

Adresse postale

C.P. 6079, succ. Centre-ville
Montréal (Québec) Canada
H3C 3A7
Téléphone : (514) 340-INFO
(4636)
Télécopieur : (514) 340-3240

Rapport sur la thèse de Doctorat de MATTHIEU NICOLAS

Adresse civique

Campus Université de Montréal
2900, boul. Édouard-Montpetit
École Polytechnique de Montréal
2500, chemin de Polytechnique
Pavillons Pierre-Lassonde et
Claudette-MacKay-Lassonde
3^e étage
Montréal (Québec) Canada
H3T 1J4

**École affiliée à
l'Université de Montréal**

Cette thèse, intitulée « Ré-identification sans coordination dans les types de données répliquées sans conflits (CRDTs) », a été réalisée sous la direction de Monsieur Olivier PERRIN, dans l'unité de recherche « LORIA - Laboratoire Lorrain de Recherche en Informatique et ses Applications » et au sein de l'école doctorale « IAEM - INFORMATIQUE - AUTOMATIQUE - ELECTRONIQUE - ELECTROTECHNIQUE - MATHEMATIQUES ».

Cette thèse s'intéresse aux systèmes d'édition collaborative décentralisés (Pair-à-Pair). Un système d'édition collaborative Pair-à-Pair permet à un groupe d'utilisateurs de partager et de modifier un document, selon le principe de copies multiples. Chaque utilisateur dispose d'une copie locale du document qu'il peut modifier librement. Le système se charge de diffuser les modifications locales aux autres sites et d'intégrer les opérations reçues des autres sites à la copie locale.

L'un des défis majeurs de tels systèmes est d'assurer, à **moindre coût**, la convergence des copies tout en préservant la causalité et les intentions des utilisateurs. Comme, en général, les opérations d'édition ne sont pas commutatives, plusieurs approches ont été proposées, dans la littérature, pour forcer la commutativité des opérations. Parmi ces approches figurent les transformées opérationnelles (OT) et les Conflict-free Replicated Data Types (CRDT) comme TreeDoc, Logoot, Logoot-Undo et WOOT.

Dans les approches OT, les opérations distantes reçues par un site, vont parfois subir des transformations afin de tenir compte des effets des opérations concurrentes exécutées localement. Chaque site maintient l'historique des opérations exécutées ainsi que le vecteur d'horloges des événements produits dans le site. Le vecteur d'horloges courant du site est ajouté à chaque opération transmise aux autres sites. Il permet aux sites distants de déterminer les opérations exécutées jusqu'à présent qui sont concurrentes à l'opération reçue et de réaliser les transformations nécessaires avant de l'exécuter.

Adresse postale

C.P. 6079, succ. Centre-ville
Montréal (Québec) Canada
H3C 3A7
Téléphone : (514) 340-INFO
(4636)
Télécopieur : (514) 340-3240

Adresse civique

Campus Université de Montréal
2900, boul. Édouard-Montpetit
École Polytechnique de Montréal
2500, chemin de Polytechnique
Pavillons Pierre-Lassonde et
Claudette-MacKay-Lassonde
3^e étage
Montréal (Québec) Canada
H3T 1J4

École affiliée à
l'Université de Montréal

Le maintien de l'historique des opérations n'est pas requis dans les approches CRDT car elles se basent sur l'identification de manière unique de chaque élément du document. Les éléments du document sont totalement ordonnés par rapport à leurs identifiants et apparaissent dans l'ordre croissant de leurs identifiants. Le principal défi de ces approches est la gestion à moindre coût des identifiants (des métadonnées) car, d'une part, leur exactitude dépend de l'unicité des identifiants attribués aux éléments du document et, d'autre part, la latence des utilisateurs, peut être directement impacté par le coût en temps et espace de cette gestion et compromettre ainsi le passage à l'échelle. Plusieurs tentatives ont été proposées dans la littérature pour réduire le surcoût de la gestion des identifiants. Cependant, elles sont moins appropriées aux systèmes collaboratifs pair-à-pair à large échelle.

La contribution principale de cette thèse est la conception et l'évaluation d'un nouveau CRDT qui étend LogootSplit par un nouvel opérateur de renommage. Cet opérateur a pour but de produire un nouvel état du document, équivalent à son état précédent, mais avec des métadonnées plus compactes. La gestion de conflits induits par ce nouvel opérateur est basée sur les priorités et les transformations d'opérations concurrentes aux opérations de renommage et ne nécessite pas de coordination synchrone entre les noeuds. Les différentes simulations de cette approche ont permis de tester la convergence au final des copies multiples en présence, notamment, d'opérations de renommage concurrentes et de mesurer l'impact positif de cet ajout sur la taille des métadonnées malgré le surcoût engendré.

Cette thèse est composée de 5 chapitres. Le premier chapitre définit le contexte de recherche, son objectif et le plan du document. Le chapitre 2 est consacré à l'état de l'art des approches CRDTs, les avantages, les limites de chacune de ces variantes, et la motivation de la problématique de recherche abordée dans cette thèse. Le chapitre 3 présente l'approche CRDT proposée, son évaluation, sa comparaison avec les autres approches CRDT ainsi que ses limites. Le chapitre 4 présente, MUTE, l'éditeur de texte collaboratif temps réel développé par l'équipe de recherche du candidat. Le chapitre 5 résume les contributions de cette thèse puis introduit quelques pistes de recherches futures.

Les problèmes de convergence et de passage à l'échelle sont toujours parmi les défis majeurs qui freinent le déploiements à grande échelle des systèmes collaboratifs décentralisés, malgré leurs nombreux avantages, relativement aux systèmes collaboratifs centralisés.

Adresse postale

C.P. 6079, succ. Centre-ville
Montréal (Québec) Canada
H3C 3A7
Téléphone : (514) 340-INFO
(4636)
Télécopieur : (514) 340-3240

Adresse civique

Campus Université de Montréal
2900, boul. Édouard-Montpetit
École Polytechnique de Montréal
2500, chemin de Polytechnique
Pavillons Pierre-Lassonde et
Claudette-MacKay-Lassonde
3^e étage
Montréal (Québec) Canada
H3T 1J4

École affiliée à
l'Université de Montréal

Les problèmes traités dans cette thèse sont très difficiles et les solutions proposées pour améliorer l'existant semblent tout à fait louables. L'évaluation et la comparaison avec l'existant a permis de valider les améliorations proposées mais aussi de dégager les limites quant au passage à l'échelle. Les contributions présentées dans cette thèse ont fait l'objet de 2 articles de conférences et 1 article de revue. Cela témoigne de la maîtrise de la problématique des systèmes collaboratifs décentralisés.

Le document couvre les principales composantes attendues : 1) État de l'art sur les approches CRDT, les caractéristiques, les avantages et les limites de ses approches ainsi que la motivation des problèmes abordés dans cette thèse. 2) Présentation d'une approche CRDT visant à améliorer l'existant. 3) Évaluation et comparaison de l'approche proposée avec les approches CRDT existantes. La qualité de la présentation est très bonne. La liste des références couvre bien les approches CRDT, leurs comparaisons et leurs limitations. Il me semble intéressant de compléter l'état de l'art et la comparaison avec les approches OT ou motiver le choix de ne pas les considérer comme une alternative aux approches CRDT.

Pour toutes ces raisons, je suis très favorable à la soutenance de cette thèse en vue de l'obtention du grade de docteur en « Informatique ».



Hanifa Boucheneb
Professeur titulaire
Département de génie informatique
et génie logiciel
Polytechnique de Montréal
Phone : (1) 514 340 4711 ext. 4101
Email : hanifa.boucheneb@polymtl.ca