



Offre n°2024-07409

## PhD Position F/M Collaboration over a distributed file system

*Le descriptif de l'offre ci-dessous est en Anglais*

Type de contrat : CDD

Niveau de diplôme exigé : Bac + 5 ou équivalent

Fonction : Doctorant

### Contexte et atouts du poste

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

### Mission confiée

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

## Principales activités

- 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

## Compétences

- 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

## Avantages

- 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

## Rémunération

2100€ gross/month the 1st year

## Informations générales

- **Thème/Domaine** : Systèmes distribués et intergiciels  
Système & réseaux (BAP E)
- **Ville** : Villers lès Nancy
- **Centre Inria** : [Centre Inria de l'Université de Lorraine](#)
- **Date de prise de fonction souhaitée** : 2024-10-01
- **Durée de contrat** : 3 ans
- **Date limite pour postuler** : 2024-07-12

## Contacts

- **Équipe Inria** : [COAST](#)
- **Directeur de thèse** :  
Ignat Claudia-lavinia / [claudia.ignat@inria.fr](mailto:claudia.ignat@inria.fr)

## A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 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 200 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.

**Attention:** Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

## Consignes pour postuler

### **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.