INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team smis

Secured and Mobile Information Systems

Paris - Rocquencourt

Theme : Knowledge and Data Representation and Management

Activity Report

2010

Table of contents

1. Team

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]
Mehdi Benzine [UVSQ, MESR, up to September 2010]
Yanli Guo [UVSQ, CORDI INRIA]
Lionel Le Folgoc [UVSQ, CORDI]
Harold van Heerde [University of Twente (joint PhD with P. Apers team), up to June 2010]
Shaoyi Yin [UVSQ, CORDI]

Visiting Scientists

Indrajit Ray [Colorado State University, Invited Professor, up to February 2010]
Indrakshi Ray [Colorado State University, Invited Professor, up to February 2010]

Administrative Assistant

Elisabeth Baque

2. Overall Objectives

2.1. Overall Objectives

Ubiquitous and pervasive computing introduces the need for embedding and managing data in ever lighter and specialized computing devices (personal digital assistants, cellular phones, sensors and chips for the ambient intelligence, transportation, healthcare, etc). In this context, the first objective of the SMIS project is to define core database technologies tackling the hardware constraints of highly specialized computing devices. Alongside, by making the information more accessible and by multiplying the transparent ways of its acquisition, ubiquitous and pervasive computing induce new threats on data confidentiality. More generally, preserving the confidentiality of personal data spread among a large variety of sources (mobiles, smart objects as well as corporate, commercial and public databases) has become a major challenge for the database community. Therefore, the second objective pursued by the SMIS project is to define access control models preserving data confidentiality and privacy as well as tamper-resistant database architectures enforcing this control. These two objectives are detailed below.

Ubiquitous/pervasive data management: Important research efforts must be undertaken (1) to capture the impact of each device's hardware constraint on database techniques and (2) to set up co-design rules helping to calibrate the hardware resources of future devices in order to match specific application's requirements. This research direction covers storage and indexing models, query execution and optimization strategies, memory management and transaction protocols matching strong hardware constraints in terms of RAM, energy and communication bandwidth consumption. Electronic stable storage technologies (EEPROM, Flash, MEMS, etc) have also a considerable impact on the organization of the data at rest. Problems related to the interaction of ultra-light devices with a larger information system deserve also a particular attention (e.g., querying data disseminated among a large population of ultra-light devices).

Data confidentiality and privacy: The increasing amount of sensitive data gathered in databases, and in particular of personal data, imposes the definition of fine-grain access control models. While access control in client-server relational database is roughly mature, new issues appear today: fine-grain access control over hierarchical and semi-structured data (e.g., XML), integration of privacy concern in the access control policies (e.g., users consent, usage control, data retention), access control administration over multiple distributed and autonomous resources. A complementary issue we are interested in is the security (i.e., tamper-resistance) of the access control itself. Cryptographic techniques can be exploited to this end. While encryption is used successfully for years to secure communications, database encryption introduces difficult theoretical and practical problems: how to execute efficiently queries over encrypted data, how to conciliate declarative and dynamic access control policies with data encryption, how to distribute encryption keys between users sharing part of the database? We aim at providing accurate answers to these questions thanks to security models based on tamper-resistant hardware to query, update and share encrypted databases.

The complementarity of these two research issues is twofold. First, ubiquitous/pervasive data management introduces new threats on data privacy that must be tackled accurately. Hence, our first research activity is expected to feed the second one with relevant motivating examples. Second, data management techniques embedded in secured devices (e.g., smart cards, secured tokens) can be the foundation for new security models. For example, remote databases can be made secure by delegating part of the data management to a secured device. Thus, a strong cross-fertilization exists between our two research areas.

Beyond the scientific objectives sketched above, which are expected to generate publications in top level database and security conferences and journals, our ambition is to develop high quality prototypes that will serve two purposes: (1) validate our results on real hardware/software platforms and (2) integrate our results on real applications where data privacy is a major concern (e.g., Electronic Health Record systems).

3. Scientific Foundations

3.1. Ubiquitous data management

The vision of the future dataspace, a physical space enhanced with digital information made available through large-scale networks of smart objects is paint in [36]. The management of data in such dataspace differs dramatically from the mainframe database setting. In this context, the data sources are moving, managed by highly constrained computing devices, might get temporarily or permanently disconnected and have at best a partial knowledge about their environment.

This setting strongly impacts the way data is managed locally. Actually, not only data but also data management techniques (e.g., querying, access control, transaction) must usually be embedded in highly constrained hardware devices. For example, sensor networks collecting weather or pollution data [29] are evolving towards real distributed databases in which each sensor acts as an active node (i.e., as a micro-data server queryable remotely) [37]. Protecting the confidentiality of portable folders (e.g., healthcare folders, user's profiles) is another motivation to embed data management techniques into tamper-resistant devices (e.g., smart cards) [9]. Embedded database techniques are also required in every context where computations have to be performed in a disconnected mode. To conceive embedded database components is however not obvious. Each target architecture is specifically designed to meet desirable properties (portability, energy consumption, tamper resistance, production cost, etc), under imposed hardware constraints (maximum silicon die size, memory technology, etc), to tackle specific application's requirements. The challenge is then twofold: (i) being able to design dedicated embedded database components and (ii) being able to set up co-design rules helping hardware manufacturers calibrating 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, indexing and query optimization models minimizing the I/O bottleneck, parallel DBMS, main memory DBMS, replication and fault tolerance, etc), few research effort has been placed so far on embedded database techniques. Light versions of popular DBMS have been designed for powerful handheld devices but DBMS vendors never addressed the more complex problem of embedding database components into chips. Recent works have been conducted on smart card databases and on data management techniques for sensor networks but this research field is still at a preliminary stage.

The dataspace setting also impacts the way queries are expressed (spatio-temporal conditions, continuous queries) and executed (decentralized control, scarce local computing resources, uncertain availability of the data sources). Distributed query management has been extensively studied for thirty years [40], considering a reduced collection of data sources managed by high-end servers. These methods are irrelevant in a context involving potentially millions of data sources managed by lightweight devices. Query management in Peer-to-Peer systems and in Data Grids address the scalability issue and the unpredictable availability of data sources but do not consider lightweight devices. The first works to consider distributed queries over lightweight devices have been conducted in the sensor network field but they are restricted to basic filtering and aggregation queries. Hence, general queries distributed over a large collection of full-fledged databases managed by lightweight devices remains an open issue.

3.2. Data confidentiality

Confidentiality, Integrity and Availability are the three fundamental properties ruling the security of any information system. Data confidentiality has recently become a major concern for individuals as well as for companies and governments. Several kinds of data are threatened: personal data gathered by visited Web sites or by smart objects used in our daily life, corporate or administrative data stored in piracy-prone servers or hosted by untrusted Database Service Providers. The CSI/FBI reports that database attacks constitute the first source of cyber-criminology and that more than fifty percents of the attacks are conducted by insiders [32]. In this context, governments are setting up more constraining legislations. The problem is then to translate law statements into technological means: authentication mechanisms, data and communication encryption protocols, access control models, intrusion detection systems, data and operation anonymization principles, privacy preserving data mining algorithms, etc. The area of investigation is extremely large. Our own research program focuses on data access, usage and retention control and on the way this control can be made secure (i.e., tamper-resistant).

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, OrBAC [33]. While access control management in relational databases is now well established and normalized, new access control models have to be defined to cope with more complex data (e.g., hierarchical and semi-structured data like XML) and new forms of data distribution (e.g., selective data dissemination). Privacy models are also emerging today [26]. Privacy distinguishes from confidentiality is the sense that the data to be protected is personal. Hence, the user's consent must be reflected in the access control policies and not only the access but also the usage of the data as well as its retention period are safeguarded by law and must be controlled carefully.

Securing the access control against different forms of tampering is a complex issue. Server-enforced access control is widely accepted [28] but remains inoperative against insider attacks. Several attempts have been made to strengthen server-based security with database encryption [30] [35]. However, the Database Administrator (or an intruder usurping her identity) has enough privilege to tamper the encryption mechanism and get the clear-text data. Client-based security approaches have been recently investigated. Encryption and decryption occur at the client side to prevent any disclosure of clear-text data at the server. Storage Service Providers proposing encrypted backups for personal data are crude representative of this approach. The management of SQL queries over encrypted data complements well this approach [34]. Client-based decryption is also used in the field of selective data dissemination (e.g., Digital Right Management). However, the sharing scenarios among users are generally coarse grain and static (i.e., pre-compiled at encryption time). Tamper-resistant hardware can help devising secured database architectures alleviating this problem. Finally, securing the usage of authorized data is becoming as important as securing the access control as far as privacy preservation is concerned. Thus, database encryption, tamper-resistant hardware and their relationships with access control and usage control constitute a tremendous field of investigation.

4. Application Domains

4.1. Application Domains

Our work on ubiquitous data management 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. Hence, fine-grain access control models and robust database security mechanisms are highly required. Portable folder on secured mass storage chips can also help reducing the risk. In 2007, we launched two projects tackling precisely this issue (cf. Section 7.1 and 8.1).

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.

Since year 2000, we developed a succession of 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]. A first prototype was demonstrated at the VLDB'01 conference [27] and then optimized thanks to a comprehensive benchmarking campaign [5] conducted on a cycle-accurate hardware simulator. PicoDBMS has been a major vehicle to develop design rules for embedded database components and to set up a long term industrial cooperation with Axalto (today Gemalto). Chip-Secured Data Access (C-SDA) embeds a reduced SQL query engine and access right controller in a secure chip and acts as an incorruptible mediator between a client and an untrusted server hosting encrypted data [31]. Chip-Secured XML Access (C-SXA) is an XML-based access rights controller embedded in a smart card [8]. C-SXA evaluates user's privileges on a queried or streaming XML encrypted document and delivers the authorized subset of this document. 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. Link: http://www-smis.inria.fr/Eprototypes.html. The next subsections details the three 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 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 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) on a USB key form factor. 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 master piece 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, protect the data at rest against piracy and tampering (thanks to cryptographic protocols), executes queries (tackling low RAM constraint) and ensure transaction atomicity. Part of the on-board data can be replicated on a server (then synchronized) and be 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 and its Flash-based indexing system has been patented by INRIA and Gemalto [38]. It will be 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 for dependent people (See Section 7.1.2).

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

5.3. uFLIP Benchmark

Participants: Luc Bouganim [correspondent], Lionel Le Folgoc.

It is amazingly easy to get 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 (250 downloads in 6 months, more than 6000 visits), has been registered at APP in 2009.

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

5.4. GhostDB

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

GhostDB is a relational database engine embedded on a secure USB key (a large Flash persistent store combined with a tamper and snoop-resistant CPU and small RAM) that allows linking private data carried on the USB Key and public data available on a public server [2]. GhostDB ensures that the only information revealed to a potential spy is the query issued and the public data accessed. Queries linking public and private data entail novel distributed processing techniques on extremely unequal devices and in which data flows in a single direction: from public to private. The GhostDB prototype has been developed in C and currently runs on a software simulator of the USB device. This simulator is I/O accurate, meaning that it delivers the exact number of pages read and written in Flash, thus allowing assessing the GhostDB performance. The GhostDB prototype has been demonstrated at the VLDB'07 and BDA'07 conferences [39].

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

6. New Results

6.1. Introduction

The work conducted this year can be separated in two areas.

The first area concerns our research activity. While most research actions are continuation of studies initiated earlier, we are now considering them in an integrated and global vision that we call Personal Data Servers. The objective of this vision is to provide a credible alternative to a systematic centralization of personal data in servers. Indeed, the benefits of the server approach come at the price of a higher exposition of personal data to piracy and abusive use and of a loose of the control of the donor over his data. The objective is to study whether a complete information system could be organized by agglomerating distributed Personal Data Servers (PDS), each PDS remaining under the donor's control and being hardware protected. A large list of challenges remains to be solved to reach this goal: (1) how to guarantee acceptable performance for large embedded databases; (2) how to restore the traditional functions of central servers (availability, durability, global queries, data publishing, auditing, etc) with high security guarantees and (3) how to provide the user with tractable access and usage control models. The different works presented in this section must be considered as preliminary joint attempts to reach this ambitious goal.

The second area concerns the experimental work which monopolized a lot of our energy in 2009 and 2010. As stated above, the PlugDB technology will be experimented in the field with about 120 practitioners and patients. This imposed us a strong investment in terms of test and optimization to reach an acceptable level of robustness and efficiency for our prototype and also a significant and growing investment in terms of communication to promote this experiment at the regional and national level.

6.2. Personal Data Servers

Participants: Tristan Allard, Nicolas Anciaux, Luc Bouganim, Yanli Guo, Lionel Le Folgoc, Philippe Pucheral, Indrajit Ray, Indrakshi Ray, Shaoyi Yin.

An increasing amount of personal data is automatically gathered and stored on servers by administrations, hospitals, insurance companies, etc. Citizen themselves often count on internet companies to store their data and make them reliable and highly available through the internet. However, these benefits must be weighed against privacy risks incurred by centralization. In this study, we draw a radically different way of considering the management of personal data. We build upon the emergence of new portable and secure devices combining the security of smart cards and the storage capacity of NAND Flash chips. By embedding a full-fledged Personal Data Server in such devices, user's control of how her sensitive data is shared by others (by whom, for how long, according to which rule, for which purpose) can be fully reestablished and convincingly enforced. To give sense to this vision, Personal Data Servers must be able to interoperate with external servers and must provide traditional database services like durability, availability, query facilities, transactions. We proposed an initial design for the Personal Data Server approach, identified the main technical challenges associated with it and sketched preliminary solutions. This initial work gave rise to a vision paper published at VLDB'10 [15].

6.3. Embedded data management

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

Implementing the PDS vision requires tackling several important challenges, among which a central one is building an embedded DBMS engine implementing the core PDS functions. The difficulty lies in a combination of severe hardware constraints. The tamper-resistant microcontroller is equipped with (i) a tiny RAM and (ii) a scarce secure stable storage. Moreover, the mass storage is (iii) insecure because external to the microcontroller and (iv) made of NAND flash, a technology badly accommodating random writes (see above). Tackling these constraints leads to contradictory design recommendations. Indeed, executing queries on gigabytes of data with a tiny RAM entails indexing massively the database. The consequence is generating several fine-grain random writes in the indexes at insertion time. This leads to unacceptable write cost and integrity checking cost in NAND Flash. Existing solutions to decrease the amount of Flash (re)writes unfortunately incurs RAM consumption. A vicious circle is established and lets little hope to build a DBMS engine by assembling state of the art solutions. The solution we are exploring consists in breaking the implication between massive indexing and fine-grain random write pattern. To this end, we propose a new paradigm called *database serialization* where the complete database (including indexes) is organized sequentially. To maintain acceptable performance when the scalability limit of database serialization is reached, we propose a complementary principle called *database stratification*. This work capitalizes on previous results related to the indexing of Flash resident data [10] and the crypto-protection of databases [30]. It has also obvious connections with the more general study we are conducting on Flash-based data management (see Section 6.4). Partial elements of this solution have been demonstrated at [17], [22].

6.4. Flash-based Data Management

Participants: Luc Bouganim, Lionel Le Folgoc.

While disks have offered a stable behavior for decades, thus guaranteeing the timelessness of many database design decisions, flash devices keep on mutating. Their behavior varies across models, across firmware updates and possibly in time for the same model. Many researchers have proposed to adapt database algorithms for existing flash devices; others have tried to capture the performance characteristics of flash devices. However, today, we neither have a reference DBMS design nor a performance model for flash devices: database researchers are running after flash memory technology. In this study, we take the reverse approach and we define how flash devices should support database management. We advocate 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 define 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. Interestingly, the guarantees of a minimal mode have been taken for granted in many articles from the database research literature. Our point is that these guarantees are not a law of nature: we must guide the evolution of flash devices so that they are enforced. An important point is that providing optimal mapping guarantees does not hinder competition between flash device manufacturers. On the contrary, they can compete to (a) bring down the cost of optimal IO patterns (e.g., using parallelism), and (b) bring down the cost of non-optimal patterns without jeopardizing DBMS design. This work [18] capitalizes on previous results obtained in cooperation with the University of Copenhagen and the ReykjavÃk University [7] and demonstrated at [17], [22]. Future work includes designing and building a bimodal FTL in collaboration with a flash device manufacturer. Note that the scope of this study is much broader than the embedded data management context and can target any Flash-based DBMS, including high-end servers.

6.5. Data confidentiality and privacy

Participants: Tristan Allard, Nicolas Anciaux, Benjamin Nguyen, Philippe Pucheral.

SMIS is interested in the protection of personal data, also called microdata. Our initial work in this area focused on access control management and tamper-resistance of this control, notably in the healthcare context [24], [23]. Beyond access control, protecting microdata requires tackling a complementary issue named *usage control*. We are currently investigating two dimensions of usage control, namely privacy preserving data publishing and limited data retention.

Privacy Preserving Data Publishing. The primary goal of Privacy Preserving Data Publishing (PPDP) is to anonymize/sanitize microdata sets before publishing them to serve statistical analysis purposes. While most of the work done so far in PPDP does the assumption of a trusted central publisher, this study advocates a fully decentralized way of publishing anonymized datasets. It capitalizes on the emergence of more and more powerful and versatile Secure Portable Tokens raising new alternatives to manage and protect personal data (see Section “Personal Data Servers” for an illustration). The proposed approach allows the delivery of sanitized datasets extracted from personal data hosted by a large population of Secure Portable Tokens. The central idea lies in distributing the trust among the data owners while deterring dishonest participants to cheat with the protocols. Deviant behaviors are deterred thanks to a combination of preventive and curative measures. The fact that each smart token contains the data of a single individual, the tamper-resistance of the smart tokens and their low availability combined with an untrusted but highly available publishing infrastructure make the problem fundamentally different from any previously studied PPDP problem we are aware of. Preliminary results have been published in [20].

Minimal data retention problem. Our daily life activity leaves digital trails in an increasing number of databases (commercial web sites, internet service providers, search engines, location tracking systems, etc). Personal digital trails are commonly exposed to accidental disclosures and ill-intentioned scrutinization resulting from negligence, piracy and abusive usages. We consider the trail disclosure privacy breach, and studied the Minimal Retention (MR) Problem that consists in keeping only a subset of the data, while maintaining the same results for the processes using it. We focused more precisely on a typical and simple business application: classification. We showed that in this case the MR problem is NP-Hard and proposed a polynomial algorithm to approximate the solution.

7. Contracts and Grants with Industry

7.1. National grants

7.1.1. 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.

7.1.2. Secure and Mobile Healthcare folder : DMSP project

Category: project funded by the Yvelines District Council (CG78)
Duration: December 2006 – December 2009 (DMSP-V1)
Duration: November 2010 – April 2012 (DMSP-V2)
Partners: INRIA-SMIS (coordinator), Gemalto, Univ. Versailles-PRiSM, Santeos (Atos Origin)
Description: Electronic Health Record (EHR) projects have been launched in most developed countries to increase the quality of care while decreasing its cost. Despite the unquestionable benefits provided by EHR systems in terms of information quality, availability and protection against failures, patients are reluctant to leave the control over highly sensitive data (e.g., data revealing a severe or shameful disease) to a distant server. This project capitalizes on emerging hardware devices, called Secure Portable Tokens (SPT), to give the control back to the patient over his medical data. Roughly speaking, a SPT associates the security of a smart card to the storage capacity of a USB key. The objective is to complement a traditional EHR server with data management techniques embedded in SPT (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 proposed solution will be experimented in the context of a medico-social network providing medical care and social services at home for elderly people. Its architecture builds upon the technology designed in the PlugDB project (see section 8.1). It has been implemented during the first phase of the project, named DMSP-V1. The experiment in the field will be conducted with a population of about 120 volunteer patients and practitioners in the Yvelines district. It will last up to mid-2012 and constitutes the second phase of the project, named DMSP-V2.

8. Other Grants and Activities

8.1. National grants

8.1.1. PlugDB project

Category: ANR-RNTL project

Duration: February 2007 - May 2010
Partners: INRIA-SMIS (coordinator), Univ. Versailles-PRiSM, Gemalto, Santeos (Atos Origin), ALDS
Description: The goal of the PlugDB project is to design and experiment new technologies dedicated to a secured and ubiquitous management of personal data. Existing solutions for sharing and manipulating personal data (medical, social, administrative, commercial, professional data, etc.) are usually server-based. These solutions suffer from two weaknesses. The first one lies in the impossibility to access the data without a permanent, reliable, secured and high bandwidth connection. The second weakness is the lack of security warranties as soon as the data leaves the security realm of the server. The PlugDB project addresses these limitations with the help of a new secured device named SPT (Secure Portable Token). A SPT combines the tamper resistance of smart cards microcontrollers with Gigabytes-sized NAND Flash stable storage and the universality of the USB protocol. The project innovation lies in the association of sophisticated data management techniques with cryptographic protocols embedded in a SPT-like device. More precisely, a specific DBMS engine must be designed to match the peculiarities of the SPT storage memory (NAND Flash) and the limited processing capacities of its microcontroller. New cryptographic protocols dedicated to the protection of the data at rest as well as to the data in transit in collaborative scenarios must also be designed. The DMSP project will serve as a testbed for the PlugDB technology. This three years project has been extended by the ANR agency up to May 2010. The impact of the project has been rewarded by ANR and OSEO in January 2010.

8.1.2. DEMOTIS project

Category: ANR-ARPEGE project
Duration: Jan 2009 - Jan 2012
Partners: SopinSpace (coordinator), INRIA (SMIS, SECRET), CECOGI
Description: 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 jurists 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.

8.2. European Initiatives

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 (see details in Section 6.5). H.J.W. van Heerde, member of P. Apers team, is doing a PhD co-supervised by P. Apers and N. Anciaux.
- P. Bonnet (Associate Professor at the University of Copenhagen, Denmark): collaboration on Flash-based data management for high-end servers (see details in Section 6.3). Luc Bouganim did a 5 months stay in this team in 2008.

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 (see details in Section 6.3). This work has been partly funded by a Franco-Chinese research program (PRA SI-05604).

8.4. Exterior research visitors

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. Scientific activity and coordination

9.1.1. Collective responsibilities within INRIA

Philippe Pucheral has served in the Bureau du Comité des Projets (EPI council) of INRIA Rocquencourt from September 2004 to September 2008 and was in charge of the Mission Formation par la Recherche (Training through Research) at Rocquencourt. He is now member of this council.

Luc Bouganim has been president of the INRIA recruiting committee for CR in 2010 and vice-president of this same committee in 2009. He was member of the Commission Délégations-Détachements of INRIA Rocquencourt from 2004 up to 2010.

Benjamin Nguyen is member of the Commission Post-Doc since 2010. He has been member of the INRIA / Université Paris Sud 2010 Chair Selection Comity.

Nicolas Anciaux serves as a mediator at Rocquencourt to help solving difficulties which may occur between PhD students and their supervisors. He has been member of the INRIA / Université Paris Sud 2010 Chair Selection Comity.

9.1.2. Collective responsibilities outside INRIA

In 2010, the SMIS members have conducted, or participated to, the following actions in the research community:

- Philippe Pucheral
  - Area Editor of the Information Systems international journal.
  - PC member of EDBT'11, MDM'10, CODAPSY'11, BDA'10.
  - co-organizer of the French Summer School "Masses de Données Distribuées"
  - Member of the French BDA Board (Bases de Données Avancées).
  - Member of the PRiSM (UVSQ-CNRS Lab.) council.
  - Responsibilities at the University of Versailles (see Teaching section).
- Luc Bouganim
  - PC member of VLDB'10, EDBT'10, EDBT'11, Financial Crypto'10, FlashDB'11.
  - PC chair of BDA 2011.
  - Reviewer for International Journal: ACM TODS (2011), ACM SIGMOD Record (2010), Information Systems (2010), ACM TKDE (2010).
  - Member of the Commission PES (Prime d'Excellence Scientifique) for computer science at University of Versailles.
  - PhD Reviewer of Wenceslao Palma: Continuous Join Query Processing in Structured P2P Networks, University of Nantes.
- *Nicolas Anciaux*
  - PC member of ICDE'10 (demo), ICDE'11
  - Member of the Editorial Board of TSI Journal (Technique et Science Informatiques).

- *Benjamin Nguyen*
  - Assistant Director of UVSQ Computer Science Department since 2006
  - Responsible for CS Graduate course (L), since 2008
  - Member of the Selection Comity of UVSQ (CS) since 2009
  - Member of the Selection Comity of Paris-X Nanterre (CS)
  - Member of the Scientific Comity of the Science UFR of UVSQ
  - Expert of the ANR from 2010
  - Member for the Advisory Comitee of the W3C for the UVSQ
  - Member of the W3C XQuery Working Group
  - Member of the W3C Social Web Interest Group
- Indrakshi Ray
  - PC member of PAIS 2010
- Indrajit Ray
  - PC member of AINA'10, ARES'10

9.1.3. Invited talks

- Philippe Pucheral, *The PlugDB project*, STIC colloquium, Paris, January 5-7th, 2010.
- Luc Bouganim, *Dossiers personnels sécurisés et mobiles*, ANSSI, March 23rd, 2010.
- Benjamin Nguyen, *L'anonymisation des données du DMP et ses alternatives*, conference in Law "Partage et secret de l'information de santé", Nancy, October 15th, 2010.
- Nicolas Anciaux & Alexei Troussov, *Dossier médico-social sécurisé*, INRIA-industry day and Connectathon, Bordeaux, April 15th, 2010.
- Tristan Allard *Safe Anonymization of Data Hosted in Smart Tokens*. LIP6, Paris, June 2010.

9.2. Teaching activity

SMIS is a joint project-team with the University of Versailles Saint-Quentin en Yvelines (UVSQ) and CNRS. The list of the main courses given by each staff member in 2009 is given below:

- P. Pucheral: 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 (92h/y).
- Benjamin Nguyen: Assistant professor at UVSQ, vice-director of the computer sciences teaching department, courses on object programing, databases, XML (192h/y)
- L. Bouganim: DBMS architecture, data security, database technology (90h/y, given at AFTI Orsay, ENST Paris).
- N. Anciaux: DBMS internal mechanisms, database Technology (90h/y, given at UVSQ and ENSTA)
- T. Allard: Database concepts, System Programming (64h/y given at UVSQ)
- L. Le Folgoc: Relational Database Concepts and SQL, system programming (64h/y given at UVSQ)
- S. Yin: Relational Database Concepts and SQL, Embedded DBMS (51h/y given at UVSQ and CNAM)

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° 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° 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° 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° 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° 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] M. BENZINE. *Combinaison sécurisée de données publiques et sensibles dans les bases de données*, University of Versailles, September 2010.

[12] H. VAN HEERDE. *Privacy-aware data management by means of data degradation -making private data less sensitive over time*, University of Twente and University of Versailles, June 2010.

Articles in International Peer-Reviewed Journal

[13] M. BJØRLING, P. BONNET, L. BOUGANIM, B. JÓNSSON. *uFLIP: Understanding the Energy Consumption of Flash Devices*, in "IEEE Data Engineering Bulletin", 2010, vol. 33, n^{o} 4.

Articles in National Peer-Reviewed Journal

[14] B. NGUYEN, A. VION, F. X. DUDOUET, D. COLAZZO, I. MANOLESCU. *WebStand, une plateforme de gestion de données Web pour applications sociologiques*, in "Technique et Science Informatiques (TSI), numéro spécial sur L'informatique Ã l'Interface des Sciences Humaines et Sociales", 2010, vol. 29, n^{o} 8-9.

International Peer-Reviewed Conference/Proceedings

[15] 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.

[16] N. ANCIAUX, L. BOUGANIM, Y. GUO, P. PUCHERAL, J.-J. VANDEWALLE, S. YIN. *Pluggable Personal Data Servers*, in "29th ACM International Conference on Management of Data (SIGMOD)", 2010, Demo paper.

[17] M. BJØRLING, L. LE FOLGOC, A. MSEDDI, P. BONNET, L. BOUGANIM, B. JÓNSSON. *Performing Sound Flash Device Measurements: The uFLIP Experience*, in "29th ACM International Conference on Management of Data (SIGMOD)", 2010, Demo paper.

[18] 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, to appear.

National Peer-Reviewed Conference/Proceedings

[19] T. ALLARD, N. ANCIAUX, L. BOUGANIM, Y. GUO, L. LE FOLGOC, B. NGUYEN, P. PUCHERAL, I. RAY, I. RAY, S. YIN. *Serveurs Personnels Sécurisés de Données*, in "26èmes journées Bases de Données Avancées (BDA)", Toulouse, 2010.

[20] T. ALLARD, B. NGUYEN, P. PUCHERAL. *Safe Anonymization of Data Hosted in Smart Tokens*, in "26èmes journées Bases de Données Avancées (BDA)", Toulouse, 2010.

[21] N. ANCIAUX, L. BOUGANIM, Y. GUO, P. PUCHERAL, J.-J. VANDEWALLE, S. YIN. *Pluggable Personal Data Servers*, in "26èmes journées Bases de Données Avancées (BDA)", Toulouse, 2010, Demo paper.

[22] M. BJØRLING, L. LE FOLGOC, A. MSEDDI, P. BONNET, L. BOUGANIM, B. JÓNSSON. *Performing Sound Flash Device Measurements: The uFLIP Experience*, in "26èmes journées Bases de Données Avancées (BDA)", Toulouse, 2010, Demo paper.

Scientific Books (or Scientific Book chapters)

[23] T. ALLARD, N. ANCIAUX, L. BOUGANIM, P. PUCHERAL, R. THION. *Concilier sécurité et ubiquité des données médicales*, Cahiers du CRID, Editions Bruylant, 2010, vol. 32.

[24] T. ALLARD, N. ANCIAUX, L. BOUGANIM, P. PUCHERAL, R. THION. *Pervasive and Smart Technologies for Healthcare: Ubiquitous Methodologies and Tools*, A. CORONATO, G. DE PIETRO (editors), Information Science Reference, 2010.

Scientific Popularization

[25] J.-F. PARGUET, P. PUCHERAL. *Protection des données médicales numérisées : questions à Jean-François Parguet et à Philippe Pucheral, propos recueillis par Dominique Chouchan*, in "La Recherche. Les Cahiers de l'Inria", March 2010, n^{o} 439 mars 2010, http://hal.inria.fr/inria-00511468/en.

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] A. BARAANI, J. PIEPRZYK, R. SAFAVI-NAINI. *Security In Databases: A Survey Study*, 1996, http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.51.4838.

[29] P. BONNET, J. GEHRKE, P. SESHADRI. *Towards Sensor Database Systems*, in "Proc. of Int. Conf. on Mobile Data Management", 2001.

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

[31] 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.

[32] COMPUTER SECURITY INSTITUTE. *CSI/FBI Computer Crime and Security Survey*, 2010, http://gocsi.com/Survey_2010.

[33] F. CUPPENS. *Modélisation Formelle de la Sécurité des Systèmes d'Informations*, 2000, Habilitation à Diriger des Recherches, Université Paul Sabatier.

[34] H. HACIGUMUS, B. IYER, C. LI, S. MEHROTRA. *Executing SQL over Encrypted Data in the Database-Service-Provider Model*, in "Proc. of the ACM SIGMOD Int. Conf. on Management of Data", 2002.

[35] J. HE, M. WANG. *Cryptography and Relational Database Management Systems*, in "Proc. of the Int. Database Engineering and Application Symposium (IDEAS)", 2001.

[36] T. IMIELINSKI, B. NATH. *Wireless Graffiti – Data, data everywhere*, in "Proc. of the Int. Conf. on Very Large Data Bases (VLDB)", 2002.

[37] S. MADDEN, M. FRANKLIN, J. HELLERSTEIN, W. HONG. *The design of an Acquisitional Query Processor for Sensor Networks*, in "Proc. of the ACM Sigmod Int. Conf. on Management of Data", 2003.

[38] P. PUCHERAL, S. YIN. *System and Method of Managing Indexation of Flash Memory*, May 2007, Dépôt par Gemalto et INRIA du brevet européen no 07290567.2.

[39] C. SALPERWYCK, N. ANCIAUX, M. BENZINE, L. BOUGANIM, P. PUCHERAL, D. SHASHA. *GhostDB: Hiding Data from Prying Eyes*, in "33th International Conference on Very Large Data Bases, (VLDB)", September 2007, Demo paper.

[40] T. ÖZSU, P. VALDURIEZ. *Principles of Distributed Database Systems*, Second Edition, Prentice Hall, 1999.