data Management

The challenge tackled is this research action is twofold: (1) to design embedded database techniques matching the hardware constraints of (current and future) smart objects and (2) to set up co-design rules helping hardware manufacturers to calibrate their future platforms to match the requirements of data driven applications. While a large body of work has been conducted on data management techniques for high-end servers (storage, indexation and query optimization models minimizing the I/O bottleneck, parallel DBMS, main memory DBMS, etc.), less research efforts have been placed on embedded database techniques. Light versions of popular DBMS have been designed for powerful handheld devices; yet DBMS vendors have never addressed the complex problem of embedding database components into chips. Proposals dedicated to databases embedded on chip usually consider small databases, stored in the non-volatile memory of the microcontroller –hundreds of kilobytes– and rely on NOR Flash or EEPROM technologies. Conversely, SMIS is pioneering the combination of microcontrollers and NAND Flash constraints to manage Gigabyte(s) size embedded databases. We present below the positioning of SMIS with respect to international teams conducting research on topics which may be connected to the addressed problem, namely work on electronic stable storage, RAM consumption and specific hardware platforms.

Major database teams are investigating data management issues related to hardware advances (EPFL: A. Ailamaki, CWI: M. Kersten, U. Of Wisconsin: J. M. Patel, Columbia: K. Ross, UCSB: A. El Abbadi, IBM Almaden: C. Mohan, etc.). While there are obvious links with our research on embedded databases, these teams target high-end computers and do not consider highly constrained architectures with non traditional hardware resources balance. At the other extreme, sensors (ultra-light computing devices) are considered by several research teams (e.g., UC Berkeley: D. Culler, ITU: P. Bonnet, Johns Hopkins University: A. Terzis, MIT: S. Madden, etc.). The focus is on the processing of continuous streams of collected data. Although the devices we consider share some hardware constraints with sensors, the objectives of both environments strongly diverge in terms of data cardinality and complexity, query complexity and data confidentiality requirements. Several teams are looking at efficient indexes on flash (HP LABS: G. Graefe, U. Minnesota: B. Debnath, U. Massachusetts: Y. Diao, Microsoft: S. Nath, etc.). Some studies try to minimize the RAM consumption, but the considered RAM/stable storage ratio is quite large compared to the constraints of the embedded context. Finally, a large number of teams have focused on the impact of flash memory on database system design (we presented an exhaustive state of the art in a VLDB tutorial [7]). The work conducted in the SMIS team on bi-modal flash devices takes the opposite direction, proposing to influence the design of flash devices by the expression of database requirements instead of running after the constantly evolving flash device technology.

3.2. Access and Usage Control Models

Access control management has been deeply studied for decades. Different models have been proposed to declare and administer access control policies, like DAC, MAC, RBAC, TMAC, and OrBAC. While access control management is well established, new models are being defined to cope with privacy requirements. Privacy management distinguishes itself from traditional access control is the sense that the data to be protected is personal. Hence, the user's consent must be reflected in the access control policies, as well as the usage of the data, its collection rules and its retention period, which are principles safeguarded by law and must be controlled carefully.

The research community working on privacy models is broad, and involves many teams worldwide including in France ENST-B, LIRIS, Inria LICIT, and LRI, and at the international level IBM Almaden, Purdue Univ., Politecnico di Milano and Univ. of Milano, George Mason Univ., Univ. of Massachusetts, Univ. of Texas and Colorado State Univ. to cite a few. Pioneer attempts towards privacy wary systems include the P3P Platform for Privacy Preservation [34] and Hippocratic databases [24]. In the last years, many other policy languages have been proposed for different application scenarios, including EPAL [38], XACML [36] and WSPL [29]. Hippocratic databases are inspired by the axiom that databases should be responsible for the privacy preservation of the data they manage. The architecture of a Hippocratic database is based on ten guiding principles derived from privacy laws.

The trend worldwide has been to propose enhanced access control policies to capture finer behavior and bridge the gap with privacy policies. To cite a few, Ardagna *et al.* (Univ. Milano) enables actions to be performed after data collection (like notification or removal), purpose binding features have been studied by Lefevre *et al.* (IBM Almaden), and Ni *et al.* (Purdue Univ.) have proposed obligations and have extended the widely used RBAC model to support privacy policies.

The positioning of the SMIS team within this broad area is rather (1) to focus on intuitive or automatic tools helping the individual to control some facets of her privacy (e.g., data retention, minimal collection) instead of increasing the expressiveness but also the complexity of privacy models and (2) to push concrete models enriched by real-case (e.g., medical) scenarios and by a joint work with researchers in Law.

3.3. Tamper-resistant Data Management

Tamper-resistance refers to the capacity of a system to defeat confidentiality and integrity attacks. This problem is complementary to access control management while being (mostly) orthogonal to the way access control policies are defined. Security surveys regularly point out the vulnerability of database servers against external (i.e., by intruders) and internal (i.e., by employees) attacks. Several attempts have been made in commercial DBMSs to strengthen server-based security, e.g., by separating the duty between DBA and DSA (Data Security Administrator), by encrypting the database footprint and by securing the cryptographic material using Hardware Security Modules (HSM) [31]. To face internal attacks, client-based security approaches have been investigated where the data is stored encrypted on the server and is decrypted only on the client side. Several contributions have been made in this direction, notably by U. of California Irvine (S. Mehrotra, Database Service Provider model), IBM Almaden (R. Agrawal, computation on encrypted data), U. of Milano (E. Damiani, encryption schemes), Purdue U. (E. Bertino, XML secure publication), U. of Washington (D. Suciu, provisional access) to cite a few seminal works. An alternative, recently promoted by Stony Brook Univ. (R. Sion), is to augment the security of the server by associating it with a tamper-resistant hardware module in charge of the security aspects. Contrary to traditional HSM, this module takes part in the query computation and performs all data decryption operations. SMIS investigates another direction based on the use of a tamper-resistant hardware module on the client side. Most of our contributions in this area are based on exploiting the tamper-resistance of secure tokens to build new data protection schemes.

While our work on Privacy-Preserving data Publishing (PPDP) is still related to tamper-resistance, a complementary positioning is required for this specific topic. The primary goal of PPDP is to anonymize/sanitize microdata sets before publishing them to serve statistical analysis purposes. PPDP (and privacy in databases in general) is a hot topic since 2000, when it was introduced by IBM Research (IBM Almaden: R. Agrawal, IBM Watson: C.C. Aggarwal), and many teams, mostly north American universities or research centres, study this topic (e.g., PORTIA DB-Privacy project regrouping universities such as Stanford with H. Garcia-Molina). Much effort has been devoted by the scientific community to the definition of privacy models exhibiting better privacy guarantees or better utility or a balance of both (such as differential privacy studied by C. Dwork: Microsoft Research or D. Kifer: Penn-State Univ and J. Gehrke: Cornell Univ) and thorough surveys exist that provide a large overview of existing PPDP models and mechanisms [35]. These works are however orthogonal to our approach in that they make the hypothesis of a trustworthy central server that can execute the anonymization process. In our work, this is not the case. We consider an architecture composed of a large population of tamper-resistant devices weakly connected to an untrusted infrastructure and study how to compute PPDP problems in this context. Hence, our work has some connections with the works done on Privacy Preserving Data Collection (Stevens Institute of Tech. / Rutgers Univ,NJ: R.N.Wright, Univ Austin Texas: V. Shmatikov), on Secure Multi-party Computing for Privacy Preserving Data Mining (Rutgers Univ: J. Vaidya, Purdue Univ: C. Clifton) and on distributed PPDP algorithms (Univ Wisconsin: D. DeWitt, Univ Michigan: K. Lefevre, Rutgers Univ: J. Vaidya, Purdue Univ: C. Clifton) while none of them share the same architectural hypothesis as us.

4. Application Domains

4.1. Application Domains

Our work addresses varied application domains. Typically, data management techniques on chip are required each time data-driven applications have to be embedded in ultra-light computing devices. This situation occurs for example in healthcare applications where medical folders are embedded into smart tokens (e.g., smart cards, secured USB keys), in telephony applications where personal data (address book, agenda, etc.) is embedded into cellular phones, in sensor networks where sensors log row measurements and perform local computation on them, in smart-home applications where a collection of smart appliances gather information about the occupants to provide them a personalized service, and more generally in most applications related to ambient intelligence.

Safeguarding data confidentiality has become a primary concern for citizens, administrations and companies, broadening the application domains of our work on access control policies definition and enforcement. The threat on data confidentiality is manifold: external and internal attacks on the data at rest, on the data on transit, on the data hosted in untrusted environments (e.g., Database Service Providers, Web-hosting companies) and subject to illegal usage, insidious gathering of personal data in an ambient intelligence surrounding. Hence, new access control models and security mechanisms are required to accurately declare and safely control who is granted access to which data and for which purpose.

While the application domain mentioned above is rather large, two applications are today more specifically targeted by the SMIS team. The first one deals with privacy preservation in EHR (Electronic Health Record) systems and PCEHR (Personally Controlled EHR). We are developing technologies tackling this issue and experiment them in the field. The second application area deals with privacy preservation in the context of personal Cloud, that is personal data hosted in dedicated servers staying under the holder's control (e.g., in a personal internet box or in a home automation box).

5. New Software and Platforms

5.1. Introduction

In our research domain, developing software prototypes is mandatory to validate research solutions and is an important vector for publications, demonstrations at conferences as well as for cooperations with industry. Our software strategy is also driven by our ambition to see our research results produce a real societal impact. To reach this goal, we integrate our prototypes in experiments in the field - notably in the healthcare domain and with scientists of other disciplines - and we recently set up educational platforms to raise students awareness of privacy protection problems and embedded programming.

This prototyping task is however difficult because it requires specialized hardware platforms, themselves sometimes at an early stage of development. For a decade, we have developed successive prototypes relying on different hardware platforms provided by Schlumberger then Gemalto, e.g., PicoDBMS a full-fledged DBMS embedded in a smart card [37] [26], Chip-Secured Data Access (C-SDA) a tamper-resistant mediator between a client and an untrusted server hosting encrypted data [32], Chip-Secured XML Access (C-SXA) an XML-based access rights controller, recipient of the e-gate open 2004 Silver Award and SIMagine 2005 Gold award [33]. Today, most of our software development efforts are organized around a unified platform named PlugDB and we are designing our own hardware platforms, that are produced by electronic SMEs. This opens up new research and experiment opportunities and we are engaged in an open-source/open hardware initiative to disseminate our results at a larger scale, both for scientific, educational and business purposes.

The next subsections detail the two prototypes we are focusing on today.

5.2. PlugDB engine

Participants: Nicolas Anciaux [correspondent], Luc Bouganim, Aydogan Ersoz, Quentin Lefebvre, Philippe Pucheral.

More than a stand-alone prototype, PlugDB is part of a complete architecture dedicated to a secure and ubiquitous management of personal data. PlugDB aims at providing an alternative to a systematic centralization of personal data. To meet this objective, the PlugDB architecture lies on a new kind of hardware device called Secure Portable Token (SPT). Roughly speaking, a SPT combines a smart-card and a micro-controller with a large external Flash memory (Gigabyte sized). The SPT can host data on Flash (e.g., a personal folder) and safely run code embedded in the micro-controller. PlugDB engine is the cornerstone of this embedded code. PlugDB engine manages the database on Flash (tackling the peculiarities of NAND Flash storage), enforces the access control policy defined on this database, protects the data at rest against piracy and tampering, executes queries (tackling low RAM constraint) and ensures transaction atomicity. Part of the on-board data can be replicated on a server (then synchronized) and shared among a restricted circle of trusted parties through crypto-protected interactions. PlugDB engine has been registered at APP (Agence de Protection des Programmes) in 2009 [27] and a new version is registered each year. PlugDB has been experimented in the field in the Yvelines District to implement a secure and portable medical-social folder helping the coordination of medical care and social services provided at home to dependent people. This field experiment is being audited by ARS-Ile de France (the Regional Healthcare Agency) and CG78 (General Council of Yvelines District), in order to envision the opportunity of a larger deployment. In parallel, we are improving the PlugDB prototype to overcome the limitations identified during the experiment. Notably, we have integrated a Bluetooth module to communicate in wireless with the token, a fingerprint module to authenticate users and a microphone to record voice messages. These are key elements in the perspective of a generalization. Link: https://project.inria.fr/plugdb/.

5.3. Eagle Tree

Participants: Matias Bjørling, Philippe Bonnet, Luc Bouganim, Niv Dayan [correspondent].

Solid State Drives (SSDs) are a moving target for system designers: they are black boxes, their internals are undocumented, and their performance characteristics vary across models. There is no appropriate analytical model and experimenting with commercial SSDs is cumbersome, as it requires a careful experimental methodology to ensure repeatability. Worse, performance results obtained on a given SSD cannot be generalized. Overall, it is impossible to explore how a given algorithm, say a hash join or LSM-tree insertions, leverages the intrinsic parallelism of a modern SSD, or how a slight change in the internals of an SSD would impact its overall performance. In this paper, we propose a new SSD simulation framework, named EagleTree, which addresses these problems, and enables a principled study of SSD-Based algorithms. EagleTree is an extensible, customizable SSD simulator designed to enable deep analyses of the interplay between the FTL, block management scheme, IO scheduling policy and application workload. It is able to generate visual illustrations of a host of performance metrics. EagleTree is available for Linux, and is licensed under GPL. EagleTree's git repository is : https://github.com/nivdayan/EagleTree.

6. New Results

6.1. Flash-Based Data Management

Participants: Nicolas Anciaux, Matias Bjørling, Philippe Bonnet, Luc Bouganim [correspondent], Niv Dayan, Saliha Lallali, Philippe Pucheral, Iulian Sandu Popa.

There is a long tradition of work around the understanding and optimization of NAND Flash memory in the team (e.g., [7], [9]). Current work in this area covers the optimization of SSD use in DBMS engines and the design of Flash-based indexing techniques for textual and spatio-temporal data. These works on Flash-Based indexing complete the work initiated in the last years on the storage and indexing engine of PlugDB (not repeated in this report but the interested reader is referred to a DAPD'14 journal publication detailing these techniques [14]).

Flash storage optimization. Solid State Drives (SSDs), based on flash chips, are now the secondary storage of choice for data intensive applications. Database systems can now rely on high performance SSDs to store log, indexes and data either on servers or in the cloud. While SSDs provide increasingly high performance out of the box, maintaining high throughput and low latency as the utilization of SSDs increases and despite abrupt changes in the workload remains a challenge. This question is central for database designers and administrators, cloud service providers, and SSD constructors. The answer depends on write-amplification, i.e., garbage collection overhead. More specifically, the answer depends on how write-amplification evolves in time. We derived a mathematical expression that relates over provisioning to write-amplification. We introduced a new block manager, called Wolf, or WOrkload Leveler for Flash. Wolf is able to detect and quickly adapt to changes in workload by pro-actively reallocating over-provisioned space among the groups based on their changing needs. It adapts better to stable workloads by measuring the update frequencies of groups instead of making assumptions about them. It uses a novel near-optimal closed-form expression to allocate over-provisioned space to groups.

Flash-based keyword indexing. As smart objects gain the capacity to acquire, store and process large volumes of data, new services emerge. However, the smart objects have to be endowed with typical data management capabilities to enable all these services. In this work, we revisit the traditional problem of information retrieval queries over large collections of files in an embedded context. A file can be any form of document, picture or data stream associated with a set of terms. A query can be any form of keyword search using a ranking function (e.g., TF-IDF) identifying the top-k most relevant files. The proposed search engine can be used in sensors to search for relevant objects in their surroundings, in cameras to search pictures by using tags, in personal smart dongles to secure the querying of documents and files hosted in an untrusted Cloud or in smart meters to perform analytic tasks (i.e., top-k queries) over sets of events (i.e., terms) captured during time windows (i.e., files) [21]. Designing such embedded search engine is however challenging due to a combination of severe and conflicting hardware constraints (e.g., a tiny RAM combined with a NAND Flash persistent storage badly adapted to random fine-grain updates). To tackle this challenge, we introduce three design principles, namely Write-Once Partitioning, Linear Pipelining and Background Linear Merging, and show how they can be combined to produce an embedded search engine reconciling high insert/delete/update rate and query scalability. We have implemented our search engine on a development board having a hardware configuration representative for smart objects. The experimental results demonstrate the scalability of the approach and its superiority compared to state of the art methods [28]. This work is part of Saliha Lallali's Ph.D. thesis.

Flash-based spatio-temporal indexing. The convergence of mobile computing, wireless communications and sensors has raised the development of many applications exploiting a massive flow of spatio-temporal data such as location-based services, participatory sensing, or traffic management [15]. Among the most active research topics in this area is the spatio-temporal data indexing. Nevertheless, since a few years a new fundamental parameter has made its entry on the database scene: the NAND flash storage. However, the peculiar characteristics of flash memory require redesigning the existing data storage and indexing techniques that were devised for magnetic hard-disks. In this study we propose TRIFL, an efficient and generic TRajectory

Index for FLash. TRIFL is designed around the key requirements of trajectory indexing and flash storage. TRIFL is generic in the sense that it is efficient for both simple flash storage devices such as the SD cards and more powerful devices such as the solid state drives. In addition, TRIFL is supplied with an online self-tuning algorithm that allows adapting the index structure to the workload and the technical specifications of the flash storage device to maximize the index performance. Moreover, TRIFL achieves good performance with relatively low memory requirements, which makes the index appropriate for many application scenarios. The experimental evaluation shows that TRIFL outperforms the representative indexing methods on magnetic disks and flash disks. This work is part of Dai-Hai Ton That Ph.D. thesis, co-supervised by Iulian Sandu Popa.

6.2. Secure Global Computing on Asymmetric Architecture

Participants: Benjamin Nguyen [correspondent], Philippe Pucheral, Quoc-Cuong To.

Current applications, from complex sensor systems (e.g. quantified self) to online e-markets acquire vast quantities of personal information which usually ends-up on central servers. Decentralized architectures, devised to help individuals keep full control of their data, hinder global treatments and queries, impeding the development of services of great interest. In this study, we promote the idea of pushing the security to the edges of applications, through the use of secure hardware devices controlling the data at the place of their acquisition. To solve this problem, we propose secure distributed querying protocols based on the use of a tangible physical element of trust, reestablishing the capacity to perform global computations without revealing any sensitive information to central servers. This leads to execute global treatments on an asymmetric architecture, composed of a powerful, available and untrusted computing infrastructure (server or cloud), and a large set of low powered, highly disconnected trusted devices. Given our large scale data centric applications (e.g. nationwide surveys), we discard solutions based on secure multi-party computation, which do not scale. We have studied two different computing paradigms on this architecture: our first contribution was to study the execution of Privacy Preserving Data Publishing (PPDP) algorithms on such an architecture, and provided generic protocols to deal with all kinds of PPDP algorithms, which are robust against honest-but-curious and malicious adversaries [12], including vulgarization aspects [25]. Our second contribution was to study general SQL queries in this same execution context. For now, we have concentrated on the subset of SQL queries without joins, but including Group By and aggregates, and show how to secure their execution in the presence of honest-but-curious attackers [19]. Cost models and experiments demonstrate that this approach can scale to nationwide infrastructures [20][16]. This work is part of Quoc-Cuong To's Ph.D. thesis started in sept. 2012, and should be extended in particular to cover joins. We also plan to extend this general framework through a collaboration with INSA Centre Val de Loire, LIFO Lab and University of Paris Nord, LIPN lab, to study the secure execution of Map/Reduce on the Asymmetric Architecture.

6.3. Personal Cloud

Participants: Nicolas Anciaux [correspondent], Luc Bouganim, Athanasia Katsouraki, Benjamin Nguyen, Philippe Pucheral, Iulian Sandu Popa, Paul Tran Van.

We are witnessing an exponential increase in the acquisition of personal data about the individuals or produced by them. Today, this information is managed using Web applications, centralizing this data in cloud data servers, under the control of few Web majors [5]. However, it has now become clear that (1) centralizing millions of personal records exposes the data to very sophisticated attacks, linked to a very high potential benefit in case of success (millions of records being revealed), and (2) delegating the management of personal records without any tangible guarantee for the individuals leads to privacy violations, the data being potentially made accessible to other organizations (e.g., governments, commercial partners) and being subject to lucrative secondary usages (not advertised to the individuals). To face this situation, many recent initiatives push towards the emergence of the Personal Cloud paradigm. A personal cloud can be viewed as a personal server, owned by a given individual, which gives to its owner the ability to store her complete digital environment, synchronize it among various devices and share it with other individuals and applications under control. Many projects and startups currently investigate this solution, like OpenPDS, CozyCloud, OwnCloud, etc. In the SMIS team, we claim the need of a Secure Personal Cloud, and promote the introduction of a secure (tamper resistant) data engine in the architecture [11]. On this basis, we investigate new data sharing and dissemination models, where usage and access control rules endorsed by the individuals could be enforced. In 2014, we have presented this vision at EDBT'14 [18]. Several underlying research problems and perspectives have been presented in [11]. We have started a cooperation with the startup CozyCloud at the end of 2014. A contract was signed at the end of 2014 to integrate PlugDB in a CozyCloud instance and the PhD of Paul Tran Van (CIFRE SMIS-CozyCloud) has just started to explore new data sharing techniques which could be enforced in the secure personal cloud model. Athanasia Katsouraki is working on privacy issues and on adoption of the secure data engine in cooperation with the economists (CERDI) in the context of the Digital Society Institute (DSI).

6.4. Folk-IS

Participants: Nicolas Anciaux [correspondent], Luc Bouganim, Philippe Pucheral.

According to many studies, IT should become a key facilitator in establishing primary education, reducing mortality or supporting commercial initiatives in Least Developed Countries. The main barrier to the development of IT services in these regions is not only the lack of communication facilities, but also the lack of consistent information systems, security procedures, economic and legal support, as well as political commitment. In [5], we proposed the vision of trusted cells, a data platform for personal data services where the shared infrastructure (typically the cloud) is untrusted, while personal devices (such as smart phones, tablets or set-top box) are trusted execution environments. We revisited this vision to the context of LDCs. We proposed a new paradigm, that we call Folk-enabled Information System (Folk-IS), based on a fully decentralized and participatory approach, where each individual implements a small subset of a complete information system without the need for a shared networked infrastructure. As trusted cells, Folk-IS builds upon the emergence of highly secure, portable and low-cost storage and computing devices, called hereafter Smart Tokens. Here however, the focus is on low-cost of ownership, deployment and maintenance, and on the absence of a networked infrastructure. With Folk-IS and thanks to their smart tokens, people will transparently and opportunistically perform data management and networking tasks as they physically move, so that IT services are truly delivered by the crowd. We have published in [17] the Folk-IS vision and main principles, and in [13] a more detailed paper including technical challenges, specific to that approach and an exploitation and feasibility analysis of the Folk-IS vision.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

The SMIS project has a long lasting cooperation with Gemalto, the world's leading providers of microprocessor cards. Gemalto provides SMIS with advanced hardware and software smart card platforms which are used to validate numbers of our research results. In return, SMIS provides Gemalto with requirements and technical feedbacks that help them adapting their future platforms towards data intensive applications. While no bilateral contract exists between Gemalto and SMIS, we are partners in several projects. Meanwhile, we are developing partnerships with SMEs capable of building ad-hoc hardware prototypes conforming to our own design.

7.2. DMSP3 Yvelines District Grant (Nov 2013 - Nov. 2014)

Partners: Inria-SMIS (coordinator), Gemalto, UVSQ, Santeos.
SMIS funding: 75k€.

Electronic Health Record (EHR) projects have been launched in most developed countries to increase the quality of care while decreasing its cost. Despite their unquestionable benefits, patients are reluctant to abandon their control of highly sensitive data to a distant server. The objective of the DMSP project is to complement a traditional EHR server with a secure and mobile personal medical folder (1) to protect and share highly sensitive data among trusted parties and (2) to provide a seamless access to the data even in disconnected mode. The DMSP architecture builds upon the technology designed in the PlugDB project. This architecture has been designed and developed under grant DMSP1 ended in 2010. It has been experimented in the context of a medical-social network providing care and services at home for elderly people. The experiment in the field, founded by grant DMSP2, lasted from September 2011 to December 2012 with volunteer patients and practitioners in the Yvelines district. The goal of grant DMSP3 (Nov 2013 - Nov 2014) is to correct the imperfections observed during DMSP2 and port our prototype in an open hardware platform with the final objective to set up a technology transfer. This project is being audited by ARS-Ile de France (the Regional Healthcare Agency) and CG78 (General Council of Yvelines District), in order to envision the opportunity of a larger deployment.

7.3. Cozy Cloud bilateral contract (Dec 2014 - Nov. 2015)

Partners: Cozy Cloud, Inria-SMIS
SMIS funding: 50k€.
Many personal data end up today on servers where they can be scrutinized by companies and governmental agencies. To face this situation, the most emblematic initiative is the Personal Cloud paradigm. Roughly speaking, the Personal Cloud is an architecture which gives users the ability to store their complete digital environment, synchronize it among various devices and share it with other users and applications under their control. It reflects the expectation of the individuals for the emergence of privacy-by-design next-generation storage and computing services. Cozy Cloud is a French startup providing such 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. Cozy defines itself as the “Android of personal servers”. While centralizing all personal data in the holder’s hand is a natural way to reestablish his control on his privacy, this represents an unprecedented threat in case of attacks by an intruder, especially for individuals who are not security experts. The objective of this bilateral contract is typically to address this issue by integrating the PlugDB solution into the Cozy stack. Roughly speaking, the Cozy data system will be modified in such a way to store only encrypted files and each file access will be intercepted and routed to PlugDB. PlugDB will act as a doorkeeper for the whole individual dataspace by managing the files’ metadata, the access control rules defined on these metadata, the decryption keys and the user/application authentication.

7.4. Cozy Cloud CIFRE contract (Oct 2014 - Sept 2017)

Partners: Cozy Cloud, Inria-SMIS
SMIS funding: 30k€.
In relation with the bilateral contract mentioned above, a CIFRE PhD thesis has been started by Paul Tran Van. The objective is to capitalize on the Cozy-PlugDB platform to devise new access and usage control models to exchange data among devices of the same user (devices may have different levels of trustworthiness) and among different users. A particular focus will be put on the enforcement of the access and usage control rules in this thesis.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR KISS (Dec. 2011 - Dec. 2015)

Partners: Inria-SMIS (coordinator), Inria-SECRET, LIRIS, Univ. of Versailles, CryptoExperts, Gemalto, Yvelines district.

SMIS funding: 230k€.

The idea promoted in KISS is to embed, in trusted devices, software components capable of acquiring, storing and managing securely various forms of personal data (e.g., salary forms, invoices, banking statements, geolocation data, depending on the applications). These software components form a Personal Data Server which can remain under the holder's control. The scientific challenges include: embedded data management issues tackling regular, streaming and spatio-temporal data (e.g., geolocation data), data provenance-based privacy models, crypto-protected distributed protocols to implement private communications and secure global computations.

8.1.2. CAPPRIS Project-Lab (Dec. 2011 - Dec. 2015)

Inria Partners: PRIVATICS (coordinator), SMIS, PLANETE, CIDRE, COMETE.
External partners: Univ. of Namur, Eurecom, LAAS.
Funding: not associated to individual project-teams.

An Inria Project Lab (IPL) is a long-term multi-disciplinary project launched by Inria to sustain large scale risky research actions in line with its own strategic plan. CAPPRIS stands for "Collaborative Action on the Protection of Privacy Rights in the Information Society". The key issues that are addressed are: (1) the identification of existing and future threats to privacy, (2) the definition of formally grounded measures to assess and quantify privacy, (3) the definition of the fundamental principles underlying privacy by design and methods to apply them in concrete situations and (4) The integration of the social and legal dimensions. To assess the relevance and significance of the research results, they are confronted to three classes of case studies CAPPRIS partners are involved in: namely Online Social Networks, Location Based Services and Electronic Health Record Systems.

8.1.3. PEPS PAIP (Pour une Approche Interdisciplinaire de la Privacy) (Sept. 2013 - Sept. 2014)

Partners: ADIS and SMIS (co-organizers), CERDI, DANTE, COMETE, GRACE, TPT, LIMSI.
Funding: 30K€ from CNRS, not associated to individual project-teams.

The Digital Society Institute (DSI) is the UPSa IDEX catalyst for multidisciplinary research on societal challenges inherent to eLife/life digitization. DSI plans to be one of the European leading institutes fostering multidisciplinary research across ICTS and SHES. In 2013 DSI already hosts two kick-off major research projects : (1) Human and Machine Coevolution and (2) Privacy/digital identities. ADIS and SMIS are co-organizing project (2) on data privacy. The PEPS PAIP is part of project (2) and aims at fostering the cooperation between lawyers, economists and computer scientists on privacy issues, through the organization of brainstorming days and workshops and a study of possible joint experiments of privacy preserving applications.

8.2. European Initiatives

8.2.1. FP7 & H2020 Projects

8.2.1.1. PDS4NRJ (Aug. 2013 - Aug. 2014)

Instrument: Marie Curie Intra-European Fellowships for Career Development

Duration: 2013 Aug. - 2014 Aug.

Inria contact: Philippe Bonnet

This project, called PDS4NRJ, is based on the insights that (a) secure personal data management can be radically improved with the advent of secure hardware embedded on personal devices at the edges of the Internet, and (b) that a secure personal data management infrastructure should be applied in the context of smart buildings. Our overall objective is to define a new form of decentralized infrastructure for sharing smart meter data with access and usage control guarantees. The PDS4NRJ project is a unique opportunity for Philippe Bonnet, currently associate professor at ITU (Denmark), to become a leading expert in the field of secure personal data management thanks to a tight cooperation with SMIS members.

8.2.2. Collaborations in European Programs, except FP7

Program: Danish Council for Independent Research (FTP call)

Project acronym: CLyDE

Project title: Cross-LaYer optimized Database Engine

Duration: 10/2011 - 10/2014

Partners: IT University of Copenhagen (Denmark), SMIS

Abstract: The goal is to explore how flash devices, operating system and database system can be designed together to improve overall performance. Such a co-design is particularly important for the next generation database appliances, or cloud-based relational database systems for which well suited flash components must be specified. More generally, our goal is to influence the evolution of flash devices and commodity database systems for the benefit of data intensive applications. The project should result in two complementary open-source software systems: (i) a bimodal flash device software component based on the idea from [30], and (ii) a database system optimized for bimodal flash devices. The project funding is managed by the IT University of Copenhagen and covers the expenses for two co-supervised PhD students (including regular visits to and from Denmark).

8.2.3. Collaborations with Major European Organizations

The SMIS members have developed tight European cooperations with the following persons/teams:

Philippe Bonnet (Associate Professor at the University of Copenhagen, Denmark)

Collaboration on Flash-based data management for high-end servers with Philippe Bonnet from IT University of Copenhagen and Björn Þór Jónsson from Reykjavík University (see Section 8.2.2). The study of flash devices started during a short sabbatical of Luc Bouganim (from April to August 2008) in Copenhagen.

Michalis Vazirgiannis (Athens University of Economics and Business)

Collaboration on Minimal Exposure in the context of Michalis' Digiteo Chair at LIX (Ecole Polytechnique).

8.3. International Research Visitors

Philippe Bonnet, associate professor at the IT University of Copenhagen, visited SMIS in the context of a Marie Curie grant from August 2013 until July 2014 (see Section 8.2.1.1).

9. Dissemination

9.1. Scientific Animation

Philippe Pucheral:

- PC member of EDBT'14, MOBIWIS'14, EDBT'15, MOBIWIS'15, DATA'15
- Member of the direction committee of the PRISM Lab.
- Member of the HDR committee of the STV doctoral school (UVSQ)
- Elected member of the ED STIC doctoral school of University Paris-Saclay, 'Data, Knowledge and Interactions' committee (about 250 PhD students)
- Co-founder of the "Masses de Données Distribuées" French summer school and co-organizer of its last three editions (2010, 2012, 2014)

Luc Bouganim:

- PC member of EDBT'15, PDA@IOT'14, HardBD'14, EUC'14, BDA'14
- President of the Inria Post-Doc and Delegation Commission
- Member of the Inria "Bureau du Comité des Projets" (BCP)
- Member of the Inria "Cordi-S" (Inria PhD grant) commission
- Member of the Inria "PES" commission
- Member of the UVSQ-PRiSM "Conseil de Laboratoire"

Nicolas Anciaux:

- PC member of BDA'14 Demo. Track
- Member of the Inria " Commission des Développements Technologiques " (CDT)

Benjamin Nguyen:

- PC member of ECML-PKDD'14, ACOMP'14, EDA'14
- Member of APVP Steering Committee
- Member of recruiting Committee of CNAM'14, Paris-Sud'14
- Member of PRiSM Steering Committee until 08/14
- Co-director (with Fabrice Le Guel) of the Institut de la Société du Numérique, Privacy WG.

Quoc-Cuong To:

- PC member of FDSE'2014
- External reviewer of EDBT'2014

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

SMIS is a joint project-team with University of Versailles St-Quentin (UVSQ) and CNRS. Hence SMIS members are naturally deeply involved in teaching.

Besides academic teaching, SMIS has set up a software and hardware platform named 'PDIS- Privacy by Design Information System' which is used to conduct programming activities with Master and Engineer students. This platform capitalizes on the PlugDB engine. It is now available at ENSIIE and INSA Centre Val de Loire engineer schools and at UVSQ. It is also being integrated in the UVSQ FabLab (VSL - Versailles Sciences Lab).

P. Pucheral:

- Full professor at UVSQ : courses on databases, DBMS architecture and security in Master1, Master2 and engineer school ISTY
- Director of the research Master COSY (UVSQ)
- Co-Director of the future Master DataScale to be launched at Univ. Paris-Saclay (UPSay) in 2015 with UVSQ, ENSIIE, Télécom SudParis and Télécom ParisTech

L. Bouganim: (90h/y)

- Courses on DBMS architecture, data security, database technology in Master1 and Master2 (AFTI, Orsay) and in engineering school (ENST Paris)

B. Nguyen:

- Courses on Databases, programming in Masters 2 and Licence
- MOOC on Introduction to relational databases (co authors Serge Abiteboul, Yannick Le Bras)
- Director of the UVSQ Computer Science Masters until august 2014
- Co-Director (with Franck Quessette, UVSQ) of the in-house training course for Computer Science high-school teachers at UVSQ for Academie de Versailles
- Member of Labex Digicosme Master's Committee until august 2014

N. Anciaux: (90h/y)

- Courses on database internal mechanisms and database security, in Master1 and Master2 (AFTI, Orsay) and in engineering school (ENSTA ParisTech, Telecom Paristech)

I. Sandu Popa:

- Associate professor at UVSQ. Courses on databases, DBMS architecture and security in Licence, Master1, Master2 and engineer school ISTY

P. Tran-Van:

- Courses on distributed and secure databases in engineering school ENSIIE.

9.2.2. Supervision

PhD in progress : Quoc-Cuong To, Secure Global Computations on Personal Data Servers, since October 2012, co-supervised by Benjamin Nguyen and Philippe Pucheral

PhD in progress : Saliha Lallali, Document Indexing for Embedded Personal Databases, since November 2012, co-supervised by Nicolas Anciaux, Iulian Sandu Popa and Philippe Pucheral

PhD in progress : Athanasia Katsouraki, Access and Usage Control for Personal Data in Trusted Cells, since October 2013, co-supervised by Luc Bouganim and Benjamin Nguyen

PhD in progress : Niv Dayan, Database Algorithms and Flash Internals, since December 2011, co-supervised by Philippe Bonnet and Luc Bouganim

PhD in progress : Dai-Hai Ton-That, Secure Management and Sharing of Private Personal Traces, since November 2012, co-supervised by Iulian Sandu Popa and Karine Zeitouni (Univ. of Versailles)

PhD in progress : Paul Tran Van, Secure Data Sharing in Personal Cloud, since October 2014, Co-supervised by Benjamin André (Cozy Cloud), Philippe Pucheral and Nicolas Anciaux.

9.2.3. Juries

P. Pucheral:

- PhD committee member (president) of Sara Hachem (UVSQ, February 2014)
- HDR reviewer of Sébastien Gambs (U. Rennes 1, June 2014)
- HDR committee member (president) of Sophie Chabridon (Télécom SudParis, November 2014)
- HDR committee member of Nicolas Anciaux (UVSQ, December 2014)
- HDR committee member of Sofian Maabout (U. Bordeaux, December 14).

L. Bouganim:

- PhD committee member of Matteo Casalino (Univ. Claude Bernard - Lyon 1, July 2014)
- PhD committee member of Jean-Pierre Lozi (UPMC, Paris 6, July 2014).
- PhD reviewer of Pierre Olivier (Université de Bretagne Sud, December 2014).
- HDR tutor of Nicolas Anciaux (UVSQ, December 2014).

9.3. Popularization

- "Les enjeux du Big Data", TV interview of P. Pucheral, Future-Mag, Arte, September 20th, 2014.
- Tutorial on experimental methodology at the BDA summer school 2014 "Masses de données distribuées". L. Bouganim
- Tutorial on managing personal data with strong privacy guarantees at EDBT 2014 [18].
- Présentation invitée au Lycée Albert Einstein sur 'Minimum Exposure', Sainte-Geneviève-des-Bois, Apr. 2014, B. Nguyen.
- Séminaire en deux parties à Luminy, pour les enseignants de CPGE sur 'les bases de données en CPGE, programme officiel et aspects transactionnels', Marseille, 2014-04-22, B. Nguyen.

- Présentation invitée aux journées "La Science Informatique pour tous les lycéens", à Rocquencourt, Apr. 2014, B. Nguyen.
- Le droit à l'oubli numérique, table ronde, Colloque 'La protection des données personnelles : approche pluridisciplinaire', Paris, Dec. 2014, B. Nguyen.
- 'Représenter et interroger efficacement de grands volumes de données : une introduction aux bases de données relationnelles', olympiades de mathématiques, Paris, May 2014, B. Nguyen
- 'Comment garantir la confidentialité des données médicales ?', Envirorisk, Atelier risques numériques, Nov. 2014. Chairman: B. Nguyen (Inria). Speakers : N. Anciaux (Inria), C. F. Viala (YesProfile).
- 'Garantir la confidentialité des données personnelles', Futur en Seine, Rencontres Inria-Industries, June 2014, N. Anciaux.
- 'Vers un modèle de gestion des données respectueux de la vie privée : application à la collecte limitée d'informations personnelles', IREP "BIG DATA" Seminar, May 2014, N. Anciaux.
- 'Les patients dans l'écosystème de santé - Enjeux d'information et questions de communication', International Seminar organized by the "Institut des sciences de la communication CNRS" with the participation of the "Groupe de réflexion avec les associations de malades (GRAM Inserm)', Dec. 2014, N. Anciaux et P. Pucheral.
- 'La protection des données personnelles - approche pluridisciplinaire', Seminar organized by the Digital Society Institute (ISN), Dec. 2014, N Anciaux and P. Pucheral
- 'Démonstration du Dossier Médico-Social Portable et sécurisé', Rencontres Inria Industrie, Télécoms du futur, Nov. 2014, Q. Lefebvre and N. Anciaux.
- 'Table ronde : Sécurité des Réseaux et Contenus', Rencontres Inria Industrie, Télécoms du futur, Nov. 2014, L. Bouganim

10. Bibliography

Major publications by the team in recent years

[1] T. ALLARD, N. ANCIAUX, L. BOUGANIM, Y. GUO, L. LE FOLGOC, B. NGUYEN, P. PUCHERAL, I. RAY, I. RAY, S. YIN. *Secure Personal Data Servers: a Vision Paper*, in "Proc. of the 36th Int. Conf. on Very Large Databases (VLDB)", 2010

[2] T. ALLARD, B. NGUYEN, P. PUCHERAL. *Safe Realization of the Generalization Privacy Mechanism*, in "Privacy, Security and Trust", Montreal, Canada, 2011, pp. 1-8, Best Paper Award, http://hal.inria.fr/hal-00624043/en

[3] N. ANCIAUX, M. BENZINE, L. BOUGANIM, P. PUCHERAL, D. SHASHA. *GhostDB: querying visible and hidden data without leaks*, in "26th International Conference on Management of Data (SIGMOD)", June 2007

[4] N. ANCIAUX, M. BENZINE, L. BOUGANIM, P. PUCHERAL, D. SHASHA. *Revelation on Demand*, in "Distributed and Parallel Database Journal (DAPD)", April 2009, vol. 25, n^{o} 1-2

[5] N. ANCIAUX, P. BONNET, L. BOUGANIM, B. NGUYEN, P. PUCHERAL, I. SANDU POPA. *Trusted Cells : A Sea Change for Personnal Data Services*, in "CIDR 2013 - 6th Biennal Conference on Innovative Database Research", Asilomar, United States, 2013, 4 p. , http://hal.inria.fr/hal-00768379

[6] N. ANCIAUX, L. BOUGANIM, P. PUCHERAL, P. VALDURIEZ. *DiSC: Benchmarking Secure Chip DBMS*, in "IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE)", October 2008, vol. 20, n^{o} 10

[7] P. Bonnet, L. Bouganim, I. Koltsidas, S. Viglas. *System Co-Design and Data Management for Flash Devices*, in "Very Large Data Bases (Tutorial)", 2011

[8] L. Bouganim, F. Dang-Ngoc, P. Pucheral. *Dynamic Access-Control Policies on XML Encrypted Data*, in "ACM Transactions on Information and System Security (ACM TISSEC)", January 2008, vol. 10, n^{o} 4

[9] L. Bouganim, B. Jónsson, P. Bonnet. *uFLIP: Understanding Flash IO Patterns*, in "4th Biennial Conference on Innovative Data Systems Research (CIDR)", Asilomar, California, USA, January 2009, best paper award

[10] S. Yin, P. Pucheral. *PBFilter: a Flash-Based Indexing Scheme for Embedded Systems*, in "Information Systems", 2012, vol. 37, n^{o} 7, pp. 634-653 [*DOI :* 10.1016/j.is.2012.02.002], http://hal.archives-ouvertes.fr/hal-00768380

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] N. Anciaux. *Gestion de données personnelles respectueuse de la vie privée*, Université de Versailles Saint-Quentin-en-Yvelines, December 2014, Habilitation à diriger des recherches, https://hal.archives-ouvertes.fr/tel-01104999

Articles in International Peer-Reviewed Journals

[12] T. Allard, B. Nguyen, P. Pucheral. *MetaP: Revisiting Privacy-Preserving Data Publishing using Secure Devices*, in "Distributed and Parallel Databases", June 2014, pp. 1-55 [*DOI :* 10.1007/s10619-013-7122-x], https://hal.archives-ouvertes.fr/hal-00934586

[13] N. Anciaux, L. Bouganim, T. Delot, S. Ilarri, L. Kloul, N. Mitton, P. Pucheral. *Opportunistic Data Services in Least Developed Countries: Benefits, Challenges and Feasibility Issues*, in "ACM SIGMOD Record", March 2014, https://hal.inria.fr/hal-00971805

[14] N. Anciaux, L. Bouganim, P. Pucheral, Y. Guo, L. Le Folgoc, S. Yin. *MILo-DB: a personal, secure and portable database machine*, in "Distributed and Parallel Databases", March 2014, vol. 32, n^{o} 1, pp. 1-27 [*DOI :* 10.1007/s10619-012-7119-x], https://hal.archives-ouvertes.fr/hal-00768355

[15] I. Sandu Popa, K. Zeitouni, V. Oria, A. Kharrat. *Spatio-temporal compression of trajectories in road networks*, in "GeoInformatica", 2014, 29 p. [*DOI :* 10.1007/s10707-014-0208-4], https://hal.inria.fr/hal-01096623

Articles in National Peer-Reviewed Journals

[16] C. Q. To, B. Nguyen, P. Pucheral. *Exécution sécurisée de requêtes avec agrégats sur des données distribuées*, in "Ingénierie des Systèmes d'Information", 2014, 26 p. [*DOI :* 10.3166/isi.19.4.118-143], https://hal.inria.fr/hal-01096637

International Conferences with Proceedings

[17] N. Anciaux, L. Bouganim, T. Delot, S. Ilarri, L. Kloul, N. Mitton, P. Pucheral. *Folk-IS: Opportunistic Data Services in Least Developed Countries*, in "40th International Conference on Very

Large Data Bases (VLDB)", Hangzhou, China, Zhejiang University, September 2014, https://hal.inria.fr/hal-00906204

[18] N. ANCIAUX, B. NGUYEN, I. SANDU POPA. *Managing Personal Data with Strong Privacy Guarantees*, in "17th International Conference on Extending Database Technology (EDBT) (Tutorial track)", Athens, Greece, March 2014 [*DOI :* 10.5441/002/EDBT.2014.71], https://hal.inria.fr/hal-01096633

[19] C. Q. TO, B. NGUYEN, P. PUCHERAL. *Privacy-Preserving Query Execution using a Decentralized Architecture and Tamper Resistant Hardware*, in "17th International Conference on Extending Database Technology (EDBT)", Athens, Greece, March 2014 [*DOI :* 10.5441/002/EDBT.2014.44], https://hal.inria.fr/hal-01096639

[20] C. Q. TO, B. NGUYEN, P. PUCHERAL. *SQL/AA: Executing SQL on an Asymmetric Architecture*, in "40th International Conference on Very Large Data Bases (VLDB) (Demonstration track)", Hangzhou, China, September 2014, https://hal.inria.fr/hal-01096642

Scientific Popularization

[21] N. ANCIAUX, P. BONNET, L. BOUGANIM, P. PUCHERAL. *Trusted Cells : Ensuring Privacy for the Citizens of Smart Cities*, in "ERCIM News", 2014, 2 p. , https://hal.inria.fr/hal-01096643

[22] N. ANCIAUX, B. NGUYEN. *Limiter la collecte des données personnelles : un problème juridique NP-difficile*, in "Tangente", 2014, vol. 52, 5 p. , https://hal.inria.fr/hal-01096645

[23] B. NGUYEN. *Techniques d'anonymisation*, in "Statistiques et Société", 2014, vol. 2, n^{O} 4, pp. 53-60, https://hal.archives-ouvertes.fr/hal-01113412

References in notes

[24] R. AGRAWAL, J. KIERNAN, R. SRIKANT, Y. XU. *Hippocratic Databases*, in "Proc. of the Int. Conf. on Very Large Data Bases (VLDB)", 2002

[25] T. ALLARD, B. NGUYEN, P. PUCHERAL. *L'anonymat, un bien fragile*, in "Le Monde - science&médecine - 08 avril 2014", 2014

[26] N. ANCIAUX, C. BOBINEAU, L. BOUGANIM, P. PUCHERAL, P. VALDURIEZ. *PicoDBMS : Validation and Experience*, in "Proc. of the Int. Conf. on Very Large Data Bases (VLDB)", 2001

[27] N. ANCIAUX, L. BOUGANIM, P. PUCHERAL, S. YIN, M. BENZINE, K. JACQUEMIN, D. SHASHA, C. SALPERWYCK, M. E. KHOLY. *Logiciel PlugDB-engine version 2, enregistré à l'Agence pour la Protection des Programmes (APP) sous le numéro IDDN.FR.001.280004.000.S.C.2008.0000.10000 en date du 27 avril 2009*, April 2009

[28] N. ANCIAUX, S. LALLALI, I. S. POPA, P. PUCHERAL. *A Scalable Search Engine for Mass Storage Smart Objects*, in "PRISM Technical Report", 2014, http://www.prism.uvsq.fr/~isap/files/RT.pdf

[29] A. ANDERSON. *An introduction to the web services policy language (WSPL)*, in "IEEE Computer Society", 2004

[30] P. BONNET, L. BOUGANIM. *Flash Device Support for Database Management*, in "5th Biennial Conference on Innovative Data Systems Research (CIDR)", Asilomar, California, USA, January 2011, pp. 1-8

[31] L. BOUGANIM, Y. GUO. *Database Encryption*, in "Encyclopedia of Cryptography and Security", S. JAJODIA, H. VAN TILBORG (editors), Springer, 2009, pp. 307-312

[32] L. BOUGANIM, P. PUCHERAL. *Chip-Secured Data Access : Confidential Data on Untrusted Servers*, in "Proc. of the 28th Int. Conf. on Very Large Data Bases (VLDB)", 2002

[33] L. BOUGANIM, F. DANG-NGOC, P. PUCHERAL. *Client-Based Access Control Management for XML Documents*, in "Proc. of the 30th Int. Conf. on Very Large Databases (VLDB)", 2004

[34] L. CRANOR. *Web Privacy with P3P*, O'Reilly Media, 2002

[35] B. FUNG, K. WANG, R. CHEN, P. YU. *Privacy-preserving data publishing: A survey of recent developments*, in "ACM Computing Surveys (CSUR)", 2010, vol. 42, n^o 4

[36] T. MOSES. *Extensible access control markup language (XACML) version 2.0*, in "Oasis Standard 200502", 2005

[37] P. PUCHERAL, L. BOUGANIM, P. VALDURIEZ, C. BOBINEAU. *PicoDBMS : Scaling down Database Techniques for the Smartcard*, in "Very Large Data Bases Journal (VLDBJ), Best Paper Award VLDB'2000", 2001, vol. 10, n^o 2-3

[38] M. SCHUNTER, C. POWERS. *Enterprise privacy authorization language (EPAL 1.1)*, in "IBM", 2003