INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team SMIS

Secured and Mobile Information Systems

Paris - Rocquencourt

THEME SYM

Table of contents

1. Team

Research Scientist

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

Faculty Member

Philippe Pucheral [PR1 - UVSQ, HdR]

Technical Staff

Kévin Jacquemin [ENSIMAG, IA, up to September 15th]
Maggy El Kholy [ENSIMAG, IA, from September 1st]

PhD Student

Mehdi Benzine [UVSQ, MESR]
Bhaskar Biswas [Ecole Polytechnique, CORDI INRIA (joint PhD with EPI SECRET)]
Harold van Heerde [University of Twente (joint PhD with P. Apers team)]
Shaoyi Yin [UVSQ, CORDI]

Administrative Assistant

Elisabeth Baque [AI - INRIA]

2. Overall Objectives

2.1. Overall Objectives

Keywords: *Database management systems*, *database security (data confidentiality and privacy)*, *mobile and embedded databases*.

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 devices 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 applications 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

Keywords: *embedded databases*, *query processing*, *secured computing platforms*, *storage and indexing models*, *transaction 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 [37] 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) [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 [48], 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

Keywords: *access control models*, *data confidentiality and privacy*, *encrypted databases*, *secure operating environments*.

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 [34]. 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 [36] but remains inoperative against insider attacks. Several attempts have been made to strengthen server-based security with database encryption [45] [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

Keywords: *ambient intelligence*, *healthcare*, *secure data dissemination*, *web-hosting databases*.

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 cooperation with industry. The Gold Award we received at the SIMagine2005 international software contest illustrates well this strategy (see Section 5.3). This prototyping task is however difficult because it requires specialized hardware platforms (e.g., smart cards), themselves sometimes at an early stage of development (see Section 7.1.1).

The following subsections present a succession of prototypes we developed on specialized hardware. These prototypes consider different application domains, address different challenges, and exploit different technical solutions, generally linked to hardware characteristics, but all capitalize on our growing experience in this field since year 2000.

5.2. PicoDBMS

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

PicoDBMS is a smart card full-fledged DBMS aiming at managing shared secured portable folders. A first prototype written in JavaCard has been demonstrated at the VLDB01 conference [35]. It showed the feasibility of the approach but exhibited disastrous performance. Since then, a second prototype has been written in C and optimized partly with the help of Axalto (their smart card OS has been modified to better support data intensive on-board applications). This prototype is now running on an experimental smart card platform and exhibits two orders of magnitude better performance than its JavaCard counterpart. A cycle-accurate hardware simulator allowed us to predict the PicoDBMS performance on future smart card platforms. Extensive experimentations have been conducted recently on this prototype and have served to validate a new benchmark dedicated to secure chip DBMSs [13]. More generally, PicoDBMS prototype has been a major vehicle to validate our results, to develop important skills in terms of design rules for embedded database components and to set up a long term industrial cooperation with Axalto. Link: http://www-smis.inria.fr/Eprototype_PicoDBMS.html.

5.3. Chip-Secured XML Access

Participants: Luc Bouganim [correspondent], Philippe Pucheral.

Chip-Secured XML Access (C-SXA) is an XML-based access rights controller embedded in a smart card. C-SXA evaluates users privileges on a queried or streaming XML encrypted document and delivers the authorized subset of this document. Compared to existing methods, C-SXA supports fine grain and dynamic access control policies by separating access control issues from encryption. Application domains cover the exchange of confidential data among a community of users (e.g., collaborative work) as well as selective data dissemination. A first C-SXA prototype has been developed on a hardware cycle-accurate simulator to assess the medium-term viability of the approach in terms of performance [7]. Then, a C-SXA engine has been developed in JavaCard on a real smart card platform and has been demonstrated at the SIGMOD05 conference [38]. An application scenario dealing with selective disseminations of multimedia content has been developed on top of this engine (MobiDiQ) and has been rewarded by the Gold award of the SIMagine2005 international software contest.

Link: http://www-smis.inria.fr/Eprototype_C-SXA.html.

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 (See Section 6.3). 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 [47].

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

5.5. PlugDB engine

Participants: Nicolas Anciaux [correspondent], Mehdi Benzine, Luc Bouganim, Kévin Jacquemin, Maggy El Kholy, Philippe Pucheral, Shaoyi Yin.

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, an 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 [31] and its Flash-based indexing system has been patented by INRIA and Gemalto [46].

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

6. New Results

6.1. Embedded data management

Keywords: *benchmarks*, *co-design*, *query processing*, *storage and indexing models for embedded databases*.

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

A first achievement in this field is the definition of a new benchmark, called DiSC, dedicated to secure chip DBMSs. Taking advantage of our experience in designing and developing PicoDBMS (see Section 5.2), the DiSC benchmark has been conceived to: (1) compare the relative performance of candidate storage and indexing data structures, (2) predict the limits of on-chip applications, and (3) provide co-design hints to help calibrating the resources of a future secure chip to meet the requirements of on-chip data intensive applications. This work, which concludes the PicoDBMS study, has been published in [13].

A second achievement concerns the definition of storage and indexing models dedicated to electronic stable storage technologies, and more precisely to NAND-Flash, the most popular persistent data storage medium for mobile and embedded devices. As on-board storage capacity increases, the need for efficient indexing techniques arises. Such techniques are very challenging to design due to a combination of NAND Flash constraints (for example the block-erase-before-page-rewrite constraint and limited number of erase cycles) and embedded system constraints (for example tiny RAM and resource consumption predictability). Previous work adapted traditional indexing methods to cope with Flash constraints by deferring index updates using a log and batching them to decrease the number of rewrite operations in Flash memory. However, these methods were not designed with embedded system constraints in mind and do not address them. In this work, we propose a new alternative for indexing Flash-resident data that specifically addresses the embedded context. This approach, called PBFilter, organizes the index structure in a purely sequential way. Key lookups are sped up thanks to two principles called Summarization and Partitioning. We instantiated these principles with data structures and algorithms based on Bloom Filters and shown the effectiveness of this approach through a comprehensive performance study. PBFilter has been patented by Gemalto and INRIA [46] and published in [23][26][24].

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. More specifically, we established what kind of IOs should be favored (or avoided) when designing algorithms and architectures for flash-based systems. We focused on flash IO patterns, that capture relevant distribution of IOs in time and space, and our goal was to quantify their

performance. We defined uFLIP, a benchmark for measuring the response time of flash IO patterns. We also set up a benchmarking methodology which takes into account the particular characteristics of flash devices. This work, published in [21] has been done in cooperation with the University of Copenhagen and the Reykjavík University.

6.2. Data confidentiality and privacy

Keywords: *access control models*, *data confidentiality and privacy*, *data retention*.

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. In this direction, a new XML access control model was defined to express authorization rules over ancestor and sibling relationships in an XML document [8]. This work was based on the assumption that relationships may reveal information as sensitive as the one carried out by the document nodes themselves. Now, the team focuses on the protection of personal data (also called micro-data) with a particular interest for the expression of usage control policies. Usage control goes beyond access control by integrating concepts like users consent, purpose declaration, limited collection, limited retention, etc. We started to study how user consent could be more easily expressed and better enforced in the context of Electronic Health Record (EHR) systems thanks to portable and secure healthcare folders [12], [17]. Conversely to (and complementary with) our initial approach, the challenge we are addressing now is to define as simple models as possible to help a user calibrating a predefined access control policy to her specific situation and sensitivity.

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, 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 in databases. In this study, we propose a 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. 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 degradation model is flexible enough to remain in line with the applications purposes, and thus favors data utility. Such a data degradation model strongly impacts database storage and indexing structures, logging and locking mechanisms, opening up several research perspectives. Preliminary results about this model, obtained in collaboration with the University of Twente, can be found in [19], [18], [25].

6.3. Tamper-resistant data management

Keywords: *access control models*, *data confidentiality*, *query processing*, *secure computing platforms*.

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

Our most recent study on tamper-resistant data management focuses on the management of database mixing public and sensitive data. People talk about privacy, but give it up very easily, especially when faced with complex security procedures that offer only conditional guarantees. This implies that for peoples sensitive data to be protected, the cost to protect it must require little physical effort and must perform well. We proposed a system whereby people carry hidden sensitive data on a tamper-resistant USB key and they plug that key into a personal computer when they need to link their hidden data with visible public data, all with the assurance that no hidden data will ever go out in the open. The principal novelties follow directly from the challenges of

implementing this mode of operation: (1) how to declare which data should be visible and hidden simply and how to query it, (2) how to index the data, and (3) which query processing strategies to use to link public and private data hosted on extremely unequal devices (standard computer and smart USB key). Our philosophy is to make the users life as easy as possible while efficiently supporting SQL queries on arbitrarily large databases. Efficiency considerations on the small RAM Secure USB key lead us to the design of generalized join indexes, Bloom filters for approximate filtering, the postponement of selections until after joins in certain cases, and algorithms that reflect the differences in read/write performance in the Secure USB key. This study has led to a first publication in SIGMOD 2007 [2]. Since then, a prototype has been implemented and demonstrated at VLDB and BDA conferences (5.4). This initial work has recently been extended to tackle the case of aggregate computations [11] and a complete performance study on a cycle-accurate hardware emulator provided by Gemalto has stated. More general work on tamper-resistant data management has been published in [27].

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 examples for their future smart card platforms as well as technical feedbacks that help them adapting their 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), 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 100 volunteer patients and 25 practitioners in the Yvelines district.

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 international cooperations with the following persons/teams (all with co-authored papers):

- Dennis Shasha (Professor at the University of New-York, USA): collaboration on tamper-resistant data management issues (see details in Section 6.3). 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 6.1). 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 6.2). 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 6.1). Luc Bouganim did a 5 months stay in this team in 2008.

9. Dissemination

9.1. Scientific activity and coordination

9.1.1. Collective responsibilities within INRIA

Philippe Pucheral has been member of the Bureau du Comité des Projets (Project Committee) of INRIA Rocquencourt from September 2004 up to September 2008. He was in charge of the Mission Formation par la Recherche (Training through Research) at Rocquencourt: relationships with the Parisian universities, funding of summer schools, annual campaign for doctoral grants, etc. Luc Bouganim 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. He is also co-responsible for the organization of the monthly scientific seminars ("Le modèle et l'Algorithme") at INRIA Rocquencourt. 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 2008, 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.
  - Member of the Scientific Board of the ARPEGE program (from Embedded Sytems to Large Infrastructure) launched by the French National Research Agency (ANR).
  - PC member of MDM'08, DEXA'08, DS2ME'08, UbiMob'08, BDA'08.
  - Member of the BDA Board (Bases de Données Avancées).
  - Member of the commission de spécialistes 27th section (recruiting committee) of Ecole Normale Supérieure (ENS Cachan antenne de Bretagne).
  - Referee for the HDR (Habilitation à Diriger des Recherches) of N. Cuppens (ENST-B) and E. Pacitti (Univ. Nantes) and of the PhD thesis of R. Thion (INSA Lyon).
- Luc Bouganim
  - PC member of CIKM'08, MDM'08, SIGMOD'08 Demo
- Nicolas Anciaux
  - PC member of ICDE08, SIGMOD08, EDBT08 demo.
  - Member of the Editorial Board of TSI Journal (Technique et Science Informatiques).

9.1.3. Invited talks

- Philippe Pucheral
  - SGBD sur puce : Challenges et applications, 4èmes journées Francophones Mobilité et Ubiquité (UbiMob), Keynote Speech, May 2008.
  - Sécurité des bases de données : dossiers médicaux personnels, Colloque PRIAM (Les technologies au service du droit), November 2008.
- Nicolas Anciaux
  - Les Yvelines département leader en Ile-de-France pour la R&D, Table ronde, C. Beley, F. Becquet, L. Montagnier, Y. Haentjens, J.-P. Arragon, N. Anciaux, Les Rendez-vous Carnot / R&D Network 2008, March 2008.

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 2008 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 and University of Copenhagen).
- N. Anciaux: DBMS internal mechanisms, database Technology (90h/y, given at UVSQ and ENSTA).
- M. Benzine: Java programming, business intelligence, database concepts (110 h 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] L. BOUGANIM, F. FABRET, F. PORTO, P. VALDURIEZ. *Processing Queries with Expensive Functions and Large Objects in Distributed Mediator Systems*, in "Proc. of the 17th Int. Conf. on Data Engineering (ICDE)", 2001.

[6] 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.

[7] 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.

[8] 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.

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

[10] P. PUCHERAL, ET AL. *Mobile Databases : a Selection of Open Issues and Research Directions*, in "ACM Sigmod Record", collective report written under the supervision of P. Pucheral, vol. 33, n^{O} 2, 2004.

Year Publications

Articles in International Peer-Reviewed Journal

[11] N. ANCIAUX, M. BENZINE, L. BOUGANIM, P. PUCHERAL, D. SHASHA. *Revelation on Demand*, in "Distributed and Parallel Database Journal (DAPD)", to appear, 2008.

[12] 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.

[13] 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.

[14] 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.

Articles in National Peer-Reviewed Journal

[15] N. ANCIAUX, L. BOUGANIM, P. PUCHERAL. *SGBD embarqué dans une puce : retour d'expérience*, in "Revue Technique et Science Informatiques (TSI)", vol. 27, n^{o} 1, April 2008.

[16] F. DANG-NGOC. *CSXA: Sécurisation du Contrôle d'Accès pour les Documents XML*, in "Revue Technique et Science Informatiques (TSI)", numéro spécial des Prix de Thèse ASTI et SPECIF, à paraître, 2008.

International Peer-Reviewed Conference/Proceedings

[17] 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.

[18] 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", short paper, October 2008.

[19] 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", short paper, April 2008.

[20] M. BERTHELOT, P. KESMARSZKY, P. PUCHERAL, J.-J. VANDEWALLE, K. ZEITOUNI. *Patient Medical Records in Secure Portable Tokens*, in "Proc. of the 9th International e-smart Conference, Sophia Antipolis, France", September 2008.

[21] 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.

[22] P. SERDYUKOV, L. FENG, A. VAN BUNNINGEN., S. EVERS, H. VAN HEERDE, P. M. G. APERS, M. FOKKINGA, D. HIEMSTRA. *The right expert at the right time and place: From expertise identification*

to expertise selection, in "Proc. of the 7th International Conference on Practical Aspects of Knowledge Management (PAKM2008), Yokohama, Japan", November 2008.

[23] S. YIN, P. PUCHERAL, X. MENG. *PBFilter: Indexing Flash-Resident Data through Partitioned Summaries*, in "Proc. of the 17th ACM International Conference on Information and Knowledge Management (ACM CIKM), Napa Valley, USA", short paper, October 2008.

[24] 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.

National Peer-Reviewed Conference/Proceedings

[25] 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.

[26] S. YIN, P. PUCHERAL, X. MENG. *PBFilter: Indexer les données résidant en Flash par résumés partitionnés*, in "24èmes journées Bases de Données Avancées (BDA)", October 2008.

Scientific Books (or Scientific Book chapters)

[27] N. ANCIAUX, L. BOUGANIM, P. PUCHERAL. *A Hardware Approach for Trusted Access and Usage Control*, in "Handbook of research on Secure Multimedia Distribution", IGI Global, 2008.

[28] L. BOUGANIM. *Data Skew*, in "Encyclopedia of Database Systems", L. LIU, T. OZSU (editors), Springer, 2008.

[29] L. BOUGANIM. *Query Load Balancing in Parallel Database Systems*, in "Encyclopedia of Database Systems", L. LIU, T. OZSU (editors), Springer, 2008.

Scientific Popularization

[30] P. PUCHERAL. *Bases de données : l'individu suivi à la trace*, Interview du dossier L'identité à l'ère numérique - Le Journal du CNRS - à paraître, October 2008.

Other Publications

[31] N. ANCIAUX, L. BOUGANIM, P. PUCHERAL, M. BENZINE, K. JACQUEMIN, S. YIN, D. SHASHA, C. SALPERWYCK. *Logiciel PlugDB-engine version 1, enregistré à l'Agence pour la Protection des Programmes (APP) sous le numéro IDDN.FR.001.280004.000.S.P.2008.0000.10000 en date du 8 juillet 2008*, July 2008.

[32] P. L. MONTAGNIER, Y. HAENTJENS, J.-P. ARRAGON, N. ANCIAUX. *Les Yvelines département leader en Ile-de-France pour la recherche*, Table ronde animée par C. Beley et F. Becquet, Les Rendez-vous Carnot, R. and D. Network, Versailles, France, March 2008.

[33] P. PUCHERAL. *SGBD sur puce : challenges et applications*, Keynote Speech, 4èmes journées Francophones Mobilité et Ubiquité (UbiMob), May 2008.

References in notes

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

[35] 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.

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

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

[38] L. BOUGANIM, C. CREMARENCO, F. DANG-NGOC, N. DIEU, P. PUCHERAL. *Safe Data Sharing and Data Dissemination on Smart Devices*, in "Proc. of the ACM Sigmod Int. Conf. on Management of Data", 2005.

[39] COMPUTER SECURITY INSTITUTE. *CSI/FBI Computer Crime and Security Survey*, 2004, http://www.crime-research.org/news/11.06.2004/423/.

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

[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] ORACLE CORPORATION. *Advanced Security Administrator Guide*, in "Release 10.1", 2003.

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

[47] 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)", Demo session, September 2007.

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