IN PARTNERSHIP WITH:
CNRS

Université Versailles Saint-Quentin

Activity Report 2011

Project-Team SMIS

Secured and Mobile Information Systems

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

RESEARCH CENTER
Paris - Rocquencourt

THEME
Knowledge and Data Representation and Management

Table of contents

Project-Team SMIS

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

1. Members

Research Scientists
Luc Bouganim [Senior Researcher - INRIA, HdR]
Nicolas Anciaux [Junior Researcher - INRIA]

Faculty Members
Philippe Pucheral [Team leader, Professor - UVSQ, HdR]
Benjamin Nguyen [Associate Professor - UVSQ]

Technical Staff
Alexei Troussov [Senior Software Engineer]

PhD Students
Tristan Allard [UVSQ, MESR]
Yanli Guo [UVSQ, CORDI INRIA]
Lionel Le Folgoc [UVSQ, CORDI]
Shaoyi Yin [UVSQ, CORDI]

Administrative Assistant
Laurence Bourcier

2. Overall Objectives

2.1. Introduction

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.

2.2. Highlights

BEST PAPER AWARD :
[15] **Privacy, Security and Trust**. T. ALLARD, B. NGUYEN, P. PUCHERAL.

3. Scientific Foundations

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 [20]). 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 [26]. In the last years, many other policy languages have been proposed for different application scenarios, including EPAL [38], XACML [36] and WSPL [30]. 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) [32]. 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 (R. Agrawal : IBM Almaden,

C.C. Aggarwal: IBM Watson), 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 (R.N.Wright : Stevens Institute of Tech. / Rutgers Univ, NJ, V. Shmatikov : Univ Austin Texas), on Secure Multi-party Computing for Privacy Preserving Data Mining (J. Vaidya : Rutgers Univ, C. Clifton : Purdue Univ) and on distributed PPDP algorithms (D. DeWitt : Univ Wisconsin, K. Lefevre : Univ Michigan, J. Vaidya : Rutgers Univ, C. Clifton : Purdue Univ) 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, one application is today more specifically targeted by the SMIS project. This application deals with privacy preservation in EHR (Electronic Health Record) systems. Several countries (including France) launched recently ambitious EHR programs where medical folders will be centralized and potentially hosted by private Database Service Providers. Centralization and hosting increase the risk of privacy violation. In 2007, we launched two projects (PlugDB and DMSP) tackling precisely this issue, with the final objective to experiment our technologies in the field. In 2011, we launched a new project (KISS) capitalizing on the previous ones and extending their scope towards the protection of any personal data delivered to individuals in an electronic form.

5. Software

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 [9] [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 [33]. Chip-Secured XML Access (C-SXA) was an XML-based access rights controller embedded in a smart card [8]. 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 details the two prototypes we are focusing on today.

5.2. PlugDB engine

Participant: Nicolas Anciaux.

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 [29] and its Flash-based indexing system has been patented by INRIA and Gemalto [37]. 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.

Link: http://www-smis.inria.fr/Econtrat_PlugDB.html .

5.3. uFLIP Benchmark

Participant: Luc Bouganim.

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 [31]. It has been demonstrated at SIGMOD.

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

6. New Results

6.1. Embedded data management

Participants: Nicolas Anciaux, Luc Bouganim, Yanli Guo, Lionel Le Folgoc, Philippe Pucheral, Shaoyi Yin.

Inspired by low cost economic models, this work draws the idea of a one-dollar database machine, with the objective to disseminate databases everywhere, up to the lightest smart objects. In contrast to traditional database machines relying on massively parallel architectures, the one-dollar database machine considers the cheapest form of computer available today: a microcontroller equipped with GBs size (external) Flash storage. Designing such a database machine is very challenging due to a combination of conflicting RAM and NAND Flash constraints. To tackle this challenge, this work proposes a new paradigm based on database serialization (managing all database structures in a pure sequential way) and stratification (restructuring them into strata when a scalability limit is reached). We show that a complete DBMS engine can be designed according to this paradigm and demonstrate the effectiveness of the approach through a performance evaluation.

This work capitalizes on previous results related to the indexing of Flash resident data [39] and has also obvious connections with the more general study we are conducting on Flash-based data management (see Section 6.2). Partial elements of this solution have been demonstrated at [28].

6.2. Flash-based Data Management

Participant: Luc Bouganim.

Bimodal flash devices. While disks have offered a stable behavior for decades, thus guaranteeing the timelessness of many database design decisions, flash devices keep on mutating. Many researchers have proposed to adapt database algorithms to existing flash devices. However, today, there is no reference DBMS design based on solid assumptions of flash devices behavior, precisely because flash device behavior varies across models, across firmware updates and possibly over time for the same model: database researchers are running after flash memory technology. In this study, we took the reverse approach and defined how flash devices should support database management. We advocated that flash devices should provide guarantees to a DBMS so that it can devise stable and efficient IO management mechanisms. Based on the characteristics of flash chips, we defined a bimodal FTL that distinguishes between a minimal mode where sequential writes, sequential reads and random reads are optimal while updates and random writes are forbidden, and a mode where updates and random writes are supported at the cost of sub-optimal IO performance. This work started at the end of 2010 and was published at CIDR'11 [19], in cooperation with the IT University of Copenhagen. DBMS/Flash device co-design considerations were the focus of a tutorial on flash devices given recently at VLDB 2011 [20].

6.3. Privacy-Preserving Data Publishing

Participants: Tristan Allard, Benjamin Nguyen, Philippe Pucheral.

While most PPDP works make the assumption of a trusted central publisher, this study advocates a decentralized way of publishing anonymized datasets. More precisely, our work concerns the proof of feasibility of adapting traditional PPDP schemes, such as k-anonymity, ℓ-diversity or differential privacy to encompass the use of secure portable devices. In the applications we consider, each secure device is a data provider with weak computing capacities and weak connectivity (frequency and duration of connections are unpredictable)[1]. Weak connectivity precludes any P2P solution to the problem. A server allowing asynchronous communications between the devices becomes necessary to implement a distributed PPDP mechanism but this server does not benefit from the same trustworthiness as the participating devices. Our work aims to provide a generic method to adapt an important subclass of PPDP algorithms to this context, using both the limited secure computation capacities of each device (but taking advantage of their number) and the powerful computation abilities of an untrusted server available 24/7. Our proposal is based on a meta algorithm divided in three phases: (1) a collection phase where encrypted data is collected by the untrusted server, (2) a construction phase where the untrusted server performs a sound computation of a given privacy mechanism to generate sanitization rules and (3) a sanitization phase where the encrypted data is decrypted then sanitized by the devices to produce a final clear-text result. The last phase can be distributed using many different devices for better efficiency.

[1] E.g., in the e-health context, patients may have their medical folder embedded in a secure device and connect it sporadically when they visit their physician or when they want to consult it at home.

In [15], [17], we showed how it is possible to transform existing anonymity mechanisms into decentralized ones using secure devices, while maintaining equivalent security guarantees against honest-but-curious and weakly malicious adversaries. In [16], we studied the (unlikely) event that some secure devices might be compromised, and can collude with the untrusted server. We provided schemes to detect the compromised devices with a probability that can be fixed as close to 1 as desired (the trade-off being the latency of the protocol).

6.4. Minimal Exposure

Participants: Nicolas Anciaux, Benjamin Nguyen.

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 currently explore a reverse approach, where service providers would publicly describe the data they require to complete their task, and where the applicants would confront those descriptions with their own data to determine *themselves* the minimal subset of information to expose. We have first investigated solutions for simplistic tasks (e.g., evaluating a decision tree to determine the loan rate and duration a given applicant can claim), and we plan to address more complex ones (e.g., building the profile of customers, mining association rules, etc.) in the short term. The work on Minimal Exposure has just started and a first paper is under evaluation.

6.5. Experiment in the medical field

Participants: Nicolas Anciaux, Luc Bouganim, Lionel Le Folgoc, Philippe Pucheral, Alexei Troussov.

The PlugDB engine is being experimented in the field since September 2011 to implement a secure and portable medical-social folder. The objective is to improve the coordination of medical care and social services provided at home for dependent people. Details related to this experiment conducted with about 120 practitioners and patients are given in Section 7.2. While this action did not generate new academic results (though it helped us validating some previous results), it imposed us a strong investment in terms of test and optimization for our prototype and in terms of communication to promote this experiment at the regional level.

7. Contracts and Grants with Industry

7.1. Industrial collaborations

The SMIS project has a long lasting cooperation with Axalto, recently merged with Gemplus to form Gemalto, the world's leading providers of microprocessor cards. Gemalto provides SMIS with advanced hardware and software smart card platforms which are essential to validate numbers of our research results. In return, SMIS provides Gemalto with application requirements and technical feedbacks that help them adapting their future platforms towards data intensive applications. SMIS has also a growing cooperation with Santeos, an Atos Origin company developing software platforms of on-line medical services. Santeos is member of the consortium selected by the French Ministry of Health to host the French DMP (the national Personal Medical Folder initiative) . This cooperation helps us tackling one of our targeted applications, namely the protection of medical folders.

7.2. DMSP Yvelines District grant (Nov 2010 - Apr. 2012)

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). It is currently experimented in the context of a medical-social network providing care and services at home for elderly people. The experiment in the field started in September 2011 with a population of 120 volunteer patients and practitioners in the Yvelines district.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR DEMOTIS (Feb. 2009 - Feb. 2012)

Partners: SopinSpace (coordinator), INRIA (SMIS, SECRET), CECOGI
SMIS funding: 85k€
http://www.demotis.org/
The design and implementation of large-scale infrastructure for sensitive and critical data (e.g., electronic health records) have to face a tangle of legal provisions, technical standards, and societal concerns and expectations. DEMOTIS project aims to understand how the intrication between legal and technical domains constrains the design of such data infrastructures. DEMOTIS consists of two interdependent facets: legal (health law, privacy law, intellectual property law) and computer science (database security, cryptographic techniques). Combining expertise of researchers in Law and computer scientists should help to better assess whether law statements can be actually put in practice, to characterize the related technological challenges when mismatches are detected and, when possible, to suggest preliminary solutions.

8.1.2. 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.

8.2. European Initiatives

8.2.1. 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

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 [19], and (ii) a database system optimized for bimodal flash devices. The project funding will be managed by the IT University of Copenhagen and will cover the expenses for two co-supervised PhD students (including regular visits to and from Denmark

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

- P.M.G. Apers (Professor at the University of Twente, The Netherlands): collaboration on data confidentiality issues.
- Michalis Vazirgiannis (Athens University of Economics and Business): collaboration on Minimal Exposure in the context of Michalis' Digiteo Chair at LIX (Ecole Polytechnique).
- P. 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 the second study on bimodal flash devices. A masters student has started a PhD thesis, co-supervised by Luc Bouganim and Philippe Bonnet on bimodal flash devices. Philippe Bonnet has planned a 1 year visit to SMIS in 2012-2013.

8.3. International Initiatives

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

- Dennis Shasha (Professor at the University of New-York, USA): collaboration on tamper-resistant data management issues. Dennis Shasha has done a one year sabbatical stay in SMIS (July 2006 to June 2007).
- Xiaofeng Meng (Professor at Renmin University, Beijing, China): collaboration on embedded data management issues, partly funded by a Franco-Chinese research program (PRA SI-05604).
- I. Ray and I.Ray (Professors at Colorado State University, USA): collaboration on data privacy and usage control (Indrajit and Indrakshi Ray have visited SMIS from September 2009 up to February 2010).

9. Dissemination

9.1. Animation of the scientific community

- Philippe Pucheral
  - Area Editor of the Information Systems international journal (2007-now).
  - Scientific evaluation for ANR (programmes Blanc and Emergence) since 2009.
  - PC member of MOBIWIS'11, CODAPSY'11, EDBT'11.
  - Member of the recruiting committees of UVSQ and ENSIMAG.
  - Co-founder of the bi-annual French Summer School "Masses de Données Distribuées" and co-organiser of this school in2010 and 2012.
  - Referee for the PhD thesis M. Jawad (U. Nantes, 2011) , M. Tlili (U. Nantes, 2011).
- Luc Bouganim
  - Tutorial on Data Management in Flash Memories, given at VLDB 2011 [20].
  - President of the INRIA Post-Doc and Delegation Commission
  - PC member of EDBT'11, FlashDB'11, MobiWIS'11, VLDB'11 PhD WS.
  - PC chair of BDA 2011 (Bases de Données Avancées).
  - Reviewer for the "Digital engineering & security" program of the French research agency (ANR) (2011).
  - Member of the Commission PES (Prime d'Excellence Scientifique) for computer science at UVSQ (since 2010).
  - Referee for the PhD thesis of Brice Chardin (INSA Lyon) and Stéphane Jacob (Univ. Paris 6) in 2012
- Nicolas Anciaux
  - PC member of ICDE'11.
  - Member of the Editorial Board of TSI Journal (Technique et Science Informatiques) (2007 – now).
  - Jury member of the PhD thesis of S. Yin (UVSQ) in 2011.
- Benjamin Nguyen
  - Co-author of a book on the introduction of Computer Science in high school [21].
  - Member of the Advisory Committee of the W3C for the UVSQ, of the W3C XQuery Working Group (Test Suite Editor) and of the W3C Social Web Interest Group.
  - PC member of ICDE 2011, EDA 2011, ECML-PKDD 2011, BDA 2011.
  - Coach of the UVSQ H.E.O. student team that reached the French finals (top 5 teams) of Microsoft Imagine Cup (Software Design Category).
  - Member of the Selection Committee of UVSQ (since 2009), of U. Paris-X Nanterre (since 2007), and of U. Paris-XI (in 2011).
  - Elected member of the Scientific Committee of the Science Faculty of UVSQ.
  - Member of the INRIA Post-Doc Commission since 2010.

9.1.1. General Audience Actions

SMIS members have an important dissemination activity, motivated both by the popularity of the addressed research domain (security/privacy) and of the targeted applications (e.g., personal medical folder), leading to

- interactions with different institutions, for instance, the French Deputy Chamber, The Parliamentary Office for Evaluation of Scientific and Technological Options (OPECST), or the French Network and Information Security Agency.
- interviews resulting in articles in large audience magazine like "La Recherche", the CNRS Journal or BBC news, as well as participation in debates [25].
- talks and demonstrations targeting industrials in wide audience conferences like JavaOne, or e-Smart or meetings with industrials like "Les rendez-vous Carnot" and others [23].
- invited talks in conference or workshops with broad audience: targeting physicians [24], Researchers in law or keynote speech at UbiMob.
- actions targeting students like the organization of the French Summer School "Masses de Données Distribuées" in 2010 and 2012 or interventions at IBM France for developer training.

9.2. 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: (120h/y)
  - Full professor at UVSQ.
  - Director of the research Master COSY (UVSQ).
  - Member of the HDR committee of the STV doctoral school.
  - Courses on databases, DBMS architecture and security in Master1, Master2 and engineer school ISTY.
- B. Nguyen: (192h/y)
  - Associate professor at UVSQ.
  - Courses on object programing, databases, XML in undergraduate and Master2.
  - Courses on Teaching CS courses for High School teachers.
- L. Bouganim: (90h/y)
  - Courses on DBMS architecture, data security, database technology in Master1 and Master2 (AFTI, Orsay) and in engineering school (ENST Paris).
- N. Anciaux: (90h/y)
  - Courses on DBMS internal mechanisms, database technology in Master1 and Master2 (UVSQ), and in engineering school (ENSTA Paris).
- T. Allard: (64h/y)
  - Courses on Database concepts, System Programming in undergraduate and Masters 1 (UVSQ).
- L. Le Folgoc: (64h/y)
  - Courses on Relational Database Concepts and SQL, system programming in Masters 1 (UVSQ).
- S. Yin: (51h/y)
  - Courses on Relational Database Concepts and SQL, Embedded DBMS in Master1 and 2 (UVSQ and CNAM).

PhD & HdR:

Shaoyi Yin. Un modèle de stockage et d'indexation pour des données embarquées en mémoire flash. PhD Thesis University of Versailles Saint-Quentin-en-Yvelines (UVSQ), June 2011, Supervized by Philippe Pucheral.

Yanli Guo.Confidentialité et intégrité de bases de données embarquées. PhD Thesis University of Versailles Saint-Quentin-en-Yvelines (UVSQ), December 2011, Co-supervized by Luc Bouganim and Nicolas Anciaux

Tristan Allard. Sanitizing Microdata Without Leak: A Decentralized Approach. PhD Thesis University of Versailles Saint-Quentin-en-Yvelines (UVSQ), December 2011, Co-supervized by Philippe Pucheral and Benjamin Nguyen

PhD in progress : Lionel Le Folgoc, October 2009, Co-supervized by Luc Bouganim and Nicolas Anciaux

PhD in progress : Matias Bjørling, December 2011, Co-supervized by Philippe Bonnet and Luc Bouganim

10. Bibliography

Major publications by the team in recent years

[1] M. ABDALLAH, R. GUERRAOUI, P. PUCHERAL. *Dictatorial Transaction Processing : Atomic Commitment without Veto Right*, in "Distributed and Parallel Database Journal (DAPD)", 2002, vol. 11, n^{o} 3.

[2] 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.

[3] 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.

[4] N. ANCIAUX, L. BOUGANIM, P. PUCHERAL. *Memory Requirements for Query Execution in Highly Constrained Devices*, in "Proc. of the 29th Int. Conf. on Very Large Data Bases (VLDB)", 2003.

[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] 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.

[7] 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.

[8] 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.

[9] 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.

[10] S. YIN, P. PUCHERAL, X. MENG. *A Sequential Indexing Scheme for Flash-Based Embedded Systems*, in "Proc. of the International Conference on Extending Database Technology (EDBT)", Saint-Petersburg, Russia, March 2009.

Publications of the year

Doctoral Dissertations and Habilitation Theses

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

[12] Y. GUO. *Confidentialité et intégrité de bases de données embarquées*, University of Versailles, 2011.

[13] S. YIN. *Un modèle de stockage et d'indexation pour des données embarquées en mémoire flash*, University of Versailles, 2011.

Articles in International Peer-Reviewed Journal

[14] B. NGUYEN, A. VION, F.-X. DUDOUET, D. COLAZZO, I. MANOLESCU, P. SENELLART. *XML content warehousing: Improving sociological studies of mailing lists and web data*, in "Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique", October 2011, vol. 112, n^{o} 1, p. 5-31 [*DOI :* 10.1177/0759106311417540], http://hal.inria.fr/hal-00616613/en.

International Conferences with Proceedings

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

[16] T. ALLARD, B. NGUYEN, P. PUCHERAL. *Sanitizing Microdata Without Leak: Combining Preventive and Curative Actions*, in "ISPEC 2011 - Information Security Practice and Experience Conference", Gangzhou, China, 2011, p. 333-342, 10 pages, http://hal.inria.fr/hal-00624047/en.

[17] T. ALLARD, B. NGUYEN, P. PUCHERAL. *Towards a Safe Realization of Privacy-Preserving Data Publishing Mechanisms*, in "Ph. D. colloquium of the 12th IEEE International Conference on Mobile Data Management (MDM)", Luleå, Sweden, 2011, p. 1–4.

[18] I. BEDINI, C. MATHEUS, P. PATEL-SCHNEIDER, A. BORAN, B. NGUYEN. *Transforming XML schema to OWL using patterns*, in "ICSC 2011 - 5th IEEE International Conference on Semantic Computing", Palo Alto, United States, 2011, p. 1-8, http://hal.inria.fr/hal-00624055/en.

[19] 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, p. 1-8.

[20] P. BONNET, L. BOUGANIM, I. KOLTSIDAS, S. VIGLAS. *System Co-Design and Data Management for Flash Devices*, in "Very Large Data Bases Tutorial", 2011.

Scientific Books (or Scientific Book chapters)

[21] J.-P. ARCHAMBAULT, E. BACCELLI, S. BOLDO, D. BOUHINEAU, P. CÉGIELSKI, T. CLAUSEN, G. DOWEK, I. GUESSARIAN, S. LOPÈS, L. MOUNIER, B. NGUYEN, F. QUESSETTE, A. RASSE, B. ROZOY, C. TIMSIT, T. VIÉVILLE, J.-M. VINCENT. *Une introduction à la science informatique: Pour les enseignants de la discipline informatique au lycée*, CNDP-CRDP Eds., 2011.

[22] I. BEDINI, G. GARDARIN, B. NGUYEN. *Semantic technologies and e-business*, in "Electronic Business Interoperability : Concepts, Opportunities and Challenges", E. KAJAN (editor), IGI Global Publishing, 2011, p. 243-278, http://hal.inria.fr/hal-00623913/en.

Other Publications

[23] L. BOUGANIM. *Serveurs portables et sécurisés de données personnelles : Application aux données médicales*, June 2011, Les Rencontres des Tuileries, INRIA Paris-Rocquencourt.

[24] L. BOUGANIM. *Serveurs portables et sécurisés de données personnelles : Application aux données médicales*, May 2011, 18 emes journées de la SFIM@R.

[25] D. CARDON, G. DESGENS-PASANAU, B. NGUYEN. *Le droit à l'oubli sur Internet est-il possible ?*, February 2011, Conférence débat au Café des techniques, Musée des arts et métiers, Paris.

References in notes

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

[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, Y. GUO, P. PUCHERAL, J.-J. VANDEWALLE, S. YIN. *Pluggable personal data servers*, in "Proceedings of the 36th International Conference on the Management of Data (SIGMOD)", 2010, p. 1235-1238.

[29] 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.

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

[31] 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.

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

[33] 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.

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

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

[39] S. YIN, P. PUCHERAL, X. MENG. *A Sequential Indexing Scheme for Flash-Based Embedded Systems*, in "Proceedings of the 12th International Conference on Extending Data Base Technology (EDBT)", Saint-Petersburg, Russia, March 2009, p. 588-599.