2021-03319 - PhD Position F/M [ALRC 2021] A Tracable Revocable Private Data Platform

Type de contrat : CDD
Niveau de diplôme exigé : Bac + 5 ou équivalent
Fonction : Doctorant

A propos du centre ou de la direction fonctionnelle

The Inria Lille - Nord Europe Research Centre was founded in 2008 and employs a staff of 320, including 280 scientists working in sixteen research teams. Recognised for its outstanding contribution to the socio-economic development of the Hauts-De-France region, the Inria Lille - Nord Europe Research Centre undertakes research in the field of computer science in collaboration with a range of academic, institutional and industrial partners.

The strategy of the Centre is to develop an internationally renowned centre of excellence with a significant impact on the City of Lille and its surrounding area. It works to achieve this by pursuing a range of ambitious research projects in such fields of computer science as the intelligence of data and adaptive software systems. Building on the synergies between research and industry, Inria is a major contributor to skills and technology transfer in the field of computer science.

Contexte et atouts du poste

Our society and industry manipulate enormous amounts of private data (with various privacy levels and properties). Medical personal data is the archetype of such data. Every day, individuals and doctors exchange and update data, but these sensitive updates are often traced nor controlled in ad-hoc fashion. In addition, the pharma industry is aggregating data from multiple sources to perform analyses on drugs and such data cohorts are consumed by different players. Within such scenarios, corrupted data is expensive because it can render completely large campaigns invalid. However, once mixed is impossible to trace and revoked corrupted data.

Such scenarios exhibit key properties that modern society is expecting from data manipulation:

- Traced. Any data should be able to request its origin, history and ownership. Even when recomposed to form new data, elementary data origin should be identifiable. In addition, the history of data changes should be accessible.
- Revocable. Invalid or corrupted data should be revocable and identifiable and this over the complete chain of composition.
- Multi-facet data. A single piece of data can have multiple facets (kind, dimension, validation, trigger...). In addition, such faceted data should also handle the fact that a piece of data can be nested and/or resulting from the composition of multiple other piece of data. Such result complex data should support revocation and origin/ownership of its subparts. Such faceted data may depend on the user and the user access rights: the same data in addition to its other properties may be invisible, read only, or writable.
- Ownership. To support ownership, researchers investigated ownership types [CNP01,BSBR03,CDN05,CSW13] for object encapsulation. [GN07] proposed different message sends to support ownership, implementation (for example of warranty of immutability), and speed.

State of the art within the laboratory

Several international efforts have been carried around elements of the proposal but none of them on its entirety and in the context of the RSPO.

- Capabilities. Works such as the ones around E (the language), [MMFO1,MS03,MY503,MI06], introduced capabilities to restrain object properties to support isolation.
- Isolation. Caja is an attempt to isolate Javascript applications [MSL+08,TEM+11].
- Ownership. To support ownership, researchers investigated ownership types [CNP01,BSBR03,CDN05,CSW13], for object encapsulation. [GN07] proposed different message sends to support dynamic ownership in a dynamically-type languages.

Work in the team

Nevertheless none of them radically rethink object-oriented programming in presence of the properties supporting traceability, ownership, history and revocation. We performed some preliminary study of the topic and developed a first prototype in the context of the CPER Data - 2 [S020]. The current proposal will revisit such background and take advantage of the knowledge accumulated by the following PhD defended in the team:

- The PhD of C. Teruel entitled "Adaptability and Encapsulation in Dynamically-Typed Languages: Taming Reflection and Extension Methods" and financed by the DGA was about the development of language constructs around ownership and how to control reflective operations that breach security concerns [Ter16, TDCD15].
- The PhD of J.B. Arnaud entitled "Towards First Class References as a Security Infrastructure in Dynamically-Typed Languages" was about the development of security constructs in dynamic languages [Arn13, ADO+10, ADOT11].
- The work around proxies as part of the PhD of M. Martinez-Peck and C. Teruel [MPBF+15, TWDN15].

Mission confiée

Ph.D. Objectives

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3500 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 180 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

Consignes pour postuler

PHD : CV + application letter + recommendation letters + school transcripts

Sécurité défense :
Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :
Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.
RGPD is a law to help citizens to control their data. However, there is a challenge to help developers build systems exhibiting by construction such properties. The easier it will be to produce systems like that the more chance we will have that our data are safely managed. The goal of this PhD is to explore and design a language and execution engine where data exhibit properties that support RGPD and Smart Data.

The PhD will develop language constructs and execution engine to support Smart Data. The objectives are:

- Revisit and design object model and execution supporting the definition of multi-facet properties (frame KR-based).
- Define a model of elementary operations to compose, aggregate object while supporting warranty of no modification, history and origin.
- Revisit capability model to support revocation and the smart data properties.
- Define publication mechanisms with non modification warranty based on a block-chain back-end.
- Validate the results on real case studies. PhD will work on:
  - State of the art on revocation, capability, ownership, and support for RGPD.
  - Identify scenarios with sensible data and complex multi player multi access rights.
  - Evaluate first class instance variables and how they can contribute to the design of faceted data.
  - Explore design of object-oriented languages supporting smart data (origin, history, ..).
- Identify a minimal change algebra supporting smart data properties.
- Explore design on the light of scalability issues.

**Principales activités**

The PhD will work on

- Identification of key scenarios extracted from real life case studies. We will contact the partners mentioned in the next section.
- State of the art on revocation, capability, ownership, and support for RGPD.
- Identify scenarios with sensible data and complex multi player multi access rights.
- Evaluate first class instance variables and how they can contribute to the design of faceted data.
- Explore design of object-oriented languages supporting smart data (origin, history, ..).
- Identify a minimal change algebra supporting smart data properties.
- Design an architecture to support the execution of such platform.
- Access the proposed solution on the light of scalability issues.

**Collaborations expected:**

At the moment of this writing we did not settle formal collaborations on the topic mainly because we want to be more advanced to be able to propose a first collaboration. Nevertheless the topic presented in this proposal has been designed based on discussions and workshops held with the following companies.

- Kertyx/Novartis Dominique Pahud is expert in medical and pharma data. These are the discussions with him that triggered this topic. We will recontact him and use his expertise to validate our prototypes.
- ZWEIDENKER GmbH is a company interested in the resulting infrastructure (execution engine and results of the PhD). They are interested in applying it in the context of IoT data.

In addition, Utocat is expert in Blockchain and we would like to see how the solution we may propose could raise an interest. We worked with Utocat during two years in the past, but Utocat is a start up and we prefer be able to propose them a first solution to launch a possible collaboration.

**Selected list of publications**

- Jean-Baptiste Arnaud, Marcus Denker, Stéphane Ducasse, Damien Pollet, Alexandre Bergel, and Mathieu Suen. Read-only execution for dynamic languages. In Proceedings of the 48th International Conference Objects, Models, Components, Patterns (TOOLS’10), Málaga, Spain, June 2010
- Ronie Salgado and Stéphane Ducasse. Towards a smart data processing and storage model. In International Workshop on Smalltalk Technologies IWST’20, August 2020
- Camille Teruel, Stéphane Ducasse, Damien Cassou, and Marcus Denker. Access control to reflection with object ownership. In Dynamic Languages Symposium (DLS’2015), 2015
- Camille Teruel, Erwann Wenli, Stéphanie Ducasse, and Oscar Nierstrasz. Propagation of behavioral variations with delegation proxies. Transactions on Aspect-Oriented Software Development (TAOSD), pages 63–95, 2015

**References:**

Jean-Baptiste Arnaud, Stéphane Ducasse, Marcus Denker, Damien Pollet, Alexandre Bergel, and Mathieu Suen. Read-only execution for dynamic languages. In Proceedings of the 48th International Conference Objects, Models, Components, Patterns (TOOLS’10), Málaga, Spain, June 2010.

Je souhaite intégrer votre équipe en tant que programmiste. Mes compétences techniques et relationnelles me permettent d’apporter un apport significatif à votre projet.

**Compétences**

- **Technical skills and level required** :
  - Languages : French and English
  - Relational skills :
    - Autonomy
    - Good communication

- Additional skills appreciated :
  - Agile programming
  - [http://www.pharo.org](http://www.pharo.org)
  - Test driven development

**Avantages**

- Partial reimbursement of public transport costs
- Subsidized meals
- Leave : 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Access to vocational training
- Possibility of French courses
- Social, cultural and sports events and activities

**Rémunération**

1st and 2nd year : 1 982€ Gross monthly salary (before taxes)

3rd year : 2 085€ gross monthly salary (before taxes)