INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Smis

Secured Mediation Systems

Rocquencourt

THEME 3A

Table of contents

1. Team **1**
2. Overall Objectives **1**
3. Scientific Foundations **2**
4. Application Domains **2**
5. Software **3**
5.1. PicoDBMS 3
5.2. C-SDA 3
5.3. DBA Companion 4
6. New Results **4**
6.1. Integration of heterogeneous resources 4
6.1.1. Data source integration 4
6.1.2. Software component integration 5
6.1.3. Data personalization 5
6.2. Mobility and data confidentiality 6
6.2.1. Embedded database components 6
6.2.2. Data confidentiality 6
7. Contracts and Grants with Industry **7**
7.1. National grants 7
7.1.1. ANVAR 7
7.1.2. Axalto (Schlumberger) 7
7.2. European grants 7
7.2.1. Ankh-Morpork (A-213 MEDEA+) 7
8. Other Grants and Activities **8**
8.1. National grants 8
8.1.1. ACI CASC 8
8.1.2. ACI Padoue 8
8.1.3. ACI MediaGrid 8
8.1.4. ACI DataGraal 8
8.1.5. CNRS Specific Actions 9
8.2. International grants 9
8.2.1. SPHIDERS Project 9
8.3. International and national cooperations 9
9. Dissemination **10**
9.1. Teaching activity 10
9.2. Scientific activity and coordination 10
10. Bibliography **12**

1. Team

Head of project-team
Philippe Pucheral [PR - UVSQ (Civil servant, on partial secondment)]

Vice-head of project team
Luc Bouganim [CR1 - INRIA]

Staff members
Mokrane Bouzeghoub [PR - UVSQ]
Béatrice Finance [MC - UVSQ]
Daniéla Grigori [MC - UVSQ]
Zoubida Kedad [MC - UVSQ (Civil servant, on partial secondment)]
Stéphane Lopes [MC - UVSQ]

Administrative assistant
Elisabeth Baque [AI - INRIA]

Ph. D. students
Nicolas Anciaux [4th year, UVSQ]
François Dang Ngoc [2nd year, UVSQ - INRIA]
Sophie Giraud [1st year, UVSQ]
Dimitre Kostadinov [1st year, UVSQ]
Aurélian Lavric [3rd year, Paris 6 - INRIA]
Saïda Medjdoub [3rd year, UVSQ]
Véronika Peralta [2nd year, Montévidéo - UVSQ]
Cristian-Augustin Saita [3rd year, UVSQ - INRIA]
Assia Soukane [4th year, UVSQ]
Xiaohui Xue [1st year, UVSQ]

Post-doctoral fellow
Chun Yuan [Univ. Tsinghua]

Project technical staff
Jean-Pierre Matsumoto [DEA Paris 6]
Nicolas Dieu [ENSEEIHT]

Graduate student intern
Lilan Wu [DEA UVSQ]
Cosmin Cremarenco [Ecole Polytechnique de Bucarest]
Adrian Dragusanu [Ecole Polytechnique de Bucarest]

Previous members
François Llirbat [CR - INRIA]
Eric Simon [DR - INRIA]
Bernadette Farias-Loscio [U. Pernambucco, Brazil]
Eric Simon and François Llirbat were members of the SMIS team from January 2003 to May 2003. During that period, they were involved in the creation of Medience, a startup aiming at organizing the technological transfer of Le Select. Le Select is a mediation system built in the Caravel INRIA project headed by Eric Simon until December 2002. Eric Simon and François Llirbat are now members of the Medience's staff. Link:http://www.medience.frhttp://www.medience.fr

2. Overall Objectives

Key words: *heterogeneous data and program integration*, *query execution and optimization*, *data confidentiality and privacy*, *mobile and embedded databases*, *data personalization and data quality*.

SMIS (Secured Mediation Systems) is a joint team between INRIA Rocquencourt and the PRiSM Laboratory (UMR CNRS 8144, University of Versailles). This team has been created in January 2003 and pursues the scientific objectives mentioned below.

The SMIS team aims at defining methods and tools that help building a uniform, consistent and secured view of a large amount of heterogeneous resources distributed over the Internet (relational, XML and multimedia data as well as application programs and Web services). This view must be able to adapt itself to different user's profiles. It must also reflect the continuous evolution of the number, content, quality and availability of the data sources participating in the mediation. Regarding program integration, data-driven workflow specification tools and execution models have to be defined.

Enforcing the confidentiality of the data stored in these resources is today a tremendous challenge. Indeed, the threats on data confidentiality and privacy increase with the permanent connection of these resources to the Internet, the resort to untrusted Web-hosting companies to store and manage them and the democratization of ubiquitous computing to access them. The SMIS project addresses these issues by devising chip-secured data access models and architectures. Embedding software components into secured chips (e.g., smart cards, tokens) gives new opportunities to control the querying, updating and sharing of encrypted data stored on untrusted repositories. More generally, SMIS contributes to the definition of database components embedded into a growing variety of lightweight computing devices used in ambient intelligence (chips for home networks, transportation, telephony, healthcare, etc.).

3. Scientific Foundations

Key words: *databases*, *integration of distributed data sources*, *large scale query processing*, *database mobility and security*.

The proliferation and the diversity of the resources made available through the Internet have strongly impacted the architecture and the usage of existing information systems. The information is today distributed across a large amount of autonomous and heterogeneous sources. By allowing a large scale sharing of these resources, the Internet deeply modifies: (i) the access patterns to the data (very high number of clients and resources, notification and personalization of the data relevant to a given user, ubiquitous access to the data thanks to a variety of mobile devices), (ii) the exploitation of the data (data warehouse, data mining, data grid) and (iii) the rules to share and disseminate the data (confidentiality, intellectual property). To tackle this strong mutation, the database community has opened new research fields during the last decade, among others: XML data management, data and program integration, large scale query processing, query optimization and personalization, multimedia data management, data warehousing and data mining, mobile databases, database security.

In the last years, the SMIS members have been strongly engaged in two of these research fields, namely data integration and mobile databases. Capitalizing on this work, the SMIS project is now addressing specific facets of three related areas: program integration, information personalization and data confidentiality. The expected contributions of the SMIS project in these areas will take the form of schema and data integration techniques, query execution and optimization techniques, storage and indexation models, data-centric algorithms and performance evaluation and measurements.

4. Application Domains

Key words: *e-commerce*, *e-learning*, *decision support systems*, *business intelligence*, *software engineering*, *ambient intelligence*, *home networks*, *sensor networks*, *healthcare*, *transportation*, *telecommunication*, *Web-hosting databases*.

By providing a transparent access to heterogeneous data sources, mediation systems meet different application needs: querying business databases, discovering software components or web services to compose them into new business applications, extracting knowledge from large data sets containing multimedia data or

disseminating information to large communities of users. Data warehousing, data mining and data grids are intermediate technologies which can be based on mediation systems to build high level applications on various domains such as e-commerce, e-learning, decision support systems and business intelligence. The work we are conducting in publish-subscribe models can impact new dissemination-based applications involving timely delivery of information to large sets of subscribers (stock exchange, auctions, small ads, service delivery). All these applications share the same needs of transparency, flexibility and uniform access to heterogeneous distributed data sources as well as scalability and evolution. They also require data personalization techniques to improve the relevance of the delivered data.

Our work on mobility and data confidentiality addresses different application domains. Typically, database components 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 complex portable folders are embedded into smart cards, in telephony applications where personal data (address book, agenda, etc.) are embedded into cellular phones, in sensor networks where sensors log row measurements and perform local computation on them, in home networks where a collection of smart objects gather information about the occupants to provide them a personalized service, and more generally in most applications related to ambient intelligence. Meanwhile, mobility and ambient intelligence introduce new threats on data confidentiality and privacy. Indeed, mobility may lead people to store confidential data on untrusted environments while ambient intelligence makes the gathering of personal data more insidious. Thus, most of the applications mentioned above require new mechanisms to protect the data confidentiality and to better control who is permitted to access these data and for which purpose.

5. Software

5.1. PicoDBMS

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

As smart cards become multi-application and more and more powerful (32 bit CPU, more than 1MB of stable storage soon), the need for database management arises. PicoDBMS is a full-fledged embedded DBMS (storage, indexation, query processing, access rights and transaction control) addressing this issue. The application domain of PicoDBMS is the management of shared secured portable folders (medical folder, user profile, agenda, etc.). Its indexation and query engines implement new strategies required to cope with the smart card hardware constraints [7]. A first prototype written in JavaCard has been demonstrated at the VLDB'01 conference [2]. Since then, a second prototype has been written in C and optimized. At the same time, Schlumberger provided us with an experimental smart card platform and modified part of its smart card OS (cf. section 7.1.2), so that the current prototype exhibits two order of magnitude better performance than its JavaCard counterpart. Experimentations are still on-going on this prototype.

5.2. C-SDA

Participants: François Dang Ngoc [correspondent], Luc Bouganim, Nicolas Dieu, Philippe Pucheral, Lilan Wu.

Several approaches have been proposed to secure databases by means of encryption. Unfortunately, server-based approaches cannot resist to insider attacks while client-based approaches make access control static and coarse grain (cf. section 6.2.2). The C-SDA (Chip-Secured Data Access) architecture allows querying encrypted data while controlling fine grain and dynamic personal privileges [3]. C-SDA is a client-based security component acting as an incorruptible mediator between a client and an encrypted database. This component is embedded into a smart card to prevent any tampering to occur on the client side. The CNRS (Centre National de la Recherche Scientifique) took out a patent on C-SDA [39]. A JavaCard prototype has been developed partly with the support of the French ANVAR agency (Agence Nationale pour la VAlorisation de la Recherche) [33] and has been demonstrated at the VLDB'03 conference [22].

5.3. DBA Companion

Participant: Stéphane Lopes [correspondent, in cooperation with the database team of LIMOS lab, University of Clermond-Ferrant].

This prototype integrates algorithms for schema analysis of relational databases. The main objective of this tool is to bring help to the logical administration (of the schema) of a relational database. From a database instance, the software can extract functional dependencies and inclusion dependencies satisfied by this instance. Several possibilities of visualization of the results are proposed which makes it possible to highlight the problems of the database (normalization, inconsistency, etc.). The prototype has been demonstrated at BDA'03 and will be demonstrated at ICDE'04 [25].

6. New Results

6.1. Integration of heterogeneous resources

Mediation systems are software infrastructures which allow transparent access to heterogeneous data sources. The main challenge of mediation systems is the integration of these data for different application goals: querying business databases, discovering software components or web services to compose them into new business applications, or extracting knowledge from images for example. Our research concern covers the definition of global schemas used by mediators, the detection of semantic mismatch between data sources, the definition of data transformations while taking into account scalability and evolution, metadata and quality management, as well as data personalization.

6.1.1. Data source integration

Participants: Mokrane Bouzeghoub, Bernadette Farias-Loscio, Zoubida Kedad, Aurélian Lavric, Stephane Lopes, Assia Soukane, Xiahui Xue.

Key words: *data integration*, *data cleaning and transformation*, *mediation query generation*, *schema evolution*, *constraints acquisition*.

The goal of data integration consists in providing a uniform view of data sources, called mediation schema, and defining a set of queries, called mediation queries, which define instances of the mediation entities. Our main research goal focuses on the generation of mediation queries in the context of heterogeneous and evolving data sources. More precisely, we are studying: (i) the evolution of data sources, (ii) the detection of semantic mismatch between data sources attributes and mediation schema attributes as well as selection of relevant transformation rules, (iii) the extraction of source description knowledge (functional and inclusion dependencies) and (iv) the definition of new data integration operators. In mediation systems, local data sources are often autonomous and may evolve independently of the mediators using them. The challenge is to maintain the mediation queries consistent with the source evolution. Our approach to this problem [23] is to isolate the subset of mediation queries that are impacted by a source change and to automatically redefine these queries by applying only the necessary changes.

The data cleaning and transformation problem (i.e., solving inconsistencies between the mediation schema and the data source schemas related to integrity constraints, object identification, data types and formats, duplicates, etc) must be carefully tackled as it impacts the execution and the semantics of the mediation queries. We have defined a new mediation query generation algorithm [5] [36] which checks mismatches between constraints and extended types and selects the appropriate queries as well as data transformations within a predefined library of procedures. Constraints such as functional dependencies can be used to select relevant mediation queries and to improve the data cleaning process. Unfortunately, several data sources may not provide these constraints. Thus, we proposed to apply data mining techniques to extract functional and inclusion dependencies that hold in a data source [12]. A prototype, called DBA companion [25], has been developed and provides several data mining algorithms devoted to the understanding of data semantics.

6.1.2. Software component integration

Participants: Mokrane Bouzeghoub, Daniéla Grigori, François Llirbat, Jean-Pierre Matsumoto.

Key words: *workflow models and languages*, *scientific workflows*, *workflow analysis*, *software components*.

As for data source integration, the integration of software components (e.g. web services) can be perceived at two levels: the specification level, similar to the schema level, and the execution level, similar to the instance level. The specification level introduces software engineering problems: how to publish, discover and compose services in order to form a new application satisfying functional and qualitative requirements of users. The execution level concerns the execution of applications in a distributed environment, under qualitative constraints such as response time, data coherence and security. We made contribution at the two levels. If current workflow models and systems apply efficiently to administrative and production business processes, they show their limits when one wants to model the subtlety of cooperative interactions as they occur in more interactive or creative processes, typically co-design and Web engineering processes. We have defined a flexible workflow technology that is based on two complementing contributions [14]: anticipation that allows succeeding activities to cooperate, and coo-transaction that allows parallel activities to exchange intermediate results.

Proper analysis of workflow management systems (WFMS) execution logs can yield important knowledge and help designers improve the quality of their processes. We have defined a set of integrated tools that supports in managing process execution by providing several features, such as analysis, prediction, monitoring, control, and optimization [13]. The next step in this research direction is to analyze process past executions in order to rank services that offer the same functionalities, with different quality criteria, for a given user.

Scientific workflows were introduced as a practical mean to specify scientific experiments. However, current solutions impose a centralized management of data and experiments around the same database and then badly adapt to the context of large-scale distribution. We have defined a new data centric workflow model, which allows the definition of distributed applications and the optimization of their execution in the context of data intensive usage. This work is conducted in the ACI Padoue (cf. section 8.1.2).

6.1.3. Data personalization

Participants: Mokrane Bouzeghoub, Dimitre Kostadinov, Véronika Peralta, François Llirbat, Cristian-Augustin Saita.

Key words: *user profiles*, *user preference languages*, *approximate queries*, *multidimensional indexing*.

As the volume and the diversity of the data increase, query systems (Web search engines, DBMS, mediation systems, etc.) deliver massive results in response to user requests, thus generating an informational overload. The definition of user (or community) profiles addresses this problem by providing a personalized information access model. The approach we follow lies in a global study about information personalization and its impact on database systems (architecture, query language, query evaluation, etc.). We are defining a generic profile metamodel based on a set of dimensions such as the domain of user interest, data customization, security, ordering of selection criteria within queries, expected quality for data, etc. We aim at achieving three complementary objectives: (i) the incremental building of a specific user profile (metamodel instantiation), (ii) the use of profiles in queries and (iii) the relationships between profiles and data source quality.

Another way for the user to get the most relevant information is provided by publish-subscribe notification systems. We are studying a general-purpose notification mechanism where the subscriptions and the events specify range intervals for their attributes. In such context, subscriptions and events can be represented as multidimensional extended objects. The matching subscriptions are retrieved based on intersection, containment, or enclosure queries (spatial queries). We have designed a multidimensional indexing method relying on a new cost-based adaptive clustering method that can cope with a large number of subscriptions (millions) and a large number of significant attributes (tens of dimensions) [28].

6.2. Mobility and data confidentiality

Ubiquitous computing and ambient intelligence introduce the need for embedding data and evaluating database queries in a growing variety of ultra-light computing devices (PDA, cellular phones, smart cards, sensors, chips integrated in home networks). The first objective of our research is thus to design embedded database components that can match highly constrained hardware resources.

Meanwhile, ubiquitous computing and ambient intelligence increase the threats on data confidentiality. Typically, which level of trust can be put on a Database Service Provider (DSP) hosting the private data of a mobile user on the Internet? In the same way, how a human being can control the increasing amount of sensitive information automatically collected about himself by many smart objects? The second objective of this research is therefore to devise new architectures that preserve data confidentiality, by combining data encryption with security software embedded in secured chips.

6.2.1. Embedded database components

Participants: Nicolas Anciaux, Luc Bouganim, Sophie Giraud, Philippe Pucheral.

Key words: *mobile database*, *embedded database*, *storage and indexation models*, *query evaluation*, *benchmarks*, *system on chip*, *co-design*.

Personal folders on chip, networks of sensors and data hosted by autonomous mobile computers are different illustrations of the need for evaluating queries confined in hardware constrained computing devices [17][16]. Preliminary studies led us to design and validate a full-fledged DBMS, called PicoDBMS, embedded in an advanced smart card platform (cf. section 5.1). Capitalizing on this work, new research actions have been undertaken [11]: (i) to better capture the impact of each device's hardware constraint on database techniques, (ii) to propose new techniques allowing to build ad-hoc embedded database components and (iii) to set up co-design rules helping to calibrate the hardware resources of future devices in order to match specific application's requirements.

In a recent paper [21], we gave a thorough analysis of the RAM consumption problem. First, we proposed a query execution model that reaches a lower bound in terms of RAM consumption. Second, we devised a new form of optimization, called iteration filter, that drastically reduces the prohibitive cost incurred by the preceding model, without hurting the RAM lower bound. Third, we analyzed how the preceding techniques can benefit from an incremental growth of RAM. The effectiveness of the approach has been measured through a performance evaluation. This work provides preliminary guidelines that can be used either to calibrate the RAM resource of a hardware platform according to data centric application's requirements or to adapt data centric applications to an existing platform.

The next steps in this research action are to analyze the energy consumption problem and to capture the impact of new stable storage technologies (FeRAM, MEMS) on database techniques.

6.2.2. Data confidentiality

Participants: Luc Bouganim, François Dang Ngoc, Béatrice Finance, Saïda Medjdoub, Philippe Pucheral, Lilan Wu, Chun Yuan.

Key words: *data confidentiality and privacy*, *database encryption*, *querying encrypted data*, *smart cards and secured computing platforms*, *access right management*.

While encryption has been used successfully for years to secure communications, database encryption introduces new theoretical and practical issues [18]: how to execute efficiently database queries over encrypted data, how to conciliate declarative (i.e., predicate based) access rights with encryption, how to distribute encryption keys between users sharing part of the database, how to take advantage of secured computing devices? This research action tries to provide some answers to these questions by devising chip-secured models for querying, updating and sharing encrypted databases.

In a previous work [3], we proposed a solution called C-SDA (Chip-Secured Data Access), which allows querying encrypted data while controlling predicate-based personal privileges. C-SDA is embedded into a smart card to prevent any tampering to occur on the client side. This cooperation of hardware and software

security components allows reestablishing the orthogonality between access right management and data encryption. The CNRS (Centre National de la Recherche Scientifique) took out a patent on C-SDA [39]. A JavaCard prototype has been developed, partly with the support of the French ANVAR agency [33], and has been demonstrated at the VLDB'03 conference [22] (cf. section 5.2). C-SDA has been designed in a relational database context with a pull interaction model in mind. We are currently designing solutions to cope with semi-structured (XML) and multimedia data and with other interaction models (push, P2P, cooperation). Preliminary results have been obtained in the context of video stream encryption and right management [19][30][45][29].Beyond database encryption, a more general issue is the definition and management of access rights over multiple distributed and heterogeneous resources accessible through the Internet. We are planning to address different facets of this issue in the context of the ACI CASC project (cf. section 8.1.1) and of the European Ankh-Morpork project (cf. section 7.2.1).

7. Contracts and Grants with Industry

7.1. National grants

7.1.1. ANVAR

Category: ANVAR innovation program
Duration: October 2001 - February 2003
Partners: e-XMLMedia, PRiSM-SMIS (correspondent : P. Pucheral)
Description: This project aimed at enforcing the security of an XML-EDI (Electronic Data Interchange) suite. To prevent any disclosure and tampering, XML documents are encrypted and signed by the issuer (according to the XMLEncryption and XMLDsig standards). On the receiver side, the document integrity is checked and the document is decrypted according to the access control policy enforced by a smart card.

7.1.2. Axalto (Schlumberger)

Category: bipartite cooperation
Duration: unbounded
Partners: Axalto, INRIA-SMIS (correspondent: P. Pucheral)
Description: The SMIS project has a long lasting cooperation with the smart card subsidiary of Schlumberger (formerly Bull-CP8, then SchlumbergerSema and now Axalto). Axalto provides SMIS with advanced smart card platforms, cycle accurate simulators and a strong technical support. SMIS exploits these tools to conduct experiments and performance measurements on different prototypes (cf. sections 5.1 and 5.2).

7.2. European grants

7.2.1. Ankh-Morpork (A-213 MEDEA+)

Category: MEDEA+ project (Eureka)
Duration: July 2003 - July 2006 (project labeled in may 2003, French funding still pending)
Partners: Philips, Thomson, Schlumberger, STM, Canal+, Esterel, Cryptolog, INRIA-SMIS (correspondent: P. Pucheral)
Description: This project aims at ensuring the safety of data contents (i.e., preventing content piracy and infringement of copyrights) and services associated with a home network towards all its connection links. The involvement of SMIS in this project concerns the preservation of user's privacy and the support of non-lucrative access right models (e.g., parental control, free access for educational purpose, etc.).

8. Other Grants and Activities

8.1. National grants

8.1.1. ACI CASC

Category: ACI Sécurité
Duration: July 2003 - July 2006
Partners: INRIA-SMIS (CASC coordinator: L. Bouganim), ENS Ulm, LRI, ENST-Bretagne, Univ. Pau
Description: The CASC project focuses on access right management and data confidentiality preservation. More precisely, the goal is to tackle three important weaknesses of existing DBMS access control models: the inability to deal with the complexity of distributed and decentralized organizations, the semantics fuzziness of access control policies when applied to semi-structured and hierarchical data, and the vulnerability of a centralized administration of access rights. The CASC project expects significant advances in the following areas: (i) abstraction of the fundamental concepts required in any access right model and formalization of the associated administration procedures, (ii) definition of a powerful and sound access right model for XML documents and (iii) definition of chip-secured data access and administration architectures. Link: http://www-smis.inria.fr/~bouganim/CASC.

8.1.2. ACI Padoue

Category: ACI GRID
Duration: June 2002 - June 2004
Partners: LIP6, INRIA-SMIS (correspondent: J-P. Matsumoto), Cemagref, LICS, UMR 3S, CDS, LIRMM, IRD
Description: The Padoue project (Partage des données pour des utilisations en environnement) aims at building a uniform and integrated view of a collection of distributed and heterogeneous environmental information sources. More precisely, the goals of the project are: (i) to design wrappers providing a better interoperability with both data and program sources, (ii) to design tools to develop scientific workflows on top of existing program resources, (iii) to archive and describe the data and programs resulting from this integration and (iv) to localize these resources on the network. Link: http://www-poleia.lip6.fr/padoue.

8.1.3. ACI MediaGrid

Category: ACI GRID
Duration: January 2003 - December 2004
Partners: LSR-IMAG, PRISM-SMIS (correspondent: Z. Kedad), LaMI
Description: The objective of the MediaGRID project (a mediation framework for a transparent access to biological data sources) is to contribute to the definition of an open mediation framework for the Grid. Mediation systems built from this framework would be able to: (i) support more and more available sources, (ii) consider sources containing weakly structured data, (iii) authorize partial results for queries in case of data sources unavailability, and (iv) support a query evaluator able to adapt itself to the execution environment. Research topics include the generation of mediation queries and the evaluation of queries in a distributed, adaptive and interactive way. Link: http://www-lsr.imag.fr/mediagrid.

8.1.4. ACI DataGraal

Category: ACI GRID
Duration: June 2002 - June 2004
Partners: INRIA-Regal, INRIA&PRiSM-SMIS (correspondent: B. Finance), IRISA, LISI, LIP, LIP6, LSR-IMAG, ID-IMAG, Hewlett-Packard Labs, N2P3, LIRMM
Description: This project brings together researchers from the distributed system community and from the database community as well as people working on GRID applications. The work done in this project takes the form of regular workshops. A summer school, named DRUIDE, is being organized and will be held at Port-aux-rocs in May 2004. Link: www-src.lip6.fr/projets/datagraal.

8.1.5. CNRS Specific Actions

The STIC department from CNRS launched recently specific actions (AS) with the aim to review the state of progress in a set of important research topics and identify major research directions for the French computer science community.

Two of these actions have been coordinated by SMIS members. Philippe Pucheral (SMIS-PRiSM) coordinated the AS "Mobilité/accès aux données" from October 2001 to March 2003. This AS, that focused on mobile databases, led to two collective publications [17][16] and allowed to launch three scientific cooperation's among the partners [38]. Detailed information about this AS can be found at: http://www-inf.int-evry.fr/~defude/ASMobilite. Mokrane Bouzeghoub (SMIS-PRiSM) coordinated the AS "Personnalisation de l'information" from January 2003 to December 2003. This AS, that focused on user profiling and data personalization, led to a collective report [34] and a proposal to follow the research within a CNRS specific team. Detailed information about this AS can be found at:
http://www.prism.uvsq.fr/recherche/themes/sial/cnrs.

In addition, SMIS members participated to the AS "Médiation d'informations via les metadonnées", that focused on the role of meta-data within mediation systems, and to the AS "Model-Driven Architecture", that studied the impact of the OMG-MDA proposal on the software engineering and database research communities.

8.2. International grants

8.2.1. SPHIDERS Project

Category: Franco-Mexican Cooperation (LAFMI)
Duration: 2003 - 2004
Partners: France: LSR-IMAG, PRiSM-SMIS (correspondent: B. Finance), Mexico: LANIA, UDLA/BD, UDLA/XALTAL
Description: The SPIDHERS project (SPatial data Integration from Distributed and HEteRogeneous Sources) aims at characterizing the problem of mediation and data integration in the context of geographical data. The main objectives are to study (1) the role of meta-data and their integration, (2) the transparent access to heterogeneous data sources and (3) the indexation of multidimensional objects. This collaboration is performed within the LAFMI (Laboratoire Franco-Mexicain d'informatique : http://lafmi.imag.fr/).

8.3. International and national cooperations

The SMIS members have developed tight cooperation with the following persons/institutions:

- Rachid Guerraoui (EPFL): collaboration on mobile transaction processing [1].
- Mike Franklin (UC Berkeley): visit at INRIA in July 2003. Collaboration on embedded database systems.
- Alejandro Gutiérrez and Raul Ruggia (INCO, Montevideo): long lasting cooperation on database systems. Veronika Peralta is doing a Ph. D. thesis on data integration co-directed by Raul Ruggia and Mokrane Bouzeghoub. Daniel Calegari will spend 3 months at INRIA (first quarter 2004) thanks to the INRIA-INCO internship program to work on data confidentiality.
- Chun Yuan (Univ. Tsinghua, China): Post-doctoral fellow (august 2003-august 2004). Cooperation on database encryption.
- J.L. Zechinelli and Genoveva Vargas-Solar (Universidad de las Américas, Puebla Mexico): collaboration on mediation systems within the context of the French and Mexican computer sciences laboratory (LAFMI: http://lafmi.imag.fr/).
- Fabio Casati (Intelligent Enterprise Technology Lab, HP Laboratories, Palo Alto) has a continuous cooperation on workflows with Daniela Grigori. He is invited for one month by the project.

- Mohamed Ben Ahmed (RIADI, Tunisia): collaboration on the medical applications of PicoDBMS.

SMIS organizes a seminar twice a month where internal members and visitors are invited to present their work. In 2003, Reind van de Riet (Free University of Amsterdam), Sihem Amer Yahia (AT&T Labs Research), Mike Franklin (UC Berkeley), Zahir Tari (RMIT University, Melbourne) and Vincent Oria (University of Alberta) participated to this seminar.

At the national level, SMIS members cooperate intensively with several French labs and university, either through national projects (see section 8.1) or through bilateral relationships. Stéphane Lopes has strong relationships with the LIMOS database group (University of Clermond-Ferrant) on data-mining techniques. Daniéla Grigori pursues regular cooperation with the ECOO project at INRIA Loraine on workflow management.

9. Dissemination

9.1. Teaching activity

SMIS is a joint team between INRIA and the University of Versailles (UVSQ). Hence, all staff members have an important teaching activity. Below is the list of the main courses given at UVSQ by each staff member:

- P. Pucheral: co-director of the DEA MISI, DBMS architecture (60h).
- L. Bouganim: DBMS architecture, data security, database technology, mediation systems (160 h)
- M. Bouzeghoub: data warehousing and data integration, software engineering, database design (220h).
- B. Finance: database technology, programming languages, mediation systems, distributed object systems (240h).
- D. Grigori: database technology, workflow systems, software engineering (230h).
- Z. Kedad: database technology, software engineering (220h).
- S. Lopes: database tuning, formal specifications, object programming (250h).
- J. P. Matsumoto: database technology, DBMS architecture (80h)
- A. Lavric: database technology, grammatical analysis (60h)
- C.-A. Saita: elements of computer science (45h)
- E. Simon : datawarehouse (50h)

9.2. Scientific activity and coordination

The SMIS members have conducted, or participated to, the following actions in the database community:

- P. Pucheral
  - Coordinator of the CNRS Specific Action "Mobilité/Accès aux données" (cf. sections 7.1 and 7.2).
  - PC member of ACM SIGMOD'04, UbiCom'04, Xsymp'04, BDA'04.
  - PC member of the special issue " Systèmes d'information Pervasifs ", ISI Journal, 2003.
  - Editor of the special issue " Bases de Données Avancées : modèles, systèmes et usages", Technique et Science Informatiques (TSI), 2003.
  - Member of the BDA Board (Bases de Données Avancées) since 2002.

  - Lecturer in DRUIDE'04, summer school on large scale data distribution architectures (DistRibUtIon de Données à grande Echelle).

- M. Bouzeghoub

  - Member of the EDBT Endowment (International Conference on Extending Database Technology) since 1996
  - Member of the ER Steering Committee (International Conference on Conceptual Modeling), from 1999 to 2003
  - Member of the Editorial Board of ISI Journal (Ingénierie des Systèmes d'Information)
  - Member of the CNRS/RTP9 Thematic Network on Distributed Information
  - Coordinator of the CNRS Specific Action on Data Personalization
  - General Chair of the Int. Conf. on Semantics of a Connected World: Semantics for Grid Databases (Paris 2004)
  - PC member of CAiSE'04, ER'03, ISE 2003, DOLAP'03, DMDW'03, DSE'03, DIWeb'04, NLDB'03.

- Luc Bouganim

  - Coordinator of the ACI "Sécurité Informatique" CASC
  - Demonstrations Co-Chair of ACM SIGMOD'04
  - PC member of ACM SIGMOD'03, IDEAS'03, IDEAS'04, BDA'04
  - Member of the Editorial Board of TSI Journal (Technique et Science Informatiques)
  - Co-Organizer of DRUIDE'04, summer school on large scale data distribution architectures (DistRibUtIon de Données à grande Echelle).

- Béatrice Finance

  - PC member of BDA'04 (Journées Bases de Données Avancées).
  - Co-Organizer of DRUIDE'04, summer school on large scale data distribution architectures (DistRibUtIon de Données à grande Echelle).

- Zoubida Kedad

  - PC member of the Int. Workshop On Applications of Natural Language To Databases (NLDB'03, NLDB'04)
  - PC member of the Int. Conf. On Natural Language in Information Systems (NLIS'04)

- Eric Simon

  - Lecturer in LipariSchool (data cleaning), 15th International School for Computer Science Researchers on Algorithmics for Data Mining and Pattern Discovery. July 2003

- Stéphane Lopes

  - PC member of BDA'03 (Journées Bases de Données Avancées).

- Véronika Peralta

  - PC member of the Int. Conf. DEXA'04.

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) », number 3, volume 11, 2002.

[2] 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.

[3] L. BOUGANIM, P. PUCHERAL. *Chip-Secured Data Access : Confidential Data on Untrusted Servers.* in « Proc. of the Int. Conf. on Very Large Data Bases (VLDB) », 2002.

[4] M. BOUZEGHOUB, Z. KEDAD. *Quality in Data Warehousing.* in « book chapter in Information and Database Quality, M.G. Piattini, C. Calero and M. Genero (eds), Kluwer Academic Publisher », 2002.

[5] M. BOUZEGHOUB, Z. KEDAD, A. SOUKANE. *Génération de requêtes de médiation intégrant le nettoyage de données.* in « Revue Ingéniérie des Systèmes d'Information (ISI) », number 3, volume 7, 2002.

[6] M. BOUZEGHOUB, M. LENZERINI. *Special Issue on Data Extraction and Data Cleaning.* in « Information Systems International Journal », number 8, volume 26, 2001.

[7] 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 », number 2-3, volume 10, 2001.

[8] D. THEODORATOS, M. BOUZEGHOUB. *Data Currency Quality Satisfaction in the Design of a Data Warehouse.* in « Cooperative Information System Journal », number 1, volume 10, 2001.

[9] P. VASSILIADIS, M. BOUZEGHOUB, C. QUIX. *Towards Quality Oriented Data Warehouse Usage and Evolution.* in « Information Systems Journal », number 2, volume 25, 2000.

Doctoral dissertations and "Habilitation" theses

[10] B. FARIAS LOSCIO. *Managing Evolution of Mediation Systems.* Ph. D. Thesis, University of Versailles & University of Pernambucco, Brazil, March, 2003.

Articles in referred journals and book chapters

[11] N. ANCIAUX, L. BOUGANIM, P. PUCHERAL. *Database Components on Chip.* in « ERCIM News », volume 54, 2003.

[12] F. DE MARCHI, S. LOPES, J. PETIT, F. TOUMANI. *Understanding existing databases at the logical level : the DBA companion project.* in « ACM SIGMOD Record », number 1, volume 32, 2003.

[13] D. GRIGORI, F. CASATI, M. CASTELLANOS, U. DAYAL, M. SAYAL, M. SHAN. *Business Process Intelligence.* in « Computers in Industry Journal, special issue on workflow mining », 2003, to appear.

[14] D. GRIGORI, F. CHAROY, C. GODART. *COO-flow : a Processes Technology to Support Cooperative Processes.* in « Int. Journal of Software Engineering and Knowledge Engineering (IJSEKE) », 2003, to appear.

[15] I. MANOLESCU, L. BOUGANIM, F. FABRET, E. SIMON. *Interrogation efficace de ressources distribuées dans des systèmes de médiation.* in « Techniques et Sciences Informatiques (TSI) », 2004, to appear.

[16] P. PUCHERAL, E. AL.. *Mobile Databases : a Selection of Open Issues and Research Directions.* in « ACM Sigmod Record, collective report written under the supervision of P. Pucheral », 2004, to appear.

[17] P. PUCHERAL, ET AL.. *Mobilité et bases de données : état de l'art et perspectives.* in « Technique et Science Informatiques (TSI) », number 3&4, volume 22, 2003.

[18] P. PUCHERAL. *Ubiquité et confidentialité des données.* in « book chapter, CNRS Edition », 2004, to appear.

[19] B. B. ZHU, C. YUAN, Y. WANG, S. LI. *Scalable Protection for MPEG-4 Fine Granularity Scalability.* in « IEEE Transaction on Multimedia », to appear.

[20] P. PUCHERAL (EDITOR). *Bases de Données Avancées : modèles, systèmes et usages.* in « Technique et Science Informatiques (TSI) », 2003, to appear.

Publications in Conferences and Workshops

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

[22] L. BOUGANIM, F. D. NGOC, P. PUCHERAL, L. WU. *Chip-Secured Data Access : Reconciling Access Right with Data Encryption.* in « Proc. of the Int. Conf. on Very Large Data Bases (VLDB), demo session », 2003.

[23] M. BOUZEGHOUB, B. F. LOSCIO, Z. KEDAD, C. SALGADO. *Managing the Evolution of Mediation Queries.* in « Proc. of the Int. Conf. on Cooperative Information Systems (Coopis) », 2003.

[24] D. GRIGORI, F. CHAROY, C. GODART. *COO-flow : a Process Technology to Support Cooperative Processes.* in « Proc. of the Int. Conf. on Software Engineering and Knowledge Engineering (SEKE) », 2003.

[25] S. LOPES, F. DE MARCHI, J. PETIT. *DBA Companion : A Tool for Logical Database Tuning (Demonstration).* in « Proc. of the Int. Conf. on Data Engineering (ICDE) », 2004, to appear.

[26] V. PERALTA, A. ILLARZE, R. RUGGIA. *On the Applicability of Rules to Automate Data Warehouse Logical Design.* in « Proc. of the Int. Workshop on Decision Systems Engineering (DSE), colocated with CAiSE », 2003.

[27] V. PERALTA, R. RUGGIA. *Using Design Guidelines to Improve Data Warehouse Logical Design.* in « Proc. of the Int. Workshop on Design and Management of Data Warehouses, colocated with VLDB », 2003.

[28] C.-A. SAITA, F. LLIRBAT. *Clustering Multidimensional Extended Objects to Speed Up Execution of Spatial Queries.* in « Proc. of the Int. Conf. on Extending Database Technology (EDBT) », 2004, to appear.

[29] C. YUAN, B. ZHU, M. SU, S. LI, X. WANG, Y. ZHONG. *Efficient and fully scalable encryption for MPEG-4 FGS.* in « Proc. of IEEE ISCAS », 2003.

[30] C. YUAN, B. ZHU, M. SU, S. LI, X. WANG, Y. ZHONG. *Layered access control for MPEG-4 FGS Video (MA-P7).* in « Proc. of IEEE ICIP », 2003.

[31] MEDIAGRID PROJECT. *A mediation framework for a transparent access to biological data sources.* in « Proc. of the Entity-relationship Conf. (ER), short paper », 2003.

Internal Reports

[32] G. BERNOT, G. BRUNO, C. COLLET, B. FINANCE, Z. KEDAD, D. LAURENT, F. TAHI, G. VARGAS-SOLAR, T. T. VU, X. XUE. *Deliverable MEDIAGRID PL-9, 'Etat de l'art'.* Technical report, 2003, http://www-lsr.imag.fr/mediagrid.

[33] L. BOUGANIM, F. D. NGOC, P. PUCHERAL. *Composant de sécurité C-SXA : spécifications techniques.* Technical report, Final deliverable TR_CSXA_0203, programme d'innovation ANVAR, 2003.

[34] M. BOUZEGHOUB. *Bilan de l'AS CNRS Personnalisation de l'Information.* Technical report, Final deliverable AS98-RTP09, short version published in BDA'03 (Journée Bases de Données Avancées), 2003.

[35] M. BOUZEGHOUB, B. FARIAS-LOSCIO, Z. KEDAD, C. SALGADO. *Design and Evolution of Large Scale Mediation Systems.* Technical report, PRiSM Technical Report 2003/41, Univ. of Versailles, 2003.

[36] M. BOUZEGHOUB, Z. KEDAD, A. SOUKANE. *Improving Mediation Queries Generation Using Metadata.* Technical report, PRiSM Technical Report 2003/42, Univ. of Versailles, 2003.

[37] B. FINANCE, Z. KEDAD. *Médiations d'informations via les méta-données.* Technical report, Contributions to the final deliverable of CNRS/STIC RTP09, also published as a short version in BDA'03 (Journées Bases de Données Avancées), 2003.

[38] P. PUCHERAL. *Bilan de l'AS CNRS Mobilité/Accès aux données.* Technical report, Final deliverable AS19_RTP09, short version published in BDA'03 (Journée Bases de Données Avancées), 2003.

Miscellaneous

[39] L. BOUGANIM, P. PUCHERAL. *Procédé de sécurisation de bases de données.* Patent deposit by the CNRS in august 2001, request for PCT (USA, Europe, Canada, Japan) in august 2002, end of the examination phase in august 2003, 2003.

[40] M. BOUZEGHOUB. *MDA : Some lessons learned from Database Technology and Design.* OMG Meeting on MDA, Paris, 2003, http://www.omg.org/docs/ad/03-06-01.pdf.

[41] C. CREMARENCO. *ActiveXML peer on mobile devices.* Proc. of the Int. Workshop on Emerging Database Research in Eastern Europe, colocated with VLDB'03, Berlin, 2003.

[42] V. PERALTA. *Data warehouse logical design from multidimensional conceptual schemas.* XXIX Conferencia

Latino-americana de Informática, winner of the 3rd Prize in the X Master Thesis Concourse of CLEI-UNESCO, La Paz, Bolivia, 2003.

[43] E. SIMON. *Specifying Scientific Applications by Means of Scientific Dataflow.* Int. workshop e-Science Workflow Services, Edinburgh, Scotland, 2003.

[44] G. VARGAS-SOLAR, J. ZECHINELLI-MARTINI, D. SOL-MARTINEZ, N. CRUZ-RAMIREZ, A. OLIART-ROS, P. VELASCO-ELIZONDO, C. COLLET, B. FINANCE, Z. KEDAD. *Spatial data integration from distributed and heterogeneous sources.* Proceedings of the Workshop on Advances in Databases and Information Retrieval, Tlaxcala, Mexico, 2003.

[45] C. YUAN, Y. ZHONG, Y. HE. *Chaos-Based Encryption Algorithm for Compressed Video.* Journal of Tsinghua University On Natural Science Version, 6, 2003.

[46] MEDIAGRID PROJECT. *A mediation framework for a transparent access to biological data sources.* European Conf. on Computational Biology, Paris, France, 2003, http://www.inra.fr/eccb2003/posters.htm.

Bibliography in notes

[47] G. AMATO, U. STRACCIA. *User Profile Modeling and Applications to Digital Libraries.* in « Proc. of the European Conf. on Research and Advanced Technology for Digital Libraries (ECDL) », 1999.

[48] E. BERTINO, S. CASTANO, E. FERRARI, M. MESITI. *Specifying and Enforcing Access Control Policies for XML Document Sources.* in « WWW journal, Baltzer Science Publisher », number 3, volume 3, 2000.

[49] H. HACIGUMUS, B. IYER, C. LI, S. MEHROTRA. *Executing SQL over Encrypted Data in the Database-Service-Provider Model.* in « Proc. of the Int. Conf. On Management of Data (ACM SIGMOD) », 2002.

[50] A. HALEVY. *Answering queries using views : A survey.* in « VLDB journal », number 4, volume 10, 2001.

[51] M. JARKE, M. LENZERINI, Y. VASSILIOU, P. VASSILIADIS. *Fundamentals of Data Warehouses.* in « VLDB journal, Springer (eds) », 1999.

[52] W. KIESSLING. *Foundations of Preferences in Database Systems.* in « Proc. of the Int. Conf. On Very Large Databases (VLDB) », 2002.

[53] M. LENZERINI. *Data Integration : a Theoretical Perspective.* in « Proc. of the Int. Conf. on Principles of Database Systems (ACM PODS) », 2002.

[54] S. MADDEN, M. FRANKLIN, J. HELLERSTEIN, W. HONG. *The Design of an Acquisitional Query Processor for Sensor Networks.* in « Proc. of the Int. Conf. On Management of Data (ACM SIGMOD) », 2003.

[55] R. MILLER, M. HERNÁNDEZ, L. HAAS, L. YAN, C. T. HOWARD HO, R. FAGIN, L. POPA. *The Clio Project : Managing Heterogeneity.* in « SIGMOD Record », number 1, volume 30, 2000.

[56] B. SCHNEIER. *Applied Cryptography.* in « book, 2nd Edition, John Wiley & Sons », 1996.

[57] G. WIEDERHOLD. *Mediators in the architecture of future information systems.* in « IEEE Computer », 1992.

[58] H. YU, D. AGRAWAL, A. EL ABBADI. *Tabular Placement of Relational Data on MEMS-based Storage Devices.* in « Proc. of Int. Conf. On Very Large Databases (VLDB) », 2003.

[59] COMPUTER SECURITY INSTITUTE. *CSI/FBI Computer Crime and Security Survey.* in « http://www.gocsi.com/forms/fbi/pdf.html », 2002.

[60] ORACLE CORP.. *Database Security in Oracle8i.* in « Oracle Technical White Paper », 1999, http://otn.oracle.com/deploy/security/oracle8i/index.html.