

## Progress Report and CSI Report – 2018-19

*This report contains 4 parts to be filled in French or English:*

- an administrative form, to be filled by the PhD student*
- the progress report of the PhD thesis, to be filled by the PhD student*
- the conclusion of the supervisors*
- the report of the Comité de Suivi Individuel (CSI), to be filled by the CSI*

*For the 4th part, the PhD student must plan a meeting with the “réfèrent scientifique” (first member of his/her CSI) in May. The goal of this meeting is to discuss the progress of the thesis. Prior to this meeting, the other parts must be filled and the full document e-mailed to the réfèrent scientifique.*

*At the end of this meeting, the report is to be signed by both members of the CSI, and sent by the CSI to the PhD student, the supervisors and the CMI Doctoral school (francoise.laurent@loria.fr). It is then to be uploaded by the PhD student to her/his ADUM profile.*

*In order to facilitate the process, the different parts can be prepared and signed (electronically) as separate documents, but they must be merged again as a single pdf file for uploading to ADUM.*

---

### **Administrative information** (to be filled by the PhD student)

**PhD student:** Matthieu Nicolas

**Current PhD year in 2018-19:** 2A

**Laboratory:** Loria

**Supervisor(s):** Olivier Perrin et Gérald Oster

**Title of the PhD thesis:** Efficient (re)naming in Conflict-free Replicated Data Types (CRDTs)

**Fundings for the PhD thesis:** Contrat doctoral

#### **Members of the CSI:**

- **scientific member:** Stephan Merz
- **auxiliary member:** Ye-Qiong Song

## Progress report (to be filled by the PhD student)

**Short description of the subject:** *general context, considered approach, scientific objectives compared to the state of the art*

Dans le cadre de ma thèse, j'étudie et travaille sur les Conflict-free Replicated Data Types (CRDTs). Les CRDTs sont de nouvelles spécifications des types abstraits de données, tels que l'*Ensemble* ou la *Séquence*. Contrairement aux spécifications traditionnelles, les CRDTs sont conçus pour supporter nativement les modifications concurrentes. Pour ce faire, ces structures de données intègrent un mécanisme de résolution de conflits directement au sein de leur spécification. Cette spécificité rend les CRDTs particulièrement adaptés pour concevoir des systèmes distribués hautement disponibles dans lesquels différents noeuds répliquent et modifient une même donnée sans aucune coordination.

Pour résoudre les conflits de manière déterministe, les CRDTs utilisent généralement des identifiants qu'ils associent aux éléments stockés au sein de la structure de données. Cependant, selon le type de CRDT, les identifiants doivent respecter un ensemble de contraintes telles qu'être unique ou appartenir à un espace dense. Dans certains cas, ces contraintes empêchent de borner la taille des identifiants. La taille des identifiants croît alors continuellement avec le nombre de modifications effectuées.

Ces identifiants représentent donc un surcoût lié à l'utilisation des CRDTs par rapport aux structures de données traditionnelles. Ce surcoût, parfois non-borné, peut décourager l'adoption des CRDTs par les systèmes distribués. Le but de cette thèse est de proposer des solutions pour pallier à ce problème. Une approche possible est d'étudier ce sujet comme étant un cas particulier du problème du renommage. Il convient alors de proposer un mécanisme permettant de renommer les identifiants pour réduire leur taille, sans introduire la nécessité d'une coordination entre les différents noeuds.

**Main results of the year:** *studies and works achieved, results obtained with respect to the objectives of the thesis; difficulties encountered*

**Travaux précédents.** Au cours de ma 1ère année, j'ai étudié principalement un CRDT donné, *LogootSplit*. Ce CRDT permet de représenter une séquence répliquée et est notamment utilisé dans des applications d'édition collaborative. *LogootSplit* souffre particulièrement du problème de croissance non-bornée des identifiants qu'il utilise.

Afin d'adresser ce problème, j'ai proposé une première version d'un nouveau CRDT, *RenamableLogootSplit*. Ce CRDT associe *LogootSplit* à un mécanisme de renommage. Ce mécanisme vise à réduire le surcoût de *LogootSplit* en réallouant ponctuellement des identifiants plus petits à l'ensemble des éléments stockés au sein de la séquence. Pour gérer les modifications concurrentes à un renommage, *RenamableLogootSplit* utilise des fonctions de transformation. Ces fonctions permettent de transformer les modifications avant de les appliquer à la copie locale pour prendre en compte l'effet du renommage.

Cette première version était néanmoins incomplète: elle n'était pas capable d'intégrer des renommages concurrents. Pour l'utiliser en état, il était donc nécessaire d'empêcher des renommages concurrents d'être générés, notamment en reposant sur un mécanisme de coordination. Les mécanismes de coordination étant coûteux et peu adaptés aux systèmes dynamiques, il s'agissait du défaut principal de ce prototype.

**Travaux en cours.** Cette année, j'ai poursuivi mes travaux sur *RenamableLogootSplit* en concevant et implémentant une nouvelle version de ce dernier. La principale évolution de cette nouvelle version

est la capacité de gérer des renommages concurrents. Pour cela, *RenamableLogootSplit* établit un ordre de priorité entre des renommages concurrents. En se basant sur cet ordre, *RenamableLogootSplit* n'applique en finalité qu'un seul renommage d'un ensemble de renommages concurrents. De nouvelles fonctions de transformation permettent d'annuler l'effet d'un renommage précédemment appliqué, s'il s'avère finalement qu'il existe un renommage concurrent prioritaire.

Cette solution présente toutefois un défaut. Afin que les différentes fonctions de transformation soient déterministes, elles reposent sur l'état sur lequel s'est appliqué le renommage à prendre en compte ou à annuler pour décider de la transformation à effectuer. Il est donc nécessaire de conserver les différents états renommés tant qu'ils peuvent être requis pour transformer une modification concurrente ou qu'un renommage concurrent prioritaire peut être reçu. J'ai donc étudié *RenamableLogootSplit* pour identifier les conditions nécessaires pour pouvoir supprimer définitivement les états renommés et concevoir un mécanisme de *garbage collection* adapté.

Avec l'aide d'un ingénieur de notre équipe, nous sommes en train de mettre en place plusieurs expériences pour valider cette contribution. La première expérience consiste à simuler une collaboration à grande échelle à l'aide de bots. Les objectifs de cette expérience sont multiples. Tout d'abord, elle nous permet d'éprouver le bon fonctionnement de l'application en vérifiant que les différents noeuds convergent à terme. Nous utilisons aussi cette expérience pour générer et collecter un grand nombre d'états différents, qu'ils s'agissent d'états intermédiaires ou finaux. Ces états sont utilisés dans le cadre d'une seconde expérience pour mesurer la consommation en mémoire du *CRDT*, mais aussi le temps d'intégration des modifications. Ceci nous permet de suivre l'évolution de ces données au fur et à mesure de la progression de la collaboration.

Afin de présenter cette contribution à la communauté des *CRDTs*, je rédige actuellement un article scientifique décrivant son fonctionnement et comparant ses performances par rapport à l'existant.

**Plan for next year:** *remaining problems to be considered, approach; precise planning (in particular for those who are at least in 3rd year, for the report writing and the defense)*

L'objectif de cette année est de finir la rédaction et de publier l'article présentant *RenamableLogootSplit*.

Afin de compléter la partie validation de cet article, nous allons poursuivre les expériences que nous menons pour effectuer des mesures de performances de *RenamableLogootSplit* sur différents aspects: temps d'intégration du renommage, temps d'intégration des modifications concurrentes à un renommage, évolution de la mémoire utilisée par la structure de données... Ces mesures seront utilisées pour comparer les performances de *RenamableLogootSplit* par rapport à l'existant.

Pour garantir la convergence à terme des différents noeuds du système mais aussi pour garantir le respect de l'intention de l'utilisateur, les fonctions de transformation doivent vérifier plusieurs propriétés. Ces fonctions étant critiques pour la correction du système, je souhaite prouver formellement leur respect de ces propriétés.

Concernant la suite de mes travaux, j'étudie actuellement la littérature sur les *CRDTs* pour identifier d'autres structures de données qui bénéficieraient d'une adaptation du mécanisme de renommage proposé pour *LogootSplit* ou, le cas échéant, pour identifier d'autres problématiques de recherche liées à l'utilisation des *CRDTs* dans les systèmes distribués.

**Publications:** *if any*

[1] Matthieu Nicolas. Efficient renaming in CRDTs. In *Middleware 2018 - 19th ACM/IFIP International Middleware Conference*, Rennes, France, December 2018.

**Project after the thesis:** *for 3rd year (and beyond) PhD students: professional project*  
À définir.

**Scientific and professional modules validated:** *list of all modules validated from the beginning of the thesis (do not forget to add them in Adum, together with your publications)*

### Scientific modules

- Réplication et cohérence des données
- Participation à l'école d'été SATIS 2018

### Professional modules

- Fi4 152 E Sauveteur Secouriste du Travail (SST)
- Fi4 162 C Formation à la communication orale et corporelle en milieu professionnel
- Fi4 282 Outils numériques pour la pédagogie (plateforme Arche, studio professeur)
- Fi4 305 Culture de l'intégrité scientifique
- PA1.1 MDD 14 - Eléments d'innovations pédagogiques

**Date and signature of the PhD student**

.....

<b>Opinion of the supervisors</b>
-----------------------------------

*(to be filled by the supervisors)*

**Opinion on this progress report:** .....

**Agreement for an additional year?** *Yes/No (if No, please justify)*

.....

**Date of the defense:** *this can be an approximation*

.....

**Date and signature of the PhD supervisors**

.....

# Report of the Comité de Suivi Individuel *(to be filled by the référent scientifique)*

*The questions below are suggestions. For each section, the report can be as short as “Yes” or more detailed according to the feeling of the CSI.*

**Is the student confident with the PhD progression?** *E.g.: How does the PhD student view the progress of her/his thesis? How often does the student meet with her/his supervisors? How is the student’s relationship with her/his supervisors? In case of problems, does the student know who to discuss these issues with?*

.....

**Is the PhD on good tracks?** *E.g.: Is the student comfortable with his/her thesis topic? Did she/he embraced the subject? Do you have any comments or advice concerning publications? Does the student have opportunities to present her/his work? Do you have noteworthy concerns about the progression of the thesis?*

.....

**Is the professional project sound?** *E.g.: What is the professional project of the PhD student for after the PhD? Is she/he aware of the various options and associated expectations (postdoc abroad or in France, industrial R&D, application periods, teaching requirements, etc.)? Have contacts been established? Do you have comments about the PhD student’s plans for her/his training courses/schools?*

.....

**Conclusion:** No problem for an additional registration / Interview with the ED recommended

**Date:** .....

## Signatures:

PhD student

.....

Two members of the CSI

.....