Privacy Risks of Aggregates in Data Centric-Workflows (M/F)

Contract type: Public service fixed-term contract
Level of qualifications required: A levels or equivalent
Function: PhD Position

About the research centre or Inria department

The Inria Lille - Nord Europe Research Centre was founded in 2008 and employs a staff of 360, including 300 scientists working in sixteen research teams. Recognised for its outstanding contribution the socio-economic development of the Nord - Pas-de-Calais 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 as 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.

Context

Job environment

The overall goal of this project is to develop a sound understanding of privacy risks in systems involving big data (and in particular data mining) at a global level, to develop new algorithms allowing users to assess and understand privacy risks and control information flows, and to develop a proof of principle software tool demonstrating the value of these contributions. In particular, while many contributions in the state-of-the-art consider isolated problems, this project will study data handling processes at a global level within data-centric workflows, in order to allow for more integrated privacy risk analysis and more human-friendly tools (offering transparency and explanation).

Outcomes include a PhD and high-quality publications describing research addressing key questions towards this goal, the developed open-source software tool, and its use in a pilot study to demonstrate the value of the research in complex, real-world settings. This project will focus on a more global level (than previous approaches focusing on isolated tasks, on data-centric workflows, that may interact with the crowd). The same private data is often used by many interacting parties, who may all perform actions correlating with the data they have observed. Even if a user was perfectly aware of all individual treatments of his data and their privacy implications, he may not be able to assess the global privacy risks. This project aims at providing better and more global transparency and explanation.

This project will involve the following key challenges. First, it will focus on a more global level (than previous approaches focusing on isolated tasks, on data-centric workflows, that may interact with the crowd). Second, we will focus in this context on aggregation aspects. We will study amongst others the following key questions:

(Q1) How can we formalize data exploitation objectives, privacy requirements and data handling policies within the framework of data-centric workflows and crowd sourcing systems?
(Q2) Given a model detailing these elements, can we construct algorithms for assessing privacy risks and explaining them to users? In particular, what kinds of aggregates of private data collected by the workflow may be published without risk?
(Q3) How can we answer aggregation queries as needed for mining data in data-centric workflows on graph-structured data efficiently?

Outcomes include a PhD and high-quality publications describing research addressing key questions towards this goal, the developed open-source software tool, and its use in a pilot study to demonstrate the value of the research in complex, real-world settings. The project starts a new activity between the two Inria teams Links and Magnet. On the site of Links, the project subscribes to the ANR project Headwork (2016-2021) on data-centric workflows for crowd sourcing systems, for collecting and exploiting data by the crowd. It is also related to the ANR project Agreg (2014-19), where aggregate queries for database systems are studied on their own right.
On the side of Magnet, the project is related to Jan Ramon’s ERC Starting Grant ERC-StG 240186 MiGraNT concerned data mining in graph-structured data. After a first ERC-PoC project in the field of proteomics, he is currently leading ERC-PoC 713626 SOM on the use of the obtained results in the domain of transportation.

Own previous publications related to the topic:
2. Johann Brault-Baron, Florent Capelli, Stefan Mengel: Understanding Model Counting for beta-acyclic CNF-formulas. STACS 2015: 143-156

International publications relevant to the topic:

Benefits package

Benefits

- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Flexible working hours
- Sports facilities

More information about Lille:
http://www.lille3000.eu/portal/
http://www.lillemetropole.fr/mel.html

Remuneration

The gross monthly salary is 1982€ for 1st and 2nd year and 2085€ gross for the 3rd year.