Offer #2024-07409

PhD Position F/M Collaboration over a distributed file system

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

Context

This PhD thesis will take place in team COAST, under the supervision of Claudia-Lavinia Ignat, HDR, CRCN Inria, Inria center of Lorraine University and Gérald Oster, MCF, Lorraine University

Assignment

File system services are essential for data sharing and collaboration among users. Most of the collaborative file system services such as GoogleDrive and Dropbox rely on a central authority and place personal information in the hands of a single large corporation which is a perceived privacy threat. Users must provide their data to the vendors of these services and trust them to preserve the privacy of their data, but they have little control over the usage of their data after sharing it with other users. Moreover, the centralisation of the platforms hosting these services makes their scalability and reliability very costly. They often limit the number of persons that can simultaneously modify shared data, they generally rely on costly infrastructures and do not allow sharing of infrastructure and administration costs, and centralisation is not suitable for collaboration among a federation of organizations that want to keep control over their data and do not want to store their data at a third party.

A collaborative file system has to support hybrid collaboration including several collaboration modes:

- connected where user modifications are immediately shared and visible to the other users

- disconnected where users are not connected to the network. User modifications will be transmitted to the other users at the reconnection

- ad-hoc collaboration where subgroups of users can work together and synchronise at a later time with other members of the group

We want to build a distributed collaborative file system where control over the data is given to users who can share it directly only with the users they trust and without having to store it at a central authority. The distributed collaborative file system has to support the mentioned collaboration modes and seamless switch from one mode to the others.

We propose to investigate the use of peer-to-peer infrastructures such as IPFS (https://ipfs.io/) and Matrix (https://matrix.org/) on which we can plug replication mechanisms for file system synchronisation.

Data replication algorithms have to be reliable (i.e. after the reception of all modifications the replicas have to converge) and explainable (i.e., the decisions taken by these algorithms have to be understood by users and their intentions have to be respected). These algorithms have to be suitable for a large community of users that produces a large number of modifications with a high frequency. As data replication mechanism we propose to use CRDTs (Conflict-free Replication Data Types) [1] that respect Strong Eventual Consistency, a property that ensures convergence as soon as every replica has integrated the same modifications without further message exchange among replicas. Several works proposed CRDTs for file systems [2,3]. However, it rests to be investigated whether the proposed merging semantics satisfy user intentions.

The proposed solution will be tested with user studies.

Bibliography:

[1] Marc Shapiro, Nuno M. Preguiça, Carlos Baquero, and Marek Zawirski. Conflict-Free Replicated Data Types. In Xavier Défago, Franck Petit, and Vincent Villain, editors, Stabilization, Safety, and Security of Distributed Systems - 13th International Symposium, SSS 2011, Grenoble, France, October 10-12, 2011. Proceedings, volume 6976 of Lecture Notes in Computer Science, pages 386–400. Springer, 2011. doi:10.1007/978-3-642-24550-3_29.

[2] Mehdi Ahmed-Nacer, Stéphane Martin, and Pascal Urso. 2012. File system on CRDT. https://arxiv.org/abs/1207.5990

[3] Vinh Tao, Marc Shapiro, Vianney Rancurel. Merging semantics for conflict updates in geo-distributed file systems. SYSTOR 2015: 10:1-10:12

Main activities

- Study of literature on CRDTs

- Study CRDTs for file systems

- Propose a file system CRDT with merging semantics that satisfy user intentions

- Implement the proposed CRDT over a peer-to-peer infrastructure such as Matrix or IPFS

- Design user studies for testing the proposed system

Skills

- Engineering and/or Master 2 degree in Computer science / Applied mathematics with an experience in computer networks.

- Theoretical expertise: distributed systems, P2P networks

- Good collaborative and networking skills, excellent written and oral communication in English
- Good programming skills
- Strong analytical skills

Benefits package

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

Remuneration

2100€ gross/month the 1st year

General Information

- **Theme/Domain :** Distributed Systems and middleware
System & Networks (BAP E)
- **Town/city :** Villers lès Nancy
- **Inria Center :** Centre Inria de l'Université de Lorraine
- **Starting date :** 2024-10-01
- **Duration of contract :** 3 years
- **Deadline to apply :** 2024-05-31

Contacts

- **Inria Team :** COAST
- **PhD Supervisor :**
Ignat Claudia-lavinia / claudia.ignat@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

Instruction to apply

Defence Security :
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :
As part of its diversity policy, all Inria positions are accessible to people with disabilities.