IN PARTNERSHIP WITH:
CNRS

Université Versailles Saint-Quentin

Activity Report 2013

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]
Luc Bouganim [Inria, Senior Researcher, HdR]

Faculty Members

Philippe Pucheral [Team leader, Univ. Versailles, Professor, HdR]
Benjamin Nguyen [Univ. Versailles, Associate Professor, HdR]
Iulian Sandu Popa [Univ. Versailles, Associate Professor]

Engineers

Quentin Lefebvre [Inria, from Sep 2013]
Alexei Troussov [Inria, until Dec 2013]

PhD Students

Soumya Chatterjee [Inria, until Nov 2013]
Athanasia Katsouraki [Inria]
Saliha Lallali [Inria, granted by ANR KISS project]
Quoc-Cuong To [Inria]
Niv Dayan [Univ. of Copenhagen, co-supervision]
Matias Bjørling [Univ. of Copenhagen, co-supervision]

Visiting Scientists

Philippe Bonnet [Marie-Curie grant, Inria, from Aug 2013 until July 2014]
Javier Gonzalez [PhD, Univ. of Copenhagen, Inria, from Aug 2013 until Dec 2013]

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 [36] and Hippocratic databases [25]. In the last years, many other policy languages have been proposed for different application scenarios, including EPAL [41], XACML [38] and WSPL [31]. 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 behaviour 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) [33]. 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 [37]. 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 (Personnally 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. 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 and exhibitions as well as for cooperations with industry. This prototyping task is however difficult because it requires specialized hardware platforms (e.g., new generations of smart tokens), themselves sometimes at an early stage of development.

For a decade, we have developed successive prototypes addressing different application domains, introducing different technical challenges and relying on different hardware platforms. PicoDBMS was our first attempt to design a full-fledged DBMS embedded in a smart card [39] [27]. Chip-Secured Data Access (C-SDA) embedded a reduced SQL query engine and access right controller in a secure chip and acted as an incorruptible mediator between a client and an untrusted server hosting encrypted data [34]. Chip-Secured XML Access (C-SXA) was an XML-based access rights controller embedded in a smart card [35]. Prototypes of C-SXA have been the recipient of the e-gate open 2004 Silver Award and SIMagine 2005 Gold award, two renowned international software contests. The next subsections detail the two prototypes we are focusing on today.

5.2. PlugDB engine

Participants: Nicolas Anciaux [correspondent], Luc Bouganim, Philippe Pucheral, Shaoyi Yin, Yanli Guo, Lionel Le Folgoc, Alexei Troussov.

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 secure microcontroller (similar to a smart card chip) 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 secure microcontroller. 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 [28] and a new version is registered each year. The underlying Flash-based indexing system has also been patented by Inria and Gemalto [40]. It has been demonstrated in a dozen of national and international events including JavaOne and SIGMOD. It is being experimented in the field to implement a secure and portable medical-social folder helping the coordination of medical care and social services provided at home to dependent people. The next step in our agenda is to put this software in open-source so that students and communities of developers can complement it and develop innovative privacy-by-design applications. In 2012, we have ported PlugDB-engine on a new hardware platform to 1) become completely independent from Gemalto, 2) have a plug-and-play implementation on Android, 3) serve as a basement to port it on other custom hardware implementations. We have already discussed with hardware companies located in "Ile-de-France" to produce new hardware tokens to host future versions of PlugDB-engine. Link: http://www-smis.inria.fr/_DMSP/home.php.

5.3. uFLIP Benchmark

Participants: Luc Bouganim [correspondent], Philippe Bonnet, Bjorn Jónsson, Lionel Le Folgoc.

It is amazingly easy to produce meaningless results when measuring flash devices, partly because of the peculiarity of flash memory, but primarily because their behavior is determined by layers of complex, proprietary, and undocumented software and hardware. uFLIP is a component benchmark for measuring the response time distribution of flash IO patterns, defined as the distribution of IOs in space and time. uFLIP includes a benchmarking methodology which takes into account the particular characteristics of flash devices. The source code of uFLIP, available on the web (700 downloads, 4000 distinct visitors), was registered at APP in 2009 [32]. It has been demonstrated at SIGMOD.

Link: http://www.uflip.org.

6. New Results

6.1. Minimum Exposure

Participants: Nicolas Anciaux, Marouane Fazouane, Benjamin Nguyen [correspondent], Michalis Vazirgiannis.

When users request a service, the service provider usually asks for personal documents to tailor its service to the specific situation of the applicant. For example, the rate and duration of consumer's loans are usually adapted depending on the risk based on the income, assets or past lines of credits of the borrower. In practice, an excessive amount of personal data is collected and stored. Indeed, a paradox is at the root of this problem: service providers require users to expose data in order to determine whether that data is needed or not to achieve the purpose of the service. We explore a reverse approach, where service providers would publicly describe the data they require to complete their task, and where software (placed, depending on the context, on the client, on the server, or in a trusted hardware component) would use those descriptions to determine a minimum subset of information to expose.

Following our 2012 seminal works on the general Minimum Exposure framework, we have pursued its general study in 2013 [15], [29]. We have also developed a prototype system, using a low powered and highly secure smartcard [21], which is used to support hidden decision rules.

6.2. Flash-Based Data Management

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

Mass-storage secure portable tokens are emerging and provide a real breakthrough in the management of sensitive data. They can embed personal data and/or metadata referencing documents stored encrypted in the Cloud and can manage them under holder's control. Mass on-board storage requires efficient embedded database techniques. These techniques are however very challenging to design due to a combination of conflicting NAND Flash constraints and scarce RAM constraint, disqualifying known state of the art solutions. To tackle this challenge, we proposed a log-only based storage organization and an appropriate indexing scheme, which (1) produce only sequential writes compatible with the Flash constraints and (2) consume a tiny amount of RAM, independent of the database size [13].

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 2013, we worked on a new SSD simulation framework, named EagleTree, which addresses these problems, and enables a principled study of SSD-Based algorithms. We published a demonstration on EagleTree at VLDB'13 [20]. The demonstration scenario illustrates the design space for algorithms based on an SSD-based IO stack, and shows how researchers and practitioners can use EagleTree to perform tractable explorations of this complex design space.

6.3. Secure Global Computing on Asymmetric Architecture

Participants: Benjamin Nguyen [correspondent], Philippe Pucheral, Cuong Quoc 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 primarily studied 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]. This work is an extension of [26]. A vulgarization paper on the scientific and societal challenges related to PPDP techniques has been published in a newspaper [24]. We are now trying to support general SQL queries in this same execution context. We concentrate first 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. Cost models and experiments demonstrate that this approach can scale to nationwide infrastructures [23], [42]. This work is part of Cuong Quoc To's Ph.D. thesis started in sept. 2012.

6.4. Trusted Cells

Participants: Nicolas Anciaux, Philippe Bonnet, Luc Bouganim, Benjamin Nguyen, Philippe Pucheral [correspondent], Iulian Sandu Popa.

With the convergence of mobile communications, sensors and online social networks technologies, we are witnessing an exponential increase in the creation and consumption of personal data. Such data is volunteered by users, automatically captured by sensors or inferred from existing data. Today, there is a wide consensus that individuals should have increased control on how their personal data is collected, managed and shared. Yet there is no appropriate technical solution to implement such personal data services: centralized solutions sacrifice security for innovative applications, while decentralized solutions sacrifice innovative applications for security. In this work, we argue that the advent of secure hardware in all personal IT devices, at the edges of the Internet, could trigger a sea change. We propose the vision of trusted cells: personal data servers running on secure smart phones, set-top boxes, secure portable tokens or smart cards to form a global, decentralized data platform that provides security yet enables innovative applications. We motivate our approach, describe the trusted cells architecture and define a range of challenges for future research in a paper published at CIDR'13 (Int. Conf on Innovative Data Systems Research). This work was based on a thorough analysis of existing and potential threats on personal data, which led to a tutorial on data privacy [18], [30].

In parallel, we revisited the Trusted Cells vision to the context of Least Developed Countries (LDCs). 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. We propose, Folk-enabled Information System (Folk-IS), a new paradigm based on a fully decentralized and participatory approach, where each individual implements a small subset of a complete information system without the need for infrastructure. As trusted cells, Folk-IS builds upon the emergence of highly secure, portable, low-cost storage and computing devices, called hereafter Smart Tokens. Here, however, the focus is on the low cost of ownership, deployment and maintenance, and on the absence of a networked infrastructure. With Folk-IS, and thanks to 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 [17].

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€.
http://www-smis.inria.fr/_DMSP/accueil.php

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 (see above). 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.

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. ARC CAPPRIS (Dec. 2011 - Dec. 2015)

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

An ARC 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 will be 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 will be 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) will be 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.

8.1.4. Digiteo LETEVONE chair (2010-2013)

Partners: LIX (Ecole Polytechnique), PRiSM (UVSQ), DBWeb (Telecom ParisTech), Exalead S.A..
Funding: Grant covers the expenses of Pr. Vazirgiannis' visits to France (hosted by LIX) and of 2 PHD students.

Participant in the DIGITEO Learning Techniques for Evolving Networks chair, held by Pr. Michalis Vazirgiannis (Athens University of Economics and Business) from 2010 to 2013. The overall objective of the proposed project is mining and learning from the large scale and dynamically evolving data and graphs generated in the Web 2.0 context. Our particular collaboration has delt with privacy protection of users' data in this context.

8.2. European Initiatives

8.2.1. FP7 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

Coordinator: Philippe Bonnet (ITU of Copenhagen)

Other 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 wellsuited 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 [6], 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. The study of flash devices started during a short sabbatical of Luc Bouganim (from April to August 2008) in Copenhagen. The uFLIP study has been conducted in close cooperation with Philippe Bonnet from IT University of Copenhagen and Björn Þór Jónsson from Reykjavík University. The cooperation with Copenhagen is very active and led to new studies on flash devices and on the Trusted Cell architecture Two PhD students are currently co-supervised by Luc Bouganim and Philippe Bonnet. Philippe Bonnet got a Marie-Curie IEF grant and will visit SMIS for one year in 2013-2014.

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

See Section 8.2.1.

9. Dissemination

9.1. Scientific Animation

Philippe Pucheral:
- Area Editor of the Information Systems international journal (2007-2013)
- General chair of BDA'2014
- PC member of EDBT'14, APVP'13, MOBIWIS'14

Luc Bouganim:
- PC member of EDBT'14, ICDE'14, HardDB'13
- 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

Nicolas Anciaux:
- Co-chair and co-organiser of APVP'2013
- PC member of BDA 2013
- Member of the Inria " Commission des Développements Technologiques " (CDT)
- Member of the recruiting committee at Université Paris-Dauphine for the assistant professor position N°0230

Benjamin Nguyen:
- chair of BDA'2013 Demonstration Track
- co-chair and co-organiser of APVP'2013

• PC member of ICCSAMA'13 and ACOMP'2013
• Editorial Committee of TSI French Journal since 2012
• Member of the recruiting committees of Paris Dauphine and Paris VI

Iulian Sandu Popa:
• PC member of MOBILWARE'2013

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.

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
• Member of the HDR committee of the STV doctoral school
• Co-founder of the "Masses de Données Distribuées" French summer school and co-organizer of its last three editions (2010, 2012, 2014)

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:
• Associate Professor at UVSQ. Courses on databases, programming in Master 2, Licence
• Responsible for the Computer Science Masters at UVSQ
• Responsible for in-house training course for Computer Science high-school teachers at UVSQ for Académie de Versailles

N. Anciaux: (90h/y)
• Courses on DBMS internal mechanisms, database technology in Master1 and Master2 (AFTI, Orsay), and in engineering school (ENSTA Paris)

I. Sandu Popa:
• Assistant professor at UVSQ. Courses on databases, DBMS architecture and security in Licence, Master1, Master2 and engineer school ISTY

9.2.2. Supervision

PhD in progress : Quoc-Cuong To, Secure Global Computations on Personal Data Servers, since November 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, supervised by Luc Bouganim

PhD in progress : Matias Bjørling, Boosting IO Subsystem Performance on NVM Devices, since December 2011, co-supervised by Philippe Bonnet and Luc Bouganim

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

PhD in progress : Javier Gonzalez, Trusted Cell X: A Client-side Reference Monitor for Usage Control on a Trusted Execution Environment, since November 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)

9.2.3. Juries

• P. Pucheral: reviewer of Adeel Anjum PhD (Univ. Nantes, June 2013).
• B. Nguyen : PhD committee member of Johann VINCENT (Univ. Caen, June 2013).
• I. Sandu Popa: PhD Proposal Defense committee member of Susan (Juan) Pan (New Jersey Institute of Technology, April 2013), PhD committee member of Susan (Juan) Pan (New Jersey Institute of Technology, December 2013).

9.3. Popularization

• "Comment préserver l'anonymat", Pour la Science, n° 433, novembre 2013.
• "Bases de données en Classe Préparatoire", séminaire à Luminy, mai 2013.
• Tutorial on experimental methodology at the BDA summer school 2014 "Masses de données distribuées".
• Participation to the "Futur en Seine" event in Paris: http://web-tech.fr/linria-nous-devoile-ses-recherches-sur-les-smart-cities/
• "Une nouvelle approche de la protection de nos données", My Science Work, July 8, 2013, by Abby Tabor http://www.mysciencework.com/news/10335/une-nouvelle-approche-de-la-protection-de-nos-donnees

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, 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

[6] 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

[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] B. NGUYEN. , *Privacy-Centric Data Management*, Université de Versailles-Saint Quentin en Yvelines, December 2013, Habilitation à Diriger des Recherches, http://hal.inria.fr/tel-00936130

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", March 2013, pp. 1-55 [*DOI :* 10.1007/S10619-013-7122-X], http://hal.inria.fr/hal-00934586

[13] 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", 2013, pp. 1-27 [*DOI :* 10.1007/S10619-012-7119-X], http://hal.inria.fr/hal-00768355

[14] J. PAN, I. SANDU POPA, K. ZEITOUNI, C. BORCEA. *Proactive Vehicular Traffic Re-routing for Lower Travel Time*, in "IEEE Transactions on Vehicular Technology", October 2013, vol. 62, n^o 8, pp. 3551-3568, http://hal.inria.fr/hal-00935601

Articles in National Peer-Reviewed Journals

[15] N. ANCIAUX, B. NGUYEN, M. VAZIRGIANNIS. *Exposition Minimum de Données pour des Applications à Base de Classifieurs*, in "Ingénierie des Systèmes d'Information", November 2013, vol. 18, n^o 4, pp. 59-85, http://hal.inria.fr/hal-00937103

International Conferences with Proceedings

[16] 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

[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 "VLDB 2014 - 40th International Conference on Very Large Data Bases", Hangzhou, China, Zhejiang University, September 2014, http://hal.inria.fr/hal-00906204

[18] N. Anciaux, B. Nguyen, I. Sandu Popa. *Personal Data Management with Secure Hardware: How to keep your Data at Hand*, in "IEEE 14th International Conference on Mobile Data Management", Milan, Italy, June 2013, vol. 2, pp. 1-2, http://hal.inria.fr/hal-00935613

[19] M. Bjorling, P. Bonnet, L. Bouganim, N. Dayan. *The Necessary Death of the Block Device Interface*, in "CIDR 2013 - 6th Biennial Conference on Innovative Data Systems Research", Asilomar, United States, 2013, 4 p. , http://hal.inria.fr/hal-00768386

[20] N. Dayan, M. K. Svendsen, M. Bjorling, P. Bonnet, L. Bouganim. *EagleTree: Exploring the Design Space of SSD-Based Algorithms*, in "39th International Conference on Very Large Data Bases (VLDB)", Trento, Italy, August 2013, http://hal.inria.fr/hal-00935797

[21] B. Nguyen, W. Bezza, N. Anciaux, M. Vazirgiannis. *MinExp-Card : Limiting Data Collection using a Smart Card*, in "EDBT/ICDT 2013 Joint Conference : 16th International Conference on Extending Database Technology", Genoa, Italy, ACM, March 2013, pp. 753-758, http://hal.inria.fr/hal-00937105

Conferences without Proceedings

[22] C. Q. To, B. Nguyen, P. Pucheral. *A Prototype for Privacy-Preserving SQL Query Execution on Distributed Data (Demonstration)*, in "APVP 2013 - 4e Atelier sur la Protection de la Vie Privée", Les Loges en Josas, France, June 2013, http://hal.inria.fr/hal-00937109

[23] C. Q. To, B. Nguyen, P. Pucheral. *Privacy-Preserving SQL Query Execution on Distributed Data*, in "BDA 2013 - 29emes Journées Bases de Données Avancées", Nantes, France, October 2013, http://hal.inria.fr/hal-00937110

Other Publications

[24] T. Allard, B. Nguyen, P. Pucheral. , *Comment préserver l'anonymat ?*, November 2013, Pour la Science, nr 433, http://hal.inria.fr/hal-00937554

References in notes

[25] R. Agrawal, J. Kiernan, R. Srikant, Y. Xu. *Hippocratic Databases*, in "Proc. of the Int. Conf. on Very Large Data Bases (VLDB)", 2002

[26] T. Allard. , *Sanitizing Microdata Without Leak: A Decentralized Approach*, University of Versailles, 2011

[27] 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

[28] 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

[29] N. ANCIAUX, D. BOUTARA, B. NGUYEN, M. VAZIRGIANNIS. *Limiting Data Exposure in Multi-Label Classification Processes*, in "Fundamenta Informaticae", 2014, to appear

[30] N. ANCIAUX, B. NGUYEN, I. S. POPA. *Managing Personal Data with Strong Privacy Guarantees*, in "Proc. of the 17th International Conference on Extending Database Technology (EDBT)", Athens, Grece, Mars 2014, tutorial

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

[32] L. BOUGANIM. , *Logiciel uFLIP version 2.1, enregistré à l'Agence pour la Protection des Programmes (APP) sous le numéro IDDN.FR.001.110020.000.S.P.2009.0000.10000 en date du 10 mars 2009*, March 2009

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

[34] 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

[35] 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

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

[37] 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

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

[39] 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

[40] P. PUCHERAL, S. YIN. , *System and Method of Managing Indexation of Flash Memory*, May 2007, Dépôt par Gemalto et Inria du brevet européen nr 07290567.2

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

[42] C. Q. TO, B. NGUYEN, P. PUCHERAL. *Privacy-Preserving Query Execution using a Decentralized Architecture and Tamper Resistant Hardware*, in "Proc. of the 17th International Conference on Extending Database Technology (EDBT)", Athens, Grece, Mars 2014