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

Table of contents

1. Team **1**
2. Overall Objectives **1**
3. Scientific Foundations **2**
3.1. Ubiquitous data management 2
3.2. Data confidentiality 3
4. Application Domains **4**
5. Software **4**
5.1. Introduction 4
5.2. GhostDB 5
5.3. PlugDB engine 5
6. New Results **5**
6.1. Introduction 5
6.2. Embedded data management 6
6.3. Data confidentiality and privacy 6
6.4. Tamper-resistant data management 7
6.5. Databases and Cryptography 8
7. Contracts and Grants with Industry **8**
7.1.1. Industrial collaborations 8
7.1.2. Secure and Mobile Healthcare folder : DMSP project 8
8. Other Grants and Activities **9**
8.1. National grants 9
8.1.1. PlugDB project 9
8.1.2. DEMOTIS project 9
8.2. International and national cooperations 10
9. Dissemination **10**
9.1. Scientific activity and coordination 10
9.1.1. Collective responsibilities within INRIA 10
9.1.2. Collective responsibilities outside INRIA 10
9.1.3. Invited talks 11
9.2. Teaching activity 11
10. Bibliography **12**

1. Team

Research Scientist

Luc Bouganim [DR2 - INRIA, HdR]
Nicolas Anciaux [CR1 - INRIA]

Faculty Member

Philippe Pucheral [Team leader, PR1 - UVSQ, HdR]

Technical Staff

Marcel Lienafa [CNAM, IE, from July 15th to November 31]
Maggy El Kholy [ENSIMAG, IA, up to August 31]

PhD Student

Mehdi Benzine [UVSQ, MESR]
Harold van Heerde [University of Twente (joint PhD with P. Apers team)]
Lionel Le Folgoc [UVSQ, CORDI, from October 1st]
Shaoyi Yin [UVSQ, CORDI]
Yanli Guo [UVSQ, CORDI INRIA]
Tristan Allard [UVSQ, MESR]

Visiting Scientist

Indrajit Ray [Colorado State Univeristy, Invited Professor, from September 1st]
Indrakshi Ray [Colorado State Univeristy, Invited Professor, from September 1st]

Administrative Assistant

Elisabeth Baque [AI - INRIA]

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 the definition of 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. Thus, the second objective pursued by the SMIS project is the definition of access control models preserving data confidentiality and privacy and the definition of tamper-resistant database architectures enforcing this control. These two objectives are detailed below.

Ubiquitous/pervasive data management: Important research efforts must be undertaken to capture the impact of each device's hardware constraints on database techniques and 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 is interested in storage and indexing models, query execution and optimization strategies, 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, defining and managing ambient databases).

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), access control administration over multiple distributed, heterogeneous 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 (i.e., predicate based) and dynamic access rights with 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 generates specific confidentiality problems that must be tackled accurately. Hence, this first area of research 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 these two research areas.

Beyond the scientific objectives detailed 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 confidentiality is a primary 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 [43]. 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 [38] are evolving towards real distributed databases in which each sensor acts as an active node (i.e., as a micro-data server queryable remotely) [44]. Protecting the confidentiality of portable folders (e.g., healthcare folders, users' profiles) is another motivation to embed data management techniques into tamper-resistant devices (e.g., smart cards) [10]. 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 [47], 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 (restricted to filters and aggregations) over lightweight devices have been conducted in the sensor network field. Hence, regular 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 [39]. 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 [40]. 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 [30]. 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 very important issue. Server-enforced access control is widely accepted [37] but remains inoperative against insider attacks. Several attempts have been made to strengthen server-based security with database encryption [27] [42]. 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 [41]. 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 and the data on transit, 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).

5. Software

5.1. Introduction

In our domain of expertise, developing software prototypes is mandatory to validate research solutions and is an important vector for research 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., smart cards), 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 [10]. A first prototype was demonstrated at the VLDB'01 conference [33] 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. 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 [7]. 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 concentrate on the two prototypes focusing our current activity.

5.2. 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 (See Section 6.4). 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 recently demonstrated at the VLDB'07 and BDA'07 conferences [46].

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

5.3. PlugDB engine

Participants: Nicolas Anciaux [correspondent], Mehdi Benzine, Luc Bouganim, Maggy El Kholy, Philippe Pucheral, Shaoyi Yin, Yanli Guo, Lionel Le Folgoc.

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 2008 [28] and its Flash-based indexing system has been patented by INRIA and Gemalto [45]. 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 for dependent people (See Section 7.1.2).

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

6. New Results

6.1. Introduction

The work conducted this year can be separated in two areas. The first area concerns the experimental work which monopolized a lot of our energy in 2009. As stated above, the PlugDB technology is being 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.

The second area concerns the research work itself. While most research actions are continuation of studies initiated earlier, we are now considering them in a more 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 works presented below must be considered as preliminary joint attempts to reach this ambitious goal.

6.2. Embedded data management

Participants: Nicolas Anciaux, Luc Bouganim, Philippe Pucheral, Shaoyi Yin.

In 2009, we pursued the work initiated on the definition of storage and indexing models dedicated to electronic stable storage technologies, and more precisely to NAND-Flash. Such techniques are very challenging to design in our context due to a combination of NAND Flash constraints (i.e., block-erase-before-page-rewrite constraint and limited number of erase cycles) and embedded system constraints (i.e., tiny RAM and resource consumption predictability). Last year, we proposed a new alternative for indexing Flash-resident data, called PBFilter, which specifically addresses the embedded context. This approach organizes the index structure in a purely sequential way and speeds up key lookups thanks to two principles called Summarization and Partitioning. PBFilter has been patented by Gemalto and INRIA [45] and published in [19]. PBFilter has been preliminary designed for primary-key indexes in an append-oriented database context. Our current work focuses on the integration of this Flash-based storage and indexing model in a complete DBMS engine supporting secondary key indexes, updates and transaction management. An efficient atomic transaction protocol has been defined based on a virtualization of all database updates (i.e., data modifications and deletions do not modify the database state; they rather apply at load time to dynamically build a consistent version of the pages in RAM).

Thanks to its excellent properties in terms of read performance, energy consumption and shock resistance, NAND Flash has become a credible competitor even for traditional disks on high-end servers. A natural extension of the aforementioned action is thus to study how database systems adapt to this new form of secondary storage. Before we can answer this question, we need to fully understand the performance characteristics of flash devices. We have designed a benchmark, called uFLIP, to cast light on all relevant usage patterns of current, as well as future, flash devices. uFLIP is a set of nine micro-benchmarks based on IO patterns (i.e., a sequences of IOs). Each micro-benchmark is a set of experiments designed around a single varying parameter, that affects either time, size, or location. Thanks to uFLIP, we established which kind of IOs should be favored (or avoided) when designing algorithms and architectures for flash-based systems. We also set up a benchmarking methodology which takes into account the particular characteristics of flash devices. This work, published in [16] and [21], was done in cooperation with the University of Copenhagen and the Reykjavík University. More recently, we have also devised a mechanism for measuring the energy consumption of flash devices. While energy consumption cannot be traced to individual IOs, we can associate energy consumption figures to IO patterns, which helps understand further the behavior of the devices. The source code of uFLIP is available on http://www.uflip.org and has been register at APP (Agence de Protection des Programmes) in 2009 [29].

6.3. Data confidentiality and privacy

Participants: Nicolas Anciaux, Luc Bouganim, Harold van Heerde, Philippe Pucheral.

SMIS has been initially involved in the definition of fine-grain access control models trying to capture the complexity of the information to be protected [9]. Now, the team focuses on the protection of personal data (also called micro-data) where concepts like user's consent, purpose declaration and limited retention play a central role. Conversely to (and complementary with) our initial approach, the challenge we are addressing now is to define models as simple as possible to help a user calibrating a predefined access control policy to her specific situation and sensitivity. We started to study how user consent could be more easily expressed in the context of Electronic Health Record (EHR) systems [32] [31]. Indeed, access control policies usually defined to regulate accesses to EHR systems are far too complex to expect collecting an enlightened consent of the patients on them, as required by the law. This is mainly due (1) to a huge number of rules (huge number of practitioners with a diversity of roles and privileges) and (2) to the intrinsic complexity of the data to be protected (which data reveals which pathology?). To tackle this issue, we are designing an Event-Based Access Control model (EBAC) helping the patient to mask sensitive records in her folder. The EBAC masking rules take priority over the default access control rules and are defined on metadata easy to manage by the user. Any document added to a folder is described by an event, events are grouped by episodes (i.e., a set of events sharing a common masking policy, like "MyAbortion", "MySecondDepression") and the participation of a practitioner to an episode is regulated by a relation of confidence. The intuition of this model has been published in [11][23] but this work is still at a preliminary stage.

In the same (usage control) line, we are tackling the limited 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. No one is sheltered because common events, like applying for a job, can suddenly make our history a precious asset. By definition, access control fails preventing trail disclosures, and anonymity techniques are often not usable in this context, motivating the integration of the Limited Data Retention principle in legislations protecting data privacy. By this principle, data is withdrawn from a database after a predefined time period. However, this principle is difficult to apply in practice, leading to retain useless sensitive information for years. To address this issue, we propose the Data Degradation Model where sensitive data undergoes a progressive and irreversible degradation from an accurate state, to degraded but still informative states, up to complete disappearance when the data becomes useless, along with suitable query and transaction semantics [36], [35]. The benefit of this model is twofold: (i) the amount of accurate data, and thus the privacy offence resulting from a trail disclosure, is drastically reduced; (ii) the model is flexible enough to remain in line with the applications purposes, and thus favors data utility. We have recently formalized those benefits, and shown (under reasonable assumptions) to which extent data degradation overcomes basic implementations of the limited data retention principle [20]. In addition, the data degradation model strongly impacts core database techniques, opening interesting research issues. We made a preliminary study into that direction, by proposing database storage and indexing structures, logging and locking mechanisms adapted to data degradation [34], to show the practical feasibility of the model.

6.4. Tamper-resistant data management

Participants: Tristan Allard, Nicolas Anciaux, Mehdi Benzine, Luc Bouganim, Philippe Pucheral.

We have extended the work done in the context of GhostDB [2] to tackle aggregate computations performed on a mix of hidden sensitive data (kept on a tamper-resistant device) and of public data (available on a public server). The goal is to produce aggregates to data warehouses for OLAP purposes, and to reveal exactly what is desired, neither more nor less. This work has been published in [12]. More general work on tamper-resistant data management has been published in [24].

In 2009, we started a new study related to Privacy Preserving Data Publishing (PPDP), where personal data sets are anonymized before being published to serve statistical analysis purpose. Most work conducted so far on PPDP considers a model where the data publisher is trusted. Considering the vulnerability of database servers against external and internal attacks, we consider in this study an untrusted PPDP model. In the proposed model, private data is not uploaded onto a central data publisher. Instead, individuals store their personal

data in smart tokens, under their control. When needed, these tokens collaborate to anonymize the data and submit the result to a data publisher. The problem becomes how to publish an anonymized version of a dataset horizontally partitioned in a large number of smart tokens. This must be done without compromising data privacy while the computing and communication infrastructure linking the smart tokens is untrusted. 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 infrastructure make the problem fundamentally different from any previously studied PPDP problem we are aware of. A first solution as been designed to solve this problem with acceptable performance but the work is still at a preliminary stage.

6.5. Databases and Cryptography

Participants: Nicolas Anciaux, Luc Bouganim, Yanli Guo, Philippe Pucheral.

We have initiated this year a cooperation with members of the SECRET project-team which focuses on the use of cryptographic techniques for ensuring the confidentiality and integrity of data stored in databases. Using cryptographic techniques "as-is" to provide the aforementioned guarantees has a large negative impact on the database size (e.g., a 20 bytes MAC is added to each encrypted attribute value in Oracle 11g TDE to ensure data authenticity) and on the database performance, thus motivating many on-going research on that topic. In a first step, we have made an exaustive study of the state of the art which reveals that many techniques devised are simply unsecure [22] [27]. This work naturally fits our new orientation on personal data servers (See Section 6.1) where we need to protect both the embedded database and the data which are stored on untrusted servers.

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 was one of the consortia selected by the French Ministry of Health to host the future DMP (the national Personal Medical Folder initiative) during its prefiguration phase. 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
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 a new hardware portable device, called SPT, associating the security of a smart card to the storage capacity of a USB key, to give the control back to the patient over his medical data. 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 architecture will be experimented in the context of a medico-social network providing medical care and social services at home for elderly people. The experiment will be conducted with a population of about 120 volunteer patients and practitioners in the Yvelines district. The experiment in the field will start in January 2010 from a 18 months period and will be funded thanks to an extension of the DMSP agreement.

8. Other Grants and Activities

8.1. National grants

8.1.1. PlugDB project

Category: ANR-RNTL project
Duration: February 2007 - February 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 intrinsic security of smart cards with the storage capacity of USB keys (several GB soon) 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.

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. International and national cooperations

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 (see details in Section 6.4). 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 6.2). This work is partly funded by a Franco-Chinese research program (PRA SI-05604).
- P.M.G. Apers (Professor at the University of Twente, The Netherlands): collaboration on data confidentiality issues (see details in Section 6.3). 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 6.2). Luc Bouganim did a 5 months stay in this team in 2008.
- I. Ray and I.Ray (Professors at Colorado State University, USA): collaboration on data privacy and usage control (Indrajit and Indrakshi Ray are visiting SMIS for 6 months starting September 1st).

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 is vice-president of the recruiting committee for INRIA CR. He is member of the Commission Délégations-Détachements of INRIA Rocquencourt since November 2004. He is the INRIA representative for the summer schools in computer science co-organized by INRIA, CEA and EDF (up to March 2009). He is also co-responsible for the organization of the monthly scientific seminars ("Le modèle et l'Algorithme") at INRIA Rocquencourt (up to March 2009).

Nicolas Anciaux serves as a mediator at Rocquencourt to help solving difficulties which may occur between PhD students and their supervisors.

9.1.2. Collective responsibilities outside INRIA

In 2009, 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'09, MDM'09, DS2ME'09, SAR-SSI'09.
  - Demonstration co-chair of VLDB'09.
  - Scientific evaluation of projects funded by the ARPEGE program (from Embedded Sytems to Large Infrastructure) launched by the French National Research Agency (ANR).
  - Member of the French BDA Board (Bases de Données Avancées).
  - Member of the PRiSM Lab (UVSQ) council.
  - Referee for the HDR (Habilitation à Diriger des Recherches) of I. Manolescu (U. Orsay).

  - Responsibilities at the University of Versailles (see Teaching section).

- Luc Bouganim
  - PC member of SIGMOD'09, SIGMOD'09 (tutorials), EDBT'09 (demo), EDBT'10, Financial Crypto'10, BDA'09, UbiMob'09
- Nicolas Anciaux
  - PC member of SIGMOD'09 (demo), ICDE'10 (demo)
  - Member of the Editorial Board of TSI Journal (Technique et Science Informatiques).
- Indrakshi Ray
  - PC member of PAIS 2010
  - Tutorial Co-Chair of ICISS'2009
- Indrajit Ray
  - Publicity Chair of ICISS'09
  - PC member of AINA'10, ARES'10

9.1.3. Invited talks

- Philippe Pucheral
  - Public hearing at the French Deputy Chamber related to the National EHR project (Dossier Médical Personnel), Assemblée Nationale, April 30th, 2009.
- Luc Bouganim
  - Dossier médico-social sécurisé, Office Parlementaire d'Evaluation des Choix Scientifiques et Technologiques (OPECST), April 1st, 2009
- Nicolas Anciaux
  - Demonstration of Electronic Health Records (EHR) on Java Card 3.0 Technology-Based Devices, JavaOne Conference, June 3rd, 2009
- Indrajit Ray and Indrakshi Ray
  - Network and information survivability - PSG College of Technology in Coimbatore, India, December 28-31, 2009

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: Professor at UVSQ, co-director of the research Master COSY (UVSQ), member of the steering committee of the SOFT doctoral school, courses on databases, DBMS architecture and security in Master1, Master2 and engineer school ISTY (92h/y).
- L. Bouganim: DBMS architecture, data security, database technology (90h/y, given at UVSQ, ENST Paris, CNAM Paris, ENSTA Paris).
- N. Anciaux: DBMS internal mechanisms, database Technology (90h/y, given at UVSQ and ENSTA).
- M. Benzine: Java programming, business intelligence, database concepts (110h/y given at UVSQ).
- T. Allard: Database concepts, System Programming (64h/y given at UVSQ)
- L. Le Folgoc: Relational Database Concepts and SQL (36h/y given at UVSQ)

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)", vol. 11, n^{o} 3, 2002.

[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, 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.

[4] N. ANCIAUX, L. BOUGANIM, P. PUCHERAL. *Future Trends in Secure Chip Data Management*, in "IEEE Data Engineering Bulletin", vol. 30, n^{o} 3, 2007.

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

[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)", vol. 10, n^{o} 4, January 2008.

[7] 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.

[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] B. FINANCE, S. MEDJDOUB, P. PUCHERAL. *The Case for Access Control on XML Relationships*, in "Proc. of the ACM Int. Conf. on Information and Knowledge Management (CIKM)", 2005.

[10] 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", vol. 10, n^{o} 2-3, 2001.

Year Publications

Articles in International Peer-Reviewed Journal

[11] T. ALLARD, N. ANCIAUX, L. BOUGANIM, P. PUCHERAL, R. THION. *Seamless Access to Healthcare Folders with Strong Privacy Guarantees*, in "special issue of the Journal of Healthcare Delivery Reform Initiatives", 2009, to appear.

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

[13] G. GEORG, K. ANASTASAKIS, B. BORDBAR, S. H. HOUMB, I. RAY, M. TOAHCHOODEE. *Verification and Trade-off Analysis of Security Properties in UML System Models*, in "IEEE Transactions on Software Engineering", 2009, to appear US .

Articles in National Peer-Reviewed Journal

[14] T. ALLARD, N. ANCIAUX, L. BOUGANIM, P. PUCHERAL, R. THION. *Concilier Ubiquité et Sécurité des Données Médicales*, in "Cahier du CRID", 2009, to appear.

[15] F. DANG-NGOC. *CSXA: Sécurisation du Contrôle d'Accès pour les Documents XML*, in "Revue Technique et Science Informatiques (TSI)", 2009, numéro spécial des Prix de Thèse ASTI et SPECIF, to appear.

International Peer-Reviewed Conference/Proceedings

[16] 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 DK IS .

[17] R. DEWRI, I. RAY, I. RAY, D. WHITLEY. *POkA: Identifying Pareto-Optimal k-Anonymous Points in a Domain Hierarchy Lattice*, in "Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM 2009), Hong Kong, China", November 2009 US .

[18] M. TOAHCHOODEE, X. XIE, I. RAY. *Towards Trustworthy Delegation in Role-Based Access Control Model*, in "Proceedings of the 12th International Information Security Conference (ISC 2009), Pisa, Italy", September 2009 US .

[19] 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 CN .

[20] H. VAN HEERDE, M. FOKKINGA, N. ANCIAUX. *A Framework to Balance Privacy and Data Usability Using Data Degradation*, in "Proc. of the IEEE International Conference on Computational Science and Engineering, Los Alamitos, CA, USA", 2009 NL .

National Peer-Reviewed Conference/Proceedings

[21] L. BOUGANIM, B. JÓNSSON, P. BONNET. *uFLIP: Un banc d'essai pour l'étude des performances de composants Flash*, in "25èmes journées Bases de Données Avancées (BDA), Namur, Belgium", October 2009 DK IS .

Workshops without Proceedings

[22] Y. GUO, S. JACOB. *Confidentialité et intégrité des grandes bases de données*, in "Journées Codage et Cryptographie, Fréjus", October 2009.

Scientific Books (or Scientific Book chapters)

[23] T. ALLARD, N. ANCIAUX, L. BOUGANIM, P. PUCHERAL, R. THION. *Trustworthiness of Pervasive Healthcare Folders*, in "Pervasive and Smart Technologies for Healthcare: Ubiquitous Methodologies and Tools", A. CORONATO, G. DE PIETRO (editors), Information Science Reference, 2009, to appear.

[24] N. ANCIAUX, L. BOUGANIM, P. PUCHERAL. *A Hardware Approach for Trusted Access and Usage Control*, in "Handbook of research on Secure Multimedia Distribution", S. LIAN, Y. ZHANG (editors), Information Science Reference, 2009.

[25] L. BOUGANIM. *Data Skew*, in "Encyclopedia of Database Systems", L. LIU, T. ÖZSU (editors), Springer, 2009.

[26] L. BOUGANIM. *Query Load Balancing in Parallel Database Systems*, in "Encyclopedia of Database Systems", L. LIU, T. ÖZSU (editors), Springer, 2009.

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

Other Publications

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

[29] 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.

References in notes

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

[31] N. ANCIAUX, M. BENZINE, L. BOUGANIM, K. JACQUEMIN, P. PUCHERAL, S. YIN. *Restoring the Patient Control over her Medical History*, in "Proc. of the 21th IEEE Int. Symposium on Computer-Based Medical Systems (IEEE CBMS), Jyväskylä, Finland", June 2008.

[32] N. ANCIAUX, M. BERTHELOT, L. BRACONNIER, L. BOUGANIM, M. DE LA BLACHE, G. GARDARIN, P. KESMARSZKY, S. LARTIGUE, J.-F. NAVARRE, P. PUCHERAL, J.-J. VANDEWALLE, K. ZEITOUNI. *A Tamper-Resistant and Portable Healthcare Folder*, in "International Journal of Telemedicine and Applications (IJTA)", vol. 2008, July 2008.

[33] 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.

[34] N. ANCIAUX, L. BOUGANIM, H. VAN HEERDE, P. PUCHERAL, P. M. G. APERS. *Dégradation progressive et irréversible des données*, in "24èmes journées Bases de Données Avancées (BDA)", October 2008.

[35] N. ANCIAUX, L. BOUGANIM, H. VAN HEERDE, P. PUCHERAL, P. M. G. APERS. *Data Degradation: Making Private Data Less Sensitive Over Time*, in "Proc. of the 17th ACM International Conference on Information and Knowledge Management (ACM CIKM), Napa Valley, USA", October 2008, short paper.

[36] N. ANCIAUX, L. BOUGANIM, H. VAN HEERDE, P. PUCHERAL, P. M. G. APERS. *InstantDB : Enforcing Timely Degradation of Sensitive Data*, in "Proc. of the 24th International Conference on Data Engineering (ICDE), Cancun, Mexico", April 2008, short paper.

[37] A. BARAANI, J. PIEPRZYK, R. SAFAVI-NAINI. *Security In Databases: A Survey Study*, 1996, http://citeseer.ist.psu.edu/baraani-dastjerdi96security.html.

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

[39] COMPUTER SECURITY INSTITUTE. *CSI/FBI Computer Crime and Security Survey*, 2009, http://www.gocsi.com/2009survey/.

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

[41] 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.

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

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

[44] 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.

[45] P. PUCHERAL, S. YIN. *System and Method of Managing Indexation of Flash Memory*, May 2007, Dépôt par Gemalto et INRIA du brevet européen n° 07290567.2.

[46] 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 session.

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